Discussion Paper

Central Bureau of Statistics

no. 85 March, 1993

Demographic Forecasting with a Dynamic Stochastic Microsimulation Model[1]

by

Leif Andreassen

Research Department
Division for Public Economics

Abstract

This paper summarizes some of the main features of demographic microsimulation modeling, including a comparison with other demographic forecasting techniques. It also presents a dynamic stochastic microsimulation model, MOSART, which projects the population, labor force, and education level. This model illustrates some of the basic principles of forecasting by means of microsimulation. The two most distinctive features of the model is that it is based on a large real sample of the population (as opposed to a synthetic sample) and that it is programmed in an object oriented language. An appendix gives a brief general survey of the use of microsimulation models in Norway.

[1] This is a revised version of a paper which was presented at the Helsinki Seminar on Microsimulation Models in Helsinki March 16-17, 1992. I would like to thank O. Bjerkholt, H. Brunborg, J.K. Dagsvik, D. Fredriksen, N. Keilman and O. Ljones for valuable comments.

1 Introduction

The purpose of this paper is to discuss some of the main features of microsimulation modeling and to give an introduction to the Norwegian dynamic microsimulation model MOSART. The main focus of the paper will be on demographic forecasting of labor supply and educational attainment. In the first part of the paper the microsimulation approach is compared with two widely used demographic forecasting techniques, the headship rate method and the transition matrix method. This comparison illustrates the benefits and shortcomings of the microsimulation approach.

Microsimulation is a technique which is used in analysis and forecasting when one has established structural relationships at the micro level in a disaggregated system, but where it is difficult or impossible to analyze analytically the whole system. In the following I will mainly discuss the use of microsimulation in demographic models incorporating a large set of structural relationships at the micro level. The paper includes a brief survey of some dynamic microsimulation models discussed in the literature. It is, however, beyond the scope of the present paper to present a detailed discussion of the programming structure and of the data which are used in these models.

In the second part of the paper I present a dynamic stochastic microsimulation model which has been developed at the Central Bureau of Statistics of Norway and which illustrates some of the concepts discussed in the first part of the paper. Microsimulation models are usually developed as tools for investigating public policy questions. The institutional setting and use of the models is therefore important in understanding their structure. In an appendix I survey the use of microsimulation models (both static and dynamic) in Norway and discuss their role in the policy process.

2 Microsimulation models

Mathematical models based on micro relationships can be so complex that it is impractical or impossible to analytically derive aggregate relationships by conventional mathematical techniques. The complexity can be due to the size and diversity of the model or to the inherent complexity of the model's micro relationships. An alternative to describing the aggregate behavior of the system analytically is to use a *microsimulation model* to imitate the behavior of the system being modeled by generating specific solutions numerically. By deriving numerical solutions under various initial conditions and for different values of the parameters and the exogenous variables, it is possible to infer general aggregate relationships from the model's micro relationships. Each such numerical solution is a simulation which can be viewed as an experiment on the model. The number of simulations required to explore all ranges of inputs and parameter values grows quickly as a model becomes more complex. Instead of doing simulations for all possible combinations of inputs and parameter values, it is often possible to design the simulation experiments in such a manner that one can infer the main characteristics of a model from a relatively small number of simulations (see Box, Hunter and Hunter (1978) for a more detailed discussion of such experiments).

Microsimulation models are often based on an empirical sample of the population being studied. The population will in the fields of demography and economics usually consist of either individuals, households, or firms. The distribution of characteristics across the micro units in the population (and the sample) can be difficult to summarize analytically, but is directly taken into account in a microsimulation model. An example is a tax model based on a random sample of individual income tax returns. Each tax rule in the model may be simple by itself, but analyzing the aggregate effects of changes in the tax rules is almost impossible to do analytically. Such analysis can, however, be fairly easily done using microsimulation.

Simulation is often used in analyzing dynamic models, since such models often are very complex. The term *dynamic* refers to phenomena that evolve over time and where the pattern of change at one time is interrelated with those at other times. We may distinguish between two types of dynamic micro models. The first type exhibiting *true state dependence* contains relationships established at the micro level, where the current realization of the dependent variables depends on the past realizations of dependent and independent variables. This type includes intertemporal behavioral models of consumption, saving, and labor supply along with behavioral models incorporating learning effects.

The second type caused by *unobserved heterogeneity* contains correlations between current and past realizations because of unobserved variables. In behavioral models these unobservables may be related to either the preferences or the choice constraints of the individual agents. Such models are dynamic relative to the analyst but not relative to the individual agents. An example is the probability of getting a job for an individual in a heterogenous group. The analyst must take into consideration that over time the average probability of getting a job for this group will decline as "high quality" individuals get jobs more quickly than "low quality" individuals. Another example of a relationship where unobserved heterogeneity may be important is the labor market behavior where individual behavior depends on the unobserved possibilities faced by the individuals.

In *dynamic microsimulation models* each micro unit in the sample is aged individually. In demographic models the number of persons in each cohort evolves over time depending on survival rates (or probabilities). New cohorts are determined by the birth rates (or probabilities) of earlier cohorts and on the sizes of these cohorts. Most dynamic microsimulation models incorporate static relationships (for example static labor supply relationships), but the underlying demographic change still leads to the system as a whole being dynamic.

Microsimulation is also used in analyzing *static models*, even though such models generally are not as complex as dynamic models. In static models the structure of the underlying sample is not modified over time. Examples of static microsimulation models are tax models which use the same sample of tax returns to simulate policy changes over several years and models with static behavioral relationships which are so complex that they can not be derived analytically but must be derived numerically. In the following we will mainly be concerned with demographic dynamic simulation models.

One can distinguish between two types of dynamic microsimulation, longitudinal and cross-sectional dynamic microsimulation. *Longitudinal microsimulation* consists of producing life histories for a single cohort, while *cross-sectional microsimulation* produces life histories for a cross-section of the population consisting of many cohorts.

Simulation models can be either deterministic or stochastic. In *stochastic simulations* the relationships are influenced by random fluctuations, while in *deterministic simulations* this is not the case. Random fluctuations in a model's relationships can either be an explicit attempt to take into account underlying uncertainty (for example taking into account the statistical uncertainty of the estimates used in a model), or it can be the result of the structure of the model (for example the simulation procedure used). Merz (1991) gives a more detailed discussion of the above concepts in his survey of recent work on microsimulation models, with an emphasis on static microsimulation models.

The substance of microsimulation models resides in the relationships which are simulated and not in the microsimulation framework itself. In models of Markov processes[1] (which are common in demographics) the main relationships are usually formulated as transition probabilities or transition intensities. The simplest models assume that these transition probabilities (or intensities) are constant throughout the simulation period. The microsimulation model MOSART, which is presented at the end of this paper, makes this assumption. More advanced microsimulation

[1]A Markov process is a stochastic process where the future behavior of the process is only determined by it's current state. The probability of any future behavior is not altered by additional knowledge concerning it's past behavior.

models incorporate transition probabilities or intensities which change over time. The simplest way of doing this is by extrapolating present time trends. Using constant probabilities or intensities usually leads to a model which is fairly easy to understand (but may be unrealistic), while complex extrapolation techniques can make it difficult to interpret the output from a model.

3 Demographic forecasting methods

Two demographic forecasting techniques which have been widely used are the headship rate method and the transition matrix method. The headship rate method makes assumptions about the distribution of characteristics across a population over time, while the transition matrix approach makes assumptions about the transitions between states. For example, at the Central Bureau of Statistics of Norway the headship rate method has been used to make forecasts of the labor force, while the transition matrix method has been used to make population forecasts and forecasts of educational attainment. In the following I will compare the demographic microsimulation method with these two much used forecasting techniques. The microsimulation approach is in principle the same as the transition matrix approach, but is based on different calculation methods. The following discussion will be fairly brief and will only consider these forecast methods in their simplest form.

The term *headship rate* derives from work on household structure and describes a much used method of forecasting household composition. The headship rate is defined as the ratio of the number of household heads (which equals the number of households) to the size of the population. An example of such a rate is the ratio of household heads among all persons in the age group under 40 years of age to the total number of persons in this category. To illustrate how forecasts are made using this method we assume that we have the following vector of headship rates

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix},$$

where a_1 is the proportion of household heads among persons under 40 years of age and a_2 is the proportion of household heads among persons 40 years and older. We assume that a population forecast tells us that at time t there will be n_{1t} persons under 40 years and n_{2t} persons 40 years and older. A headship rate forecast is then derived by doing the following vector multiplication

$$\begin{bmatrix} n_{1t}, & n_{2t} \end{bmatrix} \cdot \mathbf{A} = q_t$$

where q_t is number of households in the population at time t. This is a very rudimentary forecasting method since it does not consider the processes, such as marriage and divorce, which lead to changes in household composition. It's advantage is that it is simple and requires fairly simple data. This method is also often used in making forecasts of the labor supply, using labor force participation ratios instead of headship rates.

A more sophisticated method is the *matrix transition method*, also known as the *Leslie matrix method.* This method assumes that we have a matrix of transition probabilities $\mathbf{B} = \{b_{ij}\}$, where b_{ij} is the probability of being in state j given that state i was occupied the previous period. The demographic system is interpreted as a discrete state space Markov process and the matrix $\mathbf{B}$ as a matrix of transition probabilities. To illustrate this method we look at a model which forecasts educational attainment. The transition matrix might look like

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1K} \\ b_{21} & b_{22} & \cdots & b_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K1} & b_{K2} & \cdots & b_{KK} \end{bmatrix},$$

where b_{jk} is the probability of going from the education level j in the previous period to education level k in the present period, with $\sum_k^K b_{jk} = 1$. We assume that the distribution of educational attainment in the population at time t is given by the vector

$$\mathbf{M}_t = \left[m_{1t}, \; m_{2t}, \; \cdots, \; m_{Kt} \right],$$

where m_{jt} is the expected number of persons at time t who during their lifetime have attained education j. We now multiply this vector with the transition matrix and get

$$\begin{aligned} \mathbf{M}_t \cdot \mathbf{B} &= \left[m_{1\,t+1}, \; m_{2\,t+1}, \; \cdots, \; m_{K\,t+1} \right] \\ &= \mathbf{M}_{t+1}. \end{aligned}$$

Assuming constant transition probabilities, we can make a forecast for the period $t + 2$ by repeating the procedure,

$$\begin{aligned} \mathbf{M}_{t+1} \cdot \mathbf{B} &= \mathbf{M}_{t+2} \\ &= \mathbf{M}_t \cdot \mathbf{B}^2, \end{aligned}$$

and it is easily seen that a forecast of educational attainment for period $t + n$ is given by

$$\mathbf{M}_t \cdot \mathbf{B}^n = \mathbf{M}_{t+n}.$$

One should note that the above assumes a closed population.

This forecasting method was introduced by Leslie (1945) to make population projections. The transition matrix is in this case called a Leslie matrix and consists of survival probabilities for age groups beyond childhood and probabilities reflecting fertility and child survival. It thereby combines both the probability of an individual giving birth to a new individual and the probability of the individual surviving to the next age group. Using the Leslie matrix method to forecast the population far into the future will lead under certain weak conditions to a stable population, where the number of individuals in each age group grow at the same rate. The resulting stable age structure is independent of the initial situation.

Demographic forecasting using *stochastic microsimulation* is in principle the same as using the transition matrix approach. The difference lies in the calculation method used. Instead of matrix multiplication leading to transitions between groups of individuals, the stochastic microsimulation approach uses random drawings to simulate the transitions for all individuals in the initial population. The demographic system is also in this case interpreted as a discrete state space Markov process and the matrix **B** as a matrix of transition probabilities. These drawings are based on the same probabilities as the transition matrix approach and the expected results will be the same, but the realized results may differ because the number of drawings is finite.

To illustrate the stochastic microsimulation approach we interpret the elements of the transition matrix **B** as transition probabilities with discrete state space $[1, \cdots, K]$, where b_{ij} is, as before, the probability of entering state j given that state i was occupied at the previous point in time. On the basis of this probability we can calculate the corresponding conditional cumulative distribution function:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1K} \\ c_{21} & c_{22} & \cdots & c_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ c_{K1} & c_{K2} & \cdots & c_{KK} \end{bmatrix},$$

where $c_{jk} = \sum_{v=1}^{k} b_{jv}$. The element c_{jk} is the probability of entering one of the states $[1, \cdots, k]$ given that state j was occupied at the previous point in time. For each individual in the population at time t (or a sample of the total population at this time) we make a random draw from the uniform distribution $R(0,1)$. We let r_{it} denote the outcome of such a drawing for an individual i who at time t is in state j, and determine which c_{jk}'s it lies between. If for example we have that $c_{j3} < r_{it} < c_{j4}$ the individual goes from state j at time t to state 4 at time $t+1$. This type of drawing is done for all individuals in the population (or sample) and is commonly known as *Monte Carlo simulation.* We now let the variable I_{ijkt} be equal to 1 if and only if individual i makes the transition between state j at time t to state k at time $t+1$. This can be written as

$$I_{ijkt} = \begin{cases} 1 & \text{if } c_{j\,k-1} < r_{it} < c_{jk} \\ 0 & \text{otherwise} \end{cases} .$$

Taking the sum of this variable across all present states j and across all individuals i gives us a simulated estimate of the elements in the vector $\mathbf{M}_{t+1}$,

$$\tilde{m}_{k\,t+1} = \sum_{i}^{N}\sum_{j}^{K} I_{ijkt},$$

where $\tilde{m}_{k\,t+1}$ is an estimate of the expected number of individuals in state k at time $t+1$. The expected value of these m's will be the same as those that we get from the matrix transition method, but because of the use of random drawings they will vary between different runs of an unchanged model.

The sum $\sum_{j}^{K} I_{ijkt}$, which is equal to 1 if individual i is in state k at time $t+1$ and 0 otherwise, will be independently distributed of the same sum for other individuals. From the Central Limit Theorem we know that under certain regularity conditions a sum of independently distributed stochastic variables will converge to the normal distribution. We therefore have that if we use microsimulation on a fairly large initial population, the variable $\tilde{m}_{k\,t+1}$ will be normally distributed. From a series of microsimulations of an unchanged model (where inputs and parameters are not changed) we can therefore easily calculate the variance and a confidence interval for the sum $\tilde{m}_{k\,t+1}$. It is important to calculate such variances when using microsimulation models in order to assess how well the model performs. Large variances make a microsimulation model less useful than it otherwise would be, though uncertainty as to the expected values of the transition matrix **B** is usually much more important for the accuracy of a forecast than the variances resulting from the microsimulation procedure.

4 Why choose stochastic microsimulation?

The above discussion does not make it apparent why one might choose stochastic microsimulation instead of the transition matrix approach when making forecasts. In both cases one starts with data on the characteristics of the population in the initial period. These data will in the transition matrix approach consist of a matrix describing how many individuals have different combinations of attributes, while they in the microsimulation approach consist of a list of individuals denoting their attributes. The individuals in such a list are usually a sample of the total population being modeled, since it is impractical or impossible to do microsimulations on the whole population. If some of the initial data are not observed but imputed we speak of *synthetic data.*

The microsimulation approach leads to two extra sources of variance compared with the transition matrix approach. One source is the variance due to the use of Monte Carlo drawings and the other is the variance due to working with a sample of the population.

Microsimulation makes it possible to model a larger number of different types of events than is possible with the transition matrix approach. The reason for this is that in microsimulation models one only needs to keep track of a list of individuals with all relevant attributes. The number of elements this involves is simply the number of individuals, N, times the number of attributes considered, K. In the matrix transition approach one works with a matrix which size depends on the number of attributes, K, (for example age and education) and on the number of classes assigned to each attribute, s_k (for example the number of age groups or the number of educations). The number of elements this involves will be $\prod_{k=1}^{K} s_k$. This will be a much larger number than $N \cdot K$ when there is a large number of attributes. This also implies that it is easier to expand an already existing microsimulation model to include new attributes than it is to do so with a transition matrix model.

A main reason for using microsimulation models is that they directly give individual life histories. Individual life histories are for example important when working with pension benefits or doing distributional studies. It is also important that microsimulation automatically leads to the output of the model being consistent. Ensuring a consistent output can be a problem in aggregate models where there is interaction between individuals, such as marriage and household models. One ought, for example, to ensure that the number of males who marry is equal to the number of females who marry. In a micro model this can be taken care of relatively easily by always letting marriages take place between individuals. This ensures that the model results are internally consistent, but does not necessarily solve the problem of modeling the matching process in a theoretically consistent manner. The problem of matching couples when males and females each are exposed to different perceived opportunities has for example still not been theoretically modeled in a satisfactory manner. Individual life histories and consistency can be derived in other types of models, but with greater difficulty than in microsimulation models.

Microsimulation is an intuitive method where we follow each individual through his or her life cycle and it is relatively easy to change assumptions about the behavior of the micro units. This is important when doing experiments on the model.

Microsimulation models are also a natural framework for incorporating complex microeconometric relationships. For example behavioral dynamic relationships of the types discussed earlier in chapter 2 (caused by true state dependence or unobserved hetergeneity) can be fairly complex and can involve stochastic specifications which require simulation at the micro level. Suppose for example one has estimated, for given demographic characteristics, a probability distribution which reflects the effect unobserved heterogeneity has on the observed hours and wages of an individual. Using such a relationship in a demographic microsimulation model is fairly simple. The procedure can be as follows: First simulate the relevant demographic characteristics and then simulate labor market behavior by drawing hours worked and a corresponding wage rate from the estimated probability distribution. Such a simulation gives an example of how the estimated relationship behaves over time and can be used both for forecasting and to evaluate the relationship itself.

5 Discrete or continuous time

In discussing the above forecast models we have taken for granted that we use discrete time models. It would seem that working with continuous time models based on transition intensities would be preferable, since most of the events we study occur continuously. Stochastic microsimulation in continuous time could be done by drawing waiting times for the different events that can occur instead of drawing the transitions, as is done in discrete time models.

The choice between continuous and discrete time models is somewhat arbitrary. The choice is usually based on analytical tractability and data availability. The main reasons for using discrete time models are that they are perceived to be easier to program and use. Transition probabilities are functions of transition rates in continuous time, which are often referred to as *intensities*. The probability of being in one state at the beginning of a period and in another state at the end of the period may conceal many transitions between states which occur within the period. Estimating in discrete time may therefore disregard important structural features and cannot explicitly take into account competing risk to the same degree as a continuous time model. *Competing risk* occurs when there are many different possible causes of an event. Of course, if within-period transitions are not observed, the improvement gained by using a continuous time model may be small. In cases involving a small number of states it is often possible to calculate the relationship between probabilities and the underlying intensities using the Kolmogorov differential equations. The *Kolmogorov differential equations* give for a Markov process the relationship between the transition intensities and the transition probabilities. Except for in a few cases, it is impossible to obtain a solution of the Kolmogorov differential equations in closed form. The relationship discribed by the Kolmogorov differential equations is especially simple if the intensities are constant between periods. A problem with the parameters in a discrete time model is that they are not invariant to changes in the length of the period being studied.

Maybe the most important reason for formulating models in discrete time is that this makes it possible to do simulations one year at a time, which is computationally convenient. It also makes it easier to use aggregate variables to explain behavior (for example letting the unemployment rate in the previous period influence labor supply this period) and makes it easier to model interdependencies between individuals. Still, the development of object oriented languages and increased computing power might someday lead to continuous time models being used as much or more than discrete time models. Continuous time models have the advantage that competing risks are explicitly modeled. They also explicitly model the order in which events occur, reducing the need for simulating events in a fixed sequence.

6 A survey of some microsimulation models

During the last couple of decades the use of microsimulation models has become fairly widespread. The concepts of microsimulation were discussed as early as 1957 in Orcutt (1957), and the recent development of inexpensive and powerful computing technology has made it easier and cheaper to build such models. In the following I will describe very briefly some of the microsimulation models in use today. These models have in common that they simulate a large range of events. Most of these events are of a demographic nature, but other types of events are also simulated such as unemployment and earnings. The following description will mainly list the events simulated by each model, even though this does not adequately reveal the main differences between the models. They vary greatly in the sequence and manner of simulation, in the programming structure and language used, and in the type of data used. The models also vary in the extent to which they incorporate sophisticated econometric or statistical relationships. Orcutt et.al. (1986) and Mot (1991) give fairly comprehensive accounts of some of these models, in addition to discussing many other aspects of microsimulation modeling.

The earliest microsimulation models were of the static type associated with tax models. In the mid-1960's Orcutt developed such a model to study the U.S. welfare system. This model, RIM, applied eligibility criteria and benefit formulas to a representative sample of the population. Since then many such static models have been developed. Static microsimulation of taxes and benefits is now done on a regular basis in many countries, including the Scandinavian ones. Such models are in many of these countries important tools for formulating government policy. The appendix gives a brief review of the use of static and dynamic microsimulation models in Norway and of the institutional setting of the Norwegian models. A fairly detailed description of static microsimulation modeling is given in the microsimulation survey of Merz (1991).

One of the first dynamic microsimulation models to be developed was the Dynamic Simulation of Income Model (DYNASIM) developed by Orcutt and others at the Urban Institute between 1969 and 1976. The first version of DYNASIM simulated demographic behavior, labor force behavior, social security, other pensions, unemployment compensation, and welfare programs. Demographic behavior consisted of leaving home, death, birth, marriage, divorce, disability, education, and geographic mobility, while labor force behavior consisted of labor force participation, hours in the labor force, unemployment, and earnings. Later, a second version of the model was developed, DYNASIM2, which consists of two separate models. The Family and Earning History model simulates the demographic and labor force behavior covered in the earlier DYNASIM, but also includes welfare benefits, federal income, and payroll taxes. The resulting longitudinal life histories are used as input to the second model, the Jobs and Benefit History model. This second model simulates job change, industry, pension coverage, and social security and private pension benefits. A more detailed description of the DYNASIM2 model is given in Wertheimer II et.al. (1986).

A similar large scale microsimulation model, the Special Collaborative Programme 3 (Sfb 3), was developed during the late seventies and the early eighties in Germany. It consists of three versions, one version for dynamic cross-sectional simulation, one for dynamic longitudinal simulation (simulating a single cohort) and a static cross-section version. All three models simulate the same events. First the models simulate the demographic processes birth, death, marriage, divorce, household formation, and education. Then they simulate labor force participation, occupational status and industry, unemployment, earnings, capital income, transfer income, and taxes. At the end of a simulation the models determine disposable income, consumption, savings and wealth. The Sfb 3 models are described in more detail in Galler and Wagner (1986), and the static model is also described in Merz (1991). A description of the microsimulation of household formation and dissolution is given in Galler (1988).

The microsimulation model DEMOGEN is a Canadian model designed to produce a representative sample of complete family life histories. The model simulates death, marriage, husbands age, fertility, divorce, custody of children, child separation, remarriage, school leaving and educational attainment, labor force participation, and earnings. The model is a dynamic longitudinal microsimulation model. DEMOGEN is described in Wolfson (1988), which also discusses using the model to evaluate different types of pension reform in Canada.

Nelissen and Vossen (1989) describe a dynamic cross-sectional model, NEDYMAS, which simulates demographic events, social security benefits, labor market participation and income formation. The social security benefits which are modelled are old-age pensions, sickness benefits, family allowances, social welfare benefits, unemployment benefits, and social security payments. A labor market and income formation module simulates labor force participation, disability, unemployment, education, wage income, and taxes.

Two of the modules in NEDYMAS, the demographic and education modules, are discussed in more detail in Nelissen (1991). The demographic module simulates immigration, emigration, mortality, marriage, cohabitation, divorce, the breaking up of a cohabiting household, flows into and out of old peoples homes, fertility, and the splitting off of children. The education module simulates schooling for the population above 6 years of age. It takes into account that some students repeat a level of education and is based on transition probabilities that are influenced by social class. The model is based on an initial population of 10, 000 persons. Nelissen (1991) also presents standard deviations for the model.

In order to illustrate some of the aspects of microsimulation discussed above, the rest of this paper will discuss a microsimulation model called MOSART which we have developed for Norway. The model simulates events in a manner similar to the two modules discussed in Nelissen (1991). It does not cover as many variables as the models discussed above, but is characterized by a fairly large initial population (130, 000 individuals) and by being programmed in an object oriented language.

7 MOSART, a dynamic stochastic microsimulation model

During the last couple of years the Central Bureau of Statistics of Norway has developed a dynamic stochastic microsimulation model, MOSART[3], for forecasting population size and composition, labor force, and educational level. The present version of the model simulates life histories for a four percent sample of the Norwegian population. The life history events which are simulated are births, education, marriage, divorce, labor force participation, and death.

The events are simulated one year at a time and in a fixed order. In each year the model first simulates immigration and emigration, then mortality for all males, next demographic transitions among all women, then education for all individuals, and lastly labor force participation for all individuals.

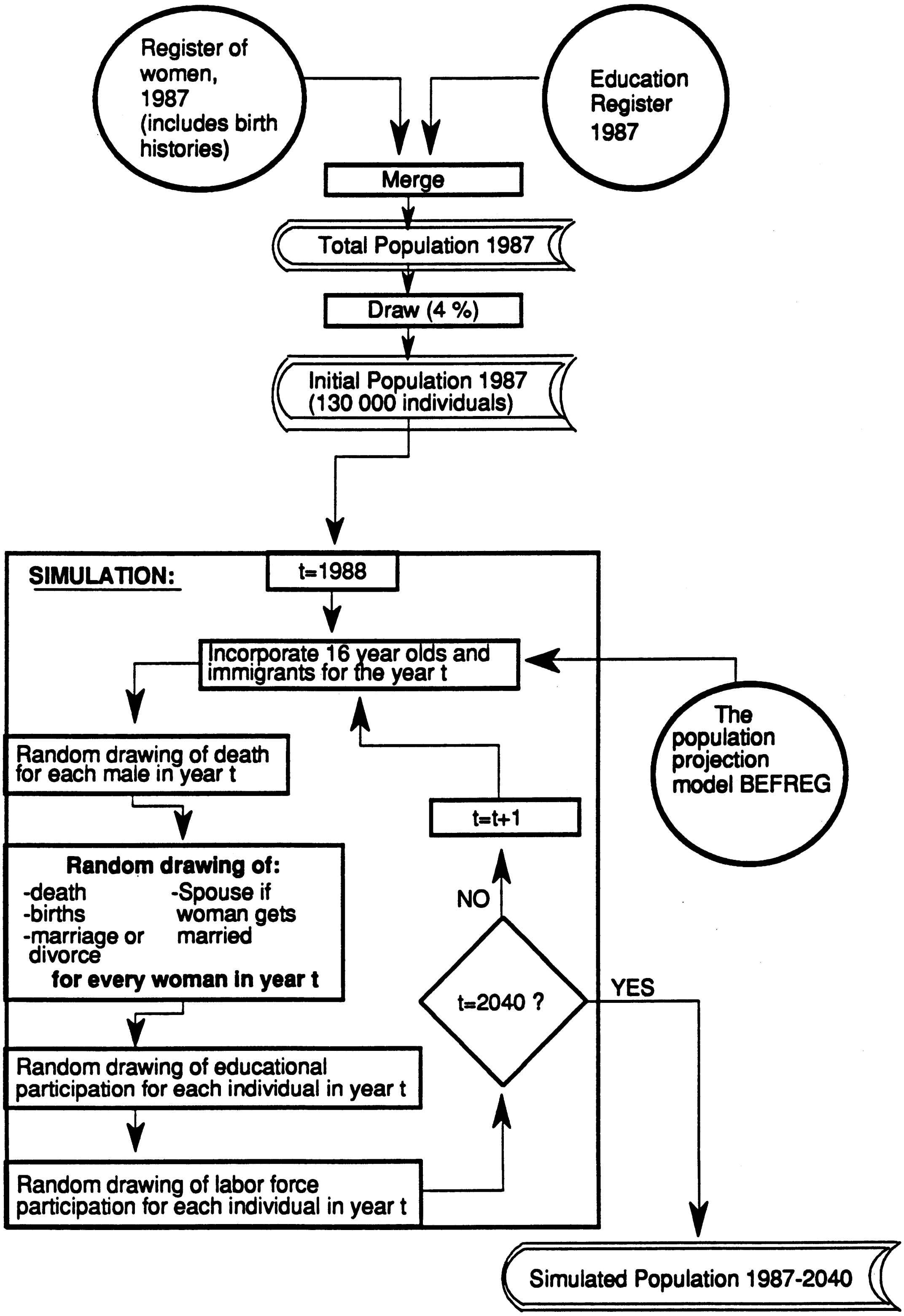
The model was originally implemented on an IBM-compatible main frame computer, but is now run on a UNIX-based SUN Sparc II work station with 16 RAM of internal memory. At present we are working on expanding the model to include the forecasting of social security pensions covering old age and disability and to incorporate household dynamics.

7.1 The initial population

A diagram describing the simulation procedure is given in figure 1. The first part of the diagram shows the merging of two data registers, a register of women with birth histories and an education register. Both registers cover the entire Norwegian population over the age of 15. The resulting merged register of the population over 15 years of age contains about 3.8 million records. Children 15 years and younger are included as attributes of their mothers but not as independent agents.

[3]MOSART is a Norwegian acronym for Model for Microsimulation of Schooling, Labor Supply and Pensions (*MO*del for mikrosimulering av *S*kolegang, *AR*beidstilbud og *T*rygd).

Figure 1. Main structure of MOSART (with exogenous input of 16 year olds)



The merged register file contains for each individual the characteristics age, sex, highest attained education, present educational activity and marital status. In addition it includes the identity of spouses and for women the number and age of children. It does not identify the father of the children.

The merging of the two registers above gives us complete knowledge of the distribution of all the model's variables in the base year, except for labor supply. From this merged data file we randomly draw a 4 percent sample of individuals 16 years and older constituting the initial population in the model. The size of this sample is about 130,000 individuals. Labor supply is simulated for each individual in the base year. A distinguishing feature of MOSART is that the only synthetic data in the initial population are the data for labor supply.

7.2 Immigration and new 16-year-olds

Before simulating the events that occur each year, it is necessary to supplement the population with new 16-year-olds and with immigrants. It is possible to determine the number of new 16- year-olds in two ways. One can either use the birth histories in the model, or one can get the number from the regional population model called BEFREG (see Central Bureau of Statistics (1986)). BEFREG is a transition matrix model which projects the population by age, sex, and region using regional death and fertility rates, and has during the last two decades been used to make official population projections for Norway. By letting the number of new 16-year-olds and net immigration be consistent with BEFREG, we get a projection which is compatible with these official population projections (MOSART uses death probabilities which are comparable with those used in BEFREG). In figure 1 and in our base line forecast we get the number of new 16-year-olds from BEFREG. When we later calculate standard deviations for the model, we determine the number of new 16-year-olds from the birth histories in the model instead of getting them from BEFREG.

MOSART models net immigration by sex and age instead of modeling gross flows. The total number of immigrants is specified exogenously in the model, and we use a very simple simulation procedure to divide into groups by age and sex. We also simulate women's birth histories and marriage among immigrants using probabilities which characterize the Norwegian population as a whole. This reflects the fact that immigrants in the model are assumed to exhibit behavior with the same probability structure as non-immigrants. This is done for simplicities sake since immigrants are very diverse and data on their behavior is limited. Educational level for immigrants is set to "unknown". If we are faced with net emigration we draw individuals from the sample who then are assumed to emigrate and therefore are removed from the model.

7.3 Death, birth and marriage

After we have supplemented the population with new 16-year-olds and immigrants we simulate demographic events. First, the model simulates death for all males using death probabilities that depend only on age. If a married man dies, his wife is made a widow. After this the model simulates the demographic events that occur for each woman; death, giving birth, and marital status. First, death is simulated in the same manner as for men. Giving birth is then simulated using probabilities which depend on the women's age, parity (the number of previous children), and the age of the youngest child. These fertility probabilities are independent of marital status, but unwed mothers are exposed to increased probabilities of marrying.

Simulation of marriage and divorce is female dominated in the sense that the model only uses the marriage and divorce probabilities for females. Marriage is simulated based on probabilities that depend on the female's age, whether she recently has born a child, and on her marital status (unmarried, divorced, or widowed). If a woman marries, the model draws the age of her spouse with probabilities depending on the woman's own age and then randomly draws a spouse among all unmarried males of this age. If a woman is drawn to divorce then her husband is automatically registered as divorced. We assume that it is only possible to change marital status once a year.

7.4 Education and labor force participation

MOSART simulates the number of students in close to 100 educational activities (24 types of education and 12 levels of education) and keeps track of the resulting educational attainment of the population. The simulation is done in two steps. The first step concerns whether an individual continues with an ongoing education or starts a new education. The probabilities used here depend on age, sex, highest attained education, and educational activity during the previous year. In the second stage the type of educational activity is drawn with probabilities that to a large degree depend on the same characteristics as above.

Labor force participation is drawn based on labor force participation ratios. This part of the model is a microsimulation variant of the headship rate model we discussed above, since it uses labor force participation ratios instead of transition probabilities. In MOSART we interpret the labor force participation ratios as estimates of the probability of being in the labor force. These ratios depend on age, sex, and education. In addition, for women they either depend on marital status or on the age of her youngest child. The base line forecast uses female labor participation probabilities which depend upon the age of the youngest child. The simulation of labor force participation takes into account that there is a large degree of state dependence in labor market transitions (Aaberge (1988)).

7.5 The computer program

Merging files, estimating probabilities, and drawing the initial population is done using SAS on an IBM-compatible mainframe. The simulation program itself is written in SIMULA[4] which is an object oriented language and is now run on a SUN Sparc II work station. Using an object oriented language makes it fairly easy to take into account interdependencies between individuals. An object oriented language defines objects and links them through the use of pointers. Two married individuals in our model are thereby two objects with two pointers pointed at each other. When something happens to one object the other object can immediately react. Such linkages are especially important when simulating complex interdependencies between individuals such as those encountered in household dynamics. Our ongoing work on a household model is based on two types of objects, individuals and households, where interdependencies are modeled through the use of a large set of pointers. The use of an object oriented language is one of the main features which distinguishes MOSART from other models. A more detailed description of the use of an object oriented language to model household formation is given in Andreassen, Spurkland, and Vogt (1992).

A simulation of 4 percent of the population from 1987 to 2040 takes about 4.5 hours of real time and about 3.6 hours of CPU-time. This includes the preparation of aggregate output tables which can be directly used in a spread sheet. During a simulation the model writes out all events to a sequential log file. After a simulation is done this file is sorted so that all records for each individual are grouped chronologically. The sorted file is read by a SAS program which produces a SAS data set of life histories. The sorting of the log file takes about a half hour, as does the SAS program (CPU-time and real time are in these cases fairly equal). From the SAS data set it is possible to produce detailed output tables.

7.6 Forecast confidence intervals

There are many factors which affect the reliability of forecasts made by a microsimulation model. The relationships in the model may often be misspecified. How serious

[4]SIMULA was developed at The Norwegian Computing Center as early as in 1967.

the consequences of misspecification are, depends on what variables one is looking at and what type of problem is being analyzed. The estimated relationships which are used in a model will, even if they are unbiased, have variances which affect the reliability of the forecasts. Such variances ought to be incorporated explicitly in a microsimulation model, but this can be difficult to do in practice.

Uncertainty with respect to the development in the model's exogenous variables will also lead to uncertainty about the model's forecasts. Scenario-based simulations can be a way of dealing with such uncertainty, but this requires that the scenarios one looks at are carefully chosen. Probably the most relevant scenario simulations are those that look at how sensitive a dependent variable is to changes in the independent variables. Later in the paper I will illustrate this approach by looking at whether fairly large increases in immigration together with smaller increases in labor force participation will reduce the tax burden of an increasingly elderly population.

All the types of uncertainty discussed above are common to most forecasting models and also constitute the most important factors affecting the reliability of microsimulation models. Two types of uncertainty which are more particular to microsimulation models are the uncertainty connected with simulating a sample of the population and the uncertainty which arises because of the Monte Carlo technique used in the model. The variances connected with using a small initial population can be fairly large, but can be reduced by stratifying or increasing the sample one uses.

The variances due to the Monte Carlo simulations done in the model are most often smaller than the other types of uncertainty discussed above, but since they only arise in microsimulation models I will discuss these in more detail below.

The variances due to the simulation process itself are in table 1 calculated on the basis of 10 simulations which have been done with the MOSART model. The model is the same in all the simulations, but different random numbers are used.

Table 1 gives means and standard deviations for different types of groups in the year 2020. The version of MOSART which is being used here recruits new

Table 1: Standard deviations for selected groups in 2020

	average number of persons in 2020	*percent stand. dev.*	*average change from last year*	*absolute stand. dev.*
population 67+	3819903	0.12	11898	1402
labor force 16-74				
total	2414763	0.13	1793	1512
men	1308060	0.18	2290	2564
women	1106703	0.14	108	1591
ages 16-19	108300	1.35	-890	1292
married couples	1009703	0.33	1890	1253
students 16+	399873	1.35	-2065	2849
nurses 16-74	101235	1.35	1075	422
40-54 year olds with 13+ years of educ.	343328	0.62	-925	973

individuals from among those that are born in the model (and not exogenously from BEFREG as is done in the base line forecast). To illustrate the variances generated by the Monte Carlo procedure, I have chosen to present standard deviations for the year 2020.

The left hand side of the table gives the mean and percent standard deviation for the number of persons in the different categories. As we see from the table, the standard deviations are very small compared with the mean.

The right hand side of the table gives the mean and absolute standard deviation for year-to-year changes. These are often important when using the model for short term forecasts. We see that the deviations are large compared to the average of the year- to-year changes. To deal with this problem, we have in the base line forecast used an average of three simulations, done on three different 4 percent samples of the population. This has led to a reduction in the standard deviations, but we still view these as being too large.

8 Results from the MOSART model

The following will present some results using the latest version of MOSART. It represents an update of the first MOSART results which were presented in Andreassen and Fredriksen (1991). The latest version of the model and the following results are described in more detail in Andreassen et.al. (1993). The following presentation aims to give an impression of how the model can be used in practice and to illustrate the use of scenario based microsimulations.

The base line forecast which will be presented below gets new 16- year-olds from the model BEFREG and uses labor force participation probabilities based on observations for 1991. The education transition probabilities are from 1987 but have been adjusted mechanically (all the probabilities have been increased proportionally) so that the model simulates the observed number of students in 1991.

Figure 2. Population 16-74 years of age
Base line forecast

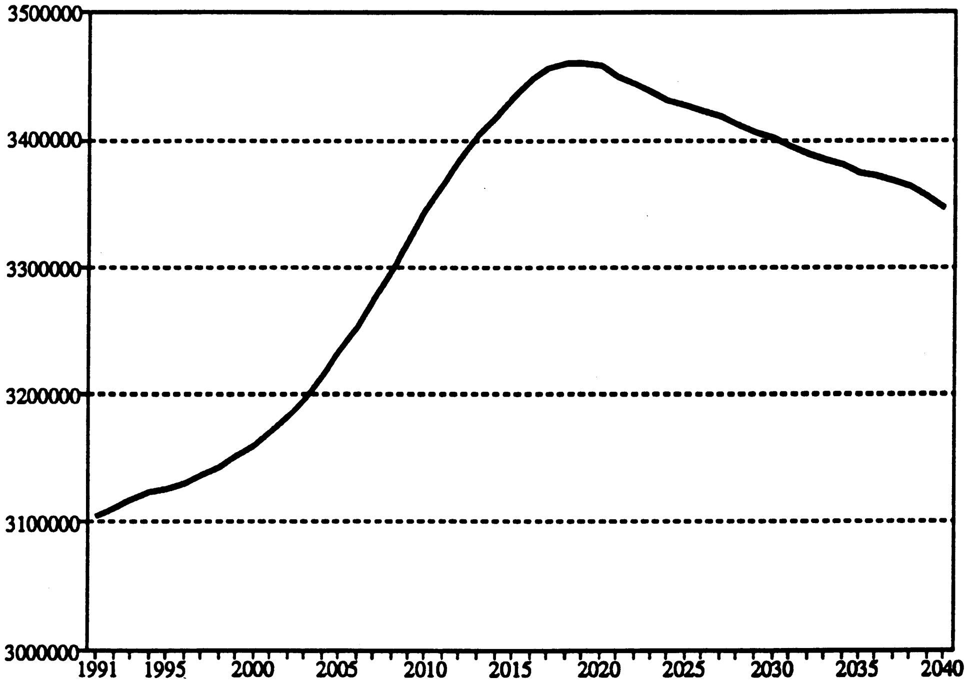
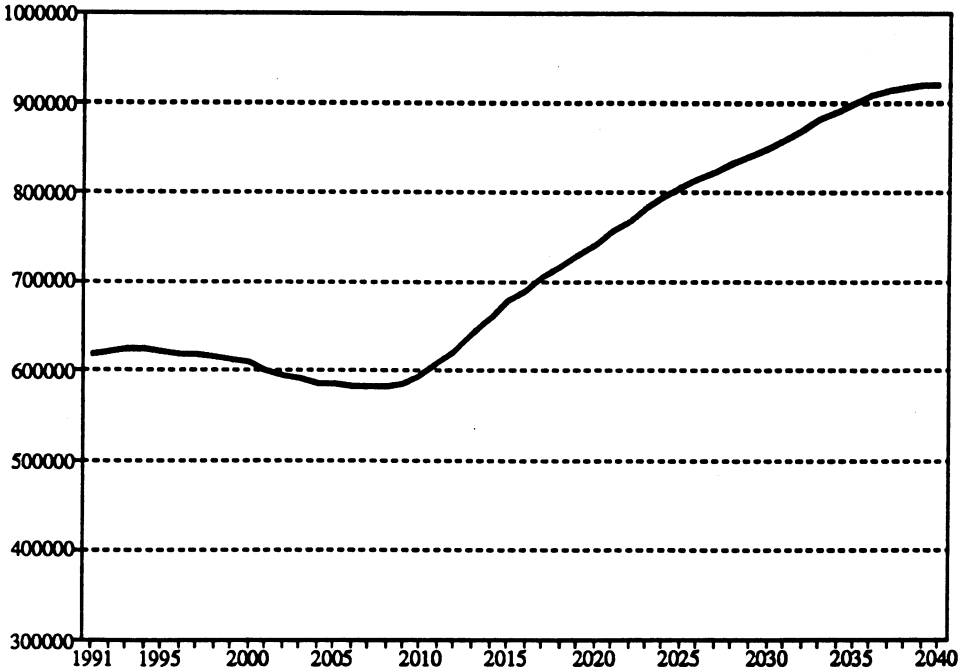


Figure 3. Population 67 years and older
Base line forecast



Female labor force participation ratios depend on the age of the youngest child and net immigration is assumed to be 5000 a year. Fertility and mortality have been estimated on 1989 data, which imply a total fertility rate of 1.8, while marriage and divorce probabilities are from 1984. The base line forecast consists of an average of three simulations, where each simulation is based on a different initial population. The simulations assume constant transition probabilities.

8.1 The population and the labor force

The Norwegian population of working age, 16-74 years of age, will in the base line forecast increase by almost 350,000 persons from 3.09 million in 1991 to 3.46 million in 2018 and then begin to fall during the rest of the period. This development is shown in figure 2. The population of retirement age, 67 years and older, will in the base line forecast fall slightly during the next 20 years, but then begin to rise fairly steeply, as shown in figure 3. From 2010 to 2040 the population 67 years and older will increase by more than 50 percent. Such an increase in the proportion of the elderly in the population is a development which Norway shares with most other industrialized countries.

Later in the paper I will discuss in more detail the consequences the above development may have for the burden of old-age pensions on the labor force and to what degree this burden can be alleviated through increased immigration. The increase in elderly will also increase the need for health care and old-age homes, but this will not be considered in the following.

Figure 4 shows the base line forecast for the total labor force (for the years 1987 to 1991 the actual observations are shown). We see that the labor force will increase by almost 300,000 persons during the next 30 years, reaching its highest point in 2019, and thereafter declining. In addition to the base line forecast the two figures also include a projection where the labor force participation rates and the original education probabilities for 1987 are used.

Figure 4. Labor force projection.
16-74 years of age

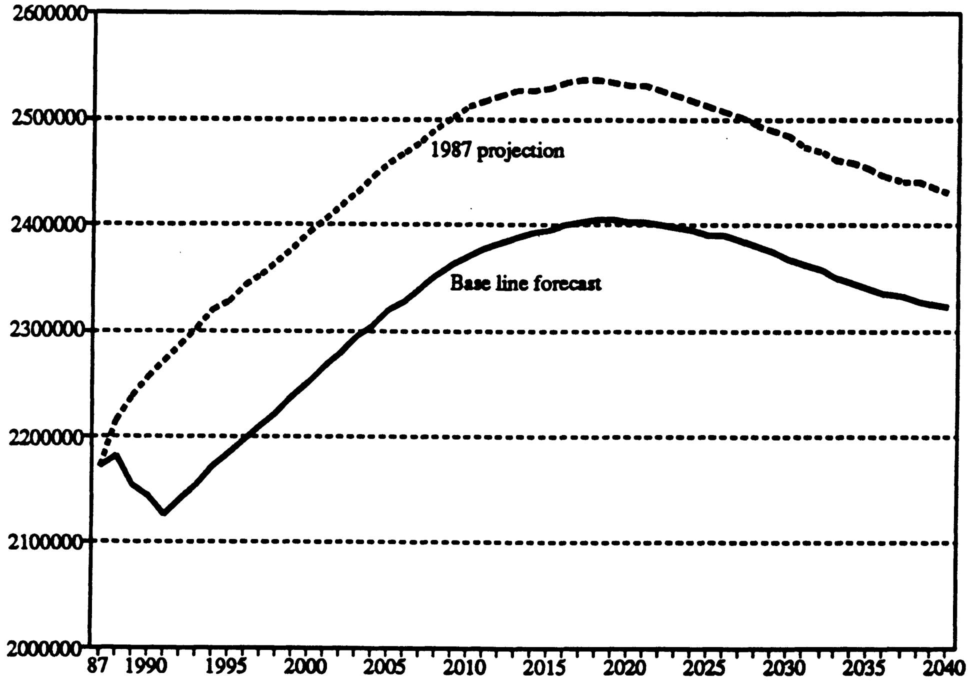
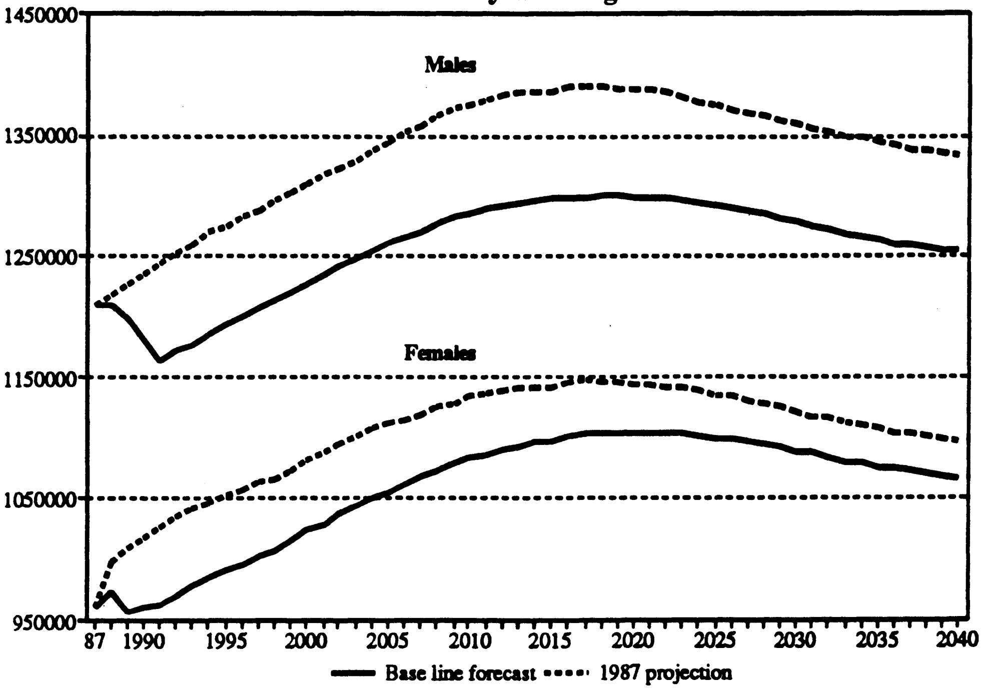


Figure 5. Male and female labor force projection
16-74 years of age



After increasing for many years the labor force participation rates peaked in 1987 in a situation with a strong economy and low unemployment. Since then unemployment has grown and labor force participation rates have fallen. The greatest fall has been among those between 16 and 20 years of age and among men over 55 years of age. In figure 5 we see this illustrated by a larger fall in male participation ratios than in female ones. Teenagers have to an increasing extent been going to school, while men over 55 are to a greater degree leaving the labor force due to disability.

Figure 4 shows that the change in labor force participation rates which has occurred during the last four years has decreased the labor force by more than a 100,000 individuals compared with the 1987 projections. Figure 4 reveals one of the drawbacks of using constant labor force participation rates. The model is not capable of modeling the decline in participation rates from 1987 to 1991. The model's projections only reflect demographic changes and do not take into account the effect the business cycle has on labor force participation. This can be accepted if one views the model mainly as a long term demographic model, but is a serious drawback when making short term forecasts. The problem can partly be alleviated by making many simulations under different assumptions about the level of the labor force participation ratios.

8.2 Students and achieved levels of education

The base line forecast uses education transition probabilities which have been estimated for 1987 and then been adjusted upwards so that the simulations give the observed number of students in the year 1991. Figure 6 shows the number of students during the next 50 years. It includes both the base line forecast and a simulation showing the number of students when the education probabilities remain unadjusted. The difference is approximately 75,000 individuals. This reflects that the number of students has increased dramatically during the last four years. This increase is thought to be closely related to the rising unemployment and falling labor force participation ratios which have occurred during the period.

Figure 6. Students 16 years and older

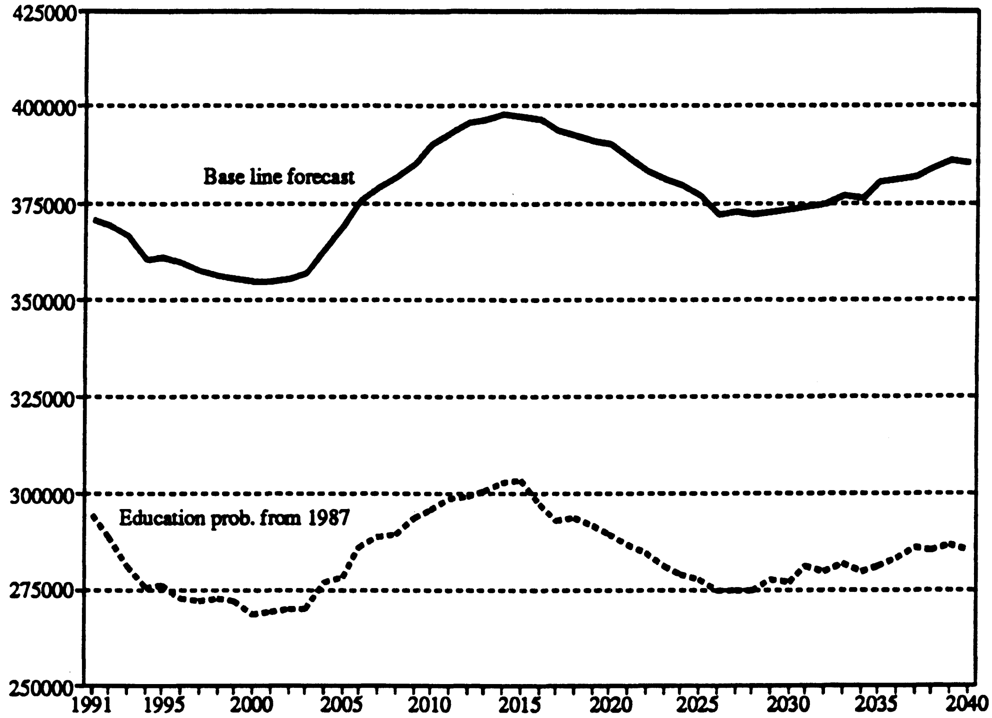
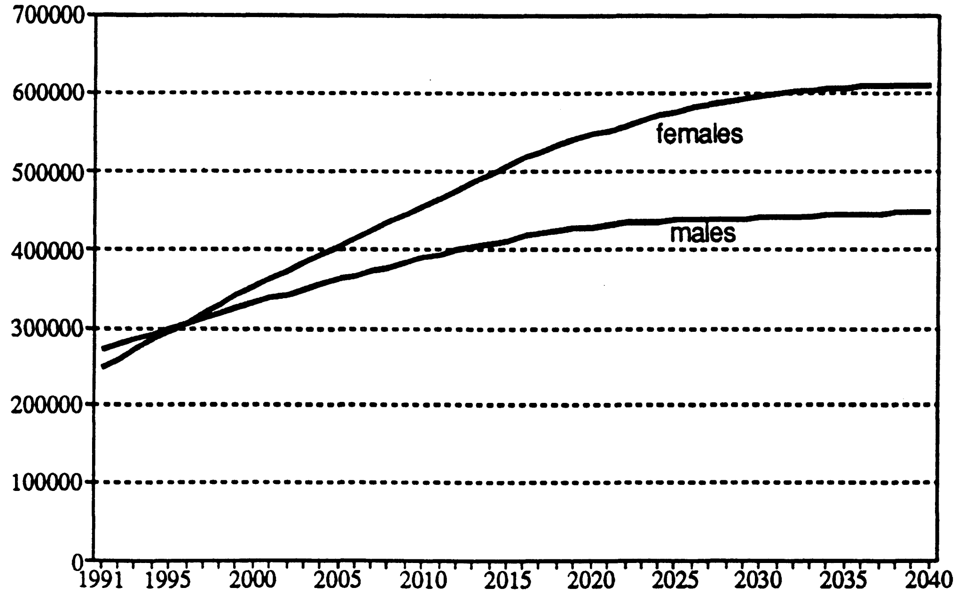


Figure 7. The population 16-74 years old with 13 years or more of education



The education transition probabilities in the model are assumed constant throughout the forecast period. During the last four years the government has increased the capacity of the educational system in an attempt to avoid a high unemployment rate among the young. The education transition probabilities simultaneously reflect both the wishes of the different agents for different types of education and the capacity constraints in the educational system. It is not possible in the present model to distinguish separately the effect of changing capacity from changes in the desire for different types of education. From figure 6 we see that if the transition probabilities remain constant the increase in students will not continue and we will experience a falling number of students during the next 10 to 15 years.

The MOSART model keeps track of the level of education achieved by all the persons in the model, and from figure 7 we see that females are set to overtake men in the length of education. Around 1995 more women than men will have 13 years or more of education. If we had looked at fields of study we would have seen that gender differences remain. The development in figure 7 partly reflects the fact that many male dominated technical educations are fairly short while educations within health care and teaching have been increasing in length. Simulations with MOSART also show that during the whole simulation period the number of males with 17 or more years of study will stay higher than the corresponding number of females.

8.3 The future cost of publicly provided old-age pensions

In 1967 Norway introduced a national public pension scheme which was such that all citizens each year were to receive entitlements to publicly provided old-age pensions based on their labor income during that year. To receive a full pension, it is necessary to have earned entitlements for 40 years. The resulting pension is then calculated on the basis of the twenty years with highest income. If a person has earned entitlements for less than 40 years she receives a partial pension.

**Figure 8. Labor force projection.
16-74 years of age**

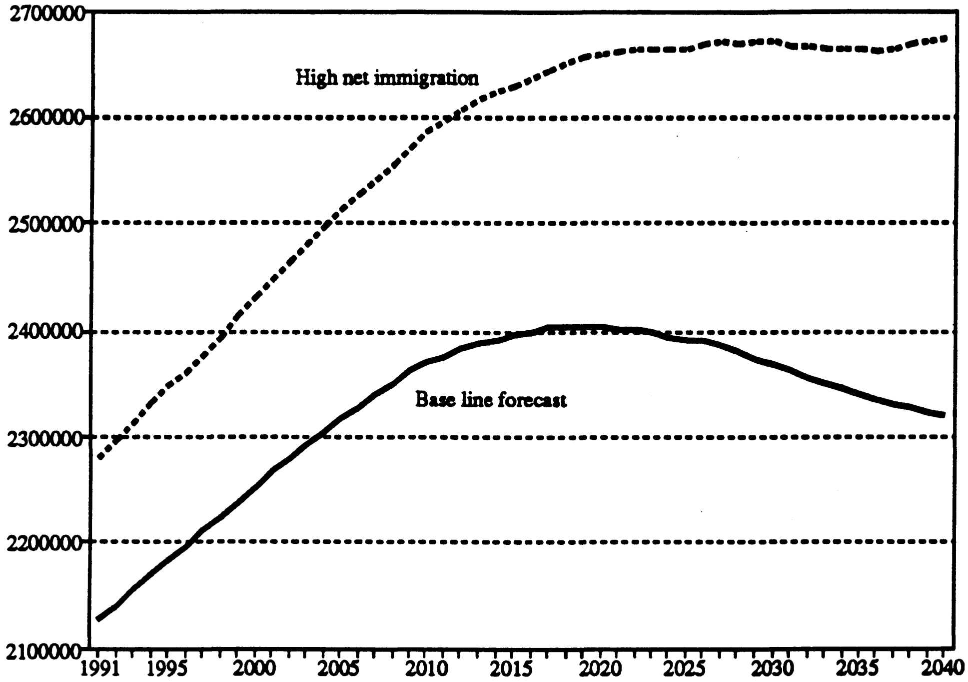
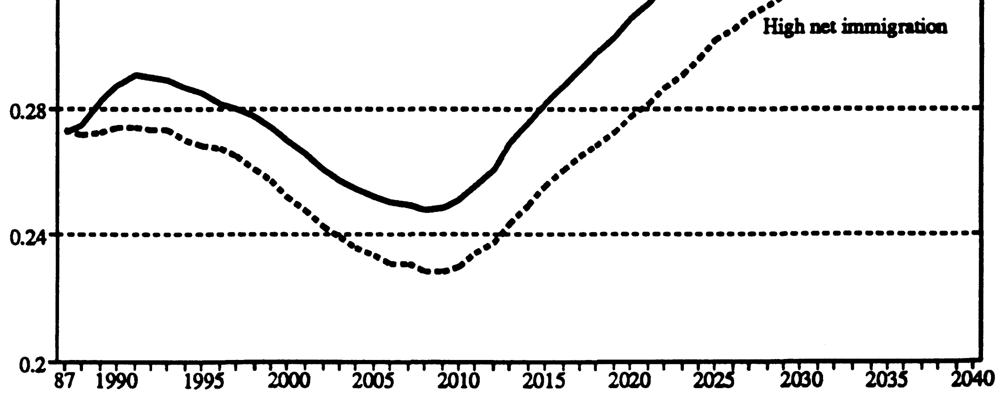


**Figure 9. The number of elderly
(persons 67 years and older) divided by the labor force**



When the system was introduced in 1967, many were too old to be able to earn entitlements over a full period of 40 years. For these there have been devised special compensation rules. It is first in 2007 that there will be retirees who have had a opportunity to earn full entitlements.

Entitlements and pensions are calculated using a basic amount (grunnbeløp) set by Parliament each year. These calculations have the interesting property that an increase in the basic amount increases current pensions while decreasing current earned entitlements and vice versa. Calculations done in Andreassen et.al (1988) reckon that average real entitlements will keep increasing by at least 1 percent a year until the year 2020, assuming that the basic amount follows the inflation rate. This increase is the result of an increasing number of persons having been able to earn full entitlements during their working careers.

The Norwegian pension system is mainly a pay-as-you-go scheme, where current old-age pensions are covered by current taxes. There is a certain building of funds but this is relatively insignificant. From the above discussion of future population developments one would expect that as the population gets older, the Norwegian public pension system will become an increasing burden on those in the labor force.

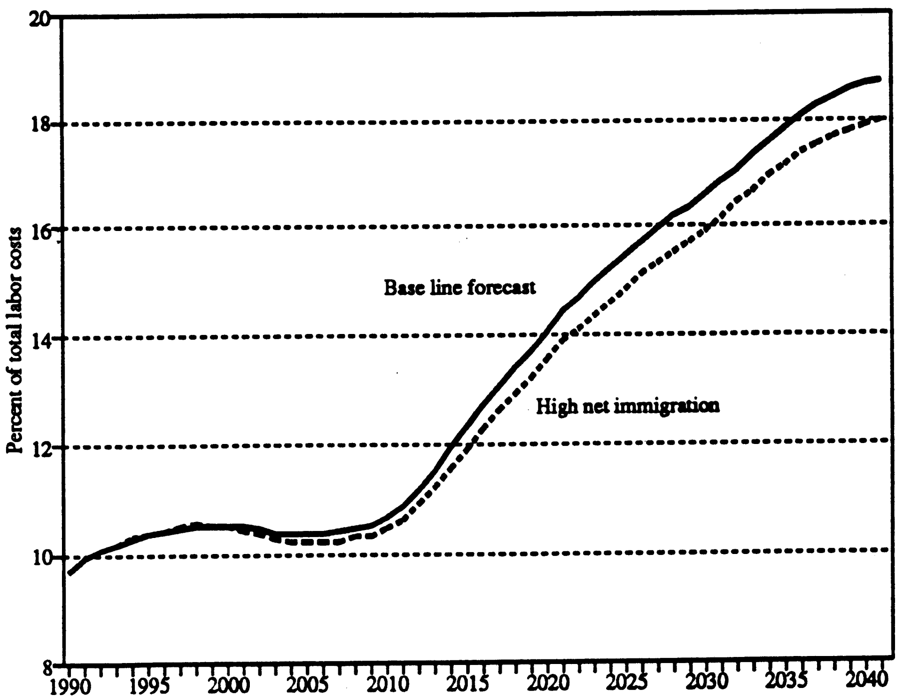
In the following we will use some simple assumptions and simulations done with MOSART to see whether a large increase in immigration can alleviate this problem. Figure 8 shows two labor force projections. The lowest is the base line forecast discussed above, and the other is a simulation which assumes that net immigration will be 7,000 individuals higher than assumed in the base line forecast. In addition to assuming a net yearly immigration of 12,000, the high net immigration alternative also assumes that labor force participation ratios and education transition probabilities will be equal to those observed in 1987. We see from figure 8 that this leads to a labor force which will increase throughout the next 50 years and which will in 2040 be more than 300,000 individuals larger than in the base line scenario. Figure 9 shows how the dependency ratio for elderly will be in the two alternatives. The high net immigration alternative leads to a lower dependency ratio than the base line forecast and the difference between the two increases as time passes.

The present version of MOSART does not simulate pension entitlements nor benefits, though we are at present working on extending the model to include this as well as disability pensions. Even so, it is possible to make rough calculations of the tax burden of publicly provided old-age pensions in the two scenarios described above. We assume that:

1. Economic growth will increase the basic amount and wage income proportionally. Pension entitlements will still increase at a slightly higher rate than wages, reflecting the increase in average entitlements calculated in Andreassen et.al. (1988), which is due to an increasing number of people earning entitlements over the whole span of their working career.

2. Pensioners face the same tax percentages on their pensions as laborers do on labor income.

Figure 10. Tax burden on the employed.



Assumption 1 enables us to abstract from any economic growth, since economic growth will according to this assumption increase pensions and wage income proportionally and therefore not affect the relative share. Assumption 2 ignores the fact that the taxes levied on pensions are lower than those levied on labor income. This will in isolation lead us to underestimate the burden faced by the employed.

Figure 10 shows the resulting tax burden on the employed. It shows the total amount of after tax pensions as a percentage of total labor costs. This is of course not an actual tax, but only a simplified illustration of the tax burden. From figure 10 we see that the tax burden is fairly constant until 2010, when it starts increasing. In 2040 the tax burden will have almost doubled in the base line forecast. The most interesting thing to notice is that the high immigration alternative does not lead to a large decrease in the tax burden. From this we can tentatively conclude that the tax burden of publicly provided old-age pensions seems robust to increases in labor force participation ratios together with large increases in immigration.

The calculation above is a very rough approximation to what might be the consequences of an increase in the proportion elderly in the population. Much more accurate calculations can be made when the MOSART model explicitly incorporates pensions and pension entitlements.

9 Summary

This paper has aimed at giving an introduction to some of the main features of demographic microsimulation modeling. Even though not all the variables discussed above are demographic, most of the simulations are of a demographic nature and the models which have been discussed are mainly driven by demographic factors. In general microsimulation is used when a model is so complex that it is impractical or impossible to derive results analytically and when comparable aggregate models are not detailed enough. The first part of the paper compared microsimulation with other demographic forecasting techniques and considered the main reasons for choosing microsimulation:

1. Microsimulation makes it possible to simulate a large number of events and thereby also making it easy to expand an existing model.

2. Microsimulation produces individual life histories which are for example needed in forecasting pension benefits and entitlements.

3. Microsimulation models makes it relatively easy to model interdependencies between individuals and to ensure consistent output from a model.

4. It is relatively easy to change assumptions in a microsimulation model, making it possible to do many types of experiments and to do scenario based studies.

5. Microsimulation models are a natural framework for incorporating microeconometric relationships which take into account unobserved heterogeneity.

6. Microsimulation is an intuitive approach which is easy to understand.

Microsimulation models also have disadvantages. They are expensive to develop and maintain and require large data bases. In addition the use of Monte Carlo simulations generates variances in the model's projections.

The paper briefly surveyed some of the microsimulation models which have been developed. In the last part of the paper the Norwegian microsimulation model MOSART was described and some results were presented. The main features which distinguish MOSART from other microsimulation models is that it is based on a large initial population and that the model is programmed in an object oriented language. The model is still being developed and will in the future include household dynamics and the simulation of old-age entitlements and pensions and disability pensions. The modeling of household dynamics will extend the model to include cohabitation of unmarried couples, children moving from home, and elderly moving to institutions.

References

Aaberge, R. (1988): Bruttostrømmer i på arbeidsmarkedet. *Naut-rapport* 1988:3, Nordisk Ministerråd, Copenhagen.

Andreassen, L., T. Andreassen, D. Fredriksen, G. Spurkland, and Y. Vogt (1993): Framskriving av arbeidsstyrke og utdanning. Mikrosimuleringsmodellen MOSART 1. *Rapporter* 93/6, Central Bureau of Statistics, Oslo.

Andreassen, L., C. Koren, J.G. de Leon and O. Ljones (1988): Inntektsoverføringer mellom aldersgrupper og befolkningsutviklingen. Appendix 4, *NOU* 1988:21, 253-291, Universitetsforlaget, Oslo.

Andreassen, L. and D. Fredriksen (1991): MOSART - En Mikrosimuleringsmodell for Utdanning og Arbeidstyrke. *Økonomisk Analyser* 2, 10-21, Central Bureau of Statistics, Oslo.

Andreassen, T., G. Spurkland, and Y. Vogt (1992): *MOSART, a microsimulation model.* Paper prepared for the 1992 Conference on Computing for the Social Sciences, May 4-7, 1992, University of Michigan Ann Arbor. Central Bureau of Statistics, Oslo.

Box, G.E.P., W.G. Hunter, and J.S. Hunter (1978): *Statistics for experimenters.* Wiley, New York.

Central Bureau of Statistics (1986): *Framskriving av Folkemengden 1990-2050*, NOS B 583, Central Bureau of Statistics, Oslo.

Galler, H.P. (1988): Microsimulation of Household Formation and Dissolution, in N. Keilman, A. Kuijsten and Ad Vossen (eds): *Modelling Household Formation and Dissolution.* Clarendon Press, Oxford, 139-159.

Galler, H.P. and G. Wagner (1986): The Microsimulation Model of the Sfb 3 for the Analysis of Economic and Social Policies, in G. Orcutt, J. Merz and H. Quinke (eds): *Microanalytic Simulation Models to Support Social and Financial Policy.* North-Holland, Amsterdam, 227-247.

Leslie, P.H. (1945): Some Further Notes on the Use of Matrices in Population Mathematics. *Biometrica* 33, 213-245.

Mertz, J. (1991): Microsimulation - A Survey of Principles, Developments and Applications. *International Journal of Forecasting* 7, 77-104.

Mot, E.S. (1991): *Survey of microsimulation models: Inventory and recommendations.* Ministerie van Sociale Zaken en Werkgelegenheid, The Hague.

Nelissen, J.H.M. (1991): Household and Education Projections by Means of a Microsimulation Model. *Economic Modelling* 9, 480-510.

Nelissen, J.H.M. and A.P. Vossen (1989): Projecting Household Dynamics - A Senario-Based Microsimulation Approach. *European Journal of Population* 5, 253-279.

Orcutt, G.H. (1957): A New Type of Socio-Economic System. *Review of Economics and Statistics* 39, 116-123.

Orcutt, G.H., J. Merz and H. Quinke (eds) (1986): *Microanalytic Simulation Models to Support Social and Financial Policy.* North-Holland, Amsterdam, 227-247.

Wertheimer II, R., S.R. Zedlewski, J. Anderson and K. Moore (1986): DYNASIM in Comparison with Other Microsimulation Models, in G. Orcutt, J. Merz and H. Quinke (eds): *Microanalytic Simulation Models to Support Social and Financial Policy.* North-Holland, Amsterdam, 227-247.

Wolfson, M.C. (1988): Homemaker Pensions and Lifetime Redistribution. *Review of Income and Wealth* 34, 221-250.

A APPENDIX

The use of microsimulation modelling in Norway

The following is a short description of the microsimulation models which are used in Norway[5]. In addition to describing the models, an attempt is made to describe how the models are used in the policy process.

A.1 Tax models

Microsimulation models are widely used in formulating tax policy in Norway. The most widely used microsimulation model is LOTTE. It is used by the Ministry of Finance and by the political parties represented in Parliament (the Storting). This arrangement whereby opposition parties in Parliament have direct access to the same tax models as the government is a unique feature of the Norwegian political system. The fact that all political parties use the same tax models to describe the consequences of their tax proposals makes it possible for the political debate to be concentrated on the tax proposals themselves rather than on whether the different parties have used reliable models. It must be remarked that this does not eliminate differences in opinion on what the consequences of a given tax policy are. Such disagreements often concern areas which are not covered by the tax models, such as behavioral and long term effects.

The tax models are developed and run by the Research Department of the Central Bureau of Statistics, which is perceived as a neutral agent by the political parties. The main purposes of the models are to evaluate different tax policies with respect to:

1. the increase or loss of revenue generated by a given tax policy;

[5]This description of the use of microsimulation models in Norway was originally prepared for the OECD Working Party on Social Policy.

2. the impact a given tax policy has on after-tax income in different income brackets for different types of households;

3. the general distributional consequences of a given tax policy.

The static nonstochastic microsimulation model LOTTE is used to evaluate tax policies with respect to 1 and 3 above. It can also do rough evaluations with respect to 2, but for this purpose the tax model ODIN, which is discussed below, is more suitable. The first version of LOTTE was constructed in 1969/1970 at the Central Bureau of Statistics. At present there is no simulation of behavioral responses to altered programs or tax incentives. The model is based on a database consisting of a sample of actual tax returns, and it estimates both the revenue and the distributional effects of changes in taxes and benefits. A simulation run gives the total amount of taxes paid and disposable income both for individuals and households. The model simulates both average tax rates and marginal tax rates. At present the model does not include public pension benefits or all benefits accruing to households with children, but we aim to include such benefits in the near future.

LOTTE is not only used to make forecasts, but also to analyze the effects of past tax policies. Such analysis is often carried out by the Central Bureau of Statistics itself. At present the Research Department of the Central Bureau of Statistics is working on a project (in close cooperation with other academic institutions and the Ministry of Finance) to analyze a major tax reform which was recently passed by Parliament.

In addition to LOTTE there is another tax model, SKATT, which is used to forecast revenues. It is not based on microsimulation but calculates tax revenue on the basis of average income in different income brackets estimated from tax statistics covering all tax payers. As a very simple, quick and fairly accurate model, it is a useful supplement to the LOTTE model. It also has the advantage that it can be run directly by the Ministry of Finance.

A third tax model, ODIN, is strictly a law simulation model and is not based on microsimulation. It is used to generate tables illustrating the consequences a tax proposal has on different groups in different income brackets (evaluation 2). ODIN calculates benefits, taxes and disposable (after-tax) income for different types of households. The households can be differentiated by income and socioeconomic variables. It covers a much wider range of benefits than LOTTE, including social security. This model is mainly used by the Ministry of Finance, but is also used by other government ministries and by special- interest groups outside of Parliament, such as those representing the handicapped.

The Ministry of Finance has received versions of the tax models SKATT and ODIN which they run themselves, while LOTTE is run only by the Central Bureau of Statistics. At present there are plans to make it possible for the Ministry of Finance to also run LOTTE. The tax models are continuously used by the Ministry of Finance in their work on government policy, while the opposition parties use the models mainly in the period just before the government budget is to be discussed in Parliament.

In conclusion the microsimulation model LOTTE plays an important part in the formulation of tax policy in Norway. It is run by the Central Bureau of Statistics for both the government and opposition parties. It is supplemented by other non-microsimulation models which can be run directly by the Ministry of Finance.

A.2 Demographic and socioeconomic forecasting using microsimulation

During the last couple of years the Research Department of the Central Bureau of Statistics has developed a dynamic stochastic microsimulation model, MOSART, for forecasting population size and composition, labor force, and educational level and activity. The present version of the model simulates life histories for a four percent sample of the Norwegian population. The life history events which are simulated are births, education, marriage, divorce, labor force participation, and death. The events are simulated one year at a time and in a fixed order. This new model replaces an earlier headship rate model, MATAUK.

The great advantage of demographic microsimulation compared to more traditional demographic models is that it makes it possible to simultaneously forecast the population over a very large set of variables. We are now working on expanding the model to include old age pensions, disability pensions, and to incorporate household supply dynamics. Microsimulation is a powerful tool in working with these topics. Pension entitlements depend on life histories, and especially work histories. These can be taken fully into account in the MOSART model, since it simulates such life (and work) histories.

Household dynamics consist of complex dependencies between different individuals (they can be parents, spouses, and children at the same time). Microsimulation makes it possible to take into account such dependencies at the individual level while automatically ensuring that consistency is acquired (that there for example are as many married men as there are married women).

MOSART makes labor force projections which are widely used in Norway, both for short term and long term economic analysis, even though the demographic nature of the model makes it most suitable for long term analysis. MOSART is mainly used by the Ministry of Finance, but is also used by other government departments. It is run only by the Central Bureau of Statistics and making new simulations can take from a couple of days to a couple of weeks according to the complexity of the simulations.

The stochastic nature of MOSART is such that the microsimulation process generates a certain variance in the output. The resulting variance is small compared to the levels being simulated but can be fairly large compared to year-to-year changes. This has lead to problems for some users of the model, especially those concerned with short term forecasts. We have therefore worked on reducing the variances in the model.

A.3 Analysis of public assistance programs

The Ministry of Social Affairs has started an ambitious data collection project called KIRUT, which follows ten percent of the population over a five year period (1989 - 1993), recording all types of public assistance the sample receives during these years. The KIRUT data base will contain the background variables: sex, age, marital status, educational status, number of children and where each individual lives. It will then, for all individuals in the sample, register the date of every change in:

1. employment/unemployment status;
2. the participation in active public assistance programs such as training schemes;
3. the take-up of disability pension;
4. the take-up of old-age pension;
5. the take-up of other types of social security benefits.

This data collection is done independently of, but in cooperation with the Central Bureau of Statistics.

The National Insurance Administration (Rikstrygdeverket), which administers the social security system in Norway, is financing the completion of a general microsimulation tool called EVENT, which is to be used on the KIRUT database. It is a general computer program for use on personal computers which will be able to analyze event history data (estimate hazard functions) and be able to do stochastic microsimulations on event history models. The National Insurance Administration wishes to use the EVENT program to estimate a model of the transitions between the different public assistance programs and to do microsimulations with this model. These microsimulation experiments will study the effect of changing the different public assistance programs. It is of special interest to find out which programs help individuals get off public welfare and which programs lock them into public assistance for the rest of their lives. These microsimulations are to be done at the National Insurance Administration and will not involve the Central Bureau of Statistics.

A.4 Microeconometric simulation

A different type of microsimulation is that which is done in connection with the estimation of microeconometric relationships. During the last decade the Research Department has worked on establishing econometric models for labor supply. This research has since 1984 focused on the joint supply decisions of married couples under a fairly detailed representation of the tax system. The tax rules imply non-convex budget sets, hence traditional marginal calculus cannot be applied. The Research Department has developed an econometric model which leads to the household's labor supply being given as a probability distribution depending on demographic characteristics and parameters related both to preferences and to the distribution of available hours and wage rates. This method is an alternative to the traditional Hausmann approach.

After the model has been estimated it is possible to run static stochastic simulation experiments on it. The model specifies a labor supply probability for each individual. The simulation procedure draws outcomes for all individuals in the sample. Typical experiments are:

1. the effects on household consumption and on the husband and wife's joint labor supply of switching from joint taxation to separate taxation;
2. the effects on consumption and labor supply of changing the tax schedule subject to the restriction that the total amount of taxes is kept constant;
3. the effects on the income distribution of changes in the tax schedule.

The results from these experiments indicate among other things that the abolition of joint taxation for married couples would lead to a fairly large increase in the labor supply of married women. These results were presented both to Parliamentary subcommittees and to the Ministry of Finance. The theoretical model has also been estimated on data for Sweden, Finland, Germany, France, Italy, and Peru.

ISSUED IN THE SERIES DISCUSSION PAPER

No. 1 *I. Aslaksen and O. Bjerkholt (1985)*: Certainty Equivalence Procedures in the Macroeconomic Planning of an Oil Economy.

No. 3 *E. Biørn (1985)*: On the Prediction of Population Totals from Sample surveys Based on Rotating Panels.

No. 4 *P. Frenger (1985)*: A Short Run Dynamic Equilibrium Model of the Norwegian Production Sectors.

No. 5 *I. Aslaksen and O. Bjerkholt (1985)*: Certainty Equivalence Procedures in Decision-Making under Uncertainty: An Empirical Application.

No. 6 *E. Biørn (1985)*: Depreciation Profiles and the User Cost of Capital.

No. 7 *P. Frenger (1985)*: A Directional Shadow Elasticity of Substitution.

No. 8 *S. Longva, L. Lorentsen and Ø. Olsen (1985)*: The Multi-Sectoral Model MSG-4, Formal Structure and Empirical Characteristics.

No. 9 *J. Fagerberg and G. Sollie (1985)*: The Method of Constant Market Shares Revisited.

No. 10 *E. Biørn (1985)*: Specification of Consumer Demand Models with Stochastic Elements in the Utility Function and the first Order Conditions.

No. 11 *E. Biørn, E. Holmøy and Ø. Olsen (1985)*: Gross and Net Capital, Productivity and the form of the Survival Function. Some Norwegian Evidence.

No. 12 *J.K. Dagsvik (1985)*: Markov Chains Generated by Maximizing Components of Multidimensional Extremal Processes.

No. 13 *E. Biørn, M. Jensen and M. Reymert (1985)*: KVARTS - A Quarterly Model of the Norwegian Economy.

No. 14 *R. Aaberge (1986)*: On the Problem of Measuring Inequality.

No. 15 *A.-M. Jensen and T. Schweder (1986)*: The Engine of Fertility - Influenced by Interbirth Employment.

No. 16 *E. Biørn (1986)*: Energy Price Changes, and Induced Scrapping and Revaluation of Capital - A Putty-Clay Model.

No. 17 *E. Biørn and P. Frenger (1986)*: Expectations, Substitution, and Scrapping in a Putty-Clay Model.

No. 18 *R. Bergan, Å. Cappelen, S. Longva and N.M. Stølen (1986)*: MODAG A - A Medium Term Annual Macroeconomic Model of the Norwegian Economy.

No. 19 *E. Biørn and H. Olsen (1986)*: A Generalized Single Equation Error Correction Model and its Application to Quarterly Data.

No. 20 *K.H. Alfsen, D.A. Hanson and S. Glomsrød (1986)*: Direct and Indirect Effects of reducing SO_2 Emissions: Experimental Calculations of the MSG-4E Model.

No. 21 *J.K. Dagsvik (1987)*: Econometric Analysis of Labor Supply in a Life Cycle Context with Uncertainty.

No. 22 *K.A. Brekke, E. Gjelsvik and B.H. Vatne (1987)*: A Dynamic Supply Side Game Applied to the European Gas Market.

No. 23 *S. Bartlett, J.K. Dagsvik, Ø. Olsen and S. Strøm (1987)*: Fuel Choice and the Demand for Natural Gas in Western European Households.

No. 24 *J.K. Dagsvik and R. Aaberge (1987)*: Stochastic Properties and Functional Forms of Life Cycle Models for Transitions into and out of Employment.

No. 25 *T.J. Klette (1987)*: Taxing or Subsidising an Exporting Industry.

No. 26 *K.J. Berger, O. Bjerkholt and Ø. Olsen (1987)*: What are the Options for non-OPEC Countries.

No. 27 *A. Aaheim (1987)*: Depletion of Large Gas Fields with Thin Oil Layers and Uncertain Stocks.

No. 28 *J.K. Dagsvik (1987)*: A Modification of Heckman's Two Stage Estimation Procedure that is Applicable when the Budget Set is Convex.

No. 29 *K. Berger, Å. Cappelen and I. Svendsen (1988)*: Investment Booms in an Oil Economy -The Norwegian Case.

No. 30 *A. Rygh Swensen (1988)*: Estimating Change in a Proportion by Combining Measurements from a True and a Fallible Classifier.

No. 31 *J.K. Dagsvik (1988)*: The Continuous Generalized Extreme Value Model with Special Reference to Static Models of Labor Supply.

No. 32 *K. Berger, M. Hoel, S. Holden and Ø. Olsen (1988)*: The Oil Market as an Oligopoly.

No. 33 *I.A.K. Anderson, J.K. Dagsvik, S. Strøm and T. Wennemo (1988)*: Non-Convex Budget Set, Hours Restrictions and Labor Supply in Sweden.

No. 34 *E. Holmøy and Ø. Olsen (1988)*: A Note on Myopic Decision Rules in the Neoclassical Theory of Producer Behaviour, 1988.

No. 35 *E. Biørn and H. Olsen (1988)*: Production - Demand Adjustment in Norwegian Manufacturing: A Quarterly Error Correction Model, 1988.

No. 36 *J.K. Dagsvik and S. Strøm (1988)*: A Labor Supply Model for Married Couples with Non-Convex Budget Sets and Latent Rationing, 1988.

No. 37 *T. Skoglund and A. Stokka (1988)*: Problems of Linking Single-Region and Multiregional Economic Models, 1988.

No. 38 *T.J. Klette (1988)*: The Norwegian Aluminium Industry, Electricity prices and Welfare, 1988.

No. 39 *I. Aslaksen, O. Bjerkholt and K.A. Brekke (1988)*: Optimal Sequencing of Hydroelectric and Thermal Power Generation under Energy Price Uncertainty and Demand Fluctuations, 1988.

No. 40 *O. Bjerkholt and K.A. Brekke (1988)*: Optimal Starting and Stopping Rules for Resource Depletion when Price is Exogenous and Stochastic, 1988.

No. 41 *J. Aasness, E. Biørn and T. Skjerpen (1988)*: Engel Functions, Panel Data and Latent Variables, 1988.

No. 42 *R. Aaberge, Ø. Kravdal and T. Wennemo (1989)*: Unobserved Heterogeneity in Models of Marriage Dissolution, 1989.

No. 43 *K.A. Mork, H.T. Mysen and Ø. Olsen (1989)*: Business Cycles and Oil Price Fluctuations: Some evidence for six OECD countries. 1989.

No. 44 *B. Bye, T. Bye and L. Lorentsen (1989)*: SIMEN. Studies of Industry, Environment and Energy towards 2000, 1989.

No. 45 *O. Bjerkholt, E. Gjelsvik and Ø. Olsen (1989)*: Gas Trade and Demand in Northwest Europe: Regulation, Bargaining and Competition.

No. 46 *L.S. Stambøl and K.Ø. Sørensen (1989)*: Migration Analysis and Regional Population Projections, 1989.

No. 47 *V. Christiansen (1990)*: A Note on the Short Run Versus Long Run Welfare Gain from a Tax Reform, 1990.

No. 48 *S. Glomsrød, H. Vennemo and T. Johnsen (1990)*: Stabilization of emissions of CO_2: A computable general equilibrium assessment, 1990.

No. 49 *J. Aasness (1990)*: Properties of demand functions for linear consumption aggregates, 1990.

No. 50 *J.G. de Leon (1990)*: Empirical EDA Models to Fit and Project Time Series of Age-Specific Mortality Rates, 1990.

No. 51 *J.G. de Leon (1990)*: Recent Developments in Parity Progression Intensities in Norway. An Analysis Based on Population Register Data.

No. 52 *R. Aaberge and T. Wennemo (1990)*: Non-Stationary Inflow and Duration of Unemployment.

No. 53 *R. Aaberge, J.K. Dagsvik and S. Strøm (1990)*: Labor Supply, Income Distribution and Excess Burden of Personal Income Taxation in Sweden.

No. 54 *R. Aaberge, J.K. Dagsvik and S. Strøm (1990)*: Labor Supply, Income Distribution and Excess Burden of Personal Income Taxation in Norway.

No. 55 *H. Vennemo (1990)*: Optimal Taxation in Applied General Equilibrium Models Adopting the Armington Assumption.

No. 56 *N.M. Stølen (1990)*: Is there a NAIRU in Norway?

No. 57 *Å. Cappelen (1991)*: Macroeconomic Modelling: The Norwegian Experience.

No. 58 *J. Dagsvik and R. Aaberge (1991)*: Household Production, Consumption and Time Allocation in Peru.

No. 59 *R. Aaberge and J. Dagsvik (1991)*: Inequality in Distribution of Hours of Work and Consumption in Peru.

No. 60 *T.J. Klette (1991)*: On the Importance of R&D and Ownership for Productivity Growth. Evidence from Norwegian Micro-Data 1976-85.

No. 61 *K.H. Alfsen (1991)*: Use of macroeconomic models in analysis of environmental problems in Norway and consequences for environmental statistics.

No. 62 *H. Vennemo (1991)*: An Applied General Equilibrium Assessment of the Marginal Cost of Public Funds in Norway.

No. 63 *H. Vennemo (1991)*: The marginal cost of public funds: A comment on the literature.

No. 64 *A. Brendemoen and H. Vennemo (1991)*: Aclimate convention and the Norwegian economy: A CGE assessment.

No. 65 *K. A. Brekke (1991)*: Net National Product as a Welfare Indicator.

No. 66 *E. Bowitz and E. Storm (1991)*: Will restrictive demand policy improve public sector balance?

No. 67 *Å. Cappelen (1991)*: MODAG. A Medium Term Macroeconomic Model of the Norwegian Economy.

No. 68 *B. Bye (1992)*: Modelling Consumers' Energy Demand.

No. 69 *K. H. Alfsen, A. Brendemoen and S. Glomsrød (1992)*: Benefits of Climate Policies: Some Tentative Calculations.

No. 70 *R. Aaberge, Xiaojie Chen, Jing Li and Xuezeng Li (1992)*: The structure of economic inequality among households living in urban Sichuan and Liaoning, 1990.

No. 71 *K.H. Alfsen, K.A. Brekke, F. Brunvoll, H. Lurås, K. Nyborg and H.W. Sæbø (1992)*: Environmental Indicators.

No. 72 *B. Bye and E. Holmøy (1992)*: Dynamic equilibrium adjustments to a terms of trade disturbance

No. 73 *O. Aukrust (1992):* The Scandinavian contribution to national accounting

No. 74 *J. Aasness, E, Eide and T. Skjerpen (1992):* A criminometric study using panel data and latent variables

No. 75 *R. Aaberge and Xuezeng Li (1992):* The trend in income inequality in urban Sichuan and Liaoning, 1986-1990

No. 76 *J.K. Dagsvik and Steinar Strøm (1992):* Labor sypply with non-convex budget sets, hours restriction and non-pecuniary job-attributes

No. 77 *J.K. Dagsvik (1992)*: Intertemporal discrete choice, random tastes and functional form

No. 78 *H. Vennemo (1993)*: Tax reforms when utility is composed of additive functions.

No. 79 *J. K. Dagsvik (1993):* Discrete and continuous choice, max-stable processes and independence from irrelevant attributes.

No. 80 *J. K. Dagsvik (1993):* How large is the class of generalized extreme value random utility models?

No. 81 *H. Birkelund, E. Gjelsvik, M. Aaserud (1993):* Carbon/energy taxes and the energy market in Western Europe

No. 82 *E. Bowitz (1993):* Unemployment and the growth in the number of recipients of disability benefits in Norway

No. 83 *L. Andreassen (1993):* Theoretical and Econometric Modeling of Disequilibrium

No. 84 *K.A. Brekke (1993):* Do Cost-Benefit Analyses favour Environmentalists?

No. 85 *L. Andreassen (1993):* Demographic Forecasting with a Dynamic Stochastic Microsimulation Model