

Klevmarken, N. Anders et al.

**Working Paper**

## Simulating the future of the Swedish baby-boom generations

Working Paper, No. 2007:26

**Provided in Cooperation with:**

Department of Economics, Uppsala University

*Suggested Citation:* Klevmarken, N. Anders et al. (2007) : Simulating the future of the Swedish baby-boom generations, Working Paper, No. 2007:26, Uppsala University, Department of Economics, Uppsala,  
<https://nbn-resolving.de/urn:nbn:se:uu:diva-11648>

This Version is available at:

<https://hdl.handle.net/10419/82694>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



UPPSALA  
UNIVERSITET

Working Paper 2007:26

*Department of Economics*

# Simulating the future of the Swedish baby-boom generations

N. Anders Klevmarken

in cooperation with

Kristian Bolin  
Matias Eklöf  
Lennart Flood  
Urban Fransson  
Daniel Hallberg  
Sören Höjgård  
Mårten Lagergren  
Björn Lindgren  
Andrea Mitrut

Department of Economics  
Uppsala University  
P.O. Box 513  
SE-751 20 Uppsala  
Sweden  
Fax: +46 18 471 14 78

Working paper 2007:26  
June 2007  
ISSN 1653-6975

## SIMULATING THE FUTURE OF THE SWEDISH BABY-BOOM GENERATIONS

N. ANDERS KLEVMARKEN

IN COOPERATION WITH

KRISTIAN BOLIN  
MATIAS EKLÖF  
LENNART FLOOD  
URBAN FRANSSON  
DANIEL HALLBERG  
SÖREN HÖJGÅRD  
MÅRTEN LAGERGREN  
BJÖRN LINDGREN  
ANDREA MITRUT

Papers in the Working Paper Series are published  
on internet in PDF formats.  
Download from <http://www.nek.uu.se>  
or from S-WoPEC <http://swopec.hhs.se/uunewp/>

# Simulating the future of the Swedish baby-boom generations\*

N. Anders Klevmarken†

Department of Economics, Uppsala University

in cooperation with

Kristian Bolin

LUCHE, Lund University

Matias Eklöf

Department of Economics, Uppsala University

Lennart Flood

Department of Economics, Göteborg University

Urban Fransson

Göteborg University and IBF

Daniel Hallberg

Department of Economics, Uppsala University

Sören Höjgård

LUCHE, Lund University

Mårten Lagergren

Stockholm Gerontology Research Center

Björn Lindgren

LUCHE, Lund University

Andrea Mitrut

Department of Economics, Göteborg University

June 15, 2007

**JEL** classifications: C15, C50, D10, D31, H24, H31, H55, I12, I32, J10, J22, J26, R21, R23

**Keywords:** Microsimulation, Ageing, Retirement, Health status, Health care, Social care, Distribution of Income, Distribution of Wealth, Poverty

---

\* Paper prepared for presentation at the 1<sup>st</sup> General conference of the International Microsimulation Association (IMA), August 20-22, 2007 in Vienna. This paper is based on a book manuscript “Simulating an ageing population. A microsimulation approach applied to Sweden”, to become published in the Elsevier series Contributions to Economic Analysis. It documents the results of the research project “The Old Baby-boomers”. Each of the above mentioned scientists have contributed to this project. This paper draws freely from their work.

Financial support from the Swedish Council for Working Life and Social Research (FAS) is gratefully acknowledged

† Corresponding author: Email [anders.klevmarken@nek.uu.se](mailto:anders.klevmarken@nek.uu.se), Mail address: Department of Economics, Uppsala University, P.O. Box 513, SE-751 20 Uppsala, Sweden.

## **Abstract**

For the purpose of studying the consequences of the ageing of the Swedish population a group of scientists have enlarged the microsimulation model SESIM - originally developed at the Swedish Ministry of Finance - with modules that simulate health status, take up of sickness benefits, retirement, the utilization of health care and social care and the dynamics of the income and wealth distributions. This paper motivates and reviews the structure of these modules with a focus on problems and solutions. It also summarizes the main results of the simulations. A complete description of the models and results are forthcoming in a volume included in the Elsevier series Contributions to Economic Analysis.

### **1. Introduction**

Many Western countries now experience the beginning of the retirement of the large baby-boom cohorts. The viability of the pension systems and the ability of society to meet the expected future high demand for health care and social care of these cohorts are urgent policy issues in these countries. For the purpose of analyzing the consequences of the aging of the baby-boom cohorts in Sweden a group of scientists from different disciplines has enlarged the dynamic micro simulation model SESIM of the Swedish Ministry of Finance by adding a number of new models capturing changes in health status, retirement, geographical mobility, use of hospital care and social care and changes in the distribution of wealth. The model has then been used to simulate alternative scenarios involving improved health, higher retirement age and higher immigration all of which are compared to a base line scenario to assess the impact on the elderly and the ability of society to provide for them. In this paper we review the structure of SESIM with a focus on the new submodels added to the simulation model, and present a few key results.

### **2. The microsimulation model SESIM**

In 1997 SESIM was developed as a tool at the Swedish ministry of finance to evaluate the Swedish system to finance higher education. Part of that work was documented in Ericson and Hussénus, (2000). We refer to this as version I of SESIM. Focus then shifted from education to pensions. SESIM was used to evaluate the financial sustainability of the new Swedish pension system. This new application implied that SESIM was developed into a general micro-simulation model that can be used for a broad set of issues. We refer to this as the

second version of SESIM and the documentation is presented in Flood et.al (2003). The present version, SESIM III - BABYBOOM, maintains the focus on pensions but extends the analyses to include health issues, regional mobility and wealth. (In the following we will usually refer to this latest version of SESIM by just using this abbreviation SESIM.)

SESIM II has recently been used in several studies: Flood (2007) calculated the replacement rates of the Swedish pension system, Pettersson and Pettersson (2003, 2007) studied income redistribution over the life-cycle and Pettersson et. al (2006) analysed inter generational transfers.

## 2.1 The structure of SESIM

SESIM is a mainstream dynamic microsimulation model in the sense that the variables (events) are updated in a sequence, and the time span of the updating processes is a year. The start year is 1999 and every individual included in the initial sample of about 100 000 individuals goes through a large number of events, reflecting real life phenomena, like education, marriage, having children, working, retirement etc. Every year individuals are assigned a status, reflecting their main occupation in that year. A status is related to a source of income, working gives earnings, retirement gives pensions etc. The tax and benefit systems are applied to simulated incomes and after tax income is calculated. If the simulations are repeated for a long time period individual life-cycle incomes can be generated.

The sequential structure of SESIM is presented in Figure 1. The first part consists of a sequence of demographic modules (mortality, adoption, migration, household formation and dissolution, disability pension, and rehabilitation). There is also a new module that simulates regional mobility. Then a module for education follows (compulsory school, high school (gymnasium), municipal adult education (komvux) and university).

The next module deals with the labor market including the retirement decision. The date of retirement can be decided according to a retirement model, but it is also possible to choose a specific age of retirement. The labor market module also includes models for sick leave and unemployment, and a model that imputes labor market sector. A sector is required for the calculations of occupational pensions.

Having gone through the sequence this far, the next step is to simulate a status for each individual. There are nine different statuses. Each individual can only have one status in a year (the status “emigrated” is an exception). These statuses reflect the main occupation in a year. This is of course a simplification, because in reality an individual can have many occupations in a year. One can be a student part of the year and work the other part, or one

can have several occupations at the same time. The list of statuses is,

- 1 child (0-15 years old)
- 2 old age retired; individuals with income from old age pension
- 3 student; individuals who study at gymnasium, adult education or university
- 4 disabled; individuals who have disability/sickness benefits
- 5 parental leave; women who give birth during the year
- 6 unemployed; individuals with income from unemployment insurance or from labor market training
- 7 miscellaneous
- 8 employed; individuals in market work
- 9 emigrated; individuals living abroad with Swedish pensions rights. (Note, this classification is not unique since they can also have income from early or old age pensions.)

Status determines income. For employed (status 8) an earnings equation (see below), is used to determine income. Unemployed get unemployment benefits, disabled get disability benefits, etc. following the statutory rules.

After the income module, a module which simulates wealth, capital income and housing is entered. After the wealth/housing module a large module applies all relevant tax, transfer and pension rules. For the old age pension system, the rules for public and occupational pensions have been implemented in all relevant details. After all incomes and taxes have been computed the household disposable income can be obtained. Next we have implemented a new health module. It simulates health status, sickness absence from work, inpatient care and the proximity of parents to their children, and social care.

Obviously, an important characteristic of the SESIM model is the notion of full time status. The model assumes that being employed, retired, student, etc is full time. This has implications for the income generation. For instance, income from work is calculated based on an estimated equation for full time earnings. Alternatively a labour supply model could have been used in order to impute yearly hours of work and then yearly income could be obtained using a model for an hourly wage rate. The advantage of this alternative approach is that earnings for part time work can be calculated. However, simplicity is the advantage of the present structure which only considers full time status. Once part time status is allowed for this has to be implemented in a consistent way. For example, if an individual is simulated to work part time, then there must be a complementary part time status, say part time pensioner

or part time student and obviously this would complicate the structure. However, there is one exception from the principle that income is generated conditional on status. Students have income from study benefits but also from earnings. Thus, even if the status is (fulltime) student we for this group allow for additional earnings from part time or temporary employment.

It is important for many purposes to get a good representation of the household composition. Many stochastic models use household information, and some benefit systems too. For instance, to compute social assistance, housing- and child allowances, one needs to know who lives in a household. In SESIM the model population lives in households. Like in reality the household composition can change. New households are formed and households split. In the base population households are real observed households (with some modifications). During the course of simulation household formations and dissolutions are simulated using the demographic models of SESIM.

## 2.2 Data sources

This section gives a brief description of the main data sources for estimation and construction of the model population. We also discuss some corrections or adjustments that have been made to this population.

### *LINDA – a panel data base<sup>‡</sup>*

LINDA is the main source of information used in SESIM. This panel data set covers about 3,5 percent of the Swedish population. For year 1999 this implies that 308 000 individuals were randomly selected. For each of them all household members were added. In total the sample size became 786 000 individuals in 1999. This is the primary database of SESIM for the estimation of models as well as the construction of the model population.

The selected individuals are followed backwards and forwards and data from a number of registers are collected. Some information, for instance pension rights, can be traced back as long as to 1960. Selected individuals who disappear from the data by death or emigration are replaced by newly selected individuals in such a way that each cross-section is a random sample from the Swedish population of the same year.

Note that the database is completely created from administrative registers. Thus

---

<sup>‡</sup> Longitudinal Individual Data for Sweden. For documentation see Edin och Fredriksson (2000).



no interview is needed and therefore a major advantage is that there are no problems of attrition. The registers used cover income and wealth, earnings, pension rights, sickness- and unemployment benefits, schooling and census data.

The base population used in SESIM is formed by a random draw of 104 000 individuals from the 1999 LINDA. To this sample 8 000 individuals have been added from a register for pensions rights at the National Social Insurance Board . This additional sample includes individuals living outside Sweden, but with Swedish pension rights.<sup>§</sup>

In the construction of the base population in SESIM two major adjustments have been made in order to obtain a model population consistent with the definitions used in SESIM. They are described below in section 2.3.

#### *Other data sources*

In addition to LINDA a few data sources have been used for estimation or imputation . They are HINK/ HEK, GEOSWEDE, The Kungsholmen study and ULF. HINK/HEK\*\* is a sample survey to which Statistics Sweden has administered telephone interviews and then merged the survey data with register information. In the interview it is possible to obtain the data needed to use a meaningful household definition for economic analysis. This is in contrast to LINDA, which essentially is limited to a household definition used for taxation purposes. Apart from being used in the creation of useful households in SESIM, HINK/HEK has also been used to estimate models of public consumption, the size of an owner occupied home and the cost of housing.

Models of regional mobility and tenure choice were based on GEOSWEDE, a research data base created from a number of administrative registers covering the whole Swedish population. The health and care module is based on data from the Kungsholmen study , ULF and HINK/HEK. The Kungsholmen Study is a small study of old age care in a parish of Stockholm. The Survey of Living Conditions in Sweden (ULF) is a sample survey covering a number of welfare components including health status.

---

<sup>§</sup> For most of our analysis in the project “The old baby-boomers” we have focused on the population resident in Sweden, while emigrants have been excluded.

\*\* The income distribution survey of Statistics Sweden, a yearly sample survey of about 30 000 individuals merged with administrative data.

### 2.3 Adjustments to a meaningful household definition.

To define a household in LINDA one uses information from the population registers (RTB) and the so called inter generational register. The latter register links parents with children. This implies that all individuals in LINDA are assumed to live in the municipality where they are recorded according to the national tax registry, and that adults living together without being legally married and having common children are considered separate households. A comparison with the HINK/HEK surveys shows that there are two problems in LINDA. First, the number of youngsters between 18 and 29 who still live in their parents' household are over reported. Many children who have moved are still tax registered in their old household. Secondly, the number of cohabitants without children is underestimated, especially for young couples.

To reduce the number of youngsters living with their parents a model for the probability to move from parents was estimated based on HINK/HEK data for those in the age bracket 18-29.<sup>††</sup> This model was used in the following way to impute in 1999 which kids had left the home of their parents. All individuals at risk were ranked according to the predicted probability of moving, and the individuals with highest probabilities were chosen such that the HINK/HEK frequencies were matched.

To increase the number of cohabiting couples without children a model was estimated for the probability of being married/cohabiting in the age group 18 to 29 among women without children using HINK/HEK.<sup>‡‡</sup> Based on the estimated model the probability of being cohabiting/married was calculated for 18-29 years old females without children in LINDA. The individuals were ranked after the predicted probabilities. A number of women with highest probabilities were matched with single males two years older than the female and at the same level of education. The number of couples created in this way was chosen such that it coincided with the HINK/HEK frequencies.

---

<sup>††</sup> The probabilities are estimated by a logit regression. The explanatory variables are: Gender, age, square of age, born in Sweden (dummy), Study allowance / 1000, earnings / 1000, highest education (dummies compulsory school and upper secondary school), interaction age and education.

<sup>‡‡</sup> The probabilities were estimated by a logit regression. The explanatory variables were: Gender, age, square of age, born in Sweden (dummy), Study allowance / 1000, earnings / 1000, highest education (dummies compulsory school and upper secondary school), interaction age and education.

## 2.4. SESIM – a stochastic simulation model

SESIM is a *stochastic simulation model*, which means that the statistical models include random components. In the simulation a *Monte Carlo* technique is used to generate a stochastic process. Consider the typical case in SESIM, and in dynamic microsimulation, with a binary dependent variable. This variable then have a Bernoulli distribution, i.e.,

$$Y_i \sim \text{Bernoulli}(\pi_i), \text{ where } \Pr[Y_i = 1] = \pi_i \text{ and } \Pr[Y_i = 0] = 1 - \pi_i .$$

As an illustration, let  $Y_i = 1$  denote that individual  $i$  is unemployed and  $Y_i = 0$  that  $i$  is in work.  $\pi_i$  denotes the probability that the individual is unemployed in a year. This event is simulated by comparing  $\pi_i$  with a uniform random number. If  $u_i < \pi_i$  the event is realized and individual  $i$  becomes unemployed.

By allowing  $\pi_i$  to be a function of individual or household attributes these attributes also determine the probability of unemployment. This is typically accomplished by a logit or probit regression. The logit model is given as  $\pi_i = [1 + \exp(-X_i\beta)]^{-1}$ , where  $X_i$  is a vector of individual or household characteristics (or any other characteristic relevant for explaining unemployment, i.e. rate of regional unemployment) and  $\beta$  is a vector of parameters.

Due to the Monte Carlo simulation, the number of generated events in repeated simulation does not have to be the same<sup>§§</sup>. Let  $T$  denote the total number of individuals in a population of size  $N$  that experience the simulated event, that is  $T = \sum_{i=1}^N Y_i$ . If the individuals are simulated independently of each other the expected number of events is  $E(T) = \sum_{i=1}^N \pi_i$  and the variance  $Var(T) = \sum_{i=1}^N \pi_i(1 - \pi_i)$ . If  $N$  is large enough, and  $\pi_i$  not too close to zero or one,  $T$  is approximately normally distributed. Assume an event with a 10 percent probability (and for simplicity that all individuals face the same probability). For a population size of 10 000 individuals and a large number of repeated simulations, the number of individuals that experience the event is between 941 and 1 059 in 95% of the replications<sup>\*\*\*</sup>.

The Monte Carlo variation can be problematic in evaluating the results from an experiment. If, for instance a change in a tax rate is evaluated, then due to the Monte Carlo

---

<sup>§§</sup> Given that the seed used for generating random numbers is changed in each simulation.

<sup>\*\*\*</sup> If the number of events is approximately normally distributed a 95 % interval is defined as:  
 $10\,000 * 0,1 \pm 1,96 * \sqrt{0,1 * 0,9 * 10\,000}$

variation, it is difficult to isolate the pure tax effect from the stochastic Monte Carlo effect. One approach could be to repeat the simulations a number of times and use the average result, since this reduces the effect of the stochastic simulation.

An alternative approach is to use methods that reduce the Monte Carlo variance. In SESIM a method is used which is directly related to calibration. Calibration is a technique used in order to predict according to an a priori defined target. In the binary model this implies that the expected number of predicted events have to be adjusted in order to coincide with a given target. This is accomplished by adjusting the predicted probabilities. A simple, and quite common technique is a proportional adjustment,  $\pi^*_i = \alpha\pi_i$ , where  $\pi^*_i$  is the adjusted probability and  $\alpha$  is the factor of adjustment. A problem with this technique is that it does not restrict  $\pi^*_i$  to be in the  $[0,1]$ -interval. If instead  $\pi^*_i = \min(1, \alpha\pi_i)$  is used, this implies that individuals with  $\pi^*_i = 1$  with certainty will experience the event. This can produce unrealistic results, for instance all individuals with the same set of attributes might die. Alternatively, the adjustments can be made using a different scale. In SESIM an additive adjustment on the logit scale is used, this is equivalent to adjusting the intercept term in the estimated logit model. Thus, SESIM uses  $\text{logit}(\pi^*_i) = \alpha + X_i\beta$ , or  $\pi^*_i = [1 + \exp(-\alpha - X_i\beta)]^{-1}$ , where the logit function is  $\text{logit}(x) = \log[x/(1-x)]$ . This implies  $\pi^*_i \in [0,1]$ .

Regardless of approach, the adjustment factor  $\alpha$  must be calculated. In the first approach this is simple (given that the frequency of truncated probabilities are low) but in the second it is a bit more complicated. Even if the techniques discussed above ensure that the expected number of events corresponds to the desired, the Monte Carlo variance can still produce discrepancies. The method for variance reduction that is used in SESIM, eliminates these discrepancies by using the  $\alpha$  that, given the random values of  $u_i$ , generates the exact number of events  $n$ . The problem of calculating  $\alpha$  is solved by the fact that this is equivalent with sorting the variable  $v_i = \text{logit}(u_i) - \text{logit}(\pi_i)$  in an ascending order and letting individuals with the lowest rank obtain a positive event.

Even if calibration is a common method in dynamic microsimulation modelling it has been criticized, see Klevmarken (2002). If the discrepancy from the expected result is due to an incorrectly specified model, then the model should be respecified and not calibrated. However, calibration can also be motivated by a desire to replicate exactly well-known statistical bench marks or predictions to avoid a discussion about the meaning of random

deviations from a targeted bench mark. Alternatively calibration can be viewed as a method of implementing different scenarios in a simulation, for instance, the effect of two different assumptions about immigration flows. An alternative to calibrate the simulated values is to calibrate the parameter estimates of the model, but this approach might be more difficult to implement if there is a complex model structure.

In SESIM calibration is used for all these purposes, primarily in the demographic block to align to official forecasts of Statistics Sweden and to introduce alternative scenarios, but also as an estimation tool in the old age care module.

Monte Carlo variance is not the only source of random variability in a dynamic microsimulation model. Since the sample used for simulation is a random sample from a population, this introduces another source of randomness. Further, because the estimated parameters are random they also introduce a source of randomness<sup>†††</sup>. Thus, a proper inference should incorporate all possible sources of randomness, but due to the complexity of the model it is quite difficult to derive analytical results for simulated entities of interest. However, methods based on replicated simulation could be used, such as the *bootstrap*<sup>†††</sup>.

## 2.5 Assumptions about exogenous variables

SESIM is not linked to any general equilibrium model or macro model which could give feed back from market adjustments to changes simulated in SESIM. The macro economic scenario is instead fed into SESIM through a number of macro indicators such as the general growth in wage rates, the rate of inflation and the return on real and financial assets. These indicators are exogenous to SESIM. From the base year 1999 until 2005 we have used already observed values on these exogenous variables. After 2010 we have assumed constant rates and in the interim period 2005-2010 we allowed the rates to adjust from the last observed rates to the assumed constant rates. There is no technical reason to assume constant rates. The model is quite flexible and can take any assumed time path, but we have limited our simulations to alternative scenarios with constant future rates to get “clean” alternatives. There is for instance one alternative with high wage rate growth and one with low growth. Through the whole analysis we have used a main or base scenario towards which new submodels are evaluated and alternative scenarios are compared. This is a scenario which is close to the medium range predictions of the Ministry of Finance. More specifically, the base scenario

---

<sup>†††</sup> However, due to the large sample size that is used for estimation in SESIM this source of error is presumably rather small.

<sup>†††</sup> For a general reference to the bootstrapping technique see, Davison and Hinkley (1997).

uses the following assumptions after 2010:

Annual rate of change in inflation (CPI)	2.00%
Annual real general increase in wage rates	2.00%
Short-term interest rate	4.00%
Long-term interest rate	5.00%
Dividends	2.50%
Rate of change in prices on stocks and shares	3.75%

The demographic changes in the base scenario are aligned to the main projection of Statistics Sweden as explained above. The properties of alternative scenarios are explained as they are introduced below.

### **3. Models of health status, care, closeness to kin, retirement, income and wealth**

The basic structure of SESIM thus gave us the means to simulate the demographic changes of the Swedish population and the incomes, benefits and taxes, but to study ageing and its consequences we also needed measures of health status, and models that could simulate retirement, and the demand for health care and social care.

An alternative to formal care is informal care provided by family and kin. An elderly person in need of care can only get support from family and relatives if the elderly person has any family or relatives living close by. Swedish society differs from the south European societies in this respect. Young people move out from their parents' household at a rather young age and migration makes the share of parents living close to their children comparatively small. To understand the future potential for informal care of elderly we thus had to study not only closeness of kin but also domestic migration. A by-product is that we also will be able to say something about regional differences in the burden of providing for the elderly.

The all dominating share of health care and social care is financed and provided through the public sector in Sweden. An interesting policy issue then is if and how much taxes have to be increased to meet the expected increased demand for care by the elderly. In the future alternatives to increased taxes are increased user fees and that the elderly to an increasing degree will have to buy the services they need from a private market. In this perspective it becomes interesting to study the income and wealth distributions of the elderly. Of those who need care how many will have the means of paying high user fees

or buying services on a private market? There is of course also a more general interest in studying the relative incomes, the wealth and poverty of the elderly. For these reasons we have added to SESIM a new module which simulates capital incomes and a completely new module to simulate the evolution of the wealth distribution.

Although SESIM is based on the rich longitudinal data set LINDA, it lacks health data, data on the utilization of care and information about closeness to kin. Models covering these aspects of behaviour thus had to be developed using alternative data sources. This constrained our choices because we could only use models which were driven by variables that already were included in SESIM and simulated by the model. In most of these cases we were able to find either good register or survey data that were using the same or similar household definitions and variable definitions as in LINDA and SESIM, but in one case – social care – there were only micro data available from a small study of a parish of Stockholm. In this particular case we have aligned to supplementary information about aggregate care levels, see below.

When simulating the future we were not only interested in simulating cross-sectional distributions but also the life-courses of birth cohorts and the corresponding cohort distributions. For this reason we could not use cross-sectional models or even models estimated from repeated cross-sections, but needed dynamic models estimated from panel data. From an efficiency point of view one would also like to have a model that uses all information available in each simulation round. For instance, simulated tenure choice should depend on current tenure choice and possibly also on when the family moved into their current home. A family which recently has moved into a new house has a low probability to move again. From a theoretical point of view it would seem quite natural that decisions depend on the current situation of the decision maker. One could for instance think of habits, costs associated with a change and decisions about durables covering more than one period. Whenever possible we have thus designed and estimated dynamic models which condition current decisions on past decisions. There is however, one exception, to start the simulations we need “data” in the base year cross-section and the only way to get these “data” is to impute start values from a cross-sectional model. In most cases we have thus estimated both a cross-sectional model and a dynamic model.

In the remainder of this section we briefly review and motivate the structure of the new models added to SESIM, while the reader is referred to Klevmarken & Lindgren (2007) for estimation results and their interpretation.

### 3.1 A model for the health status of the population

Our measure of health status is a combination of self-assessed health, self-assessed mobility, longstanding illness and working capacity. Measures of these four dimensions are combined into four states of health: full health, not full health, some illness and severe illness.

Intuitively, there are strong inter-temporal correlations between (true) health states in subsequent time-periods, which should be accounted for in the specification of the evolution of health. Although the correlation relates to the unobserved health state, most econometric models actually introduce correlation by including lagged observable measures as determinants of current health. One could argue, however, that this is an incorrect way of introducing inter-temporal dependence, when the observable proxies are imperfect measures of true health states.

In our model of health and its determinants, we explicitly introduce inter-temporal correlation in the latent health variable as follows:

$$h_{it}^* = \rho h_{it-1}^* + \mathbf{x}_{it}' \beta + \varepsilon_{it} = V_{it} + \varepsilon_{it}, \quad (1)$$

where  $h_{it}^*$  denotes latent health of individual  $i$  in period  $t$ ,  $\rho$  is the autoregressive coefficient,  $\mathbf{x}_{it}$  is a vector of observable variables influencing health with associated weights  $\beta$ , and  $\varepsilon_{it}$  reflects stochastic health shocks in period  $t$ .

The interpretation of the model parameters here is slightly different compared to “standard” models. As we have an autoregressive component in the evolution of health, the model parameter  $\beta$  reflects marginal effect on health conditional on lagged health  $h_{it-1}^*$ , i.e.  $dE(h_{it}^* | h_{it-1}^*, \mathbf{x}_{it}) / d\mathbf{x}_{it} = \beta$ . However, the marginal effect of the unconditional expectation is inflated by  $1/(1-\rho)$  such that  $dE(h_{it}^* | \mathbf{x}_{it}) / d\mathbf{x}_{it} = \beta/(1-\rho)$ . The autoregressive coefficient  $\rho$  reflects the persistence of the idiosyncratic shocks captured by  $\varepsilon_{it}$ ; the higher  $\rho$ , the higher persistence in shocks. Note that in the case of finite life, there is no reason to assume that  $\rho < 1$ . Furthermore, note that the coefficients  $\beta$  and  $\rho$  are dependent on the frequency chosen for the model, i.e.  $\beta$  and  $\rho$  will depend on whether the model is “running” on annual or bi-annual frequency.



Our data are a two year panel from the HILDA data base for the years 1988/89 and 1996/97.<sup>§§§</sup> In addition to the lagged endogenous health variable explanatory variables are: age in the form of a spline, the ratio between the respondent's income and mean income, schooling, marital status, number of children below 7 in the household, if the respondent was born in Sweden, and gender.

The income variable might deserve a comment. It is well-known that there is a positive correlation between income and health status. The causal relations might go both ways. A healthy person is more likely to have a well paid job, and someone with a good income is able to invest more in health. Our health relation above can perhaps be seen as an investment relation, conditional on past health status those who have a relatively high income will have slower deterioration of their health than those who are poor. It was, however, estimated as if relative income is exogenous to health and should then best be interpreted as a predictive relation. In SESIM health status will indirectly influence income through the sickness benefits and disability pension. But why use relative income rather than the income level as explanatory variable? If we had used the level of income the implication is that the general increase in real income eventually would have brought everyone into the full health category, which does not seem very realistic. The assumption that the relative position in the income distribution determines investments in health is more realistic. The same principal problem comes up also in other relations when income is an explanatory variable and we have usually chosen a relative income measure.

The following is a brief outline of the estimation procedure.<sup>\*\*\*\*</sup> Two distinct problems need to be handled: (a) lagged latent variables imply unobserved variables in both the left- and right-hand side variables, and (b) available health data is only collected every 8 year, while the simulation model, nevertheless, presumes that the health module be able to predict health profiles on a 1-year frequency.

*First*, the estimation of a model with lagged latent variables such as (1) is complicated by the fact that one does not have access to the latent health  $h_{it}^*$  as this implies that we have unobserved variables in both the left- and right-hand side variables. The unobserved left-hand side variable is handled by assuming a link function between the

---

<sup>§§§</sup> HILDA (Health and Individuals. Longitudinal Data Analysis) combines data from the Swedish surveys of living conditions ULF of Statistics Sweden with register data for the surveyed individuals.

<sup>\*\*\*\*</sup> For a detailed description see Ch. 4, Appendix 2 in Klevmarken and Lindgren (2007)

unobserved latent variable  $h_{it}^*$  and some observed function of the variable denoted by  $h_{it}$ . Here we assume that  $h_{it}$  represents a categorization of  $h_{it}^*$  such that

$$h_{it} = j \text{ iff } \tau_{j-1} < h_{it}^* \leq \tau_j, \quad (2)$$

where  $h_{it}$  denotes the discrete health index with  $J$  possible outcomes and  $\tau_j$  denotes the upper threshold of category  $j$  with  $\tau_0 = -\infty$  and  $\tau_J = \infty$ . With the additional assumption that the error term  $\varepsilon_{it}$  is normally distributed, an ordered-probit model with lagged latent variables emerges.

In a standard setting, i.e. without lagged latent variables, the estimation would be completely straightforward using, e.g., maximum likelihood techniques. In this case, however, it is more complex as the lagged latent variable enters the right-hand side. This means that we need to integrate out the lagged latent variable in order to calculate the sample likelihood contribution for an individual given a candidate vector  $(\beta, \rho, \tau)$ . Using the normality assumption of the error term, this implies that we need to integrate a multidimensional normal distribution for which there are no closed forms. Furthermore, the recursive structure implied by the lagged latent variable also means that the multidimensional integral will not collapse to a product of univariate integrals, even if the error terms are serially uncorrelated. We, therefore, need to retreat to either numerical or Monte Carlo techniques to evaluate the sample likelihood contribution. Unless the dimensionality is very small ( $<4$ ), the standard numerical integration techniques (e.g. Gaussian quadrature) are very time consuming and, hence, impractical. Monte Carlo techniques are feasible, though, if we could simulate the probability to observe a given sequence of health states conditional on a candidate vector. The so called GHK simulator gives us exactly this possibility.

For a given candidate vector of model parameters  $(\beta, \rho, \tau, \sigma_0)$  we can use the GHK simulator to simulate a sequence of latent health observations  $(h_{i1}^{r*}, \dots, h_{iT}^{r*})$  that translates to a sequence of health indices  $(h_{i1}^r, \dots, h_{iT}^r)$  via the link function. In principle, we could make draws from the assumed normal error distribution, calculate the sequence of latent health states and the corresponding sequence of health indicators and check if the observed sequence is identical to the simulated one. Repeating this procedure a large number of times would enable us to calculate the probability of the observed sequence as the average of replications where the simulated sequence conforms to the observed. However, if the number of time

periods is larger than 3, this would be very time consuming. The GHK simulator offers a more efficient procedure by focusing on the probabilities rather than the outcomes. In the GHK simulator we draw only sequences of error terms that guarantee conformity between simulated and observed sequences of health indices. The GHK simulator hence produces the probability  $p^r$  to observe a sequence such as  $(h_{i1}^*, \dots, h_{iT}^*)$ . Averaging over  $R$  such draws produces an estimate of the true contribution of the individual and this estimate will serve as the sample likelihood contribution

$$\Pr((h_{i1}, \dots, h_{iT}) | \beta, \rho, \tau, \sigma_0, \mathbf{X}_i, h_{i0}^*) = \frac{1}{R} \sum_{r=1}^R p_i^r, \quad (3)$$

where  $\mathbf{X}_i$  denotes the matrix of independent variables for individual  $i$  over periods  $t = 1, \dots, T$ . As the process including lagged latent variables requires an initial value  $h_{i0}^*$ , we cannot use this dynamic specification for the first period. Here, we estimate a standard ordered-probit model, where the slope coefficients  $\beta_0$  and the error variance  $\sigma_0^2$  are allowed to vary freely. The initial-period order-probit model is linked to the dynamic model by the threshold parameters  $(\tau_0, \dots, \tau_j)$ , which are assumed time invariant. Thus, the initial-period model and the dynamic model are estimated simultaneously.<sup>††††</sup>

*Second*, data is collected in 8-years intervals, whereas the required model should run on an annual frequency. In principle, one could specify the model, using the 8-year intervals,  $h_{it}^* = \rho h_{it-8}^* + \mathbf{x}_{it}' \beta + \varepsilon_{it}$ , and then translate the estimated model parameters  $(\beta, \rho, \sigma_\varepsilon)$  to an annual frequency. Another approach is to keep the annual frequency in the model specification and explicitly account for the fact that the latent variable is unobserved in intermediate periods. The only modification required in the estimation strategy is that the latent variable is allowed to vary freely in the periods where it is not observed. In order to estimate the model on annual frequencies we also need to impute values of the independent variables in the intermediate years. Here we use linear interpolation of the  $\mathbf{x}$ -variables for which additional information is lacking.

---

<sup>††††</sup> There are also some technical issues related to the assumption about the initial value  $h_{i0}$  that are discussed in Klevmarken & Lindgren(2007), Chapter 4, Appendix 2.

### 3.2 Sickness absence from work

We model sickness absence from work primarily because long spells of sickness is a good predictor of disability pension, but also because public expenditures for sickness benefits, which in recent years have taken a large share of the public budget, compete with expenditures for the elderly.

We need a model which captures take up of sickness benefits and the amount obtained. We will also need one imputation model for the first year and then a dynamic model which conditions on past sickness history. The benefit system is designed such that there is a waiting period of 14 days before benefits are paid out. We will thus only capture sickness spells that last for at least 14 days, while temporary short problems are not part of our analysis.

The model used is a negative binomial model with correction for sample selection. The dependent variable is the number of compensated days in a year. Explanatory variables are: if the individual had any compensated days last year, health status, age, if children below 7, marital status, earnings forgone net of benefits relative to the median of the same variable, gender, interaction between gender and earnings foregone, and if the family had a newborn baby.

### 3.3 Early retirement

In SESIM everyone is assumed to be retired before a certain age. The default is 65 years. Until recently this was a rather good description of reality. Most union contracts had an upper mandatory retirement age of 65, and very few actually worked after the age of 65. A few years ago Parliament, however, passed a law which stipulated that no contract could have a mandatory retirement age below 67, and in recent years one can observe that the number of people above 65 who work has started to increase. What we needed was thus a model which could simulate when, before the age of 67, people decide to retire.

There is more than one way out of the labour market before 65. One is through disability pension which requires a medical condition, another is early pension through the negotiated occupational pensions, a third is through the social security pensions, and a fourth is by using private means. The occupational pensions open for the possibility that employers can pay more in pension than the contract stipulates. In the 1990s and beginning of the current century employers found it profitable to give relatively large groups of white collar workers and employees in the public sector such golden handshakes. In modelling retirement behaviour in Sweden one has to take this into account. There were almost no data to support a

study of retirement behaviour between 65 and 67, so any simulation of the consequences of lifting the upper mandatory retirement age would have to be based on an extension of behaviour in the age bracket 60-64.

Although there is not a very strict boarder line between disability pension and old-age pension in practice, we have chosen to model take up of disability pension as driven by health problems, while early take up of old-age pension is at least partly driven by economic incentives. The model for the probability to get disability pension is just a simple logit model. Explanatory variables are health status, if sick more than 15 days in the previous year, schooling, labour market sector, income quartile, age, if born in Sweden and gender.

For early retirement via the old-age pension systems, the situation is more complicated. The individual's choice to retire is assumed to be given by the discrete choice model

$$\begin{aligned} y_{it} &= 1 \text{ if } y_{it}^* = \mathbf{x}_{it}'\beta + \varepsilon_{it} > 0 \\ y_{it} &= 0 \text{ otherwise} \end{aligned} \quad (4)$$

where  $y_{it} = 1$  indicates that the individual exit into retirement in period  $t$ ,  $\mathbf{x}_{it}$  is a vector of individual characteristics including rescaled net present values and accruals with weights  $\beta$  and  $\varepsilon_{it}$  reflects unobserved variables influencing the retirement decision.

The retirement via old-age pension model includes financial incentives, which are derived from the future stream of pension benefits. Old-age benefits are generally determined by the public old-age pension and the collectively agreed occupational-pension systems. However, we also indicated the incidence of individual contracts between the employer and the employee close to normal retirement age, denoted as early-retirement pension offers. This type of pension is not available to all individuals at all times, but it is the outcome of an unobserved negotiation process between the employee and the employer. Ideally, we would like to model the process by constructing a structural model including the main determinants of early-retirement pension offers and the level of early-retirement pension benefits. However, that is not possible with the data at hand and we are forced to retreat to a reduced form, capturing the main individual variables in the process. Here we assume that the probability of receiving an early retirement pension is determined as a discrete choice model such that

$$v_{it} = 1 \text{ if } v_{it}^* = \mathbf{w}'_{it}\delta + \xi_{it} > 0 \quad (5)$$

where  $v_{it} = 1$  indicates that the individual has access to an early retirement pension,  $\mathbf{w}_{it}$  denotes a vector of individual characteristics with weights  $\delta$ , and  $\xi_{it}$  represents unobserved variables reflecting the probability to receive an offer.<sup>\*\*\*\*</sup> There are some econometric complications as the early retirement pension is only indirectly observable for those who have retired.

First, as the observed frequencies of transitions from retirement back to work are negligible in practice, we consider retirement via old-age pension as an absorbing state. Second, access to early-retirement pension is not observable for non-retirees. This complicates the situation as the dependent variable in eq. (5) becomes only partially observable, and some of the independent variables in eq. (4) become error prone and correlated to the error term. The proposed model (Eklöf & Hallberg, 2006) is similar to the bivariate probit models with partial observability proposed by Abowd and Farber (1982) and Poirier (1982), but in this case there is an additional problem, since the independent variables are also only partially observed.

In eq. (4), the problem is that  $\mathbf{x}_{it}$  is observed with error for non-retirees, i.e. when  $\varepsilon_{it} < -\mathbf{x}'_{it}\beta$ . This implies that we have an endogenous variable problem since the measurement error in  $\mathbf{x}_{it}$  is related to the error term  $\varepsilon_{it}$ . Furthermore, since the accessibility of early retirement program is observed for retirees only, this implies that the data is censored with respect to  $v_{it}$  and we have a sample selection problem in eq. (5).

Assume that the utility parameters are not affected by the accessibility of an early retirement pension. Let  $\mathbf{x}_{it} = v_{it}\mathbf{x}_{it}^{ERP} + (1-v_{it})\mathbf{x}_{it}^{STD}$  where the superscript *ERP* refers to variables relevant if the individual has access to an early retirement pension and *STD* refers to the standard contracts. Then we can construct a system of simultaneous equations as

$$\begin{cases} y_{it}^* = (v_{it}\mathbf{x}_{it}^{ERP} + (1-v_{it})\mathbf{x}_{it}^{STD})' \beta + \varepsilon_{it} \\ v_{it}^* = \mathbf{w}'_{it}\delta + \xi_{it} \end{cases} \quad (6)$$

where the simultaneity stems from that  $\mathbf{x}_{it}$  is a function of  $v_{it}$  and  $v_{it}$  is only observed if

---

<sup>\*\*\*\*</sup> As the accrual value of social security wealth includes also its future value there is an issue on the individual's expectation of future access to early retirement programs. In this analysis we assume that an individual do not anticipate to receive an offer next year.

$y_{it} = 1$ , i.e. if  $(v_{it}\mathbf{x}_{it}^{ERP} + (1-v_{it})\mathbf{x}_{it}^{STD})'\boldsymbol{\beta} + \varepsilon_{it} > 0$ . The probability to observe the event  $(y_{it}, v_{it})$  is thus

$$\Pr(y_{it}, v_{it}) = \int_{\underline{\xi}_{it}}^{\bar{\xi}_{it}} \int_{\underline{\varepsilon}_{it}}^{\bar{\varepsilon}_{it}} f(\varepsilon, \xi) d\varepsilon d\xi \quad (7)$$

where  $f(\varepsilon, \xi)$  denotes the joint density of  $(\varepsilon_{it}, \xi_{it})$  and with integration limits defined as

$$\begin{aligned} \bar{\varepsilon}_{it} &= \begin{cases} (v_{it}\mathbf{x}_{it}^{ERP} + (1-v_{it})\mathbf{x}_{it}^{STD})'\boldsymbol{\beta} & \text{if } y_{it} = 1 \\ \infty & \text{otherwise.} \end{cases} \\ \underline{\varepsilon}_{it} &= \begin{cases} -\infty & \text{if } y_{it} = 1 \\ (v_{it}\mathbf{x}_{it}^{ERP} + (1-v_{it})\mathbf{x}_{it}^{STD})'\boldsymbol{\beta} & \text{otherwise.} \end{cases} \end{aligned} \quad (8)$$

and

$$\begin{aligned} \bar{\xi}_{it} &= \begin{cases} \mathbf{w}'_{it}\boldsymbol{\delta} & \text{if } v_{it} = 1 \\ \infty & \text{otherwise.} \end{cases} \\ \underline{\xi}_{it} &= \begin{cases} -\infty & \text{if } v_{it} = 1 \\ \mathbf{w}'_{it}\boldsymbol{\delta} & \text{otherwise.} \end{cases} \end{aligned} \quad (9)$$

Furthermore, as the propensity to retire is potentially related to unobserved individual characteristics, we control for individual specific time invariant effects. Assuming that the unobserved time invariant individual effects are uncorrelated with the idiosyncratic random error and with the observable independent variables allows us to estimate a random-effect model. The probability to observe a sequence of outcomes  $(\mathbf{y}_i, \mathbf{v}_i) = (y_{i1}, \dots, y_{iT_i}, v_{i1}, \dots, v_{iT_i})$  is

$$\Pr(\mathbf{y}_i, \mathbf{v}_i) = \int_{\underline{\xi}_{i1}}^{\bar{\xi}_{i1}} \dots \int_{\underline{\xi}_{iT_i}}^{\bar{\xi}_{iT_i}} \int_{\underline{\varepsilon}_{i1}}^{\bar{\varepsilon}_{i1}} \dots \int_{\underline{\varepsilon}_{iT_i}}^{\bar{\varepsilon}_{iT_i}} f(\varepsilon_{i1}, \dots, \varepsilon_{iT_i}, \xi_{i1}, \dots, \xi_{iT_i}) d\varepsilon_{iT_i} \dots d\xi_{i1} \quad (10)$$

To simplify, we assume that the error term in the retirement decision are independent of the error term in the early retirement program equation. Let  $\varepsilon_{it} = u_i + e_{it}$  where  $u_i \sim iidN(0, \sigma_u^2)$  and  $e_{it} \sim iidN(0, 1)$ . This implies that  $\varepsilon_{it}$  is uncorrelated with  $\varepsilon_{is}$  for  $t \neq s$  conditional on  $u_i$ . As  $\mathbf{x}_{it}$  is observed only when  $y_{it} = 1$  we join the outcomes  $(y_{it}, v_{it}) = (0, 0)$  and  $(y_{it}, v_{it}) = (0, 1)$ . The probability conditional on  $u_i$  can thus be written as

$$\Pr(\mathbf{y}_i, \mathbf{v}_i | u_i) = \prod_{t=1}^{T_i} \left[ \Phi \left( \mathbf{x}_{it}^{ERP'} \beta + u_i, \mathbf{w}'_{it} \delta, \rho \right)^{y_{it} v_{it}} \left\{ \Phi \left( \mathbf{x}_{it}^{STD'} \beta + u_i, \infty, \rho \right) - \Phi \left( \mathbf{x}_{it}^{STD'} \beta + u_i, \mathbf{w}'_{it} \delta, \rho \right) \right\}^{y_{it}(1-v_{it})} \right. \\ \left. \times \left[ 1 - \Phi \left( \mathbf{x}_{it}^{ERP'} \beta + u_i, \mathbf{w}'_{it} \delta, \rho \right) - \left[ \Phi \left( \mathbf{x}_{it}^{STD'} \beta + u_i, \infty, \rho \right) - \Phi \left( \mathbf{x}_{it}^{STD'} \beta + u_i, \mathbf{w}'_{it} \delta, \rho \right) \right]^{1-y_{it}} \right] \right] \quad (11)$$

Integrating over  $u_i$  gives the unconditional probabilities as

$$\Pr(\mathbf{y}_i, \mathbf{v}_i) = \int_{-\infty}^{\infty} \Pr(\mathbf{y}_i, \mathbf{v}_i | u) f_u(u) du \quad (12)$$

which can be rewritten on a form suitable for numerical integration routines (e.g. Gauss-Hermite quadrature) as  $\Pr(\mathbf{y}_i, \mathbf{v}_i) = \int_{-\infty}^{\infty} \exp(-r^2) g(r) dr$  where  $g(r) = \frac{1}{\sqrt{\pi}} \Pr(\mathbf{y}_i, \mathbf{v}_i | \sqrt{2\sigma^2} r)$ .

The explanatory variables in the equation that determines access to early retirement are: Wage rate, schooling, if born abroad, labour market sector, age within the bracket 60-64, and dummy variables for year. The time dummies are inconvenient in the microsimulation application because in principle one would have to forecast exogenously the time effects into the future. In practice we have chosen to use the 1999 estimate for all future years.

The decision to retire given access to early retirement is determined by the same variables except for labour market sector, and in addition by the net present value of future pension benefits (NPV), the net present value accrual - which measures the effect on NPV when postponing retirement one year - marital status, and if the spouse is not working. Including the economic incentive variables implies that SESIM in every simulation round will have to forecast future earnings for every individual and access the income tax module to obtain earnings net of income tax. To speed up this process a somewhat simplified tax code was used.

### 3.4 Geographical mobility and tenure choice

In modelling geographical mobility Sweden was divided into nine regions: The three major metropolitan areas, three rural areas - one in the south, one in the north and one in middle Sweden - , and three more urban areas. Our modelling strategy was to have a model for the probability to leave the current region, one model which determines the region of destination, another model for intra-regional migration, and finally a model for tenure choice. These models are simulated sequentially.

The probability to leave a region is modelled by a logistic regression for each combination of current region and household type. Explanatory variables are: age of



household head, country of birth, schooling, and recent family changes such as a separation, a marriage, the death of a partner and the birth a child. Additional variables were if unemployed, previous migration experience, quintile of disposable income, current tenure and gender if single.

For each region of origin a multinomial model is used to model the probability to move to a specific region given that the household leaves its current region. Explanatory variables are relative income, schooling, if born in Sweden, age in the form of a spline, and the unemployment rate and relative price of one family houses in the region of destination.

The model used to simulate intra-regional moves is very similar to the model for the probability to leave a region. It is a logistic regression for each household type with the same explanatory variables.

The tenure choice model, finally, only distinguishes between owning and renting. We use one set of logit models for households that stay within their current region - one model for each household type – and one set of logit models for households that move to another region. Explanatory variables are age of head, recent family changes, quintile of disposable income, previous tenure, if born in Sweden, schooling, gender and region of origin.

All these models were estimated from the longitudinal register data GEOSWEDE covering almost the entire Swedish population.

### 3.5. Modelling the income distribution

Income is decomposed into earnings, pensions and other benefits, and capital income. Earnings is simulated by a stochastic panel data earnings equation, pensions and benefits are computed following the rules of eligibility and compensation for each benefit, while capital income is simulated using models for the incomes of interest and dividends, capital gains from own home, capital gains from financial assets, and interest paid.

#### *Earnings*

The estimated earnings model in SESIM is a random parameter model estimated on panel data, i.e. the same individual is observed repeatedly in the data. The model is given as:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \gamma_i + \varepsilon_{it}, \text{ where } \gamma_i \sim N(0, \tau^2) \text{ and } \varepsilon_{it} \sim N(0, \sigma^2). \quad (13)$$

The error components  $\gamma_i$  and  $\varepsilon_{it}$  are assumed to be independent. The random intercept  $\gamma_i$  is designed to represent unobserved heterogeneity

The earnings equation is estimated on a four year panel and includes in the X-vector variables such as; experience, highest level of education, marital status and nationality. Separate models are estimated for each occupational sector and by gender. The dependent variable is the logarithm of annual earnings.

The simulations of the earnings equation is based on the individual attributes in  $X_{it}$ , the estimated parameters  $\hat{\beta}$  and the random numbers  $\tilde{\gamma}_i$  and  $\tilde{\varepsilon}_{it}$ . The random numbers are drawn from two independent normal distributions with variance  $\hat{\tau}^2$  and  $\hat{\sigma}^2$  respectively. The simulated earnings are calculated as  $\tilde{Y}_{it} = X_{it}\hat{\beta} + \tilde{\gamma}_i + \tilde{\varepsilon}_{it}$ . Since  $\tilde{\gamma}_i$  is specific for each individual and constant over time, only one draw at the start of the simulation is need, but draws for  $\tilde{\varepsilon}_{it}$  have to be repeated for each year (and new individual). The individual variation,  $\hat{\tau}^2$ , is larger than the random component,  $\hat{\sigma}^2$ , in all sectors. There is a large cross-sector difference in these estimates. The self employed have the largest individual variation, indicating the large heterogeneity in this sector. Self employed includes everyone from low skilled low paid job to highly paid consultants. The lowest individual variation is for blue collar males and for females in the local governmental sector.

### *Pension income and other benefits*

Social security pensions are computed using the rules that apply each year jointly with simulated income histories and eligibility status. Each worker is simulated to belong to one of the four major contract areas and will then receive occupational pensions accordingly following the rules of each contract.

Regarding the income generated from the stock of private pension annuities many different options are open for an individual such as a limited time annuity, a lifetime annuity etc. In all the simulations reported here a five year period after retirement has been used. This kind of savings is a relatively new phenomenon in Sweden and the generated stock is still rather small. However, the average amount for the 4.8 per cent who had an income from this source in year 2000 was as much as 32 196 SEK, and its relative importance is expected to increase.

Benefits other than pensions such as housing allowances, child allowances and social relief are computed using the rules of the benefit systems.

*Income from capital.*

*Interest and dividends.*

It is natural to model interest and dividend income as a rate of return on financial assets. The distribution of the sum of interest and dividends to the end of year stock of financial assets is however non-standard. It is censored at zero and positively skewed with a high kurtosis. We tried to estimate a Tobit model using schooling, age, a dummy variable for marital status and the mean value of assets as explanatory variables.

Although the estimates of the model parameter made sense and gave interesting interpretations, the model explained relatively little of the variability in returns. Most of the simulation variability came from the residual that had a rather large variance. For this reason we found that we could as well simulate the rate of return by drawing randomly from the empirical distribution. In the simulations we have interpolated linearly between the percentiles and also imposed a maximum return of 30 percent. This distribution applies to the year 2000. If the average return in the financial markets changes cyclically or trend wise these estimates have to change accordingly. When simulating we thus multiply the rate of return drawn from this distribution by the ratio of a market rate for the current year and the year 2000.

*Capital gains from real property*

In SESIM we simulate both purchases and sales of owner occupied homes and record the price at which a transaction takes place (see below). The difference between the sales price and the purchase price of a given property is the capital gain, 80 per cent of which is added to the taxable income of capital<sup>§§§§</sup>. If a household sells their home and then buys another one within one year, and the gain exceeds a threshold, adding 80 per cent of the capital gain to the taxable income of capital is partly or completely deferred. If the new house is more expensive than the old the whole amount is deferred, if it is less expensive a share of the amount is deferred equal to the ratio of the market value of the new house to that of the old. This

---

<sup>§§§§</sup> This applies symmetrically to losses.

procedure is repeated for every new sale and purchase until the owner moves to a rented home or dies. In this way it is in principle possible to defer the tax on the capital gains of own homes until the owners die. Then the survivors have to pay the tax out of the deceased's estate. For each home owner who has previously sold a house or an apartment and used the opportunity to defer tax there is a cumulated deferred capital gain registered with the tax authorities. SESIM also keeps track of these amounts and applies the rules of deferred taxation.\*\*\*\*

### *Capital gains from financial assets*

Capital gains taxation also applies to gains from other properties and from financial assets. Register data only include information about total capital gains (losses) and it is not possible to distinguish gains by source. This makes it difficult to estimate a model for the gains accruing from other assets than own home. Register data, however, include information about the tax assessed value and an estimated market value of own homes. We used these data to eliminate from the sample all households that had reduced their investments in own homes in order to get a cleaner sample of households with capital gains from other sources. A potential source of error is the fact that we cannot eliminate those who sold their property and bought a replacement at the same or a higher market value. But this group is likely to defer the taxation of their capital gains from their old home, and the corresponding amount will not be included in the reported capital gain figure.

The distribution of net capital gains (the sum of gains and losses) is rather strange. The center of the distribution is essentially zero while there is a left tail with large negative values and a right tail with even larger (positive) values. The distribution thus has a high kurtosis and is positively skewed. The median is 0 and the mean 28467 SEK. It became difficult to find a model that could reproduce this distribution. Instead we have opted in favour of separate models for capital gains and losses. We first estimated a probit for the probability of having a nonzero gain, then a probit for having a loss conditional on no gain and finally a model explaining the size of the gain (loss) if the household had one.††††

We were not very successful in formulating and estimating a model for the size of capital losses. Instead we have chosen to draw randomly from the empirical distribution of

---

\*\*\*\* There is one simplification. If sales and purchases are not done in the same year then the household is assumed to move to a rented home and no deferral is allowed. In SESIM we thus do not strictly apply the one year rule.

†††† This model was estimated as a robust regression model.

the variable as such. Random draws are obtained from a distribution with linear interpolation between percentiles. Capital gains and losses are indexed by the CPI to capture changes in the general price level.

### *Interest deductions*

Almost 25 per cent of all households had deductions for interest paid on mortgages and loans. The median deduction was just above 14000 SEK, while the mean was close to 26000 SEK. The distribution is positively skewed and the largest deductions amount to a few millions.

It might be natural to relate the interest deducted to the total debt and simulate the corresponding rate. The problem with this approach is that about 25 per cent of all households have deducted interest although they did not have any debt at the end of the same year and about 10 per cent of all households have deductions without having any debt neither in the beginning of the year nor in the end. It is of course possible to pay interest on a debt that was repaid before the end of the year, and it is also possible to take a loan after the beginning of a year and repay it before the end and still have interest to pay. Similarly we find households in the data set with large deductions but rather small debts.

In spite of these difficulties the advantages of modelling a rate rather than an amount are so great that we preferred this approach even if we will not be able to simulate a positive amount for households with no debt at the end of a year. The interest rate on debts was defined as the ratio of the interest paid by the household to the sum of all debts at the end of the year if this sum exceeded 1000 SEK. This truncation floor was used to avoid excessively high rates. Households with a debt less than 1000 even if they had paid interest were dropped from the sample. With this definition 83 per cent of all households had a positive rate. The median was 5.6 per cent and the mean 7.8 per cent. The 95<sup>th</sup> percentile was 14.1 per cent and there were a few observations exceeding 1.

The sample was split into two groups, one that did not pay any interest in the previous year and one that did pay. For each group a two phase model was estimated: a random effects probit model for the probability to have paid interest and a model for the amount.

### 3.6 The distribution of wealth

In SESIM we have chosen to model financial wealth, savings in private pension annuities, the market value of owner occupied homes, wealth invested in other real estate, and debts.

Financial wealth includes a number of different assets such as bank accounts, bonds, mutual

funds, stocks and shares, and life insurances, but not private pension annuities. The latter asset is modelled separately for two reasons, first because this kind of savings is designated life-cycle savings with the purpose of complementing public and occupational pensions, and second because investments in this asset are deductible from income at income taxation. We thus need this deduction to compute the income tax. A further break down of financial wealth by risk level would have been of interest, but it had required a completely different set of models and we also would have to model -within or outside SESIM - the returns to each of these assets, a major task well outside our project.

Investments in real estate have been divided into two components, owner occupied homes and other real estate, because the major asset of many Swedish households is just their home. This component includes both one and two family houses as well as condominiums. In 2002 about 43 per cent of Swedish households owned a house and 14 per cent a condominium.<sup>\*\*\*\*</sup> Condominiums are most common in major cities. There are no direct data on market values of owner occupied houses and condominiums in the registers of Statistics Sweden, but they have been estimated using the product of the tax assessed value of each property and so called purchase coefficients. These coefficients are the annual mean ratios of the price to the tax assessed value of each sold unit in a relatively small area. Comparisons with self-reported survey data show that these estimates give good mean levels. They might though underestimate the dispersion of house values a little. These estimated market values form the dependent variable in our model for market values, see below.

Other real estate is a mixture of different assets. One large component is secondary homes. About 13 per cent of Swedish households had a secondary home in 2002, some of which represented a major investment. Included in the aggregate Other real estate are also commercial apartment complexes, farm land and forests and other property owned by private households. There are rather few owners of these properties, but they represent large values for the owners. The corresponding distribution is thus strongly positively skewed.

Debts are modelled as a single category and it includes all kinds of debts such as mortgages on homes and other real estate, regular bank loans and consumer credit. It might have been of interest to separate these different types of debts, but such detailed information is not available in the register data of Statistics Sweden, and it is not obvious that a separation is analytically meaningful. A household can increase the mortgages on their house not only to invest more in the house but also, for instance, to buy a car, a boat or to go on a holiday trip.

---

<sup>\*\*\*\*</sup> These estimates are based on LINDA using the family concept of this source.

Thus, the legal form a loan takes does not necessarily say much about the uses of the borrowed money. Register data on mortgages and loans originate from banks and other credit institutes, which have to supply this information to the tax authorities for taxation purposes. Tax payers also have an interest to declare their loans because interest paid is deductible from incomes. Register data on mortgages and loans are thus considered being of good quality.

Figure 2 gives a view of the model structure and simulation path of the wealth model. The simulation starts with financial wealth. Different models are used depending on if the household had financial wealth or not in the previous year. Then follows the simulation of Other real wealth, again the choice of model depends on the household having Other real wealth or not in the previous year. In the third major step private pension wealth is simulated and in the fourth the value of any owner occupied home. Ownership might change if the household moves and decides to buy a (new) house after the move. SESIM thus simulates geographical mobility and tenure choice before the market value of a house is determined. Finally the debt of each household is updated and the cost of housing is simulated. The latter entity is of interest in its own right, but also needed for the computations of housing benefits.

### *Financial wealth*

In modeling financial wealth (other than tax-deferred pension savings) we have chosen to work with separate models for households which previously respectively had and did not have these kinds of assets. In the first case we use a dynamic panel model and in the second the combination of a logit model which simulates the transition from not having to having financial assets, and a regression model which simulates the amount. All three have been estimated using the Linda panel data. One might note that the period for which data are available, 1999-2003, is a period of exceptional changes in the stock market, which might have resulted in estimates that are not typical for other periods. Furthermore, our short panel does not allow any elaborated dynamic specification, nor is it possible to identify and estimate cohort effects separately from period effects.

The model for those who did not have any financial wealth previously is a so called two-part model. That is, the model for the probability to acquire financial wealth was estimated independently of the model that determines the amount of financial assets acquired. The reason for using the two-part model compared to, for instance, a generalized tobit model or a Heckit type of approach, is that we focus on obtaining good robust predictions rather than on explaining selectivity. Manning et. et al. (1987) showed that the two-part model performs at least as well as the tobit type 2 model. Flood & Gråsjö (2001) demonstrated the sensitivity

of the generalized tobit model to errors in the specification of the selection equation, which produce bias in all the estimated parameters.

For households that did not have any financial wealth the probability to acquire some increases with increasing age. This could be the result of increased financial saving in middle age when mortgages have been reduced and children have left home, and of decreased investments in own home and other real estate after retirement. The relative position in the income distribution also determines the probability to acquire financial assets. The higher incomes the higher probability.<sup>§§§§§</sup>

Although not uniformly and with the exception of the very old the amount acquired increases with increasing age. The differences due to age are though relatively small. Those who are in the top right tail of the income distribution acquire more financial wealth than most people, about 25 percent more than those who have incomes below the 90th percentile.

In the dynamic random effects model the estimated effect of the lagged stock of financial assets shows that there is a strong persistence in the investments of households. It is a little smaller among young people than among old, and among rich people compared to poor. The relative position in the income distribution has the expected effect, high income households invest more. Price changes in the stock market have a strong influence on the stock of assets held by households. Finally we might note that the variance of the purely random component is much larger than that of the unexplained household specific effects.

#### *Tax-deferred pension savings*

Because there are no register data on tax-deferred pension savings we first need estimates of these stocks as of 1999, then a model which forwards the stocks after 1999.

Because register data include information about how much each individual has paid into pension policies and claimed deduction from income each year, the simple idea is to construct accumulated savings by using Linda panels. Individual savings are summed up over years and the resulting stock is increased each year by applying the average return given by life insurance companies. In order to reduce the starting value problem, we started as early as 1980, at which time private tax-deferred pension savings were rather unusual.

---

<sup>§§§§§</sup> We have here used a relative measure of income, i.e. the percentile of the income distribution, rather than income as such in order to avoid that a general increase in income level will drive the probability towards 1 as income increases over long periods. This is a general problem in simulation models such as SESIM.



Given the accumulated stock of pension savings in 1999 we assume that those who claimed deductions in 1999 continue to do so in the following years until the age of 64 by the same amount increased by the CPI. For those who did not save anything in 1999 and were in the age range 18-64 we applied a two-part model estimated from LINDA data. The simulated amount saved in 2000 was then also applied to later years but increased by the CPI. For each year the probability of pension saving is simulated. If an individual is predicted to be a pension saver, the amount is also predicted. Again it is assumed that the individual continues to save this amount (adjusted by CPI) until he retires. Thus, for those individuals who do not save the probability of saving is simulated every year. Note, that the yearly amount saved is indexed by the CPI, while the stock of pension savings is increased by an interest rate for long term bonds.

### *Real wealth*

Household real wealth is decomposed into two components; Own home and other real wealth. Since the probability of owning a home is modeled in the regional mobility module (see above) only the model that determines the market value of a home and the model that simulates other real wealth is discussed here.

The market value of a home is primarily determined by its location, size and qualities. Changes in values depend on factors that influence demand and supply, such as changes in income and wealth and in the cost of borrowing. We do not try to formulate and estimate a model of the market value in this sense. We need a model which predicts the market value of the home of a particular family. In addition to some of the variables mentioned we will thus also use properties of the family as predictors.

The estimation of the market value model for own homes is based on both Linda and HEK data. Since information about house area is missing in Linda, HEK data from 1999 have been used to estimate a model of floor area (area in  $m^2/100$ ) in order to impute this variable. The results are not reported here but the most important findings are that the age of the owner matters and that the size reaches a maximum in the age bracket 45-49. Marital status and number of children have strong effects. Income also has a strong effect. Those who belong to the first quartile of the income distribution have a house area  $46 m^2$  smaller than those in the highest income quartile. There is a large negative Stockholm effect, dwellings are  $12m^2$  smaller compared to areas outside Stockholm, Gothenburg and Malmö for otherwise comparable houses and families.

Using the imputed value of floor area, jointly with the other covariates<sup>\*\*\*\*\*</sup>, a model was estimated on data from 1999. The sample was limited to house owners and owners of condominiums having a property value between 50 tkr and 10 mkr. There is a clear age effect and again an inverted-U relation, the maximum value is about 1.1 mkr for households in their mid forties to mid fifties. There are strong and significant effects of marital status, region, house area, financial wealth and nationality. The market value of a house in the Stockholm region is 141% higher than a house in the reference region (rural region). If a household has financial wealth it also tends to have a bigger house.

The mixture of large and rather small properties in the aggregate Other real wealth makes it difficult to estimate good models. The distribution is heavily skewed. In the simulations we distinguish between households that have this asset and those who do not have it. In the first case we use a simple random walk. In the second case a logit model was estimated for the probability to buy property in the next year and a robust regression to simulate the amount. The estimates show that the probability to invest in other properties reaches a peak at middle age. Couples have a higher probability to invest than singles and there is a rather strong income effect. High income people have a much higher probability to invest than low income people. Income also determines the amount invested.

### *Debts*

In SESIM we distinguish between study debts and other debts. Other debts include all debts but the study loans college and university students are offered by the government. It is assumed that the take up rate is 100 per cent for these loans and that students borrow as much as they are allowed to. Study debts are increased by an interest rate determined by the government. Repayments of principal and accumulated interest are proportional to the taxable income of the borrower according to certain rules, which are followed in SESIM. The remainder of this section deals with other debts than study loans.

Assuming that most households do not decrease or increase their debts much from one year to another, we need models that simulate debts at the end of next year conditional on current debts. We also need to account for any major investment a household might do in the coming year that might influence their decisions to take up new loans, such as buying a new home.

---

\*\*\*\*\* Age, marital status, number of children, percentile of taxable income, percentile of financial wealth and region.

The analysis of the distribution of debts and the dynamics of debts was limited to stable households in the period 1999-2002. New households that have been formed through marriages, separations and deaths will have their assets updated by adding the wealth of new household members, by following standard rules for bequests and in the case of separation by dividing the assets between the newly formed households using common rules. Within stable households 77 per cent have debts. The median debt in this period was 156 000 SEK while the mean debt was 355 000 SEK. The distribution is thus positively skewed. The largest registered debt was 176 millions SEK.

A few households show major changes in debts in a year. The largest observed increase was 77 millions and the largest decrease was 97 millions. The changes observed for the majority of the households are, however, much smaller. The mean change was an increase of 13 600 and the median change was 0. Of those who had no debt in a year 11 per cent had one the following year. Of those who had a debt in a year about 92 per cent also had one the following year.

The purchase of a house or other property is usually partly financed by a mortgage or loan. One might thus expect that households that buy or sell property would increase and decrease their debts respectively unless they owned property before or bought new property after having sold their old property. Data show that households that had no real estate or real estate at a value of less than 10 000 SEK in a year but owned more than 10 000 SEK worth of real estate in the following year in the mean increase their debts by 450 000 SEK while the median change was a decrease of 6000!

A similar pattern emerges if one selects households that owned real estate at a value of at least 10 000 in year, but had no such assets or at least less than a value of 10 000 in the following year. This group of households decreased their mean debts by 669 000 while the median decrease was only 15 000. There were, however, also households that increased their debts with large amounts.

We do find that changes in real estate investments influence the amount of debt a household has, but also that there is much heterogeneity in behaviour suggesting that other factors than investments in real estate might sometimes have a dominating influence on the decisions to take up loans.

Lets first consider the group of households with no debts at the end of a given year  $t-1$ . We have first estimated a random effects probit model for the event of having debt at the end of the following year  $t$ . Explanatory variables were: the age of the oldest household member, if single, the change in real estate investments and in financial assets, the change in

the sum of taxable income from work for all household members and last years disposable income. The current value of disposable income cannot be used in this equation because it will become simulated after the debt variables in SESIM. For this reason taxable income from work had to be used. The cost of borrowing is captured by a real rate of interest on short assets.

The probability to go into debt decreases with increasing age. Singles have a smaller probability to take up loans than couples, while those who have increased their investments in real estate during the year have a higher probability. If financial assets have increased since last year the probability to take a loan is smaller, but this effect is relatively weak. The higher income the easier the household has to pay interest and reduce the principle, and thus also easier to get into debt.

The model estimated to simulate the size of any loan is a random effect panel data model. The observations were restricted to households with no debt in the previous year. The explanatory variables were the same as in the probit equation except for the variable change in taxable income from work, which was dropped, and the change in the real rate of interest, which was added. The results show that not only does the probability to take up a loan decrease with increasing age but so does the size of the loan taken. Singles borrow less than couples. The larger increase in real estate investments the larger loan, and the higher disposable income the larger loan can the household afford. The real rate of interest influences strongly the size of the loan, the higher rate the smaller loan. We might finally note that unmeasured heterogeneity among households amount to a little more than 30 per cent of the total residual variation. The properties of the estimated residuals suggest that a normal approximation is not too bad. It implies that normal random numbers can be drawn when the amount of debt is simulated.

Let's now turn to the larger group of households who already are in debt and analyze how their debt changes. Also in this case a two-part model was estimated. A random effects probit model was estimated for the probability to stay in debt and a random effects regression model for the amount borrowed. The probability to stay in debt decreases with increasing age independently of the size of the debt. Independently of age the probability to remain in debt increases with increasing debt. This relation is almost the same through the whole age range, possibly with a somewhat smaller factor of proportionality above the age of 75. Singles have a smaller probability to stay in debt than couples, and households that have increased their investments in real estate have a higher probability. The effect of changes in the stock of financial assets is negligible. A high disposable income decreases the probability

somewhat, but this effect is small. The cost of borrowing is important though, the higher cost the smaller probability to stay in debt.

It was not easy to find a satisfactory model for the size of the debt given that the household remained in debt. We have finally selected a random effects regression model that was estimated using a sample constrained to households with a lagged debt exceeding 10 000 SEK. Without this constraint, i.e. using all observations with a debt, the right tail of the distribution of the residuals became very thick. In the simulations this resulted in a few households having excessively large debts. Although the sample was restricted to households with at least 10 000 in debts, the model will in SESIM be applied to all households with a debt.<sup>†††††</sup> Lagged debt is an important variable explaining current debt. The more indebted a household is the more will it reduce its debts. The estimates imply elasticities that vary from -0.2 among the youngest households to -0.04 among the oldest. So young people with high debts tend to decrease their debts more than elderly with high debts do, but independently of the size of the debt elderly generally reduce their debts more. Households which have increased their investments in real estate accumulate more debts than other households do, and the higher income a household has the more debts will it get. The cost of borrowing does not only influence the probability to stay in debt but also the size of the debt. The higher cost the less borrowed.

Examination of the residuals, both the household unique component and the general residual, shows that both distributions are negatively skewed and have a rather high kurtosis. It is thus not advisable to simulate using random draws from normal distributions. Instead we have drawn random numbers from the empirical distributions.

### 3.7 Utilization of inpatient care

We estimated two different models of the utilisation of inpatient care: (a) an imputation model and (b) a dynamic model for updating. These models were estimated separately for two different age groups: those older than 49 years of age, and those younger than 50 years of age. The reason for this is that we expected that the health-related behaviour, and in particular the utilisation of inpatient care, to differ between older and younger individuals. The dependent variable is the number of bed days. Both models were formulated as negative binomial models corrected for sample selection. Explanatory variables were health status, age, schooling, if divorced, relative income, region, if born in Sweden, gender, and if there was a

---

<sup>†††††</sup> The restriction of the sample reduced its size by about 5 per cent.

new born child in the household. The dynamic model in addition included the lagged dependent variable.

The qualitative content of the results obtained from the estimated marginal effects are almost identical between the imputation equation and the dynamic year-to-year equation. The effect of health status is positive and significant, i.e. people in bad health utilise more inpatient days. For instance, in the age group 50+ a person with bad health is expected to stay in hospital about two more days compared to someone with full health. The general average was 1.5 days in hospital. In the dynamic model the estimated marginal effect of the lagged dependent variable was positive and significant. Thus, the more inpatient care that was utilised in the previous year the more inpatient care the respondent is expected to use in the current year. The effect of being divorced differs between the young and the old age groups. In the old age group being divorced increases the expected number of days of inpatient care, while in the young age group being divorced decreases the expected number of days of inpatient care. Relative income was found to affect the expected number of days of inpatient care negatively – the higher relative income the lower the expected number of days of inpatient care. Being born in Sweden significantly increases the expected number of days of inpatient care. Being a male increases the expected number of days of inpatient care for the older age group and decreases the expected number of days of inpatient care for the younger age group.

### 3.8 The utilization of old age care

In Sweden old age care is provided by the local municipality usually in one of two forms, either as home help, people come home to the elderly and help with shopping, food, cleaning and washing, or as all day care in special housing for elderly. As a complement to these forms of formal care there is also informal care provided by children, relatives and neighbours. The SESIM data base LINDA does not have information about social care, so we had to use alternative data sources to model utilization of care. For this reason we needed an imputation model which would give us start values for the base year 1999 and a transition model to forward these start values. In both cases we used a multinomial choice model. Explanatory variables were gender, age, if a spouse was present, the spell length of care status and ADL (Activities of Daily Living). It is well-known that the ADL (PADL)<sup>\*\*\*\*\*</sup> status of a person is a good predictor of care needed, but there are no ADL-data in LINDA. For this reason we

---

\*\*\*\*\* Physical Activites of Daily Living

simulated ADL-scores using health, age, and gender as explanatory variables. These models were estimated from a data set obtained from a parish in Stockholm, which was the only data set available including the information needed. One could thus expect difficulties in an inference to the whole country. For this reason the estimates of the imputation model were calibrated so the model was able to replicate national estimates of the care distribution by age and gender.

If elderly are to receive assistance from children and kin they have to live in the same neighbourhood. The data set we used to estimate the multinomial choice models did not have any information about children (kin), so we could not use closeness to children as an explanatory variable, but we have nevertheless estimated a model which explains the probability to have children in the same neighbourhood to evaluate if the potential for informal care increases or decreases. A logistic regression model was estimated for each of single mothers, single fathers and parent couples. Explanatory variables were region, age, relative disposable income, schooling, tenure and if lived in the same neighbourhood as an adult child in the previous year.

It is well-known that people who live in special housing for elderly have excess mortality because of poor health. To account for this we estimated for each care status a duration model with a piece-wise constant hazard and gamma distributed heterogeneity. Explanatory variables were ADL-score, age, gender and marital status. The resulting simulated death frequencies by age and gender were balanced against those given by Statistics Sweden, so the frequencies for the entire population did not change.

#### **4. Simulation results**

The predicted increase in the population share of elderly in Sweden is rather modest compared to some of the central and south European countries (Figure 3). The share of 65+ will in our base scenario increase from 17.5 to 23.9 percent in the period 2000-2040. Yet, this implies a major increase in the number of elderly. The number of 65+ will increase by 58 percent. The very old and care intensive group 80+ will increase even more, by 75%, and their share of the population will increase from 5.1 to 7.9 percent in forty years. This is likely to put an increased pressure on the political system to match the expected increased demand for health care and social care by an increased supply. In some countries the increase in the

number of elderly will become balanced by a decrease in the number of children, and thus a natural reallocation of resources from children to elderly is possible. This is according to our simulations not the case in Sweden. The population share of those below 18 will stay rather stable between 21 and 22 percent.

When evaluating the future need for health care and social care of the elderly and how it can be met, we first have to look at the demand side and consider any competing demand, then to the size of the working population and its productivity, and finally to the relative incomes and wealth of the elderly to be able to discuss who can bear the increased burden of financing more care.

In international comparisons we have been used to find the Swedish population of elderly in better health than in most other countries, with less life-style health risks and more time for physical exercise. We have no reason to believe that this relative advantage would change in the nearest future, but we have found that the historical trend towards an ever more healthy elderly population might be broken and we simulate a minor decrease in the health status of the population, in particular among the young old (Figures 4 and 5). The explanation seems to be an increase of allergy, asthma, diabetes, and obesity and its consequences among the current middle aged. According to this scenario there is thus no help to find in a further improved health status of the elderly population.

However, if the decrease in health status is just a temporary set back, or if the technology of medicine improves to balance the observed changes, people might continue to improve their health relative to previous generations and the expected survival time will continue to increase. It is not obvious that this would decrease the demand for care; it might just postpone it for a few years. In our scenario with increased health status among the oldest the effect on the demand for care was marginal. More important were changes in the death risks. If improved life-styles, nutrition and medical technology continues to decrease death risk among young old and increase expected life time, we might well at the same time find an increased share of the oldest old in need of extensive care.

Our simulations suggest that the number of inpatient hospital days for people 65+ will have to increase by 70 percent (Table 1), the number of elderly in all day care by 83 percent and the number of elderly with home help by about 50 percent (Figures 6 and 7). If survival increases these numbers will become even higher. These estimates primarily reflect the increase in the number of elderly while we only predict a modest increase in utilization. This is a natural consequence of our methodology through which we take forward current



levels and trends in utilization into the future. In this sense we assume that the current standard of health care and old age care is maintained in the future.

Children and the working population compete with the elderly about public resources for health care and social care. We have already noted that the share of children will remain more or less stable, which implies that no resources will be freed from child care, schooling and health care of children, if the standard of these services will remain unchanged, but on the other hand there is no need to increase resources either to maintain the current standard. Our simulation model does not allow us to simulate all uses of public services by the working population, but for three relatively large items: sickness benefits, disability pensions and in patient hospital care. According to our simulations the number of compensated sickness days reached a maximum by 2003 and then gradually decline as the large baby-boom cohorts retire out of the sickness benefit system. The result is that the number of compensated days in 2040 will become approximately as many as in 2000. The number of inpatient hospital days for people younger than 65 only increase by a modest 10 percent, and the number on disability pension are simulated to increase by 8 percent. These findings suggest that any increase in total resources for health related purposes can be allocated towards the elderly.

Using the assumption that the cost of care will increase at the same rate as an average wage rate and that there is no productivity gain in these labour intensive activities, we found that total cost in fixed prices for inpatient care, home help and all day care will increase by a multiple of 3.7 in forty years (Table 2). Ideally one would like to compare this increase to the increase in the total budget of the public sector and to competing public expenditures. Within our model this is not feasible. All we have been able to do is to compare to total direct taxes paid by Swedish households, and we found that the expenditures for these three items that goes to the population 65+ as a share of total direct taxes will almost double, from 20 percent in 2000 to 36 percent in 2040 (Table 2). Judging from this comparison the increased demand for care of the elderly is so large that adjustments in one form or another will be needed: A reallocation of public expenditures towards the elderly, increased taxes or increased user fees, or decreased supply of public services. As suggested above the changes in the age distribution give some leeway for a reallocation within the public budget, but this will most likely not become enough. We therefore turn to focus on alternatives.

One alternative is to have spouses; children and relatives help to care for the elderly. Parents and adult children do not live as closely in Sweden as in some of the south European countries, but still much of the care elderly need is provided by spouses, children,

relatives and neighbours. Spouses help each other. Usually it is a wife, who typically is younger than her husband, who helps him. Singles are more dependent on children and other relatives. A necessary requirement for a child to be able to help its parents is that they live rather close. There are clear regional differences in this respect. In the rural and in particular northern part of the country there are fewer elderly living close to their children compared to the large metropolitan areas. One explanation is that distances are larger in the rural areas and another that children have tended to migrate to urban areas in the southern part of the country. The share of singles, in particular women, without children at commuting distance is higher than that of couples. One reason is of course that some of them never had a child, but also that they tend to be older and have survived their children. Old-old single women will find it most difficult to get help from children. Our simulations show that the share of people 65+ with children at commuting distance will decrease from about 60 percent in 2000 to 53 percent in 2040 (Figure 8). Most of this decrease is due to an increasing share of singles in high ages (Figure 9). This tendency holds for the whole country but it is a little stronger in the large metropolitan areas, in which the share of singles increases most. These simulations thus suggest that the future large groups of very old cannot count on their children but will have to find help and care elsewhere.

The continued migration from rural to metropolitan areas will contribute to the mismatch between generations. Around year 2000 about one third of the Swedish population lived in one of the three large metropolitan areas, while we now simulate that this share will increase to 45 percent in 2040 (Figure 10). Migration from urban to metropolitan areas is predominantly by young people, which implies that the aging of the population goes much faster in the rural areas. The dependency ratio is already higher in the rural areas but it will also increase more than average in the future. The rural communities, in particular in the northern part of the country, will thus face a much more difficult problem in caring for their elderly than metropolitan Sweden.

Our studies of the distributions of incomes and wealth have its intrinsic interest as an analysis of the standard of living of the elderly. They also tell us something about the means elderly will have to pay for care services.

A most striking property of the new Swedish pension system is what we have called “front loading”, i.e. young pensioners receive a higher pension than an actuarially fair annual rate would suggest. To compensate for this, future pensions are not indexed by the average annual growth rate in nominal wages, as many believe, but with this average less 1.6 percent. The implication is that pensions will lag behind the earnings of the working cohorts,

and if the real increase in wages is less than 1.6 percent, pensions will not maintain their real value. While the replacement rate, defined as the age specific mean taxable income of retired to the mean taxable income of people 20-64 years old, is 75-80 percent for someone who is 65, it drops to below 50 percent at the age of 90 (Figure 11). Because pensions make such a large share of the incomes of retired, this decrease in relative incomes with increasing age also shows up in the disposable income. Median disposable income of people 65+ relative to the median disposable income of those in the age bracket 20-64 was just above 70 percent in 2000 (Table 3). In our simulations the ratio decreases to 63 percent in 2020 and to 55 in 2040. Most of this decrease is a result of the aging of the pensioners, but also a result of younger cohorts getting lower replacement rates because they are assumed to live longer.

Another result of the design of the Swedish pension system is that the share of poor pensioners will increase. While poverty, defined as having a disposable income less than 50 percent of the median disposable income of the adult population, is less than 5 percent before people start to retire, and about 10 percent the first years after retirement, it then increases as people age. For the birth cohorts we have analysed poverty varies between 20 and 35 percent when they reach the age of 90 (Figure 12).

When considering increasing user fees (or taxes) to cover the increased costs for health care and social care it is of interest to know how high incomes the elderly in need of care will have. In year 2000 we found that 65-70 percent of those who needed home help or all day surveillance had disposable incomes in the first quartile of the income distribution, while only 47 percent of those who did not have any social care and also were 65+ were in the same quartile (Table 4). By 2040 these shares will increase to just above 75 percent and 53 percent respectively. There is thus a large and increasing share of elderly in need of care that will have so low incomes that they hardly will be able to pay any major increase in user fees.

Future changes in the wealth distribution will to a large extent depend on the growth in real estate prices and in the prices of stocks and shares, and on their relative magnitude. In our base scenario own homes and other real property will take a larger share of total wealth and the inequality of the wealth distribution will decrease. But if we instead allow the value of financial assets to grow relatively faster, inequality will increase. We have found that the large baby-boom cohorts are relatively wealthy. At the time of retirement the baby-boom cohorts will have wealth amounting to about eight annual incomes. The wealth buffer is relatively more important for the income poor than for those who have high incomes.

Our simulations then suggest that they will keep their wealth as they age (Figure 13). By the end of our simulation period the very old will have about the same wealth as those in the age

bracket 60-69. The share of elderly living in owned homes decreases with increasing age, and they thus liquidize their wealth when they sell their houses. But still much of this wealth will be tied up in owned homes and thus not very liquid. Compared to year 2000 the share of elderly in owned homes is simulated to increase with a peak in the middle of the simulation period around 2020. A recently developed market for reversed mortgages might, however, help to liquidize this kind of wealth.

When we correlate wealth and the utilization of social care we find no strong correlation. In year 2000 about 30 percent of those who had home help or all day surveillance belonged to the fourth quartile of the wealth distribution, while about 40 percent of those who had no social care belonged to the same quartile (Table 5). Those who needed care were thus not quite as rich as those who had no needed but there was no major difference in distribution of wealth, and our simulations suggest that this difference will become eliminated completely by end of the simulation period. In year 2040 about 50 percent of all 65+, independently of their care status, will belong to the fourth quartile of the wealth distribution. A large share of the elderly thus has a wealth buffer that could be used to finance health care and social care, but at the same time there will be a fair share of elderly who are both income and wealth poor. In 2000 only about 2 percent of those who had social care were poor in this sense, but by 2040 this share is simulated to increase to between 15 and 16 percent.

Which policy changes could reduce the need for care and ease the burden to provide for the elderly? We have analyzed the effects of improved health, reduced death risks, higher immigration, delayed withdrawal from the labour market and changes in the growth rates of the economy. We find no major effects on the utilization of care of improved health of the elderly. One might have expected that the demand for care would have decreased but we only find very marginal effects. This might be a result of our model specification not allowing health improvements to work through strongly enough into reductions in care, see below, but it is also possible that it takes much stronger improvements in health than we have assumed.

When we exogenously reduced the death rates the result was a major increase in the utilization of care. Our model simulated that more people with poor health and various limitations survived, and thus that the demand for health care and social care increased. It is possible that we exaggerate this effect because there is no direct functional relation in our model between health and survival with the exception of the excess mortality of those who have around-the-clock surveillance. Even so these results illustrate the importance of healthy

ageing. If the expected life span continues to increase but without improvement in age specific health the result will become ever more elderly in need of care.

Our simulations with assumed higher immigration to Sweden confirms what demographers previously have told us, namely that immigration is not the solution to the dependency burden. Although immigrants are younger than the average Swede, they have poorer health and will more frequently experience unemployment and go into disability pension. Initially they have higher fertility, but tend to adjust towards Swedish behaviour when they stay in the country. They will of course also age and eventually become retired. Immigration will only have a long-term effect on the dependency ratio if the immigration rate continuously increases.

Of the alternative scenarios we have tried postponed retirement is the most effective mean to reduce the dependency ratio and improve the income standard of the retired. When we lifted the upper retirement age from 65 to 70 the labour force participation rate of the age group 60-69 gradually increased. As a result the labour incomes of these age groups increased and they also increased their pensions. At the end of the simulation period the relative income of 65+ increased to 62 percent compared to 55 percent in the base scenario. All groups of elderly, independently of care status, will experience this increase in relative incomes but most of it will go to those who do not need any social assistance. Among those who use home help or all day surveillance there will be an overrepresentation of people who will not be able to take advantage of the opportunity to postpone retirement.

The incidence of poverty becomes much lower when people work longer. This is primarily a result of a reduction in the number of income poor, because the wealth distribution is not much influenced by people postponing retirement.

Changing the assumptions about the growth in real wage rates and in the general unemployment rates primarily influences the incomes of people. We also find changes in the labour force participation rates among those above 60, higher wage growth implies higher participation and lower growth lower participation rates, but the effects are smaller than those in the postponed retirement scenario. There are no effects on labour force participation in the age bracket 20-59. It is possible that our model underestimates the effect of increases in wage rates on labour supply, because there is no proper labour supply function in the model. Assuming a constantly higher growth in wage rates will of course in about thirty years result in much higher incomes, primarily for those who are in the labour market, but also for retired. In our high growth alternative the relative income of 65+ decreases because the incomes of the working cohorts increase faster than those of retired, while in the opposite scenario with a

lower growth the relative income of retired increases. To this result contributes that relatively more pensioners now receive the guaranteed basic pension and that the capital incomes of retired becomes relatively more important. The latter effect also implies that income inequality among 65+ increases in the low growth scenario.

In a final evaluation of the ageing of the Swedish population we suggest that the consequences of the ageing are not alarming, but adjustments are needed to meet the expected increase in demand for health care and social care. A policy which gives people incentives to postpone retirement is important. We also expect resources within the public sector to become reallocated from the working population towards the elderly. Our methodology does not allow a proper analysis of the need for increased user fees or tax increases, but we cannot exclude it. Perhaps even more important than the issue of financing future care is our demonstration of the relatively low future income standard of the elderly. It is a good prediction that relatively wealthy elderly will run down their stocks of wealth to compensate for low incomes when they age, and that young generations will increase their life-cycle savings when they discover the consequences of the pension system design. Unfortunately our simulations tell us that we will get an increasing share of both income and wealth poor elderly.

## **5. Methodology issues**

In view of the large resources need for the development of a completely new microsimulation model we decided to use an already existing model and within our resources add the new modules needed to analyze the problems of population ageing. Aside from the issue of resources this approach had its pros and cons. We got an already functioning model structure with a good interface, a demographic model, and most importantly, the whole tax, benefit and pension system included. The model is also a working tool at the Ministry of Finance which implies that there is someone who is updating it and someone who might be interested in using the modules we have developed after our project is completed. The other side of the coin is that the already existing structure put constraints on what we could do. For instance, the status concept of SESIM (see section 2.1) made it difficult to allow for part-time work and part-time retirement. We might have preferred to have a proper labour supply model.

With the experience we now have, we should probably have made the death rates health dependent so we had been able to endogenously capture the effects of health changes on survival and then simulated survival without the current close alignment to the

official predictions of Statistics Sweden. These predictions embed assumptions about an increased expected life duration which might contradict our own simulated mild decrease in health status.

SESIM is a household sector model and it does not include any markets or mechanisms for market clearance. There is thus no automatic feed back on changes that takes place in the household sector. We found this particularly troublesome when analysing domestic migration and its consequences for house prices. The historical and predicted future migration to the metropolitan areas have and should in the future drive up prices of owned homes in these areas more than in other areas in the country, and the resulting increase in the cost of housing should restrain migration. We considered introducing a price mechanism along the lines suggested in Creedy and Kalb(2005), but it turned out that relative housing prices had an insignificant effect on the choice of region to migrate to, and housing prices were also insignificant in the tenure choice model. This result might cause thoughts about the model specification, but we decided to avoid the hazard of introducing a price mechanism. House prices are now indexed with a nation wide index which implies that we maintain current regional differences in house prices in our simulations into the future.

There are almost no private markets for health care and social care in Sweden and thus no price mechanism that clears the market. Supply of the publicly provided services is determined in a complex political process and excess demand results in queues and rationing. Our simulations are based on the observed utilization of publicly provided services in the 1990s and the beginning of the current century. This implies that observed levels and changes in this period are carried forward in the simulations. We thus simulate the consequences of population ageing assuming that the frequency of utilization and in this sense the standard of care is maintained unchanged. This is probably the only practically feasible approach. The alternative, to model how the political and administrative processes react to changes in demand would become more adventurous.

There is no macro economic feed back in the model either. Indicators like CPI changes, returns on financial assets and unemployment rates are fed into the model exogenously. The model thus does not generate any business cycles but it is possible to make the model react to exogenously imposed cycles and it is of course possible to adjust the exogenous indicators to the outcome of the simulations in an interactive simulation strategy which switches back and forth between simulations and changes in the exogenous indicators. A disadvantage with such a strategy is that it is labour intensive and time consuming. It might also be difficult to know exactly which changes one should make in each iteration. We have

chosen just to run a few distinct alternatives without attempting to replicate business cycles. We leave a study of the sensitivity of our results to cycles in the exogenous variables to future research.

Data problems are legio in microsimulation. In our study we have been forced to make stochastic imputations to the start data set in a number of cases because LINDA does not have all the data needed. For the same reason we have also found it necessary to estimate models using other data sets than LINDA. In all these cases one has to make sure that the models, both for imputation and for simulation, are estimated so it is possible to draw inference to the LINDA population. As long as we use register data this is not a great problem because all these data sets are based on the same population register. The samples of ULF and HINK/HEK were drawn from the same register too. The only sample we have used which does not permit a statistical inference to the whole Swedish population is the Kungsholmen study, which only covers a parish in the down town area of Stockholm. At the time when we estimated our models for social care there was no other data set which included both ADL scores and information about care. By calibrating the estimated imputation model to national estimates of care frequencies by age and gender from the HINK/HEK survey we made sure that the start values were at the right level. (HINK/HEK does not include information about ADL.) The Kungsholmen study might also deviate from a national sample in its transition frequencies between modes of care but it is not as obvious as the difference in initial levels.

The calibration techniques have in this study also been used to make the model simulate demographic changes which coincide with the official demographic forecasts of Statistics Sweden. The demographic models of SESIM then primarily serve the purpose of assigning events such as marriage, separation and death to single individuals. We have also used calibration to introduce alternative scenarios with respect to death risks and immigration frequencies, and finally also, as explained in chapter 3, to reduce simulation uncertainty.

In a simulation model like SESIM there are both direct causal links and indirect links. For instance, the utilization of social care is primarily determined by ADL, and ADL is determined by health status in addition to age and gender. A change in the health status of an individual or in a group of individuals will thus work on the demand for care indirectly via the ADL. Because all models are stochastic and they have unexplained residuals the variance of which is not unimportant, the relation between health and social care becomes “smoothed”. In retrospect we now suspect that the effect of changes in health status on social care is too weak. There is a similar problem with respect to the relation between health and wealth. Our model does not include any direct causal effects of health on wealth or vice versa, but any



correlation would have to be generated by indirect relations through age, unemployment, disability, income etc., and it turns out to be rather weak. Contributing to this result is of course also that we had to impute initial health status rather than observe it.

The greatest advantage of microsimulation is probably that it is such a good tool for studying distributional issues and comparing subpopulations to each other. We have used our model for this purpose, in particular when we have analyzed poverty and the income standard of people in need of social care. This kind of analysis requires however that the model is designed such that it replicates well the conditional distributions of interest. As suggested above one cannot always trust a microsimulation model to do so if it has not been designed for this purpose from the beginning.

For a few variables we have observed that the model in the first few years moves rather quickly towards a new (dynamic) equilibrium. This initial adjustment might be the result of observed relatively large changes in the exogenous variables in the first years of simulation, and then the adjustment has an observable equivalence in reality. But it can also be an artefact arising from bad imputed start values or lack of coherence between start values and the dynamic model. If our start year 1999 was extreme the model tries to bring the simulated path back to a more normal course, but depending on model structure it might not react strongly enough to do it in one or two years but needs more time. A simple Markov type model with just one estimated parameter for the effect of the lagged dependent variable might not be fully realistic. To estimate a richer dynamic structure one will, however, need longer panels of data than we have been able to use.

To build and to keep a microsimulation model alive is not one man's job. It requires a team. We have been fortunate in having a team representing different disciplines. Given the topic of our project the complementing competences have been very useful, although we initially had to bridge differences in "research culture". In a project like ours it is very important to have a skilled programmer who is not only interested in technical solutions but also in the scientific problems the simulation model is designed to solve and understands the environment in which the model is to be used. A programmer who does not have these skills cannot communicate with the scientific team efficiently and will most likely find technical solutions which later will be found suboptimal. Our project has been fortunate in this respect. Even if some of our project members have been able to program Visual Basic, the language of SESIM, there is need of coordination and of having someone who has both complete overview and detailed knowledge of what the model contains and does.

Microsimulations is data demanding and needs relatively large resources in terms of scientific competence. It used to need the best computers one could get, but modern standard PCs have become so powerful that computing capacity is no longer a great problem. Most of the simulations in this project have been done on good PCs and sometimes even on notebooks. Running SESIM, even if one used the whole sample, was no great problem. Turn around time was acceptable, but to write out large data sets of simulation output on a hard disk or external media was rather time consuming. The size of these data sets also meant that conversion to, for instance STATA or SAS data sets and analysis within these packages took time and needed some planning.

Fifty years after Guy Orcut presented his path breaking ideas of mirroring natural experiments in a microsimulation model, microsimulation has become an accepted approach. In particular government agencies have seen how useful this approach is in policy making and an increasing number of models, both static and dynamic, are now in use around the world. Resistance have been stronger in the academic community, in particular among economists, who favour small neat *ceteris paribus* type models which focus on a few conditional mean relations. About ten years ago the Panel of Retirement Income Modeling of the U.S. National Research Council (Citro & Hanushek, 1997) expressed the view that our theoretical and empirical understanding of how society works was not good enough to motivate a microsimulation approach to study the problems of an ageing population. We hope that we in our project and through this book have demonstrated that microsimulation is a useful and viable approach for this and similar purposes. There is no reason why social scientists should refrain from using and developing microsimulation when this is a standard tool in science and technical science. In doing so we must of course maintain our high standards of model building and rigorously test our models against data using conventional methods of statistical inference. Plugging in desk-top models into large microsimulation structures without good empirical foundation will just give critics of large black box models of doubtful scientific quality new wind and not produce useful tools for policy analysis.

## References

- Abowd, J. and Farber, H. (1982): "Job Queues and the Union of Status of Workers," *Industrial and Labor Relations Review*, 35, p. 354-367.
- Citro, C.F. and E.A. Hanushek, (1997), Assessing Policies for Retirement Income. Needs for Data, Research and Models, National Research Council, National Academy Press, Washington D.C.
- Creedy, J. and G. Kalb, (2005), "Evaluating Policy Reforms in Behaviour Tax Microsimulation Models", Paper presented at the 34<sup>th</sup> meeting with the Economic Society of Australia, Melbourn, September 26-28, 2005
- Davison, A.. and D.V. Hinkley (1997), Bootstrap Methods and Their Applications, Cambridge University Press
- Edin, P-A. and P. Fredriksson (2000), "LINDA – Longitudinal Individual DATA for Sweden", Working Paper 2000:19, Department of Economics, Uppsala University.
- Eklöf, M., and D. Hallberg (2006): "Estimating early retirement programs with special retirement offers," mimeo, Dept. of Economics, Uppsala University, Sweden.
- Flood, L. (2003), "Formation of wealth, income of capital and cost of housing in SESIM". SESIM working paper, [www.sesim.org](http://www.sesim.org)
- Flood, L. (2007), "Can we afford the future? An evaluation of the new Swedish pension system", in Modeling our Future, eds. A. Harding and A. Gupta, Elsevier.
- Flood, L. and Gråsjö (2001), "A Monte Carlo simulation study of Tobit models", Applied Economic Letters, 8, 581-584
- Klevmarken, N. A. (2002), "Statistical inference in micro-simulation models: Incorporating external information", Mathematics and Computers in Simulation 59 (2002) 255-265
- Klevmarken, N.A: and B. Lindgren (eds.), 2007, Simulating and ageing population. A microsimulation approach applied to Sweden, book manuscript
- Manning, W. G., N. Duan and Rogers, (1987), "Monte Carlo evidence on the choice between sample selection and two-part models", J. of Econometrics 35:59-82
- Pettersson, Thomas and Pettersson, Tomas, (2003), "Fördelning ur ett livscykelperspektiv", Appendix 9, Long Term Survey, Ministry of Finance, Stockholm
- Pettersson, Thomas and Pettersson, Tomas, (2007), "Lifetime redistribution through taxes, transfers and non-cash benefits", in Modelling our Future, eds. A. Harding and A. Gupta, Elsevier.

Pettersson, Thomas, Pettersson, Tomas and A. Westerberg, (2006) “Generationsanalyser – Omfördelning mellan generationer i en växande välfärdsstat”, ESS 2006:6, The Swedish Ministry of Finance

Poirier, D. (1980): “Partial Observability in Bivariate Probit Models,” *Journal of Econometrics*, 12, pp. 209-217.

Table 1. Number of bed days in inpatient hospital care for people 65+ , by simulation scenario .

Year	Average no of days			Total no of days (1000)		
	Base scenario	Improved health	Improved health & lower death risks	Base scenario	Improved health	Improved health & lower death risks
2000	3.2	3.3	3.2	4954	5068	4990
2020	3.2	3.1	3.5	6495	6270	8117
2040	3.4	2.9	4.0	8396	7241	12087

Table 2. Total costs for inpatient care, home help and all day care in special housing for those who are 65+.

År	Total cost (Millions SEK, fixed prices)	Share of total cost			Total cost as share (%) of total direct taxes paid by households
		Hospital care	Home help	All day care	
2000	92231	0.365	0.240	0.395	20.5
2020	166719	0.396	0.204	0.400	24.9
2040	339297	0.363	0.211	0.426	36.2

Table 3. Median equivalized disposable income of 65+ relative to the median income of 20-64 years old

<b>Year</b>	<b>Basic scenario</b>	<b>High pension age</b>	<b>High immigration</b>
<b>2000</b>	<b>0.709</b>	<b>0.711</b>	<b>0.712</b>
<b>2020</b>	<b>0.634</b>	<b>0.693</b>	<b>0.634</b>
<b>2040</b>	<b>0.552</b>	<b>0.623</b>	<b>0.558</b>

Table 4. Distribution of disposable income by form of care compared to that of all households.

(Percent of all 65+ by form of care)

<b>Year/Care</b>	<b>Quartiles of disposable income</b>				
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>All</b>
<b><u>2000</u></b>					
<b>No care</b>	<b>46.7</b>	<b>30.8</b>	<b>14.1</b>	<b>8.4</b>	<b>100</b>
<b>Home help</b>	<b>68.6</b>	<b>18.7</b>	<b>7.1</b>	<b>5.6</b>	<b>100</b>
<b>All day care</b>	<b>65.5</b>	<b>22.2</b>	<b>8.0</b>	<b>4.3</b>	<b>100</b>
<b><u>2040</u></b>					
<b>No care</b>	<b>53.1</b>	<b>30.7</b>	<b>10.9</b>	<b>5.3</b>	<b>100</b>
<b>Home help</b>	<b>75.4</b>	<b>17.6</b>	<b>3.8</b>	<b>3.2</b>	<b>100</b>
<b>All day care</b>	<b>76.4</b>	<b>14.7</b>	<b>4.6</b>	<b>4.2</b>	<b>100</b>

**Table 5. Distribution of net worth by form of care compared to that of all households.**

(Percent of all 65+ by form of care)

Year/Care	Quartiles of net worth				
	1	2	3	4	All
<b><u>2000</u></b>					
No care	4.7	28.0	27.4	39.9	100
Home help	2.5	40.2	27.4	29.8	100
All day care	2.0	39.9	25.8	32.3	100
<b><u>2040</u></b>					
No care	12.0	17.4	20.8	49.9	100
Home help	12.4	20.1	18.2	49.3	100
All day care	11.1	21.6	19.9	47.4	100

Figure 1. Structure of SESIM

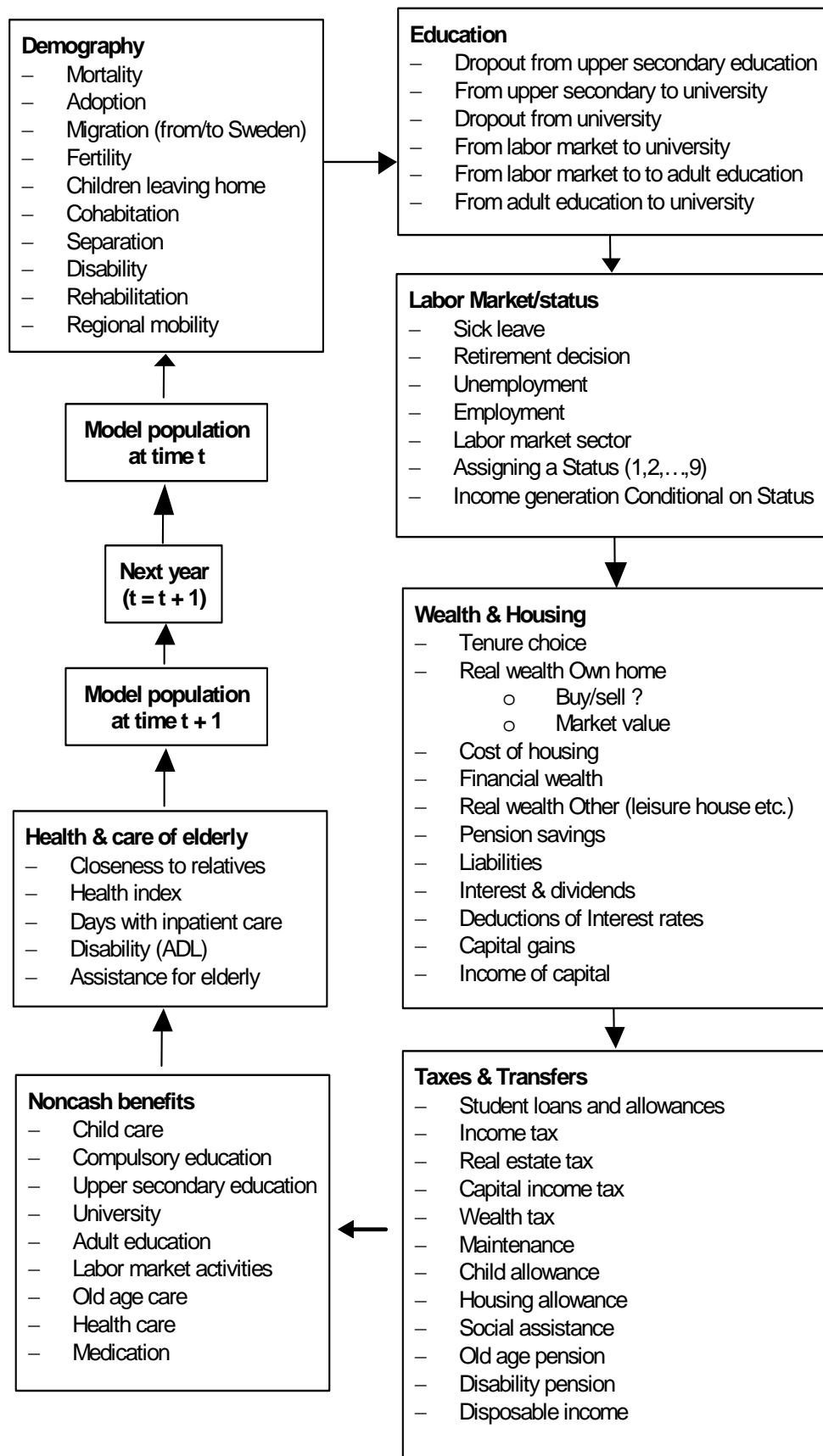
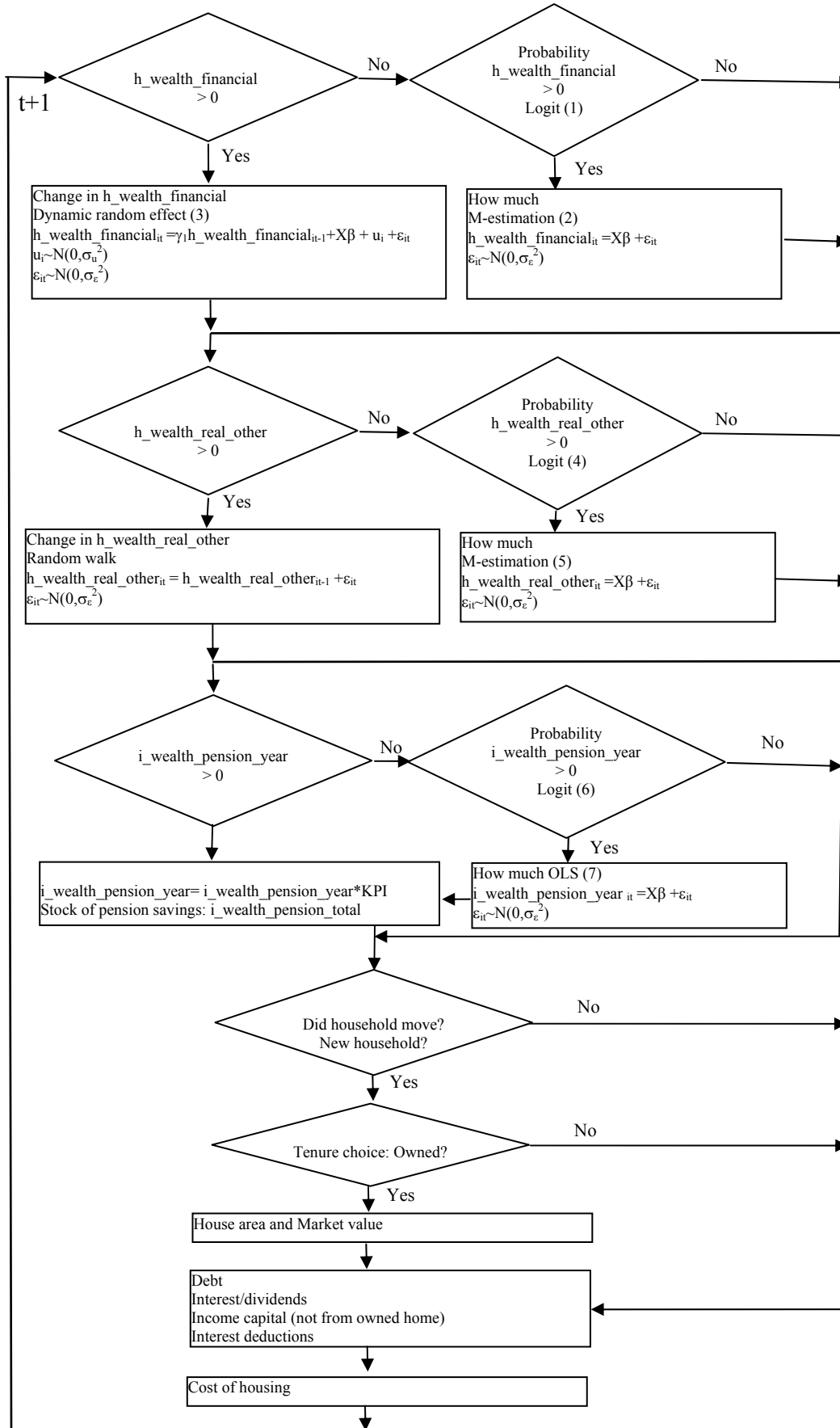
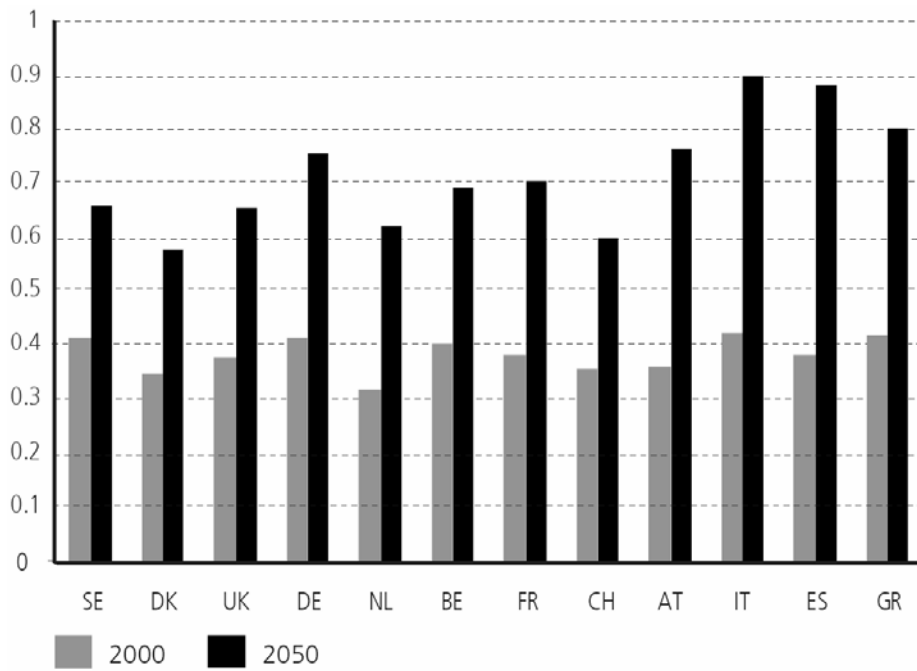




Figure 2. Financial and real wealth and cost of housing in SESIM



**Figure 3.** Old age dependency ratio, 2000 and 2050 (population 60 and over to population 20 to 59 years).



Source: Eurostat: Population by sex and age on 1st January of each year.

**Figure 4. Simulated health status 2000-2040 for people aged 50-74. Population shares by health status.**



Figure 5. Simulated health status 2000-2040 for people 74+. Population shares by health status.

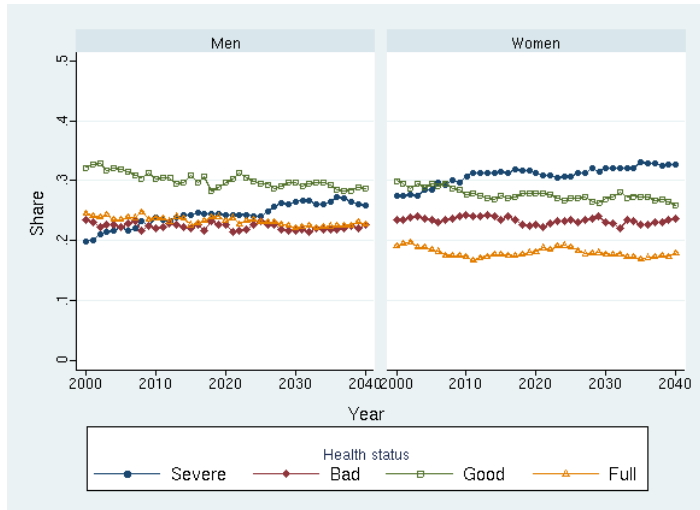


Figure 6. Number of persons with all day care in special housing, by age (1000nd)

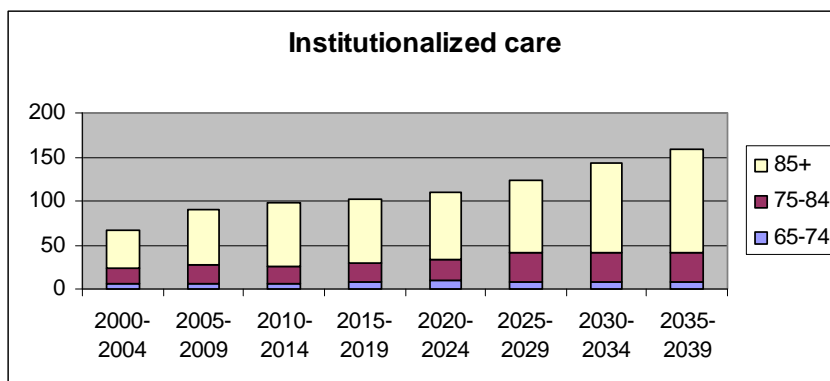


Figure 7. Number of persons with home help, by age (1000nd)

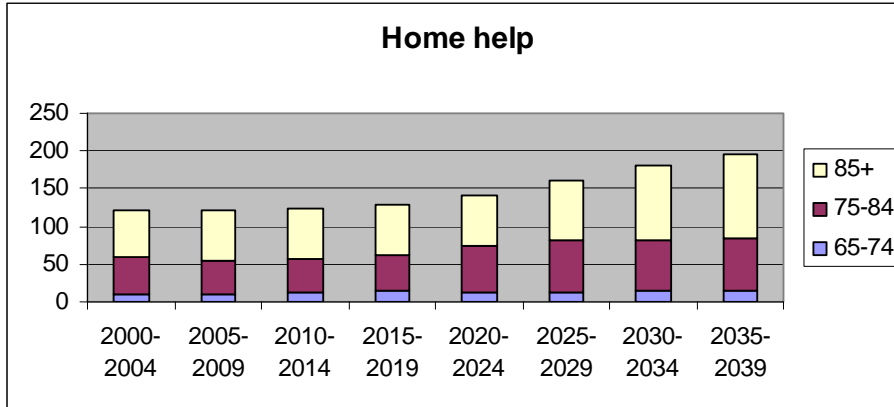


Figure 8. Share of people 65+ who has at least one child living in the neighbourhood

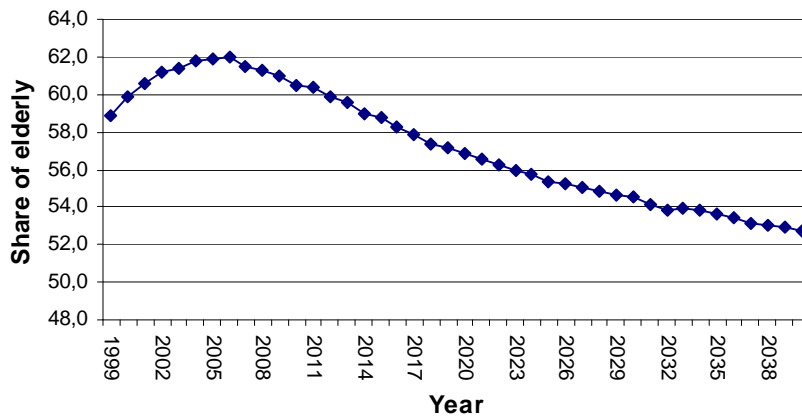


Figure 9. Share of people 65+ who live in the same neighbourhood as a child, by marital status and age

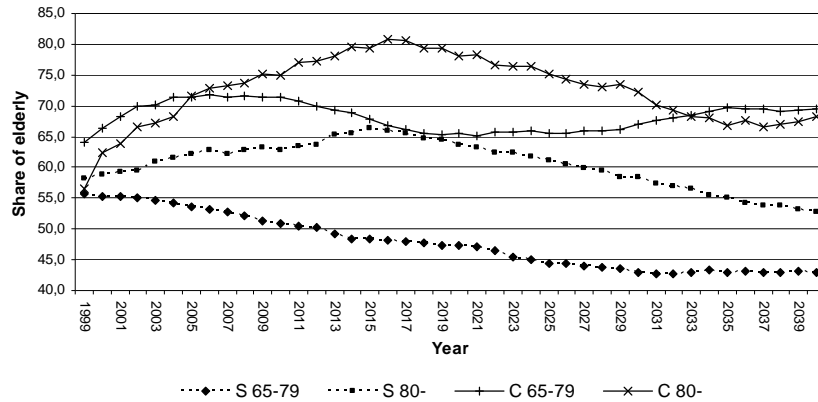


Figure 10. Population shares by region 2000-2040

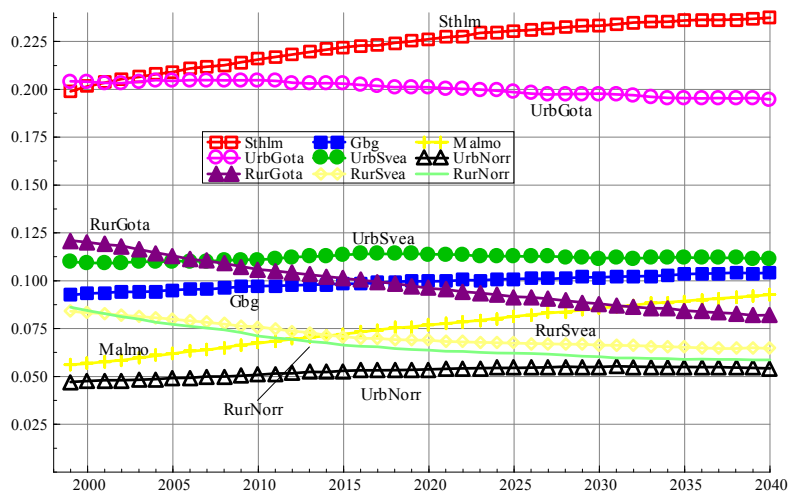


Figure 11. Mean tax-assessed income for retired by age and birth cohort relative to the mean tax assessed income for people aged 20-64

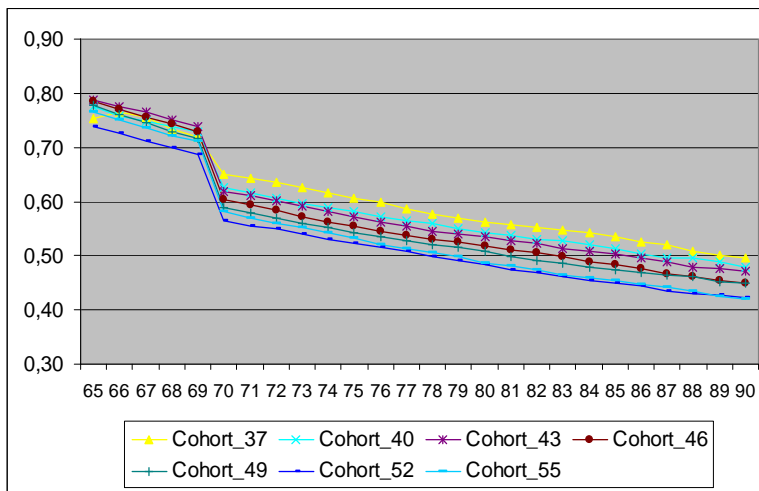


Figure 12. Share of poor (disposable income less than 50% of median for the adult population) by birth cohort and age

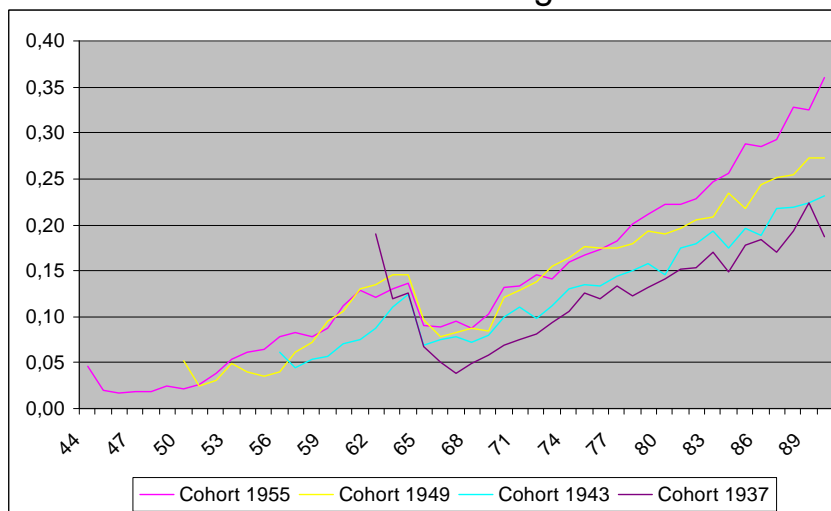
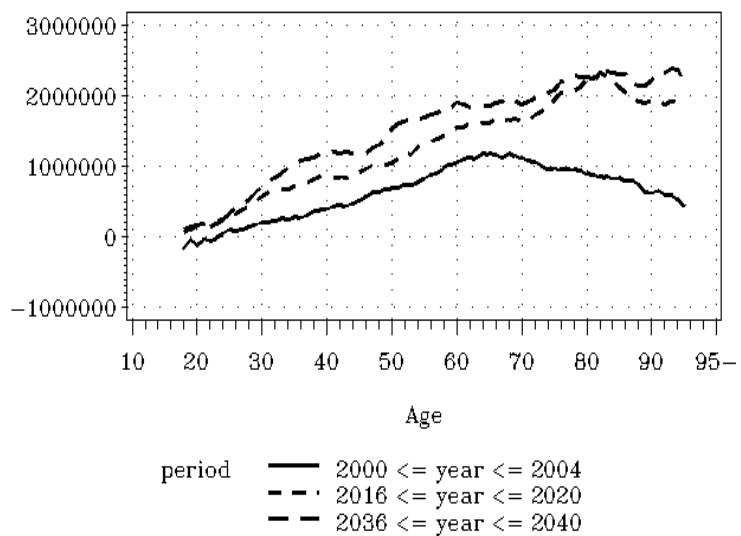


Figure 13. Mean net wealth per household in the 1999 price level, by age and period



## WORKING PAPERS\*

Editor: Nils Gottfries

- 2006:27 Henrik Jordahl and Che-Yuan Liang, Merged Municipalities, Higher Debt: On Free-riding and the Common Pool Problem in Politics. 34pp.
- 2006:28 Per Engström, Ann-Sofie Kolm and Che-Yuan Liang, Maternal Addiction to Parental Leave. 18pp.
- 2006:29 Jonas Björnerstedt and Andreas Westermark, Delay in Bargaining with Externalities. 27pp.
- 2006:30 Pär Österholm, Incorporating Judgement in Fan Charts. 36pp.
- 2006:31 Mikael Carlsson and Andreas Westermark, Monetary Policy and Staggered Wage Bargaining when Prices are Sticky. 26pp.
- 2007:1 Mikael Elinder, Local Economies and General Elections. 26pp.
- 2007:2 Ouarda Merrouche, The Long Term Impact of French Settlement on Education in Algeria. 19pp.
- 2007:3 Ouarda Merrouche, The Long Term Effect of Education Spending Decentralization on Human Capital in Spain. 15pp.
- 2007:4 Erik Post, Macroeconomic imbalances and exchange rate regime shifts. 38pp.
- 2007:5 Christian Andersson, Teacher density and student achievement in Swedish compulsory schools. 31pp.
- 2007:6 Thomas Aronsson, Sören Blomquist and Luca Micheletto, Where Should the Elderly Live and Who Should Pay for their Care? A Study in Demographics and Geographical Economics. 22pp.
- 2007:7 Sören Blomquist and Vidar Christiansen, Public Provision of Private Goods and Nondistortionary Marginal Tax Rates. 17pp.
- 2007:8 Marcus Eliason and Henry Ohlsson, Living to Save Taxes. 13pp.
- 2007:9 Åsa Ahlin and Eva Mörk, Effects of decentralization on school resources: Sweden 1989-2002. 31pp.
- 2007:10 Henry Ohlsson, The equal division puzzle – empirical evidence on intergenerational transfers in Sweden. 20pp.
- 2007:11 Daniel Hallberg and Mårten Lagergren, Moving in and out of public geriatric care in Sweden. 26pp.

---

\* A list of papers in this series from earlier years will be sent on request by the department.



- 2007:12 Per Engström, Wage Formation and Redistribution. 22pp.
- 2007:13 Henry Ohlsson, Tax avoidance – a natural experiment. 21pp.
- 2007:14 David Kjellberg and Erik Post, A Critical Look at Measures of Macroeconomic Uncertainty. 27pp.
- 2007:15 Mikael Carlsson and Andreas Westermark, Optimal Monetary Policy under Downward Nominal Wage Rigidity. 52pp.
- 2007:16 Robin Douhan and Anders Nordberg, Is the elephant stepping on its trunk? The problem of India's unbalanced growth. 33pp.
- 2007:17 Annika Alexius and Bertil Holmlund, Monetary Policy and Swedish Unemployment Fluctuations. 27pp.
- 2007:18 Meredith Beechey and Pär Österholm, The Rise and Fall of U.S. Inflation Persistence. 23pp.
- 2007:19 Henry Ohlsson and Donald Storrie, Long term effects of public policy for displaced workers in Sweden – shipyard workers in the West and miners in the North. 26pp.
- 2007:20 Niklas Bengtsson, How responsive is body weight to transitory income changes? Evidence from rural Tanzania. 38pp.
- 2007:21 Karin Edmark, Strategic Competition in Swedish Local Spending on Childcare, Schooling and Care for the Elderly. 38pp.
- 2007:22 Fredrik Johansson, How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML? 25pp.
- 2007:23 Henry Ohlsson, The legacy of the Swedish gift and inheritance tax, 1884–2004. 25pp.
- 2007:24 Ranjula Bali Swain and Fan Yang Wallentin, DOES MICROFINANCE EMPOWER WOMEN? Evidence from Self Help Groups in India. 26pp.
- 2007:25 Bertil Holmlund and Martin Söderström, Estimating Income Responses to Tax Changes: A Dynamic Panel Data Approach. 34pp.
- 2007:26 N. Anders Klevmarcken, Simulating the future of the Swedish baby-boom generations. 60pp.

See also working papers published by the Office of Labour Market Policy Evaluation  
<http://www.ifau.se/>