Building a spatial microsimulation model

Coomaren P Vencatasawmy, Einar Holm, Terry Rephann, Johan Esko, Neil Swan, Marianne Öhman, Magnus Åström, Eva Alfredsson, Kerstin Holme, and Jonny Siikavaara

Spatial Modelling Centre
Box 839
S 981 28 Kiruna, Sweden
http://www.smc.kiruna.se

Paper to be presented at the 11th European Colloquium on Quantitative and Theoretical Geography in Durham, England, on September 3-7, 1999.

Keywords: spatial, microsimulation, Sweden

Abstract: Microsimulation was introduced over four decades ago and has experienced somewhat of a revival in the social sciences lately. As a result many microsimulation models currently exist with a wide range of specialities. The Spatial Modelling Centre in Kiruna, Sweden, has contributed to this field by constructing a spatial dynamic microsimulation model, called *SVERIGE* or *System for Visualising Economic and Regional Influences in Governing the Environment*. This model is unique; it is the first national level spatial model of such kind, which makes it possible to analyse the spatial consequences of various public policies. The "space" in spatial microsimulation is achieved by incorporating regional attributes in modelling the various socio-economic modules of the model and constructing from scratch a module for modelling internal migration in Sweden. Furthermore, the results of the model can be illustrated via maps and Geographic Information Systems. This paper describes the effort in building the model and its different components.

1.0 Introduction

Microsimulation was introduced over four decades ago by Orcutt (1957) and has experienced somewhat of a revival in the social sciences over the past decade (Merz, 1991; Graham, 1996; Isard et al., 1998). It has been used to study various social phenomena such as population growth and development, the effect of ageing and pension formulas on social insurance funding, and the effect of various tax regimes on fiscal budgeting.

The spatial dynamic microsimulation model (called *SVERIGE* or *System for Visualising Economic and Regional Influences Governing the Environment*) built at the Spatial Modelling Centre in Kiruna, Sweden, is unique. It is the first national-level spatial microsimulation model to be constructed, thereby permitting analysts to study the spatial consequences of various national, regional, and local-level public policies. Assisting the model building effort is a unique database comprising longitudinal socio-economic information on every resident of Sweden for the years 1985 to 1995. The locations of the individuals in this database are given in co-ordinates accurate to the level of 100 metres. It is, therefore, possible to estimate behavioural equations on various geographical scales and to describe complex dynamic spatial relationships.

This paper describes efforts in building the model and explains the different components of SVERIGE. The paper is divided into two sections. The first section describes the model's history and its unique characteristics. The second section depicts the model structure in detail.

2.0 Model history and unique characteristics

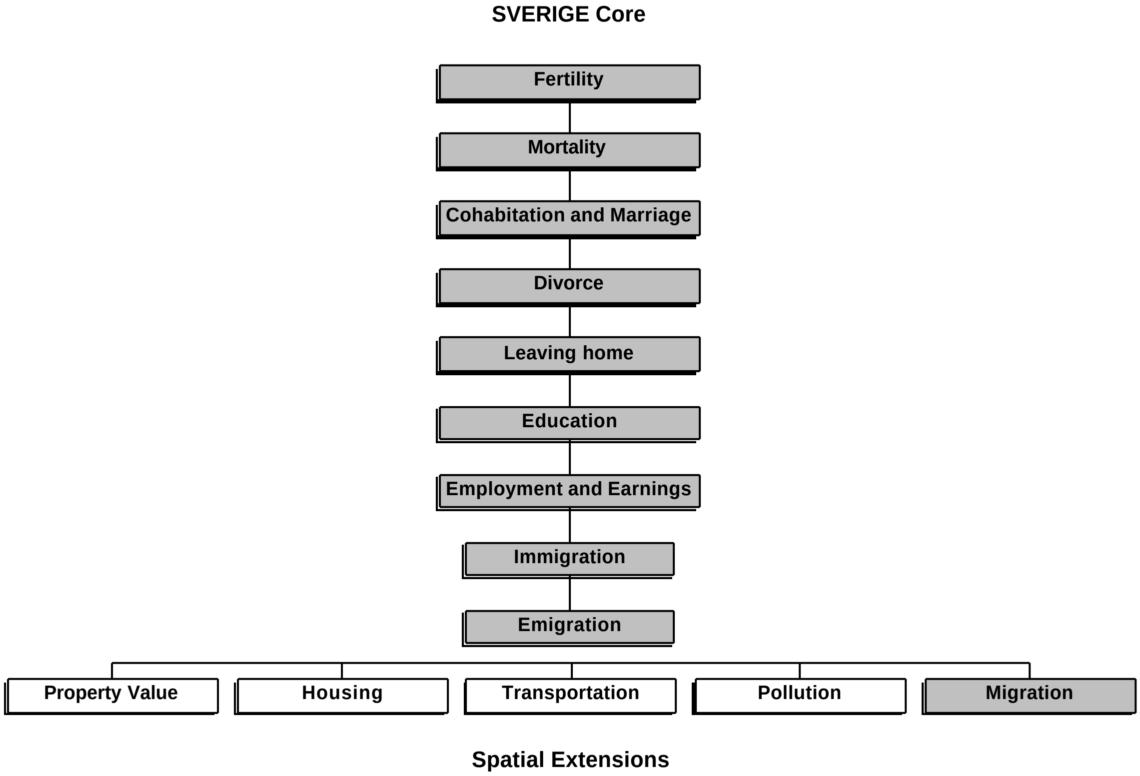
SVERIGE is a dynamic spatial microsimulation model for Sweden based on households. It generates events for individuals through the interplay of deterministic models of individual behaviour and a Monte Carlo simulation. The behaviours are functions of individual, household, and regional socio-economic characteristics, usually included as independent variables in discrete choice models or simply as categories used to estimate transition matrices that describe the probability of moving from one state to another. The model is dynamic as the evolution and development of the individuals occurs in chronological order, with initial conditions being changed for subsequent periods by counters and simulation.

The model core is based upon CORSIM (Cornell Microsimulation Model) (Caldwell et al. 1998), which itself is a modification of Guy Orcutt's DYNASIM (Dynamic Microsimulation Model), the first dynamic microsimulation model (Caldwell and Keister 1996). CORSIM has since sired other children as well, including a Canadian model named DYNACAN (Dynamic Microsimulation Model for Canada) (Morrison 1997). Both models are North American.

SVERIGE differs in several important respects from CORSIM and DYNACAN. First, SVERIGE is a Swedish model and thus explains behaviour in a different institutional context. Although it is based on the same social science as the others, differences in cultural and institutional peculiarities were too big to be ignored. Examples of such differences are the power relations between men and women, the degree of class equity, the elaborateness of social support mechanisms, and the diverse types of recognised family groupings (e.g., marriage versus cohabitation, referred to as "Sambo"). As these define the social context in which individual decisions are made and constrain the ways in which microactors interact, the equations used in SVERIGE are different but the life-cycle modular structure is the same as CORSIM. Figure 2.1 illustrates this modular structure.

Second, SVERIGE is a spatial model while CORSIM is not. In fact, SVERIGE is the first national-level spatial microsimulation model. SVERIGE not only makes life-cycle transitions dependent on a spatial context but also models individual spatial transitions such as internal migration. In the terminology of object-oriented programming geographical "objects" such as neighbourhoods and labour markets have attributes that influence the behaviour of "objects" such as individuals, households, and homes (see figure 2.2 and appendix A. for a full listing of objects and their attributes). As a result, the model is capable of generating a variety of geographically detailed reports that may interest regional scientists and policymakers.

Figure 2.1 The SVERIGE Model



Grey shading indicates that the module is operational in SVERIGE v. 1.0.

Figure 2.2 Relations between objects and attributes

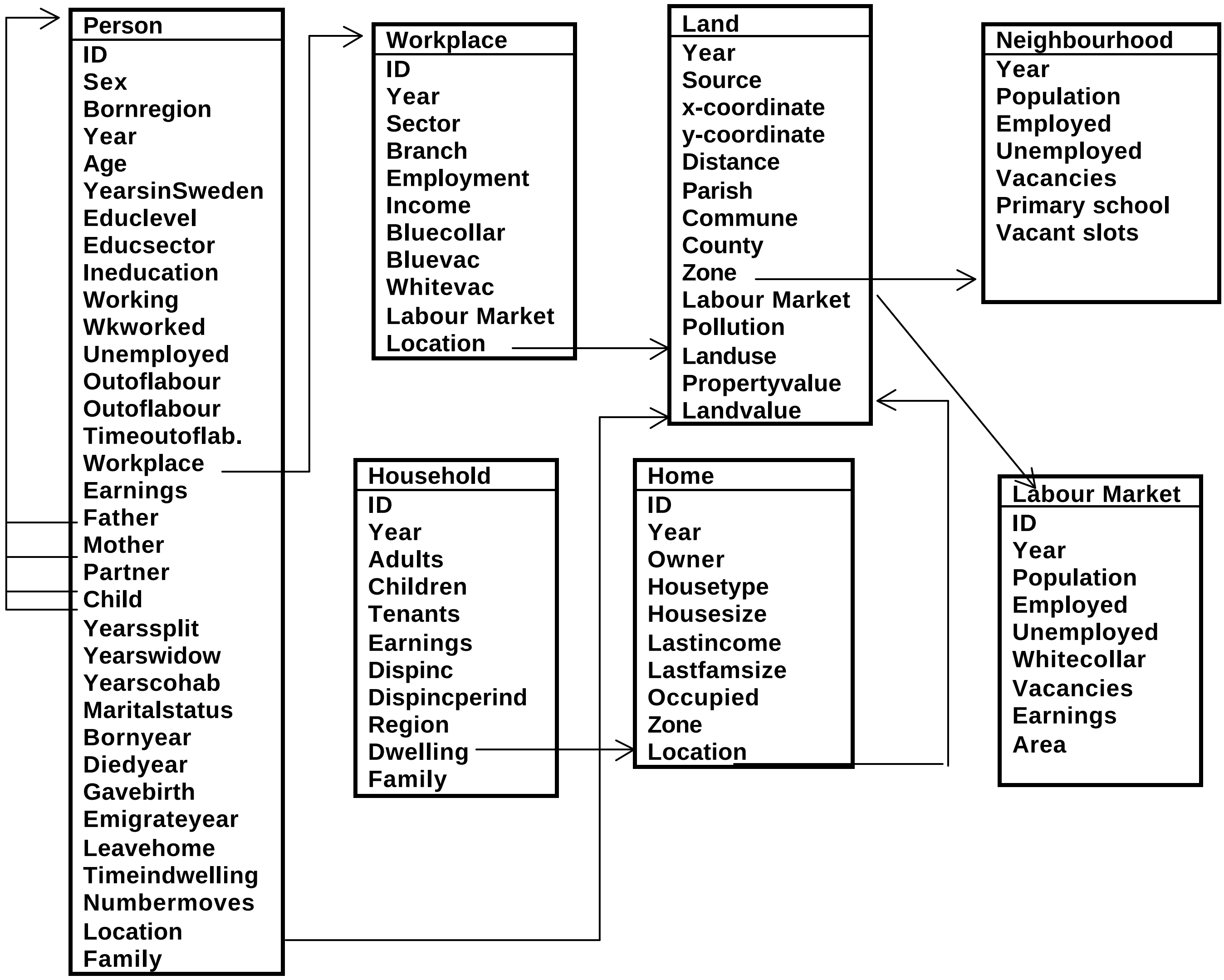


Figure 3.1 Pathways for updating object attributes

New attribute value ← event ← old attribute value, same person, same attribute
old attribute value, same person, other attributes
old attribute value, other person, same attribute
old attribute value, other persons, other attribute
other events
attributes of other objects
aggregates (of object attributes)

3.0 Structure of the microsimulation model

SVERIGE 1.0 contains 10 modules. Each module consists of a series of discrete or continuous variable equations, transition matrices, or rules that determine the occurrence of specific events in a person's life. When individual events occur, personal attributes are updated. In this version of SVERIGE, all the events occur at yearly intervals.

In SVERIGE there are three ways in which personal attributes can be changed. In the fertility module, for example, when an event is decided for a particular person, the attribute "Gavebirth" is then updated. This is the most common type of change. Changes in attributes can also be triggered by the occurrence of other significant events. For example, if two persons are chosen to cohabit then the attribute "location" of the woman is changed to that of the man. Finally, attributes such as "earnings" are influenced to a large degree by their previous values. These mechanisms are summarised in figure 3.1.

The occurrence of an event is determined by Monte Carlo simulation. Each individual is exposed to the possibility that a certain event will occur based on a simple probability. This probability is estimated from tables or by an estimated logistic regression equation. The latter is a function of personal attributes and/or the attributes of other units (such as partners, households, neighbourhoods, and labour markets). Thus each person at risk of an event is assigned a probability and a random risk. If the probability is greater than the risk, then the event occurs.

At the end of each event, the personal attributes are updated as well as the attributes of various other aggregates such as families, neighbourhood, and regions. As some modules occur before others, the attributes used in estimating the transition probabilities have different time subscripts. The model currently implements these modules in the following order: mortality, fertility, education, employment and earnings, marriage, leaving home, divorce, immigration, emigration, and migration. The equations, counters, and pointers used in the model are summarised in table 3.1 and Appendix A. They are further briefly described in the subheadings that appear below.

3.1 Mortality module

The mortality module is used to terminate lives in the model (Rephann 1999a). There are two sets of mortality equations. The first set is for individuals under the age of 25 years and is estimated as an exponential time trend from historical mortality rates grouped by age and sex. The second set is for individuals between the ages of 25 and 127 years. These equations are logistic with socio-economic attributes used as independent variables. If a person is 127 years or older, his/her life is terminated. The first set is further divided into 12 groups based on sex and age groups of 0-1, 1-4, 5-9, 10-14, 15-19, 20-24. The second set is divided into three age groups: 25-59, 60-74 and 75-127. The groupings were decided based on model fit and regression diagnostic statistics.

This module is very similar in specification to that of CORSIM although different groupings are used. These choices were supported by Himes (1994) who found that the United States has a cause-of-death structure similar to that of Sweden but that the age-specific death rates vary between the two countries. Furthermore, Wallace et al. (1985) note that Sweden has much lower infant mortality rates than the US. Other differences such as the influence of the social support system described by Minder (1991) is recognised by estimating new equations.

The second set of equations was estimated using a sample of 458 154 individuals whose survival was followed for a period of five years. According to epidemiological studies, attributes such as education level, employment status, and marital status influence an individual's probability of dying in any year[1]. No regional variables were found to be significant. Figure A.1

[1] Note, however, that the variable earnings is not used in equations for the groups of age above 60 years as such people derive their income primarily from pensions. Pension income could not be used in the equations because it is not produced in SVERIGE 1.0. Education

illustrates the overall structure of this module and the variables used in the equations are shown in equation (2) of table 3.1.

In addition to removing individuals from the simulation, deaths directly triggers both personal and household changes. For instance, when a death occurs, the marital status of the surviving spouse changes from “married” or “sambo” to “widow.” In addition, if the deceased is a single or parents die, then the under-aged children (less than 15 years) are assigned to an orphan queue and allocated to a family in the same labour market region at the end of the year. Children older than 15 years old automatically become a heads of household and leave home.

3.2 Fertility module

The primary role of the fertility module is to create new lives for simulation in the microsimulation model (Rephann 1999b). Upon birth, each infant is assigned a sex based on a fixed probability of 0.5134 of being a male. Only woman aged 15 to 44 years old are at risk of giving birth. Equations are estimated for two groups: married and unmarried woman. The group “unmarried” includes singles, sambos, widows, and divorcees. All births produce a single baby.

This module is very similar to that of CORSIM as the fertility equations are simplified versions of the CORSIM ones. They were re-estimated using a Swedish sample of 458 154 individuals and 1990 data. There is some evidence that US and Swedish fertility behaviours are different. For instance, while CORSIM finds employment apparently has an inhibiting effect on American fertility decisions, the opposite has been observed in Sweden (Hoem and Hoem 1997; Sundstrom and Stafford 1992). The influence of earnings (Tasiran 1995) and spacing between births (Hoem 1995) are also different. It was, however, decided to maintain the same structure as CORSIM because these factors are addressed in the estimation.

Fertility behaviours in SVERIGE are influenced directly by individual and household attributes such as employment, marital status, age, family earnings, and education. Figure A.2 gives a broad overview of the structure of the model and equation (1) in table 3.1 shows the variables used.

3.3 Education

This module was re-thought completely because the education system in Sweden is different from the US (Rephann 1999c). Compulsory school is called Grundskola and consists of nine years of basic preparation. Ninety percent of the students advance to the next stage, Gymnasium. This level offers vocational and college preparatory programs and typically takes three years to complete. After Gymnasium comes college (called Högskola) for around twenty-five percent of Swedish gymnasium students. A typical undergraduate school at Högskola lasts three years. Graduate or professional school follows college or university. Students in higher education receive free tuition and monthly stipends for covering monthly expenses and are eligible for student loans. There is also an adult education program called Komvux, which makes up approximately 16% of the student population.

SVERIGE uses a series of logistic regression equations to determine completion of Grundskola, completion of Gymnasium, entry to Komvux, persistence through Komvux, entry to Högskola, persistence through Högskola, entry to graduate school and persistence through graduate school. The probability to enter gymnasium was kept constant as a value of 0,9. The equations were estimated using data extracted from the SMC database for the years 1993-1995. As no information on graduation was available, this was assumed to occur when the educational level of a person changes. Only full-time students are modelled, but both traditional and adult students are eligible to participate. At any time, students may be selected to discontinue education but they are eligible to rejoin education later on. Note that the boys usually start university one year later because of compulsory military service.

---

variables were not used for the group age 75 years and above because this variable is missing from the database for elderly individuals.

Much research has been done on the influence of different factors on educational achievement. Family attributes appear to play a dominant role as pointed out by Haveman and Wolfe (1995). They also pointed out that educational levels of mothers have greater effects than those of the fathers. Other important indicators are parental income and class status, divorce and family dissolution; and size and composition of the family (Lichter et al., 1993). Equations (9)-(11) show the variables used in the module and Figures A.3 and A.4 show the overall structure of this module.

3.4 Employment and earnings

The basic structure of this module is derived from other microsimulation models such as CORSIM (Caldwell 1997), DYNACAN (Chenard 1996) and MICROHUS (Klevmarken). Its aim is to estimate the amount of time each individual between the ages of 16 and 65 is employed during the year and his/her wage rate (Alfredsson and Åström 1998).

The module consists currently of four sub-modules, sequentially structured as illustrated in Figure A.5. Initially, a logistic regression equation determines the likelihood that a given individual is employed at all during the year. For those who are simulated as being employed, the next two sub-modules determine the amount of weeks worked utilising a logistic regression equation to determine full-time workers and a transition matrix to determine the number of weeks worked by part-time workers. The final module estimates the average relative wage rate (the ratio of wage rate to average wage rate) for each employed individual. Annual earnings are computed as the product of weeks worked and wage rate. The variables in these modules are summarised in equations (12)-(14) in Table 3.1.

For estimating all four equations the HINK[2] database for the year 1995 (SCB 1995) was used. Because the HINK data used is cross-sectional rather than time-series, it was necessary to link the cross-section with the comprehensive micro database at SMC in order to obtain lagged income and employment used in the models.

The employment equation for the first sub-module is estimated for three groups: 16-20 years, 21-53 and 54-56. These groupings were decided in order to select groups of individuals who are relatively homogeneous in terms of participation rates and to account for non-linearity in the response to age. Although there is a clear regional disparity in unemployment rates, with higher unemployment occurring in regions away from the three largest metropolitan areas (Borgegård and Lundin 1993), no regional variables were found to be significant. Another major omission is the unemployment history or duration of the individual, which according to Sheehan (1998) is strongly negatively correlated with finding employment. This is not available in either SMC database outside census years or the HINK sample.

The second sub-module uses two groups, males and females, because the average workload varies by gender. For example, on average 84% of men work full-time whereas only 52% of women work full time. Furthermore, some variables such as the presence of a child age 0-6 in the household affects male and female workers differently. Two geographic variables were significant: unemployment rate in labour market region and distance of place of residence to centre of labour market. The latter variable was statistically significant only for women, and had a negative coefficient.

The third module assigns the number of weeks worked (0-48) within a year because it was impossible to predict weeks worked satisfactorily using logistic regressions. Note that this distribution has three nodes.

The fourth sub-module estimates the logarithm of relative wage rate per month. The relative wage rate is defined as the ratio of the wage rate to the average wage rate for everyone in the age group 16-65 working in Sweden. There are two groups in this sub-module: full-time workers and part-time workers. This dichotomy was used because part-timers earn less on average for a number of reasons. For instance, part-time employment is mostly prevalent in lower paying trade and service sectors. However, variables such as occupation or employment sector are not modelled in SVERIGE 1.0.

---

[2] The aims of the HINK survey is to map the disposable income distribution among households and also shed light on income structures and distribution in

3.5 Cohabitation and marriage

According to a survey conducted by Statistics Sweden (SCB 1995) the factors that influence the likelihood of starting a relationship differ between teenagers and adults. For example, the propensity to start a relationship in the teens is highest for those who have many brothers and sisters, those who did not grow up with both parents, those with blue collar parents, those with non-religious parents, and those who are pregnant or are studying. Propensity to form a relationship during young adulthood (age of 20-28 years) depends on the same personal attributes such as pregnancy or student status, but only the profession of the father matters and coming from a two parent household matters only for males. In another survey (SCB 1994), it was found that the tendency to start a relationship depends on educational level.

Every year around 90 000 new families are formed. Among these, 40% of them have experience from earlier relationships. Many couples separate when they are young, increasing the number of people living alone but also increasing the pool of people who are eligible to form new relationships. More than half of those who separated start a new relationship within 2-3 years. For women, the childless ones are more likely to start all over again. Women who grew up in large cities have a lower propensity to start a new relationship than those who grew up in smaller cities or in the countryside. For men, the pattern is opposite.

The marriage module creates common-law marriage (sambo) for selected unmarried individuals over the age fifteen and Christian marriage partners for sambo couples. The module consists of three sub-modules. The first sub-module (sambo decision) determines whether a person is eligible for sambo or not (Esko 1999a). The second sub-module (cohabitation compatibility) computes an index of compatibility for pairs of eligible singles based on their attributes and matches pairs on the basis of the compatibility index using a heuristic matching algorithm (Esko 1999b). The final sub-module (marriage decision) determines whether cohabiting individuals will get married (Esko 1999c).

The first sub-module uses logistic regression to determine who will enter the sambo market. Eight groups are used: The age groups 16-25, 25-65 and 66-127 years old were chosen for males and the age groups 16-25, 26-55 and 56-127 years were chosen for females. These groups are based on empirical analysis of the response curve by age and the discriminating power of the models. The people at risk are those that are not already married to someone else or who are already in a sambo relationship. They should also be more than 16 years old. A list of the variables used in these modules is provided in equations (3)-(5) in Table 3.1.

The first sub-module produced a pool of suitors that are going to form a sambo relationship. This is determined using a logistic regression with the variables described in the previous paragraphs. The second sub-module matches them into couples. This process can be computer intensive and it is important to reduce the number of computations. In CORSIM, each man is compared to every woman and the number of computations required is FM where F is the number of females in the pool and M is the number of males in the pool. However, because it is computationally burdensome an alternative routine was adopted here. Each woman is assigned an index based on her age, education level and country of birth. The indexes are then sorted from highest to lowest. A man is then drawn at random from the pool and an index is computed using lookup tables. The two indexes are compared and the nearest available woman is found.

The final module determines whether the sambo couple will marry or continue cohabiting. In Sweden it is very uncommon for people who move together to marry immediately soon after. Marriage usually happens after a long period of cohabitation. Couples who marry directly make up less than 5% of the total couples moving together in a given year. Because of this fact, only cohabiting couples are eligible to marry in the model. Seven groups are used: couples without children (age groups 15-32, 33-45, 46-127), couples with children (age groups 15-30, 31-45, 46-127) and couples in which at least one spouse immigrated in 1990 (age group 15-127). These groupings were chosen based on empirical studies that show the propensity to marry based on age, immigration status, and whether or not the couple has children, varies. Although empirical studies show big regional differences in the propensity for sambo pairs to get married, initial attempts at finding an appropriate explanatory regional variable have not been successful.

An overall schematic description of the module is given in Figures A.6 and A.7. Cohabitation and marriage may require several adjustments in personal and household attributes, including change in marital status (from “single,” “widowed,” or “divorced,” to “cohabiting” or “married), adjustment of household earnings for two-income households, and aggregation of children. Moreover, cohabitation triggers the movement of female partners to the male partner’s home.

3.6 Leaving home

Almost all young people will sooner or later leave home to form their own household. At the age of 16, young Swedes leave elementary school (Grundskola), which is compulsory, and have the option of enrolling in Gymnasium or beginning employment. Only a small percentage will actually leave home at the age of 16 or earlier. A small number of studies have looked at the factors that influence an individual to leave the parental home to form a new household. One finding is that girls tend to leave home earlier than boys. According to Fransson (1997), the housing market is the most important factor because lack of small and inexpensive apartments can hinder youngsters from leaving home. Unfortunately, because a housing module has not been constructed for SVERIGE 1.0, this cannot be modelled.

The leaving home module determines whether a person will leave the parental home to start a new household (Öhman, 1999). They are eligible to leave home, beginning with the age of 16. If someone has still not left home, then he or she is forced to do so at the age of 30. This module follows closely the leaving home module of CORSIM.

Logistic regression equations are estimated using two cohorts. The cohorts consist of individuals born in 1968 (128 500 individuals) and 1973 (118 197 individuals) and they are extracted from the SMC database. These two cohorts were pooled for estimating the regression equations. The dependent variable was whether an individual left home in a given year, pooling the data for the entire period. Sometimes, an individual, who has already left home, moves back into the parental household. This occurrence was common in the early 1990’s because of the economic recession. Because this amounted to less than 3%, their eventual moving from home is considered as if it were the first. No groupings were used in this module and they was believed to be unnecessary.

The variables used in this module are shown in equation (8) in Table 3.1. The overall structure of the module is illustrated in Figure A.8.

3.7 Divorce

At the beginning of this century divorces were rare but they have now become commonplace. In 1991 there were approximately 370 000 sambo pairs and the separation frequency was 79 per 1000. For married couples the frequency is much lower, only 14 per 1000.

The divorce module dissolves sambo and marital relationships (Esko 1999d). Seven groups are used: couples without children (age groups 15-20, 21-30, 31-127), couples with children (age groups 15-20, 21-30, 31-127) and couples in which one spouse immigrated in 1990 (age group 15-127), where the age group refers to age of the female spouse. These groupings were based on empirical studies with the marriage sub-module. The variables used in this module are illustrated in equation (6) Table 3.1.

A schematic diagram illustrating the structure of the model is given in Figure A.9. Divorce results in persons being assigned new marital statuses (from “married” to “divorced” or from “sambo” to “single”) and makes them eligible for remarriage. Also, it triggers a number of other microsimulation events, including movement of the former husband out of the marital dwelling, re-allocation of minor children to each new household, and de-coupling of household earnings. Currently, minor children are assigned to the female partner.

3.8 Immigration

In 1954 Sweden signed the Geneva conventions on refugees, and with its Nordic neighbours, entered into a common labour market, which allowed unhindered movement within

Scandinavia (Lundh and Olsson 1994). As a consequence of Sweden's enormous industrial success, many Finnish citizens flocked to its industrial regions (Häggstrom et al. 1990). This was welcomed because labour was in short supply in the industrial communities in central Sweden. Subsequent recruitment of labour from southern Europe, especially Yugoslavia and Greece, but also Italy and Turkey, occurred in a very liberal atmosphere. The distribution of the population, including that of immigrants, was not steered by political instruments but regulated by labour market forces.

In 1967, responding to pressure from the labour union movement, Sweden introduced 'regulated immigration' (Lundh and Olsson 1994). The need for labour remained but social changes gave rise to the demand that labour should be recruited mainly from the ranks of married woman, disabled people and the elderly. The introduction of regulated immigration and the establishment of the National Immigration and the Naturalisation Board were manifestations of the change from a non-regulating immigration policy, and set a stop to large-scale immigration (Hammmar 1992). The family-related immigration followed in the footsteps of labour force immigration. The already existing regional patterns were reinforced (Borgegård et al., 1996). From 1984 to 1994, the 'whole of Sweden strategy' determined the distribution of the 1980s refugee immigrants over the entire country. The advantage with this policy was that the financial and social burden for individual municipalities was mitigated.

Because the number of refugees that enter the country every year fluctuates due to internal and external economic and political factors, it is very difficult to design a model that will capture changes in the magnitude and settlement patterns of immigrants. As a result this module relies entirely on hard core probabilities and external data. The module is used to create new lives that are derived from outside of Sweden (Vencatasawmy 1999a). In many ways this module resembles very closely that of CORSIM, except for some minor changes, as described below.

Immigrants are clustered into ten groups based on their geographical and cultural origins. Given an initial stock of immigrants, look-up tables of historical probabilities are used to assign the age, sex and marital status of the head of the household. Given these three characteristics, a comparable individual from a pool of immigrants is chosen to clone this new immigrant family. The economic and family characteristics of this clone are thus assigned to the new immigrant household head. The labour market region where the immigrant will settle is then estimated using transition matrix based on immigrant settlement characteristics for the different immigrant groups. The family is then placed into a queue of families looking for residential locations (100 metre square land tracts). They are assigned a particular location in the migration module, which is discussed below.

The household head demographic and labour market settlement transition matrices were estimated using all new immigrants in 1989 and 1990, more than 100 000 immigrants in total. The pool of immigrants used to clone the new ones consist of all immigrants who came to Sweden in 1990 and 1991, 70 000 individuals in total. As 1990 was a census year, it was easier to create families using the data from that year. As we wanted a short history of the immigrants, the years before and after that year were also used. A schematic diagram of module is given in Figure A.10.

3.9 Emigration

This module is entirely new; no counterpart exists in CORSIM. The module determines who will leave Sweden as emigrants. Statistics Sweden defines an emigrant as anyone who intends to settle abroad for at least one year. This same definition was used here (Vencatasawmy 1999b). This module only distinguishes emigrants from non-emigrants. It does not attempt to model the country of destination.

Emigration is a fairly important source of population change in Sweden. The number of emigrants varied from 25 269 in 1985 to 20 673 in 1987 and back up to 25 196 in 1990. In those three years, the number of emigrants with Swedish citizenship was 7 883, 9 044 and 8 957 respectively. On average 1 in every 3 emigrants is a Swede.

Glavac (1998) modelled migration from Australia using two frameworks: one that relies on human capital theory and another that emphasises the role of assimilation and ties to the origin country. She argued that age, gender, marital status, occupation and number of years since arrival are personal characteristics that affect assimilation and thus directly affect return migration. She observed that, for example, older migrants tend to go back because they have stronger ties to their country of origin. Using a Cox model, she found that sex, age, profession, years since immigration and a number of economic variables are highly significant.

In the absence of similar literature on emigrants in Sweden, a more inductive approach was taken to build this module. Graphs were drawn, contingency tables examined and regression analyses were tested on a variety of variables. Four groups were used, based on this empirical analysis: everyone between the ages of 16 and 21 years; single individuals between the ages of 22 and 58 years; couples for which the head is of age between 22 and 58 years, and everyone of age greater than 59 years. The unit of observation in the third group was the head of household. In the remaining groups the unit of observation was an individual. The logistic regressions were estimated using a sample of 458 844 individuals.

A schematic diagram of the module is given in Figure A.11. The variables used in the module are listed in equation (17) of table 3.1.

3.10 Internal migration

The internal migration module (Vencatasawmy 1999c) is the other module that does not have a counterpart in CORSIM. It makes the microsimulation a fully interregional model that links events in one region with occurrences in another. This migration module tracks the movement of households and individuals to an accuracy of 100 metres. Movement can be either inter-regional or intra-regional. Sweden can be divided into 108 labour market regions where commuting between labour markets is minimised while commuting within labour market region is maximised (Finansdepartementet 1994). A move is called intra-regional when it happens within a labour market and is called inter-regional when it happens between labour markets. In this version emphasis has been given to inter-regional migration. Intra-migration is modelled by heuristic rules.

More than 80% of the change in population size and composition on the community level is related to migration. Therefore, for the microsimulation model to be effective, the real life regional and local migration patterns should be modelled accurately. The design of this module is based on the structure set out by Holm and Malmberg (1997). A description of how this module is designed is given in Vencatasawmy and Swan (1998). One hundred and eight labour market regions were chosen because it was thought that someone living in any region would prefer to commute rather than migrate. The inter-regional migration decision is divided into three stages: Decision to move, choosing a labour market, and choosing a 100-metre square. The reasons for such assumptions reside in the way people make choices and for modelling reasons.

In discriminating between movers and non-movers in stage one, both personal and regional attributes are used. In the second stage only regional attributes are used and in the third stage an attraction measure based on personal characteristics is used. In intra-regional migration, the first two steps are ignored.

For the first step a sample of 458 844 individuals alive in 1990 was used. For the second stage, all individuals who moved between labour markets from 1985 to 1995, were used. This data set contained 2 114 861 individuals. For the first group, nine groups were used based on three age groups and three household characteristics. The former consisted of age groups of 16-21, 22-58 and 59+. The three household characteristics are: single (unmarried, widowed or divorced), households without children (both spouses live together), households with children (spouses can live together or may be separated). Logistic regressions were estimated for this stage. The variables used are depicted in equation (18) of table 3.1.

The second stage is estimated using a conditional logit model also known as the McFadden's conditional Logit model. When the data consists of choice specific attributes instead of individual-specific characteristics, this is a more appropriate model than the Multinomial logit model. The model is otherwise essentially the same as the multinomial logit (Maddala, 1983). Although this model requires "the independence of irrelevent alternatives (IIA)" which is violated in such studies (Fotheringham, 1991), this model was chosen for its qualities. There is one equation for each origin as according to Fortheringham (1991) who found a link between the parameter values and the origins in similar studies.

In the third stage squares are allocated to people based on a similarity index. There are 706 307 inhabited 100-metre squares in 1990. The variables used to determine the similarity index are: average family size for the residents living on the squares, average earnings for the residents on the squares, average education level of the square inhabitants and average age of inhabitants. The overall structure of the model is given in Figure A.12.

3.11 Future modules (property value, housing, transportation, pollution)

The next version of SVERIGE will include at least four additional modules: housing, property values, transportation, and environmental pollution. The housing module will determine housing tenure choices and match households with specific apartments and single family homes. Also, the existing housing stock will be modified through depreciation, renovation, and new construction. Although the details of the module are not yet available, it should bear some resemblance to current housing microanalytic models such as Fransson (1997), and Oskamp (1997). The housing choice sub-module should eventually replace the crude local mobility sub-module in the internal migration module for identifying local location choices. The property value module, currently under development (Rephann 1998), could determine property values based on hedonic price theory. Each property parcel price is modelled as a function of housing attributes, property features, property location, and labour market demand. The transportation module will determine commuting distances to employment. Initially it will model Euclidean distances, but may eventually take into account varied transportation modes, congestion effects, and traffic assignment. The environmental module will compute levels of airborne emissions, solid waste, and water pollution. Airborne emissions are proportionate to commuting distances and housing size and heating characteristics. Solid waste and water pollution should be determined by household characteristics and regional settlement patterns.

Table 3.1 Determinants of person attributes in SVERIGE version 1.0

1. Gavebirth – (age, maritalstatus, earnings(family), working), sex of new birth is random draw
2. Diedyear – (age, maritalstatus, earnings(family), educlevel, sex, working)
3. Cohabit - (age, children, earnings, educlevel, maritalstatus, sex, working)
4. Married- (age(female partner), age(youngest child), children, children(female partner)(t-yearscohab), maritalstatus, earnings(household), educlevel, (earnings(female partner)-earnings(male partner)), (edlevel(female partner)-edlevel(male partner)), (age(female partner)-age(male partner)), bornregion(female partner), bornregion(male partner))
5. Partner – pointer to partner, mating algorithm(sex, age, educlevel, bornregion)
6. Divorced- (age(female partner), age(child), children, children(female partner)(t-yearscohab), earnings(household), educlevel, maritalstatus, bornregion(female partner), bornregion(male partner))
7. Widowed- spouse diedyear
8. LeaveHome- (age, earnings, educlevel(mother), educlevel(father), sex)
9. Ineducation- (age, earnings(family), educlevel(mother), educlevel(father), maritalstatus, working, bornregion, ineducation(t-1), sex, location)
10. Educlevel- (educlevel(t-1), ineducation)
11. Educsector - exogenous
12. Working- (age(t-1), age(youngest child), children, civilstatus, educlevel, sex, working(t-1), ineducation(t-1), bornregion, yearsinSweden, location)
13. Wkworked- (age(t-1), age(youngest child), children, civilstatus, educlevel, ineducation(t-1), working(t-1), yearsinSweden, bornregion, location)
14. Earnings- (age(t-1), earnings(t-1), educlevel, ineducation, sex, location, wkworked)
15. Bornregion- Exogenous for immigrants or parent's location for Swedes
16. YearsinSweden- New immigrants or newly born = 0, (YearsSweden(t-1), Emigrateyear)
17. Emigrateyear-(sex, maritalstatus, yearsinSweden, numbermoves, bornregion, working, educlevel, location)
18. Location- (age, age(oldest child), age(youngest child), children, sex, educlevel, yearsinSweden, bornregion, working, earnings(family), timeindwelling, numbermoves, location).
19. Age – Age(t-1)
20. Mother- Natural mother or adoption
21. Father- Natural father

Appendix A. Objects and attribute definitions.

Two types of attributes are listed: “static” (normally “one-ended” attributes like age, sex and income) and relational attributes (like mother or family). The relational attributes connect object instances. Mother is an attribute of a person object, which points at another instance of the same object type. Family is another attribute of an individual, which is a pointer to an instance of another object type: household. Relational attributes are implemented with the help of pointers or references. Their main advantage is that they give access not only to the object but also to all its attributes. For example, the mother’s education could easily be retrieved and used as one determinant of the daughter’s education, employment and income.

The following convention regarding time notation is used. Current year is denoted by *t*, but that time index is often omitted. The year before the current year is denoted by *t-1* and next year is denoted *t+1*. The new values for the attributes are calculated for the current year.

The person object contains the following properties/attributes:

Label	Type	Values	Comment
ID	Longint		Identifier
Sex	Boolean	Male, Female	
Bornregion	Byte	0-100	Country group, county in Sweden
Year	Date	1900-2100	
Age	Byte	0-120	
YearsinSweden	Byte	0-120	
Educlevel	Byte	0-7	
Educsector	Integer	SUN 3-digit	Educational discipline
Ineducation	Boolean	Yes/no	Enrolled in school or university
Working	Boolean	Yes/no	Employed
Wkworked	Integer	0-52	Weeks employed
Unemployed	Integer	0-52	Weeks unemployed
Outoflabour	Integer	0-52	Weeks out of labour force
Timeoutoflab	Integer	0-50	Years not at all in labour force
Workplace	Workplace	Pointer	Pointer to workplace
Earnings	Integer	100 SEK	Annual earnings from employment
Father	Person	Pointer	Pointer to father or adopted father
Mother	Person	Pointer	Pointer to mother of adopted mother
Partner	Person	Pointer	Pointer to partner/wife/husband
Child	Person	Pointer	Pointer to child
Yearssplit	Byte	0-100	Years since split or divorce
Yearswidow	Byte	0-100	Years since partner died
Yearscohab	Byte	0-100	Cohabitation + married
Maritalstatus	Byte	0,1,2,3,4	0=single, 1=cohab, 2=married, 3=widowed, 4=divorced
Bornyear	Boolean	Yes/no	
Diedyear	Boolean	Yes/no	
Gavebirth	Boolean	Yes/no	
Emigrateyear	Boolean	Yes/no	
Leavehome	Boolean	Yes/no	
Timeindwelling	Byte	0-	Months since move to present dwelling
Numbermoves	Integer	0-	Number of previous moves
Location	Land	Pointer	Pointer to land (redundant)
Household	Household	Pointer	Pointer to household

The household object contains the following properties/attributes:

Label	Type	Values	Comment
ID	Longint		Identifier
Year	Date	1900-2100	
Adults	Byte	1-	Number of adults in household
Children	Byte	1-	Number of children in household
Tenants	Byte	1-	Number of non-family members in household
Earnings	Integer	100 SEK	Annual earnings from employment
Dispinc	Integer	100 SEK	Disposable income for household
Dispincperind	Integer	100 SEK	Disposable income/person for household
Region	LA-region	1-108	Labour market region (redundant)
Dwelling	Home	Pointer	Pointer to home
Family	Boolean	Yes/No	

The home object contains the following properties/attributes:

Label	Type	Values	Comment
ID	Longint		Identifier
Year	Date	1900-2100	
Owner	Byte		Category of owner
Housetype	Byte		Type of housing, age of house
Housesize	Integer	m2	Size of living area
Lastincome	Integer		Disposal income of last tenants family
Lastfamsize	Byte		No. of persons in last tenants family
Occupied(t)	Boolean	Yes/no	If no, the dwelling "slot" is vacant
Occupied(t-1)	Boolean	Yes/no	
Zone	Neighbour hood	Pointer	Pointer to Neighbourhood (redundant)
Location	Land	Pointer	Pointer to land square

The land object contains the following properties/attributes:

Label	Type	Values	Comment
Year	Date	1900-2100	
Source	Byte	O,1,2,3	0=empty,1=houses,2=workplaces,3=1+2
x-coordinate	Longint		
y-coordinate	Longint		
Distance	Integer	Km	Distance to Labour Market (LA) region centre
Parish	Integer		
Commune	Integer		
County	Integer		
Zone	Neighbour hood	Pointer	Pointer to Neighbourhood
Labour Market	LA-region	Pointer	Pointer to Labour Market (LA) region
Pollution	Integer		Pollution levels
Landuse	Byte		Crude land use estimate
Propertyvalue	integer	100 SEK/m2	Total estimated value of land unit
Landvalue	Integer	100 SEK/m2	Estimated land rent

The workplace object contains the following properties/attributes:

Label	Type	Values	Comment
ID			Identifier
Year	Date	1900-2100	
Sector	Byte		Institutional sector
Branch	Integer		Standard industrial classification
Employment	Integer	0-	Number of employed at workplace
Income	Longint	100 SEK	Total salary for employed
Bluecollar	Integer	0-	Number of employed blue collar (0-3)
Bluevac	Integer	0-	No. of vacancies for educ level 0-3
Whitevac	Integer	0-	No. of vacancies for educ level 4-7
Labour Market	LA-region	Pointer	Pointer to LA-region (redundant)
Location	Land	Pointer	Pointer to land square (static only 1994)

The Labour Market (LA) region contains the following properties/attributes:

Label	Type	Values	Comment
ID	Byte		LA-region number
Year	Date	1900-2100	
Population	Longint		Total population number
Employed	Longint		Total employed number
Unemployed	Longint		Total unemployed number
Whitecollar	Longint		Total number of residents edlevel 4-7
Vacancies	Longint		Total number of vacancies
Earnings	Integer		Total earnings
Area	Integer	km2	Total land area

The Neighbourhood object contains the following properties/attributes:

Label	Type	Values	Comment
ID	Byte		Neighbourhood zone number
Year	Date	1900-2100	
Population	Integer		Total population number
Employed	Integer		Total employed number
Unemployed	Integer		Total unemployed number
Vacancies	Integer		Total number of vacancies
Primary school	Boolean	Yes/no	
Vacant slots	Integer		Number of vacant slots,

Figure A.1

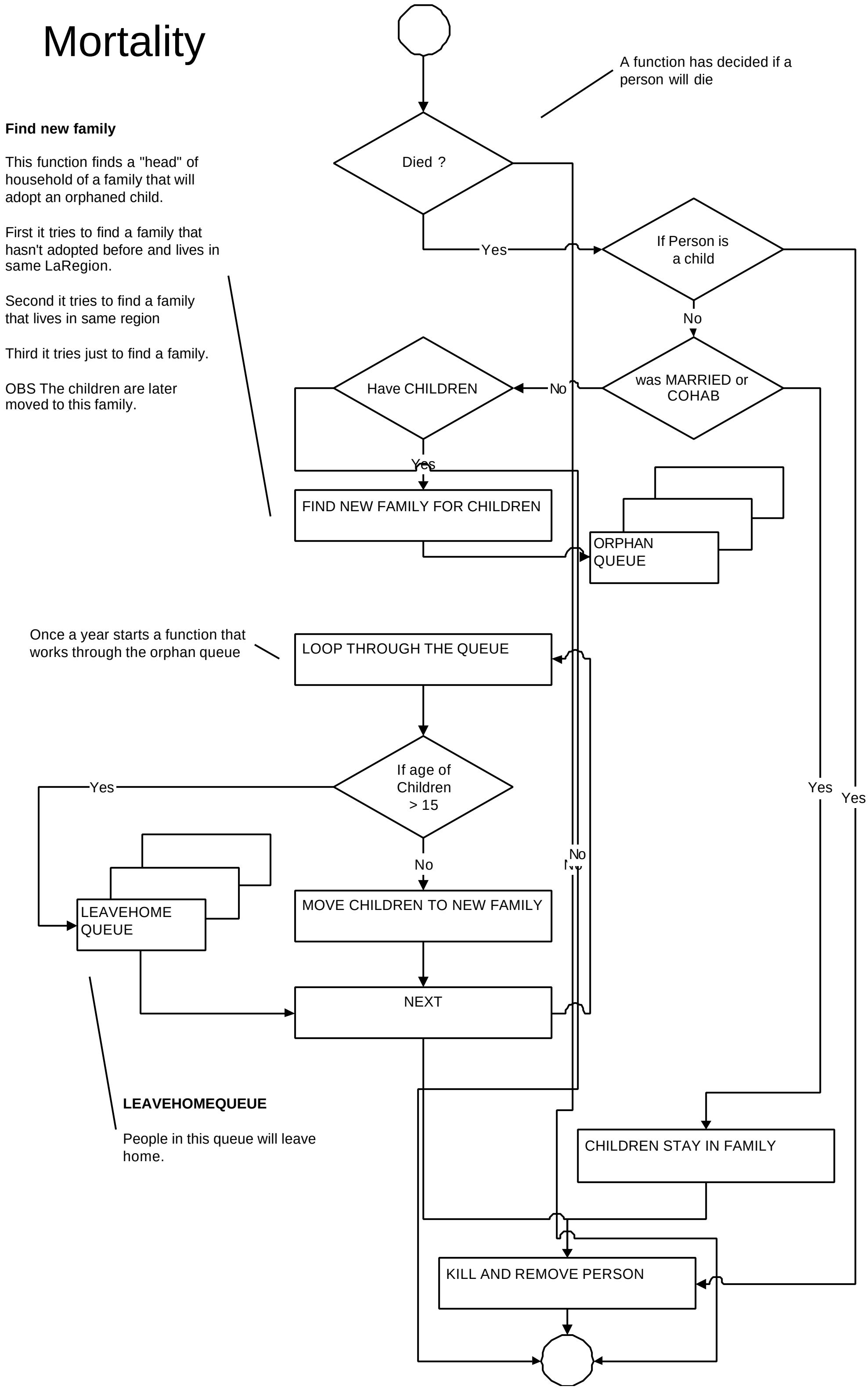

Mortality
Find new family
This function finds a "head" of household of a family that will adopt an orphaned child.
First it tries to find a family that hasn't adopted before and lives in same LaRegion.
Second it tries to find a family that lives in same region
Third it tries just to find a family.
OBS The children are later moved to this family.
A function has decided if a person will die
Died ?
Yes
If Person is a child
No
was MARRIED or COHAB
No
Have CHILDREN
Yes
FIND NEW FAMILY FOR CHILDREN
ORPHAN QUEUE
Once a year starts a function that works through the orphan queue
LOOP THROUGH THE QUEUE
If age of Children > 15
Yes
No
LEAVEHOME QUEUE
MOVE CHILDREN TO NEW FAMILY
NEXT
Yes
Yes
No
LEAVEHOMEQUEUE
People in this queue will leave home.
CHILDREN STAY IN FAMILY
KILL AND REMOVE PERSON

Figure A.2

Fertility

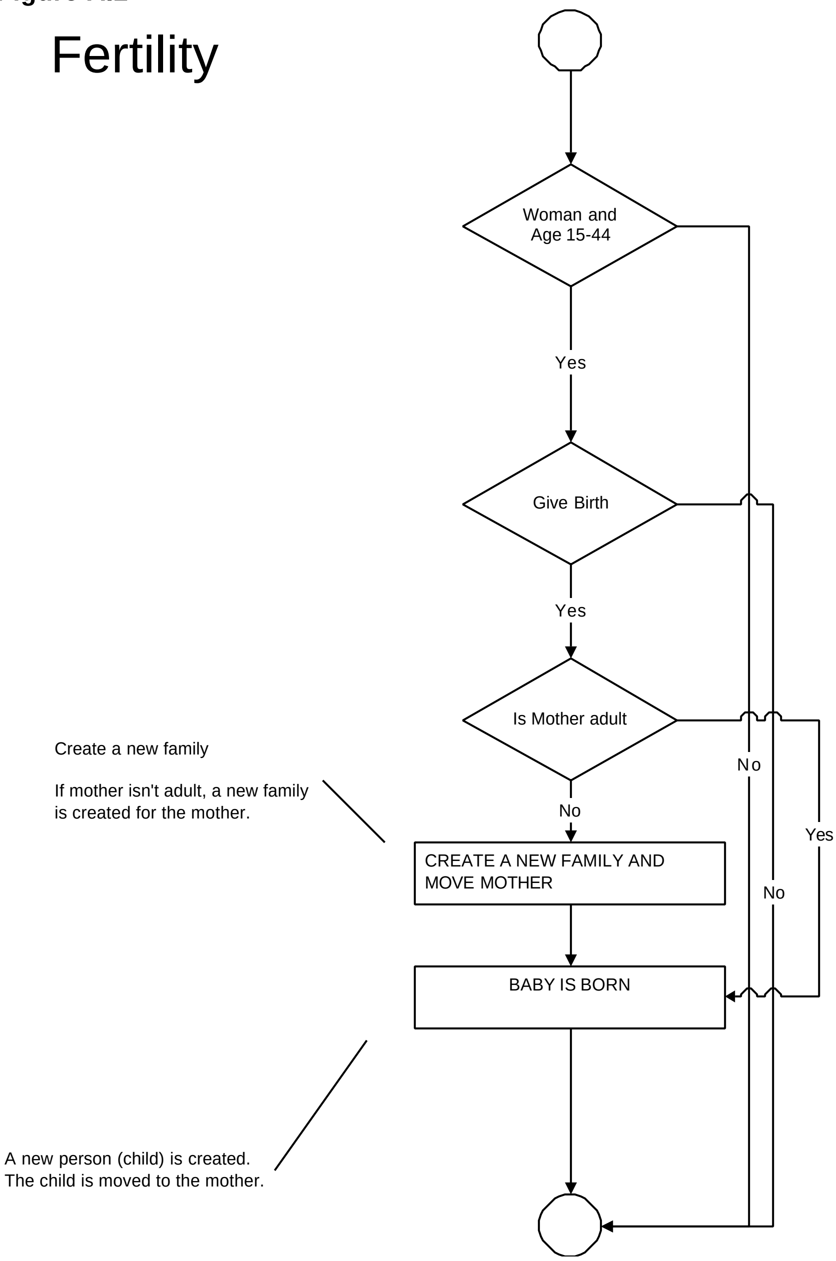


Figure A.3

Education
Grundskola - Gymnasium

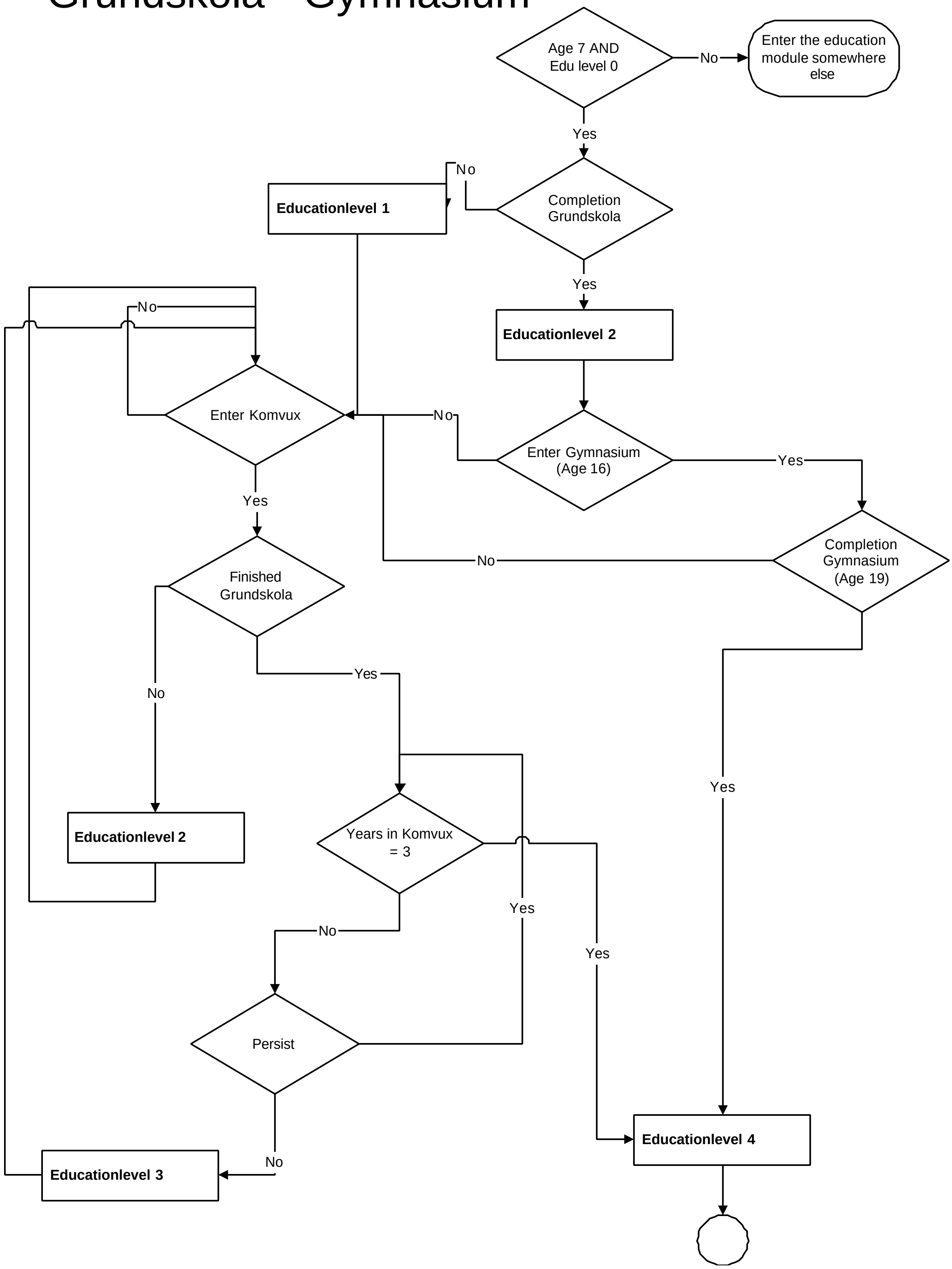


Figure A.4

Education
University - Postgraduate

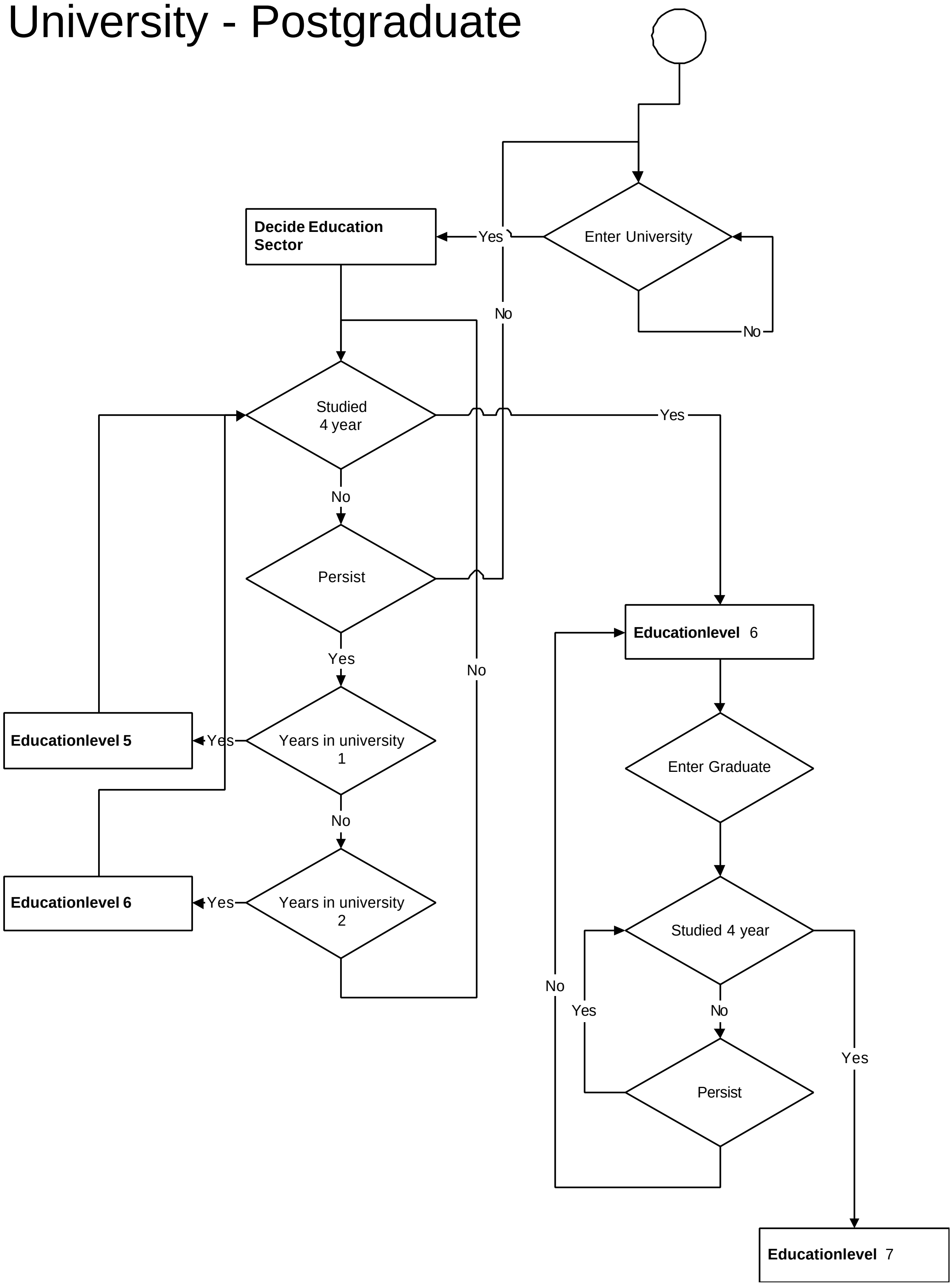
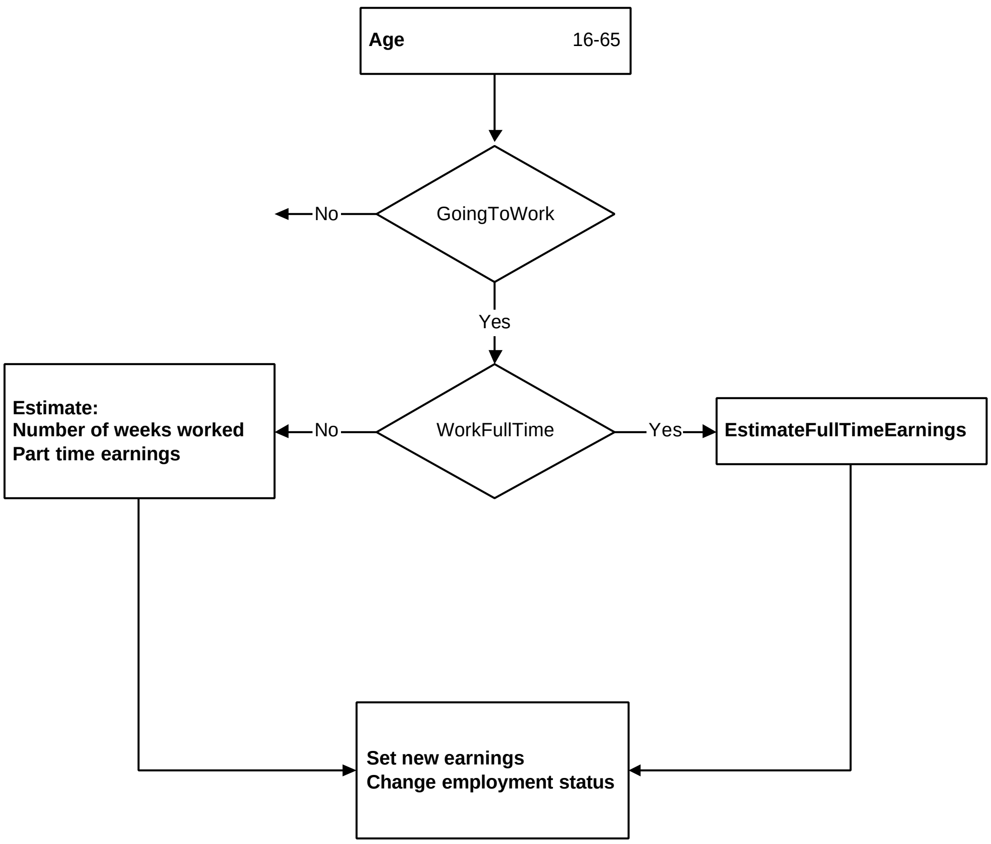


Figure A.5

Employment and earning



This graph describes the employment and earning module. The alignmnet is not described in this graph.

Figure A.6

Sambo

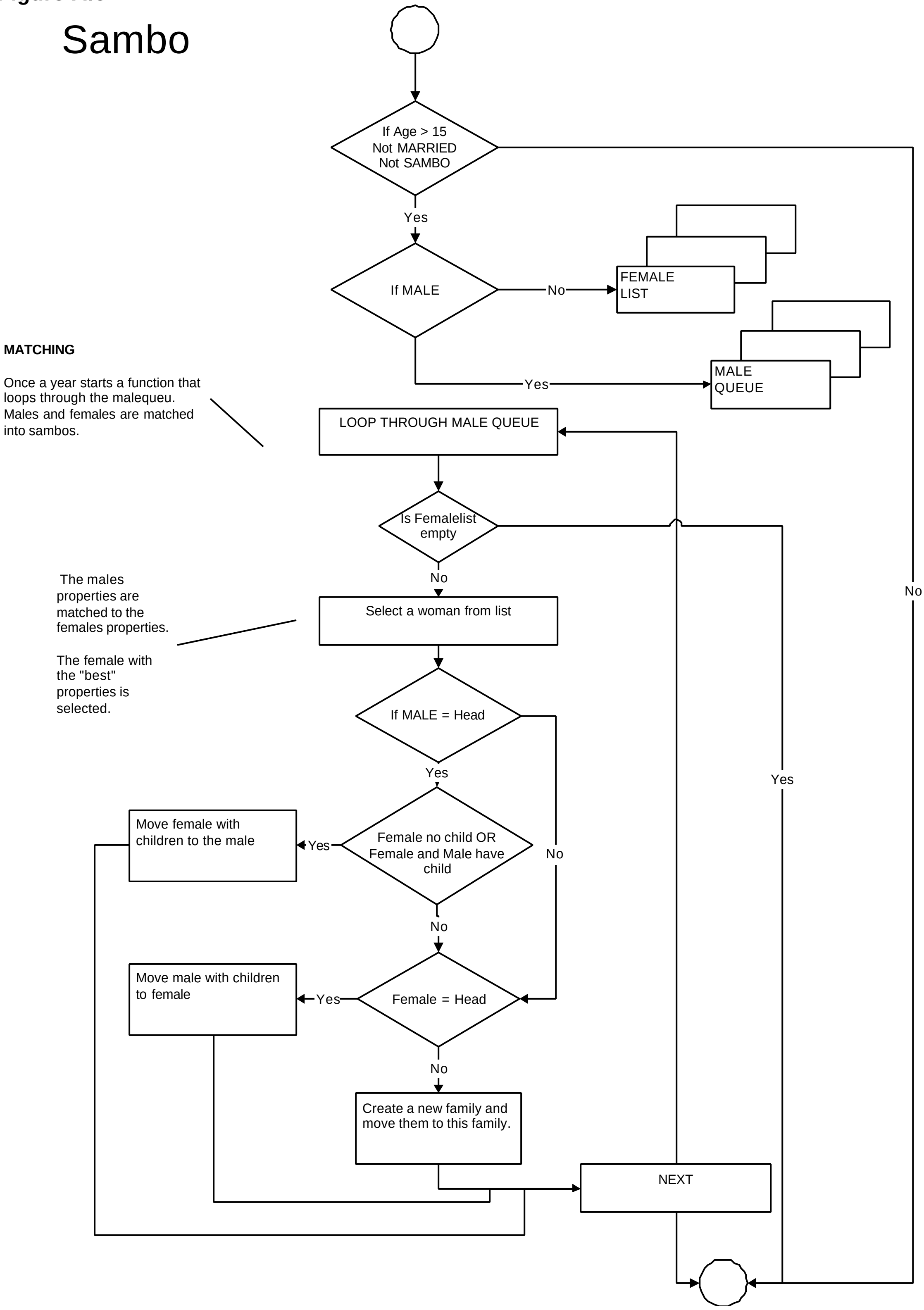


Figure A.7

Marriage

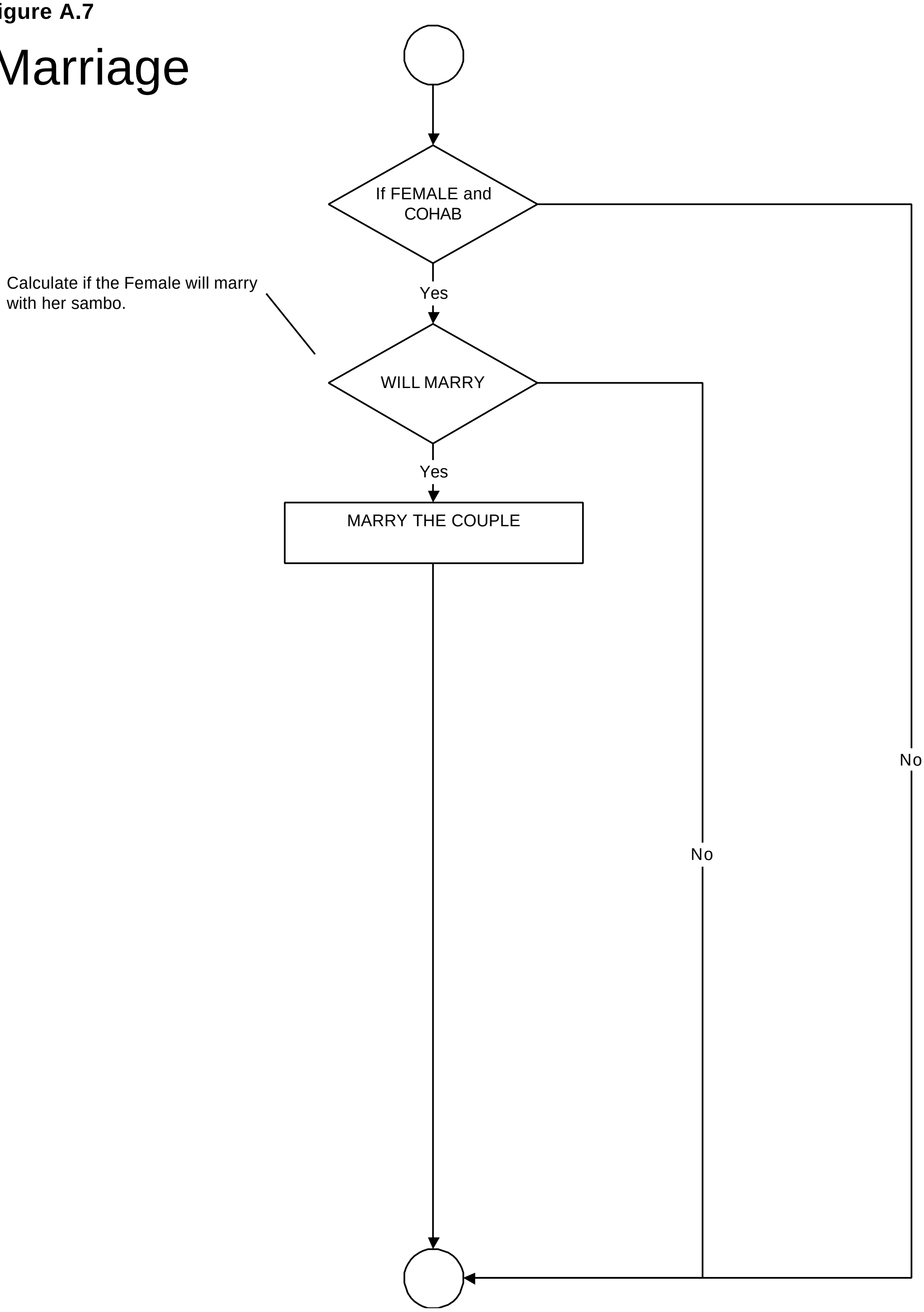


Figure A.8

Leaving home

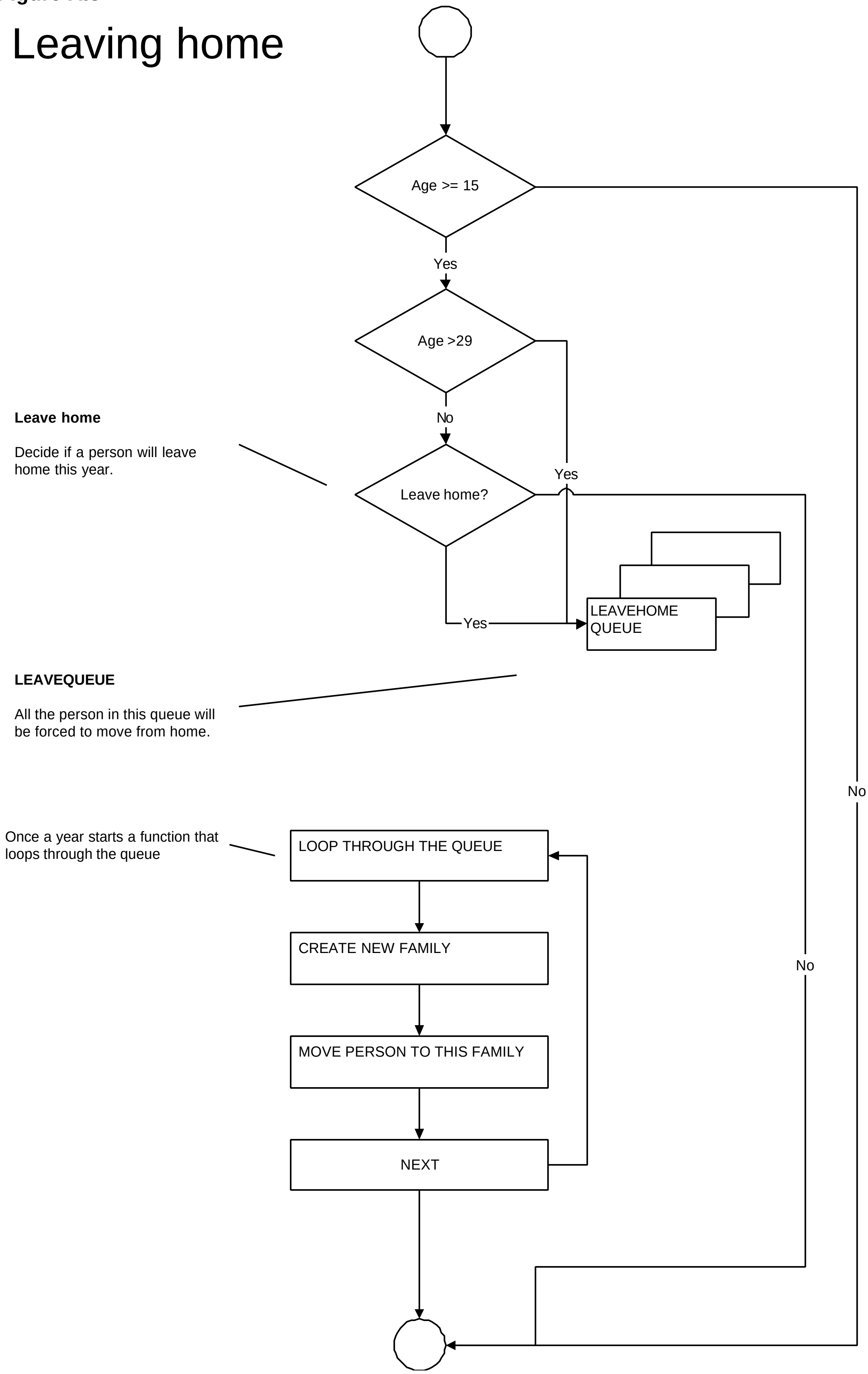


Figure A.9

Divorce

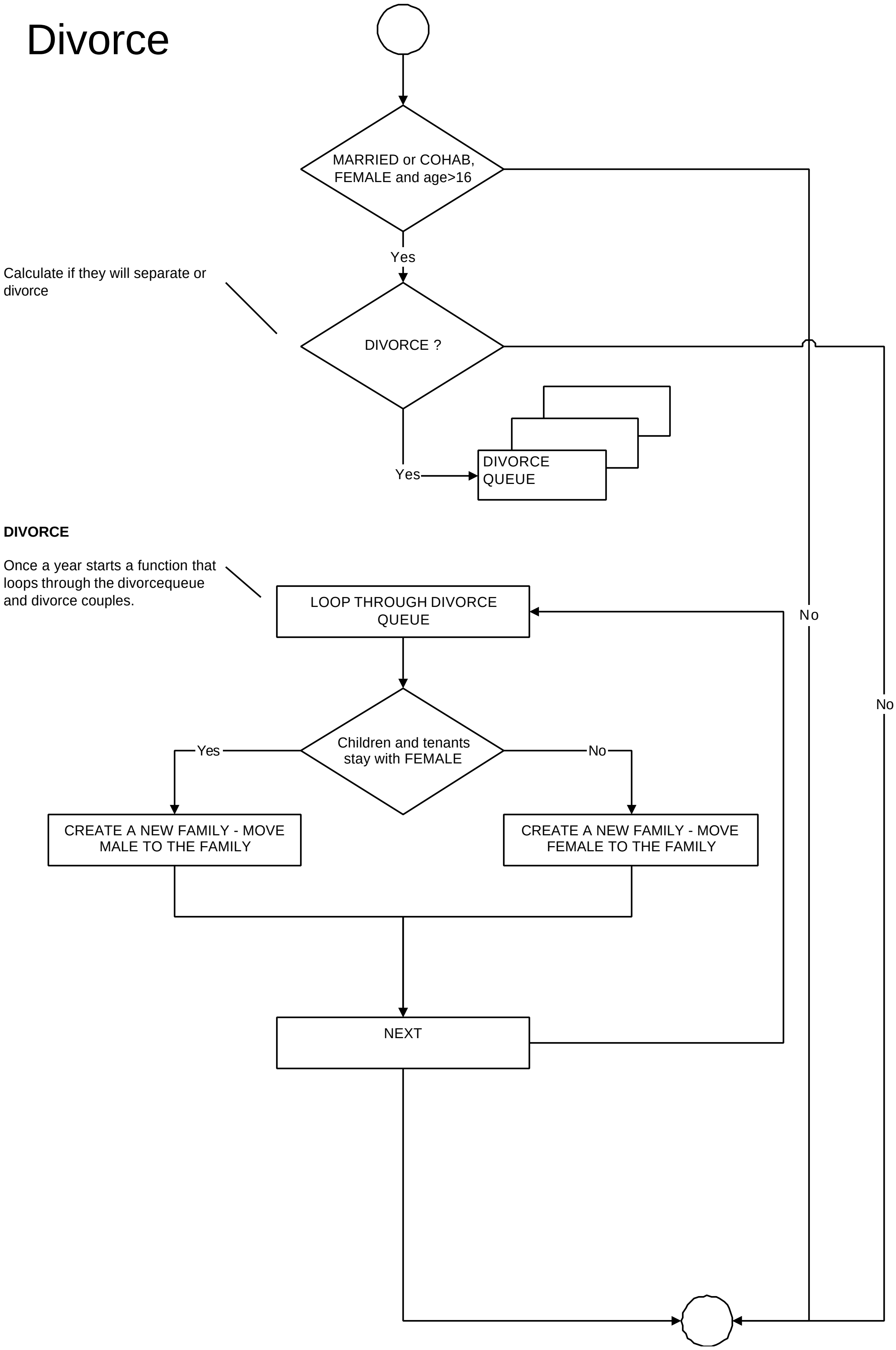


Figure A.10

Immigration

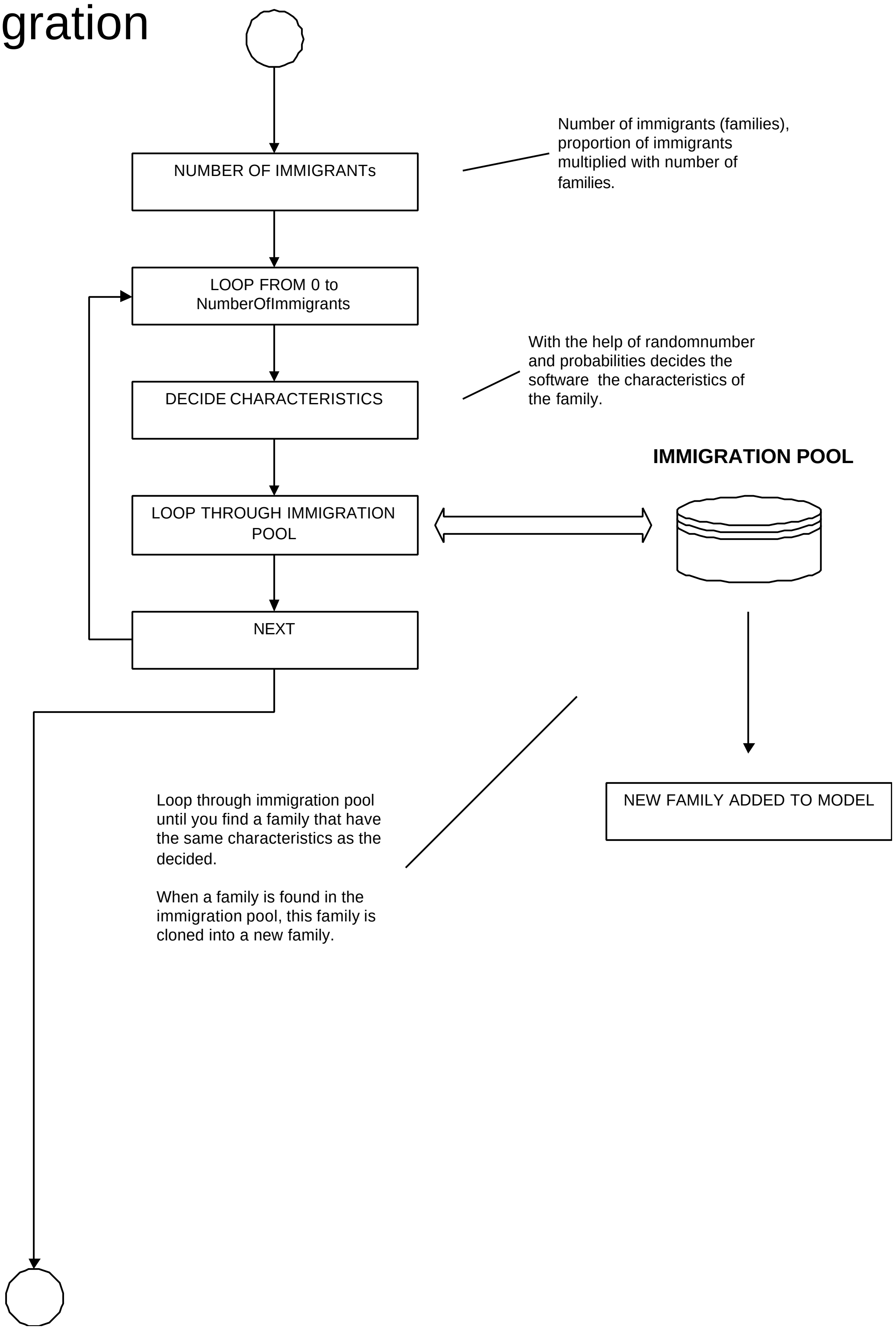


Figure A.11

Emigration

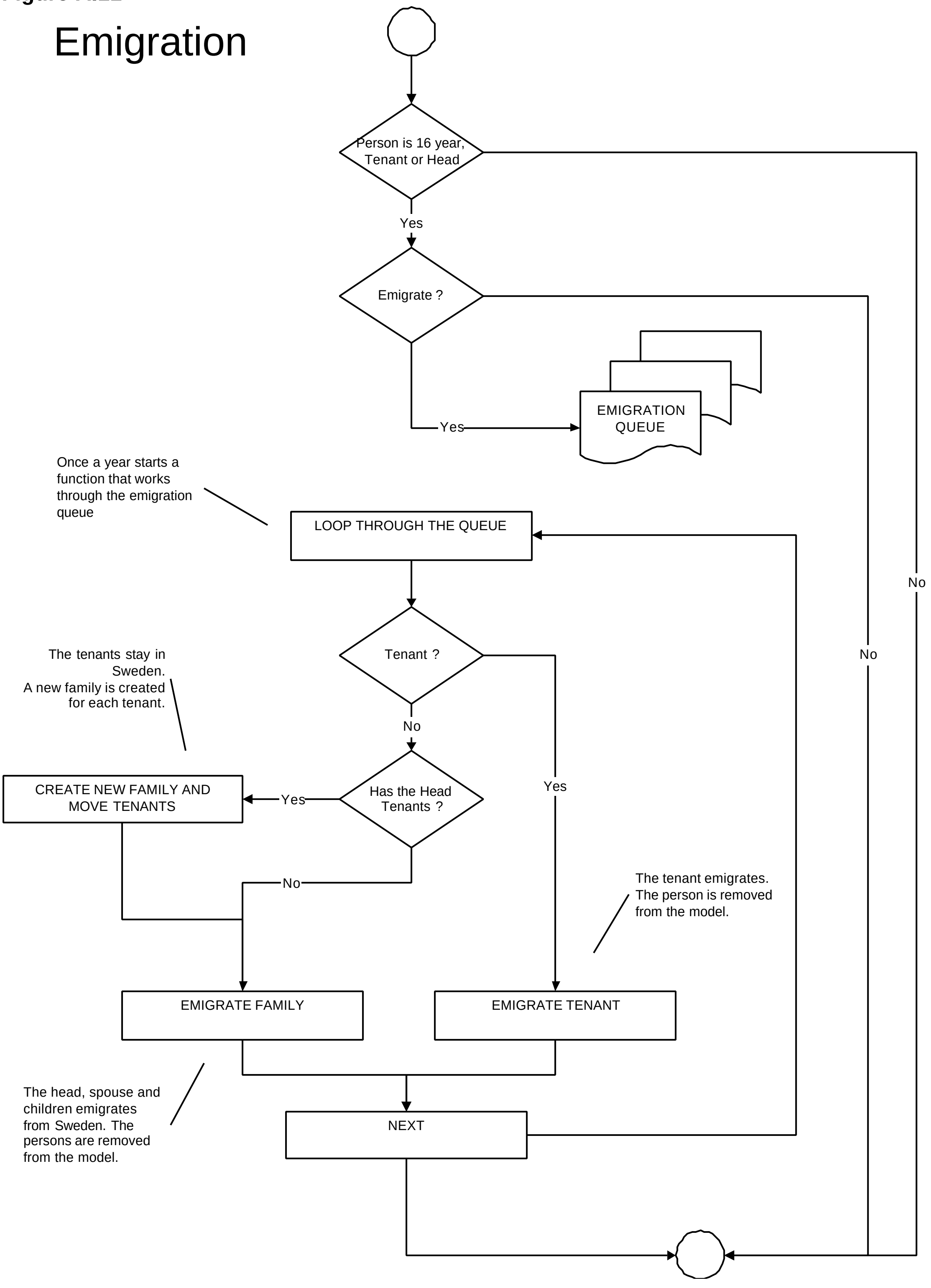
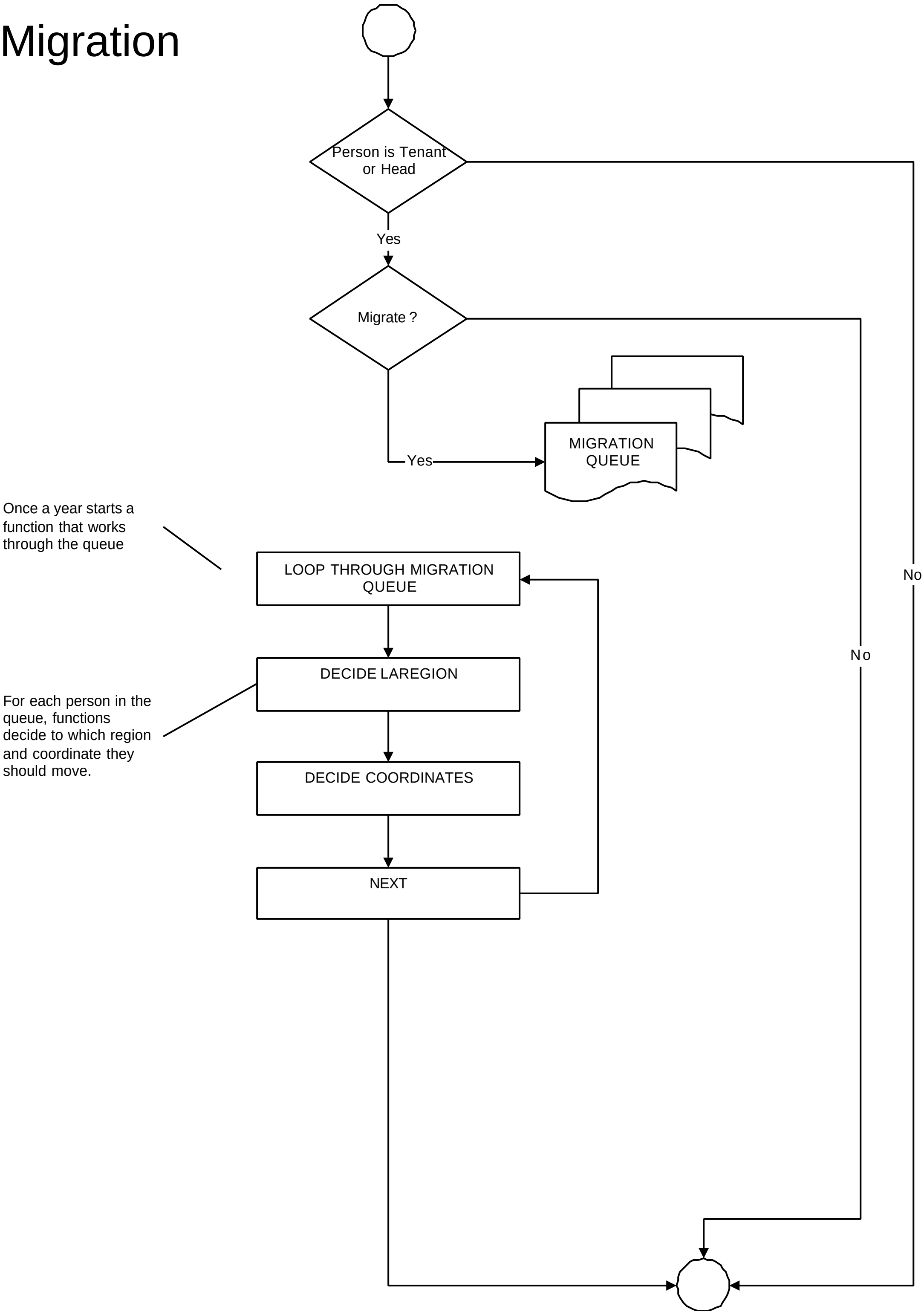


Figure A.12

Migration



References

Alfredsson, Eva and Magnus Åström. 1998. Employment and earnings module for SVERIGE: Documentation v. 1.0. Kiruna, Sweden: Spatial Modelling Centre.

Borgegård, Lars-Erik and Lars Lundin. 1993. Work and leisure. In *National Atlas of Sweden (*Ed. Kurt Viking Abrahamsson*)*. Stockholm, SNA Publishing.

Borgegård, Lars-Erik, Johan Håkansson, and Dieter K. Müller. 1996. The changing residential patterns of immigrants: The case of Sweden 1973-1992. *Yearbook of Population Research in Finland* 33: 173-183.

Caldwell, S. B. and L. A. Keister. 1996. Wealth in America: family stock ownership and accumulation, 1960-1995. In *Microsimulation for Urban and Regional Policy Analysis* (Ed. Graham P. Clarke) London: Pion.

Caldwell, Stephen B. 1997. *CORSIM 3.0 User and Technical Documentation*. Ithaca, NY: Cornell University.

Chénard, Denis. 1996. *Dynamic micro-simulation model for Canada*. Unpublished manuscript.

Clarke, G. P. 1996. *Microsimulation for Urban and Regional Policy Analysis* London: Pion.

Esko, Johan. 1999a. *Sambo Market*. Kiruna, Sweden: Spatial Modelling Centre.

Esko, Johan. 1999b. *Matching*. Kiruna, Sweden: Spatial Modelling Centre.

Esko, Johan. 1999c. *Marriage*. Kiruna, Sweden: Spatial Modelling Centre.

Esko, Johan. 1999d. *Divorce*. Kiruna, Sweden: Spatial Modelling Centre.

Finansdepartementet. 1994. *Sveriges ekonomiska geografi: Bilaga till långtidsutredningen*. 1995. Stockholm.

Fotheringham, S. 1991. Migration and spatial structure: The development of the competing destinations model. In *Migration models: Macro and micro approaches*, (Ed. J. Stillwell and P. Congdon), pp. 57-72. London: Belhaven Press.

Fransson, U. 1997. *Ungdomars hushållsbildning: Processer på en lokal bostadsmarknad*. (Young people's household formation: processes within a local housing market). Geografiska regionstudier 33, Uppsala, Sweden: Uppsala University.

Glava, Sonya. 1998. To return or not: The propensity for return migration from Australia. Paper presented at the 45th North American Meetings of the North American Regional Science Association in Santa Fe, New Mexico, on November 5, 1998.

Hammar, Tomas. 1992. Laws and policies regulating population movements: a European perspective. In *International migration systems: A global approach* (Ed. Mary M. Kritz, Lin L. Lim, and Hania Zlotnik). Oxford, UK: Clarendon Press. Pp 1-16..

Havemann, Robert and Barbara Wolfe. 1995. The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature* 33: 1829-1878.

Himes, Christine L. 1994. Age patterns of mortality and cause-of-death structures in Sweden, Japan, and the United States. *Demography* 31, 4: 633-50.

Holm, Einar and Gunnar Malmberg. 1997. Micro representation of spatial mobility. Kiruna, Sweden: Spatial Modelling Centre.

Hoem, Jan M. 1993. Public policy as the fuel of fertility: effects of a policy reform on the pace of childbearing in Sweden in the 1980s. *Acta Sociologica* 36, 1: 19-31.

Hoem, Britta; Hoem, Jan M. 1997. Fertility trends in Sweden up to 1996. Stockholm Research Reports in Demography. Stockholm, Sweden: Stockholm University, Demography Unit.

Isard, Walter, Iwan Azis, Matthew Drennan, Ronald Miller, Sidney Saltzman, and Erik Thorbecke. 1998. *Methods of Interregional and Regional Analysis*. Brookfield, Vermont: Ashgate.

Klevmarken, A. *Labor market – Model documentation*. University of Uppsala. Unpublished Manuscript.

Lichter, Daniel T., Gretchen T. Cornwell, and David J. Eggebeen. 1993. Harvesting human capital: Family structure and education among rural youth. *Rural Sociology* 58, 1: 53-75.

Lundh, Christer and Rolf Ohlsson. 1994. Från arbetskraftsimport till flyktinginvandring (From labor to refugee migration). Stockholm: SNS.

Maddala, G. S. 1983. *Limited dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

Merz, J. 1991. Microsimulation: A survey of principles, developments and applications. *International Journal of Forecasting* 7: 77-104.

Minder, C. E. 1991. A conceptual and methodological discussion of transnational comparisons of socioeconomic mortality differences and an example. In *Socio-economic differential mortality in industrialized societies*. 7: 278-94.

Morrison, R. J. 1997. DYNACAN, the Canada Pension Plan Policy Model: Demographics and Earnings Components. Microsimulation in Government Policy and Forecasting International Conference on Combinatorics, Information Theory and Statistics, Portland, Maine, July 18-20.

Öhman, Marianne. 1999. Leaving home documentation: The new module for SVERIGE. Kiruna, Sweden: Spatial Modelling Centre.

Oskamp, Anton. 1997. *Local housing market simulation: A MICRO approach*. Amsterdam: Thesis publishers.

Orcutt, G. 1957. A new type of socio-economic system. *Review of Economics and Statistics* 58: 773-797.

Rephann, Terance. 1998. Explaining property values: Quantitative evidence from Sweden. Paper presented at the 45th annual meeting of the North American Regional Science Association in Santa Fe, NM, on November 11-14, 1998.

Rephann, Terance. 1999a. The mortality module in SVERIGE. Documentation v. 1.0. Kiruna, Sweden: Spatial Modelling Centre.

Rephann, Terance. 1999b. The fertility module in SVERIGE. Documentation v. 1.0. Kiruna, Sweden: Spatial Modelling Centre.

Rephann, Terance. 1999c. The education module in SVERIGE. Documentation v. 1.0. Kiruna, Sweden: Spatial Modelling Centre.

Sheehan, Maura. 1998. Unemployment duration in an unemployment blackspot. *Labour* 12: 643-673.

Statistiska centralbyrån (SCB). 1995. Produktbeskrivning med kvalitetsdeklaration – Hushållsundersökningen, även benämnd Inkomstfördelningsundersökningen (HINK). Stockholm, SCB.

Statistiska centralbyrån (SCB). 1994. *Fakta om den svenska familjen*. Demografiska rapporter 1994: 2.

Statistiska centralbyrån (SCB). 1995. *Kvinnors och mäns liv – Del 2 Parbildning och separationer*. Demografiska rapporter 1995: 2.2. Stockholm: SCB.

Sundstrom, Marianne; Stafford, Frank P. 1992. Female labour force participation, fertility and public policy in Sweden*. European Journal of Population/Revue Europeenne de Demographie* 8, 3: 199-215.

Tasiran, A. C. 1995. *Fertility dynamics: spacing and timing of births in Sweden and the United States.* Contributions to Economic Analysis 1995. Amsterdam, Netherlands: Elsevier Science Publishers.

Vencatasawmy, Coomaren and Einar Holm. 1998. Modules, objects, attributes, driving forces, and preliminary causal structure. Kiruna, Sweden: Spatial Modelling Centre.

Vencatasawmy, Coomaren and Neil Swan. 1998. Using a unique database for building a spatial microsimulation model. Kiruna, Sweden: Spatial Modelling Centre.

Vencatasawmy, Coomaren P. 1999a. Immigration module: Version 1.1 documentation. Kiruna, Sweden: Spatial Modelling Centre.

Vencatasawmy, Coomaren P. 1999b. Emigration module: Version 1.0 documentation. Kiruna, Sweden: Spatial Modelling Centre.

Vencatasawmy, Coomaren P. 1999c. Migration module: Version 1.1 documentation. Kiruna, Sweden: Spatial Modelling Centre.

Wallace, Helen M., Jae Woong Hong, and Anders Ericson. 1995. Comparison of infant mortality in the United States and Sweden. *Journal of Tropical Pediatrics*. 31, 4: 223-228.