The education module for SVERIGE: Documentation V 1.0

Terance J. Rephann
Spatial Modelling Centre
Box 839
S 981 28 Kiruna, Sweden
http://www.smc.kiruna.se
tre@smc.kiruna.se

March 4, 1999

1.0 Microsimulation model history and structure

SVERIGE is a dynamic economic-demographic-environmental spatial microsimulation model for Sweden. By microsimulation is meant that the model represents lifetime events and choices of individual units (or objects) as a combination of structural factors (usually included in discrete choice models as independent variables or used to organise transition matrices) and random disturbance (a Monte Carlo randomisation component). By dynamic is meant that microunit ageing and development occurs in a life cycle pattern, with initial microunit conditions being changed for subsequent periods by counters and sequenced model equations. Its core is based upon CORSIM (Cornell Microsimulation Model), which itself is a modification of Guy Orcutt's DYNASIM (Dynamic Microsimulation Model), the first dynamic microsimulation model of its kind (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).

SVERIGE will differ in several important respects from its CORSIM parent and DYNACAN sibling. First, SVERIGE is a Swedish model and thus must explain behaviour in a different institutional context than either the CORSIM and DYNACAN North American models. The model core of CORSIM consists of nine modules (mortality, fertility, marriage, divorce, re-marriage, leaving home, education, employment and earnings, and immigration) that describe the human life cycle. Each module consists of equations that describe the behavioural responses of individuals as a function of their demographic and socio-economic characteristics. Although these module equations are informed by economic theory, such as Becker's theories of marriage, divorce, education, fertility, and labour force participation, they are quite sensitive to cultural and institutional peculiarities. Thus, one cannot always transport the specifications and parameters used by CORSIM to Sweden. For instance, power relations between men and women, the degree of class distributional equity, the elaborateness of social support mechanisms, and the varieties of social groupings (e.g., married couples, cohabitants, families, households) define the social context in which individual decisions are made and constrain the ways in which microactors interact. Therefore, while the life-cycle model that constitutes the CORSIM core will remain basically the same (see figure1.1), the equations that explain transitions between various stages of the life-cycle may vary considerably between the North American and Swedish contexts.

Second, SVERIGE is a spatial model while CORSIM is not. In fact, SVERIGE will be the first national-level spatial microsimulation model. Geographical environment and distance play no role in aspatial models. However, SVERIGE will model individual spatial transitions (such as internal migration) and model life-cycle transitions described by the model core within a spatial context. In addition, certain geographical units (including land parcels, neighbourhoods, and labour markets) will have attributes that influence the attributes of units such as individuals, households, employers, and homes (see figure 1.2. for a full listing of these units (i.e., "objects) and their attributes) and vice versa. For instance, property values, pollution levels, and housing characteristics will change and, in turn, modify choices made by other microactors (or objects) within the microsimulation model. Furthermore, because objects have geographical attributes, the model will be capable of generating geographically detailed reports that may interest regional scientists and policymakers.

Third, SVERIGE will emphasise environmental applications. An important premise of the model is that non-production, non-point, household consumption activities generate many unsafe emissions such as heavy metals, carbon monoxide, and sewage. This orientation arose for both empirical and practical reasons. The empirical justification is that if current trends are continued into the future, consumer generated pollution will make up a substantial proportion of overall pollution levels. This is expected to occur because point pollution is technologically and financially easier to reduce than non-point emissions (Tietenberg 1988). There are also practical reasons for not extending the model to production point emissions sources, because to develop modules that explain large firm behaviour would introduce unmanageable complexity and require proprietary firm-level data that are unavailable to the project.

Figure 1.1 The Sverige Model

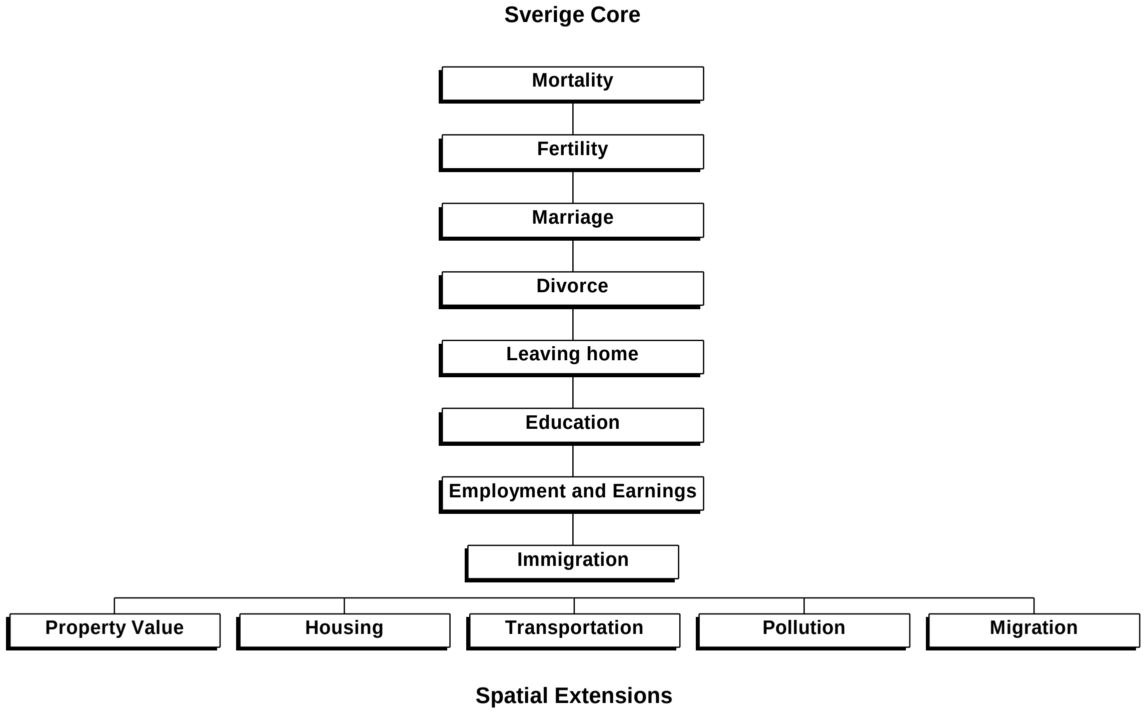
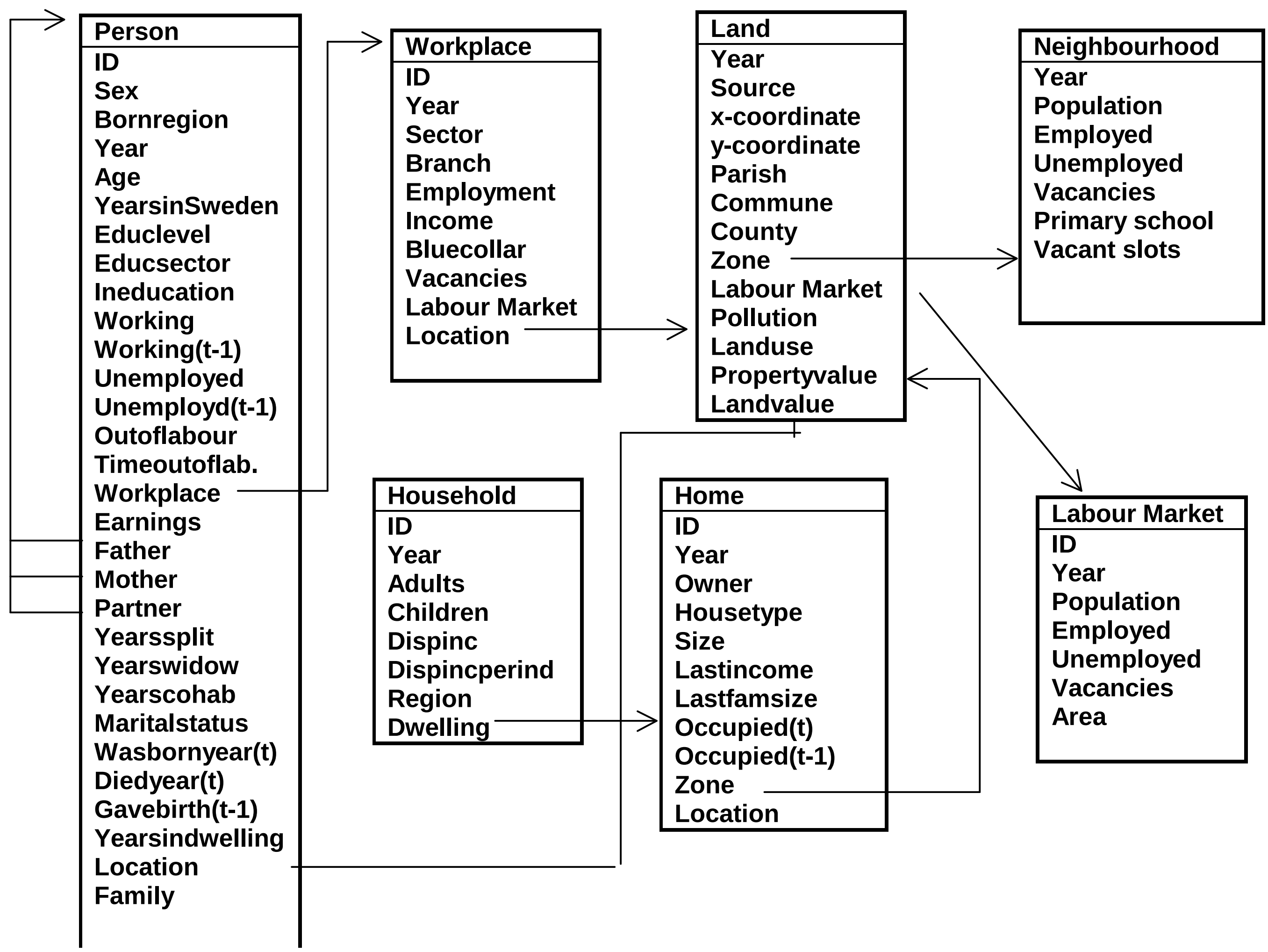


Figure 1.2 Relations between objects and attributes



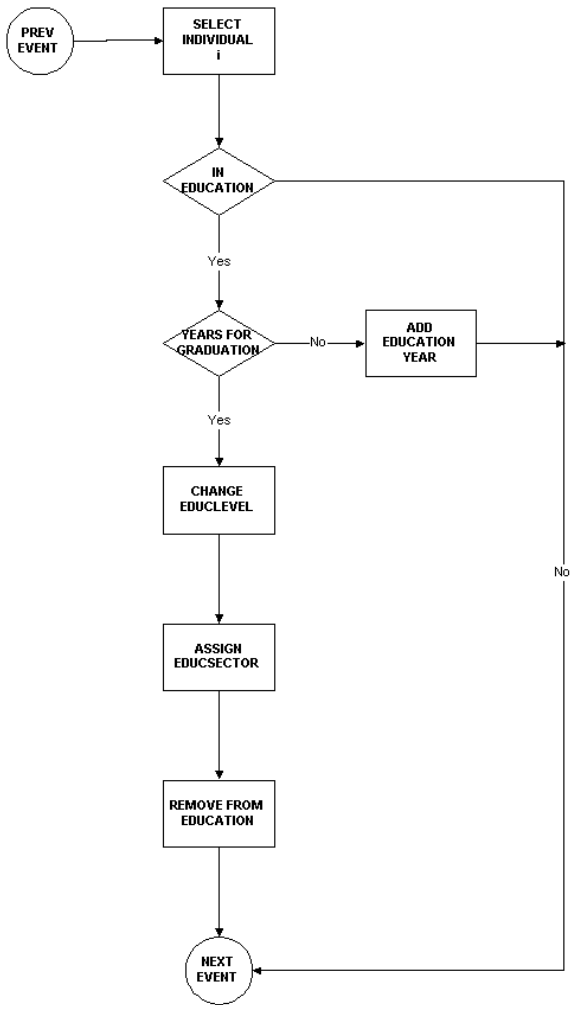
2.0 Rationale and role for an educational module

The educational module determines educational entry, persistence, and graduation. Education is a key determinant of social and economic well-being for individuals. As the microsimulation model is now constituted, educational achievement appears in each of the module equations. It reduces the likelihood of mortality, decreases fertility, increases likelihood of subsequent employment and income, increases likelihood of leaving home and interregional migration, and enhances the likelihood of divorce.

SVERIGE uses logistic regression equations and hard-coded probabilities to determine completion of compulsory school, entrance into gymnasium, completion of gymnasium, entry to adult education, persistence through adult education, entry to college, persistence through college, entry to graduate/professional school, and persistence through graduate/professional school. These equations are estimated using Swedish longitudinal data obtained from the TOPSWING (*Total Population of Sweden Individual and Geographical database*) database housed at SMC. Only full-time students are modelled, but both traditional and adult students may participate. At any time, students may be selected to discontinue education but they are eligible to rejoin education later on. The routine invoked for determining education entry and persistence is exhibited in figure 2.1 below.

The education module plays two additional roles in SVERIGE. First, it identifies educational participants, who are subsequently moved to college and university towns using a special migration destination transition matrix based on the number of annual college vacancies available. Second, participation operates in tandem with employment as a life-history place-keeper that marks what activity a given individual engaged in during each year of the productive years of his/her life. Those enrolled in education will not be employed full-time and vice-versa.

Figure 2.1 Education module



3.0 Structure of CORSIM educational module

CORSIM uses a series of logistic regression equations to determine progression through the final years of high school, graduation, entry to college education, persistence through college, graduation from college, entry to graduate school, and persistence through graduate school. The CORSIM equations were estimated using U.S. longitudinal survey data on high school students as they progressed from 10th grade through their educations and careers. These equations are listed in Appendix A.

The equation specifications give heavy emphasis to family and household influences on educational achievement. As discussed in section 6.0, this emphasis can be justified by the literature in social science. However, since SVERIGE is a spatial model, it may be constructive to introduce neighbourhood and regional influences on educational achievement as well. In CORSIM, the probability for being in school (INSCH) is computed as follows:

(1) $$\Pr\mathbf{ob}(\mathbf{INSCH}=1)=\frac{\mathbf{e}^{\mathbf{ß'x}}}{1+\mathbf{e}^{\mathbf{ß'x}}}$$

$$b'\mathbf{x}=\beta_0+\beta_1\mathbf{GRDCOMP}+\beta_2\mathbf{FEMALE}+\beta_3\mathbf{AGE}+\beta_4\mathbf{AGE2}+\beta_5\mathbf{KIDS}+\beta_6\mathbf{COMP16}+\beta_7\mathbf{DCOMPHS}$$
$$+\beta_8\mathbf{MCOMPHS}+\beta_9\mathbf{DSOMECOL}+\beta_{10}\mathbf{MSOMECOL}+\beta_{11}\mathbf{DCOMPCOL}+\beta_{12}\mathbf{MCOMPCOL}$$
$$+\beta_{12}\mathbf{MARRY}++\beta_{12}\mathbf{HOME}$$

where:

β_0	constant
GRDCOMP	number of grades completed
FEMALE	dummy variable for sex
AGE	age in years
AGE2	age squared
KIDS	number of kids
COMP16	dummy variable for completing the 16the grade level
DCOMPHS	dummy variable for the father completing high school
MCOMPHS	dummy variable for the mother completing high school
DSOMECOL	dummy variable for the father completing some college
MSOMECOL	dummy variable for the mother completing some college
DCOMPCOL	dummy variable for the father completing college
MCOMPCOL	dummy variable for the mother completing college
MARRY	dummy variable for married
HOME	dummy variable if parent(s) own(s) home

Because the CORSIM documentation is fairly sketchy, three important questions about the module operation cannot be answered conclusively without first examining the C Code used in CORSIM. It may be useful to review these questions and their answers to determine if they offer valuable insights into how to design the SVERIGE educational modules. These questions are: (1) Who are the “at-risk” populations for transitions to the various levels of educational achievement? (2) What happens to students who persist but do not graduate? and (3) How does attending school affect ones labour force participation during the period of study? Are they mutually exclusive activities, compatible activities, or does study merely affect the probability of searching for work or being employed and/or vice-versa.

First, it is not entirely clear from the current model documentation who the “at-risk” population is for transition to college/university studies and graduate studies. Presumably only secondary school graduates are eligible for college studies and equation A.3 in Appendix A. is applied only to them. However, it is not clear whether this equation is applied to the entire pool of high school graduates, a specific sub-class of these graduates, or simply the graduating cohort from the previous year. Caldwell (1993) suggests that there is a routine that feeds a certain proportion of unsuccessful students (i.e., students who dropped out of college) within the age group 17-21 back into the educational queue. However, this routine and its arguments (if there are any) are not described. The documentation should be clearer on this issue.

Second, there is no indication given about what sequences of choices are available to entrants who progress through the third year of college education but do not graduate from college. Equation A.7 is used to compute the probability of graduation but it is somewhat unusual in being the first equation to introduce age as a determinant of success at this stage. If the equation were being applied to the same high-school cohort, AGE would not be necessary since everyone in the cohort would have approximately the same age. However, it might be useful in differentiating achievers from non-achievers if members of graduating cohorts who failed to obtain their diploma were put back in the educational queue to try once again. The quadratic form in which AGE is entered in equation A.7 seems to indicate that they are more likely to be successful as they grow older and are forced to repeat. According to Caldwell (1993), this is, indeed, what happens, but the manner in which this routine operates is not described.

The third question is how participation in education affects participation in the work force. On this issue, Caldwell (1993) is much clearer. He argues that it is not an unreasonable simplification in the U.S. context to make the activities mutually exclusive. That is to say, if an individual is enrolled in school, he/she is not a participant in the labour market, either as a searcher or as an employee.

4.0 Reasons for re-estimating the equations using Swedish data

Education in Sweden is somewhat different than in the U.S. This issue needs to be taken into consideration when adapting the CORSIM structure to Sweden. In the U.S., there are two levels of compulsory education, primary and secondary. Primary school consists of six years of education during which basic reading, writing, and arithmetic skills are learned. Secondary school consists of junior-high (typically grade levels seven and eight) and high school (grade levels nine through twelve) during which more specialised career preparatory coursework is taken. Most school districts offer both vocational and academic (college preparatory) curricula at the secondary school level. With parental permission, students may “drop-out” of high school after completing a minimum of eight years in many U.S. states. Most primary and secondary schooling is funded and administered at the local and state levels. Private education is also an alternative in some communities. Because of this, there is a considerable amount of heterogeneity in educational resources, school curricula, and educational outcomes.

Post-secondary education is similarly divided into several levels in the U.S. Students have a number of options. Junior or community colleges (or technical institutes) offer one year certificate or two-year associate degrees. They are typically partly funded by city, county, and state governments and offer terminal occupational programs of study (e.g., nursing, electronic technician) as well as transfer programs for continuing studies at the college or university level. Colleges often offer only four year bachelor degree programs of study. Universities offer both undergraduate (Associates, Bachelors) and graduate/professional programs (e.g., Masters, Ph.D., LL.D., MD). Most funding for colleges and universities is derived from state sources. However, there are also hundreds of private and denominational colleges or universities in the United States. Most U.S. college students, even those enrolled in public institutions, pay a portion of their expenses in tuition and fees (though the largest portion of the cost of education is subsidy). In addition, low and middle income students are eligible for various types of financial aid, in the form of state and federal grants and subsidised loans, for covering educational expenses. As a result, access to higher education is fairly widespread. Approximately, one-half of U.S. high school students advance to post-secondary education after graduating.

The Swedish situation is markedly different in several ways. Compulsory school is called Grundskola and consists of nine years of basic preparation. Approximately ninety percent of Swedish students advance to the next level, Gymnasium, which offers vocational and college preparatory programs. Gymnasium had in the past offered two, three, and four year educational tracks in both vocational and academic concentrations. However, in the past few years this has been standardised to three years. After Gymnasium comes college (i.e., Högskola) for approximately twenty-five percent of Swedish high school graduates. Högskola offers programs of different duration, including two, three, and four years. However, the typical undergraduate education is three years rather than four as it is in the U.S. College or university is followed by graduate or professional school. Student financial access is less “patchy” than in the U.S. Swedish students in higher education receive free tuition and monthly stipends for covering monthly expenses and are eligible for student loans. Therefore, socioeconomic and financial factors may be less important in distinguishing college attendees from non-attendees and may play a smaller role in student attrition.

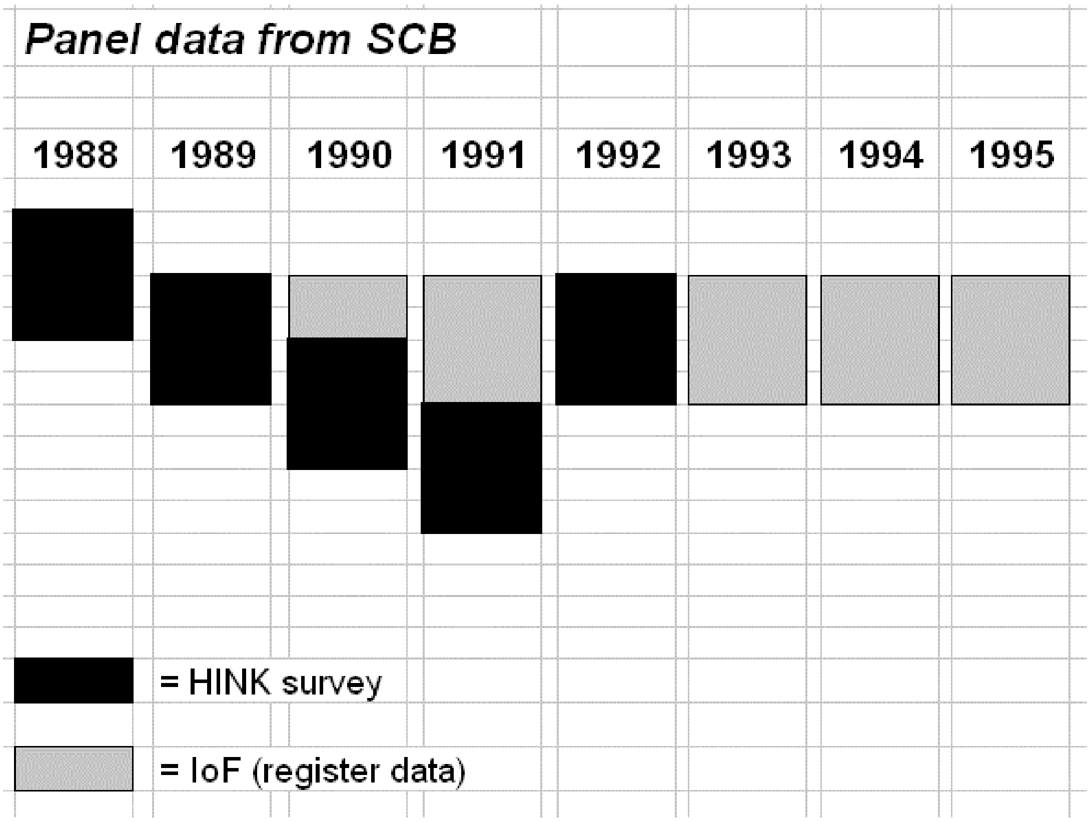
5.0 Data

The TOPSWING database contains two variables that can be used to examine entrance, persistence, and graduation in education. These variables are: (1) UTB – a code indicating whether an individual studied at the gymnasium-college/university level during the fall or not and (2) HSUN – a code that indicates both the highest level of educational achievement and the field of specialisation (if any) that he/she achieved. HSUN is available for every year in the database 1985-95. However, UTB is available for only the final three years (i.e., 1993-95).

There is no information on graduation per se in the database. However, it is possible to infer graduation by examining changes in HSUN. For instance, if HSUN changes from a code indicating that a person had a high school vocational degree as his/her highest degree in 1994 and this changed to a code indicating a two year occupational degree in 1995, we can infer that he/she graduated in 1995. However, we cannot infer when he/she actually began to study without additional information. For selected cohorts of high school graduates, on the other hand, it should be possible to examine entrance and persistence through post-secondary education.

Another promising source of demographic and socio-economic data is HINK and income register (IoF) sample data, which contains information on characteristics of approximately 25,000-30,000 residents. This data was obtained from the Finance Ministry for the years 1988-95 along with two educational files that contain a number of additional variables that allow us to distinguish educational entry and persistence for those enrolled in adult education (i.e., “Komvux” or municipal adult education) and higher education. Data regarding Komvux attendance is contained in a file named KOMVUX. Information concerning college attendance is contained in a series of files labelled HS01-HS35, making it possible to identify first-time attendees and their year and semester of attendance. There is substantial (but not complete) overlap between the annual samples. A panel of approximately 10,000 individuals can be created for the purpose of examining entry through completion (see figure 5.1).

Figure 5.1 HINK and IOF panel data from SCB



Unfortunately, each attempt to construct longitudinal samples for the school-going population using the HINK survey data resulted in biased samples. University attendees, for instance, were much older and more likely to be female than official survey data indicated. Moreover, the resulting samples were fairly small, much smaller than could be obtained with samples drawn from the TOPSWING database. Therefore, despite its limitations, it was ultimately decided to use the TOPSWING database alone. Instead of processing all of the individuals in the database, a sample of 458,911 individuals representing approximately 5% of the total Swedish population was selected. This sample was obtained using a stratified random sampling method. This sample was used in order to reduce the amount of time required for processing and analysing the data.

6.0 Determinants of Educational Achievement

It seems reasonable to expect that educational achievement is influenced by individual, past family, present family, neighbourhood, and regional characteristics. Among these categories, however, past family attributes appear to play the dominant role. For instance, intergenerational transmission of educational achievement is a continuing area of study in economics and recent research suggests that the transmission is very strong (Haveman and Wolfe 1995) and that educational achievement levels of mothers have a greater effect than fathers. Other indicators of parental maturity and class status may also be associated with achievement, for instance, the age of mother at marriage and first birth (Lichter et al. 1993). Presumably, younger mothers are less experienced and have accumulated less human capital to pass on to their children. In addition, mothers who are employed during the initial years of child's life have less time to devote to nurturing their children and this has been found to have a negative effect on child development, though employment later on does not hinder achievement.

Evidence has been found that disruptive family events and limited family resources lower educational achievement. For instance, divorce and family dissolution contribute to increased childhood stress and remove potential positive role models from childrens' lives. (Haveman 1990; Lichter et al. 1993). Two parent families (Wenk and Hardesty 1991; Lichter et al. 1993) are more likely to rear children with high educational achievement than single parent families, and two-parent families in which the original father is present are more successful still than step-families. Families with a religious orientation appear to be more successful (Haveman 1990). Family class, income, and poverty affect are also important. Families with lower incomes have fewer resources for nurturing children, maintaining a secure environment, and exposing them to stimulating experiences. Frequent moving and changing of jobs by parents, which requires the children to change schools, creates disruption and stress which may hinder subsequent child achievement (Haveman 1990).

The size and composition of the family, because it affects the amount of attention that is available to any one child, may have an affect on his/her achievement. Smaller families appear to be more successful than larger families in fostering higher educational achievement (Haveman et al 1990; Lichter et al. 1993). Birth order may matter also (Travis and Kohli 1995). First born children, perhaps because they command a greater share of full parental attention than their brothers and sisters, are higher achievers than subsequent births. Some researchers have investigated the role of sibling sex composition on educational achievement (Kaestner 1996).

Some individual level demographic and socio-economic characteristics may affect educational attainment. For instance, gender, race, and national origin may affect an individuals educational experiences. If institutionalised education systematically discriminates against individuals on the basis of certain observable features, they will have more obstacles to overcome and will exhibit lower educational achievement. If discrimination is prevalent only in the labour market then Becker's model of discrimination predicts that they are less likely to invest in human capital and should also exhibit lower educational achievement. Immigrants may face certain barriers because of discrimination or, perhaps, they have not accumulated enough cultural know-how or language familiarity to facilitate educational achievement. Individual decisions may also be affected by tastes and preferences and life events. For instance, teenage mothers significantly reduce their educational achievement potential as do those who elect to start early careers.

The decision to start a new family may have some bearing on an individual's educational aspirations and achievement. The decision to have children, as mentioned above, should increase the costs of investing in educational advancement. Moreover, the background of one's spouse may have an influence on educational aspirations and resources available for pursuing additional educational or training. For instance, a more educated spouse may increase pressures to achieve parity. Having an employed spouse and additional resources in the form of greater earnings, non-labour income, or wealth may increase financial access to higher education.

Neighbourhood or community characteristics may influence educational outcomes directly or indirectly by shaping the quality of neighbourhood schools. The quality of schools is of significant interest to policymakers because it is a variable that is most amenable to public policy. The effect of various parameters on educational achievement have been investigated including the effect of teacher quality (Kain and Rivkin 1998), class size, school size (Sares 1992), private versus public control (Neal 1997), and the socio-economic characteristics of school pupils (e.g.,, race, academic achievement, and family background) (Haveman and Wolfe 1995). In the U.S., these parameters are influenced primarily by administrative decisionmaking at the local level and funding decisions of the electorate. Since children choose many of their peers from the neighbourhood, neighbourhood socio-economic characteristics may exert a role independent of school choice. Educational levels of the area (Lichter et al. 1993; Nordhaug 1990) may be important. In a study of predominantly minority inner-city neighbourhoods, high crime rates were found to hinder school attendance and educational achievement (Grogger 1997)

The role of regional labour markets in educational achievement has received little attention in the literature. Yet, regional labour market opportunities should influence residents' demand for education. Industry structure may be a good indicator of the educational demands of regional employers. Smith (1989) argues that residents of rural areas are at a distinct disadvantage relative to urban residents because the types of industries predominant in rural areas require employees with relatively low levels of formal education. Rural residents, therefore, have less incentive to invest in education because it will not significantly enhance their regional employment opportunities or earnings. That is to say, they will reap lower expected rates of return from the investment and rationally choose to invest less. Nord and Luloff (1993) have noted that counties with large mining sectors have residents with lower educational achievement. Mumim (1988) argues that the presence of the defence industry (which requires technical training for many entry-level positions) is a boost to educational achievement. Urbanisation may also increase demand for education because it is associated with more varied and specialised occupations, and therefore there may be an incentive for residents to acquire additional education in order to facilitate lateral occupational mobility and reduce unemployment risk. Even if an individual is employed, regional unemployment rates may act as a signal of the risk of unemployment for regional residents. Higher risks of unemployment would be expected to encourage additional investment in education.
Finally, access to regional educational resources may shape individual achievement. Being in close proximity to a college or university should have three effects. First, it should reduce the out-of-pocket costs of enrolling in higher education. If an individual chooses to reside at home, he/she would incur lower costs of travel and transportation. Second, it would reduce the psychic costs of additional education. Those who have a strong attachment to place or have significant "insider advantages" (Fischer and Malmberg 1997) would be more willing to enrol in higher education. Third, information about post-secondary educational opportunities would be expected to be greater in the vicinity of the institution. Visibility and outreach efforts are likely to be more intensive in close proximity and information about vacancies and programs offered will diffuse to peripheral areas only slowly. The quality of educational institutions may also affect student achievement. Institutional performance as measured by, say, inputs such as teacher-student ratios, quality of classroom equipment and physical plant, professional credentials of faculty, and the aptitude of fellow students may have an influence on student persistence.

7.0 Data contingencies in constructing a SVERIGE education module

The most fundamental problem encountered in constructing an education module for SVERIGE was lack of quality data. CORSIM module equations were estimated using U.S. survey longitudinal data which made it easy to track the persistence of various cohorts through their entire educational careers. While the TOPSWING database is longitudinal, it does not provide information on entry and persistence through education, except for the final three years of the database (i.e., 1993-95). Unfortunately, these three years of data are not enough to track a single cohort from school entry to school completion, since most programs require at least three years to complete. Moreover, some concerns have been raised about the quality of the educational indicators, particularly UTB (enrolled in education) which is not adequately defined in the TOPSWING documentation. Because of these limitations, various options were for constructing a Swedish educational module for SVERIGE were discussed. They are listed below:

Option #1. Estimate only educational achievement. Forget about entry and persistence determinants.

The first and easiest option is to estimate only degree attainment and to forget about the factor which influence the propensity to enter and continue in education. However, since time in education affects labour force participation, a rule is needed to compute the amount of time spent out of the labour force and to incorporate this into each person's biography.

Option #2 Estimate entry and persistence by using database.

If it is assumed, say, that every individual who graduated from gymnasium (HSUN code indicates Gymnasium degree) and subsequently enrolled in college (UTB=1 in 1994) was a full-time college student, one could estimate an equation for first-time entry. Persistence to the second year of study could be estimated for the same cohort for 1995 (UTB=1 in 1995).

Option #3 Estimate entry and use transition matrices

This option would be simple but preserve the structure of CORSIM. Basically, we would estimate an entry equation and transition probabilities based on persistence rates for certain demographic characteristics would be used to determine progression through education. This option assumes that statistics on persistence rates are available from Högskolverkat or some other government agency.

Option #4 Obtain additional data

Statistics Sweden and Högskolverkat collect survey data on educational choices, entry, and persistence. If this data were to be to be combined with the TOPSWING database, it would be possible to estimate the various educational stages. Another possibility is to obtain additional years of HINK data (e.g., 1996 data) in order to be able to construct a four-year longitudinal sample for estimating educational persistence and completion.

Option selected

During discussions about these alternatives, it was learned that the Finance Ministry had obtained data that combined information from HINK data samples with additional educational information obtained from the Högskolverkat and Skolverkat for an eight-year period. This data contains all of the information needed to distinguish entry year, persistence, and degree year for individuals attending adult educational courses (KOMVUX) and college/university (HÖGSKOLA). It also contains family background information for minors and young adults who reside in the parental household. This data allows each of the educational transitions to be modelled and for more detailed models to be specified. Unfortunately, on further exploration, the educational cohorts turned out to be biased in ways that seriously undermined their usefulness for statistical analysis. For instance, the sample severely overrepresented adult and female students. Therefore, it was decided to rely exclusively on the TOPSWING database and to employ a mixture of options 1-3 described above. The details of this procedure are described in section 9.0.

8.0 Conceptual extensions for SVERIGE

The CORSIM educational module makes some fairly simple assumptions. Though they can be used to operationalize the model, there are a number of modifications that would make it more realistic. These are: (1) to allow adult students (i.e., returning students who are older than 21 years of age), (2) to assign particular educational fields to educational achievers, and (3) to distinguish between full-time and part-time students. Each of these issues are described briefly below:

Issue #1: Allowing for returning adult students

Returning students are individuals who had previously studied but may not have completed their degree requirements. For example, a secondary student who graduated and encountered difficulties during his/her first year of study and subsequently dropped out may elect at a later date to return to study. In CORSIM, these students are eligible for re-entry into the educational queue provided they are 17-21 years of age. If they are older, however, they never encounter another educational opportunity. This restriction is removed in the SVERIGE educational module by explicitly modelling adult education and allowing older individuals to enter college/university education.

Issue #2: Assigning curricula for educational achievers

It has been proposed that the education module generate the fields in which individuals achieved their education (**Educsector**) as well as the number of years completed (**Educlevel**). This revision is recommended because curriculum/field is often an important factor in obtaining employment and level of earnings. Although education field is not currently used in any of the modules, it might be included later in the next version of the employment and earning module. Therefore, the current educational module will generate fields of study for college/university graduates. It distinguishes between five basic fields used by Statistics Sweden: (1) tekniska yrken (Engineering and technical fields), (2) administrativa, ekonomiska och sociala yrken (Business and social sciences), (3) vårdyrken (allied health), (4) undervisningsyrken (education), and (5) kultur och informationsyrken (arts and humanities). It is unlikely that a deterministic model based on the data available in TOPSWING could ever be devised that predicts curriculum selection. Therefore, educational fields will be assigned on the basis of a simple matrix of transition probabilities. This procedure is described in section 9.0.

Issue #3: Distinguishing between full-time and part-time students

CORSIM models only full-time student attendance. This is an accurate assumption for primary and secondary school attendance but significantly less so for post-secondary attendance. Approximately ½ of U.S. undergraduates are enrolled in community colleges, and most of these students attend on a part-time basis. Most part-time students are "non-traditional," which is to so that they are drawn from demographic groups (older, female, and full-time employment) which have traditionally been underrepresented on college campuses. These part-time students take much longer to complete their degrees for two reasons: (1) they are drawn from backgrounds that will have more difficulty persisting through higher education, and (2) when a student takes a part-time credit load, it takes much longer to complete degree requirements.

Part-time attendance is apparently a fact-of-life in Swedish higher education also. Therefore, it may be useful to draw a distinction between part-time and full-time students at some point in model development. It is my impression that Swedish part-time students tend to be older, female, and employed, much as they are in the United States. Therefore, the model might be modified to allow for the possibility of joint employment and education decisions.

9.0 Structure of the education module

Ideally, the education module would explain the achievement of several educational levels within the Swedish system SUN (SCB 1996). These are : (1) less than Grundskola, (2) Grundskola. (3) Gymnasium (2 year program), (4) Gymnasium (3 year program). (5) Högskola (2 year program), (6) Högskola (3 year program), and (7) Forskarutbildning. However, it is possible to collapse these seven levels into four or five levels without losing needed detail. Gymnasium education has now been standardised by law to conform to a three-year program. In addition, there is no need to distinguish between graduate-level and undergraduate education because no distinction between the two is made for educational variables used in the other microsimulation model modules. For instance, Rephann (1999a, 1999b) uses dummy variables ED1 (SUN levels 1 and 9 (unspecified)), ED2 (SUN levels 2-4), and ED3 (SUN levels 5-7). ED1 is basically similar to a high-school drop-out in the U.S., ED2 is comparable to a high-school graduate, and ED3 is equivalent to a U.S. college graduate.

The Swedish levels will be handled in the following fashion within the model:

Grundskola completion. Using the TOPSWING database, it is possible to determine Grundskola completion but not year of entry or persistence. The strategy used here is to compel each individual to complete the first eight years of compulsory school, beginning at age seven. During the final year of compulsory education, a compulsory school completion likelihood equation (specified as a function of family and individual attributes) is invoked to determine completers and non-completers (See appendix B, equation (1)). Completion of compulsory school is a function of parental educational levels, sex, and country of origin. This equation is invoked at age sixteen (16) for all individuals. Non-completers are assigned educational level one (1) and will be recorded as having completed eight (8) years toward completion of a Grundskola diploma but will not be eligible for entry to Gymnasium. Completers are assigned educational level two (2) and will become eligible for entry to Gymnasium. Non-completers are eligible to re-enter the educational queue later on, at the age of nineteen (19) and afterwards, to complete a Grudskola degree through adult education (KOMVUX) programs. They need only complete one additional year of education in these programs to obtain their basic education diploma.

Gymnasium entry, persistence, and completion. The TOPSWING database does not allow one to accurately determine Gymnasium entry. Using UTB and HSUN as indicators, attempts to estimate the numbers of Gymnasium educational entrants were unsuccessful. Calculations indicated that over 99% of Grundskola completers enrolled in Gymnasium within two years. Yet, according to *Befolkningen: Sveriges Nationalatlas* (1993), only 90% of Grundskola completers directly enter Gymnasium after completing compulsory school. Since the accuracy of the TOPSWING data must be called into question in this instance, it is recommended that a hard-coded probability of .9 be used to determine Gymnasium entrants. Gymnasium education will be three-years in length. It is assumed that each entrant completes the first two years of the program. Completion of the final year and graduation is determined by a likelihood equation that is a function of personal attributes (see Appendix B, equation (2)). This equation is invoked at age nineteen (19) for all Gymnasium entrants. Completers of Gymnasium will be assigned the educational level of (4) and non-completers will be assigned (3). Only completers of level (4) will be eligible for college/university admissions. However, Gymnasium non-completers can meet college entry requirements by attending adult school for an additional year.

Komvux and Folketskolan entry and persistence. Komvux and Folketskolan are primarily adult-education institutions that permit adults to return and receive credits towards Grundskola and Gymnasium diplomas. Although some adults take coursework on a part-time basis and in areas of special interest, the vast majority are enrolled on a full-time basis and intend to complete diploma requirements, in part because eligibility for educational stipends and financial assistance are tied to full-time attendance. Since many Swedes take advantage of adult education, including it would make the educational module more realistic.

Data from TOPSWING make it possible to estimate equations for both entry and persistence. It will be assumed that those who are eligible for entry to Komvux must be between the ages of nineteen (19) and sixty-five (65) and have achieved an educational level of 3 (Gymnasium – 2 years) or less. No distinction will be made between entry to different levels of education and completion. For instance, if an individual enters Komvux, he/she will be assumed to have completed that initial year. Likelihood of educational entry is determined by a logistic regression equation which is the function of various personal attributes (see Appendix B, equation (3)). A persistence equation is estimated for the second year of educational progression, again based on personal attributes (see Appendix B, equation (4)). This equation will be used to determine persistence for all years after entry. In addition, if an individual has entered Komvux, it will be assumed that he/she is no longer a member of the labour force for the duration of study. Therefore, the persistence equation does not include economic attributes such as employment status or earnings level.

Four years of educational continuity are possible at the Komvux level. Internal educational counters will record the number of years of pre-university education completed and Komvux years completed will be added to the tally. An individual who has completed the number of years needed for an additional diploma will have his educational level changed. If the requirements for a three-year Gymnasium degree are completed, he/she will be removed automatically from Komvux. Those who had not previously completed Grundskola may persist for up to three years after entry in order to obtain a Gymnasium degree.

Högskola entry and persistence. Högskola is a maximum four-year program. Those eligible for admission to högskola must be between the ages of 19 and 65 and have achieved a three year Gymnasium degree. Entry is determined by a likelihood equation based on personal demographic and socio-economic characteristics (see Appendix B, equation (5)). Persistence through three subsequent years is determined by a persistence likelihood equation at each stage (see Appendix B, equation (6). This equation is used to determine each year of succession. Entrants to higher education are removed from the labour force. Those who enter högskola are assumed to successfully complete this initial year. Those who persist to each year are also assumed to be successful. Those who persist until the second year are assigned educational level five (5) and those persisting to a third year are assigned level (6). A fourth year of persistence does not affect the educational level but is modelled here primarily because many students do enrol in four-year programs and is needed to generate more realistic biographies. Internal counters will keep tally of the number of years completed in case the individual drops out and decides to enter later in life.

Forskarutbildning. Forskarutbilding is a multi-year professional or graduate program. Eligible participants must be between the ages of 19 and 65 and have achieved at least a three year Högskola degree. Graduate program entry and persistence are determined in a fashion similar to högskola and the program is four years in length. Those who reach the fourth year are awarded a forskarutbilding diploma and assigned educational level seven (7). The equations used for entry and persistence are exhibited in Appendix B, equations (7) and (8). There was some bias in the sample used to estimate Forskarutbildning, with entrants being somewhat older than indicated by data obtained from Högskoleverket. This sub-module can be disabled if necessary. Since each of the other modules used a categorical educational variable that aggregates undergraduate and graduate education, it is will not affect an individual's life-course, except to the extent that it removes an individual from the labour force.

Treatment of drop-outs. "Drop outs" are individuals who enter but do not persist at various levels of education. "Drop outs" may occur at the Grundskola, Gymnasium, and Högskola levels. Grundskola and Gymnasium drop-outs are eligible for educational re-entry when they enter adulthood at age nineteen (19). At that time, they become eligible for Komvux education. If they enter and fail to persist in Komvux, they may later re-enter education and their previous level of achievement will have been recorded by counters. The same is true of Högskola and Forskarutbildning. Individuals who discontinue college studies are eligible to re-enter later in life and continue where they left off. Because it will be possible for students to enter and exit education multiple times, the counters must be able to aggregate the number of years accumulated at different stages. This treatment of "drop outs" is much more flexible than CORSIM. For instance, in CORSIM only individual graduating cohorts are eligible for college entry. Drop-outs from college in the age group 17-21 are fed back into the educational queue based on a Monte Carlo experiment using a hard-coded probability.

Educational curricula. Although educsector is currently not used in the model, the module can be constructed to generate it for future use. At this stage, a vector of hard-coded probabilities will be used. Graduates will be assigned a discipline in one of five basic categories, including: (1) tekniska yrken (Engineering and technical fields), (2) administrativa, ekonomiska och sociala yrken (Business and social sciences), (3) vårdyrken (allied health), (4) undervisningsyrken (education), and (5) kultur och informationsyrken (arts and humanities). The probabilities for selecting these curricula are computed for each gender. Females have a greater tendency to choose allied health, education, and arts and humanities fields, while males are more likely to select engineering and technical fields. The transition probabilities (exhibited in a table in Appendix C.) are computed on the basis of 1995 graduates from table 426 contained in SCB (1997).

10.0 Results

The specifications for equations to determine educational entry, persistence, and completion are similar in some respects to those employed in CORSIM. They are based on individual and family demographic and socio-economic characteristics and include many of the variables identified as important to educational achievement and decisions in section 6.0. Although four geographical variables were tested for inclusion in the equations here, only one variable was used (percentage college education). At a later stage, it may be possible to test for the importance of additional geographical variables, such as neighbourhood characteristics and distance to educational facilities. However, this work was beyond the scope of what was possible in the time allocated to this project.

The models employed here are estimated via logistic regression and contain subsets of variables listed below. The dependent variable indicates whether or not an individual enters, persists, or completes education at a given level (i.e., ED_1=Completed Grundskola, ED_2=Completed Gymnasium, etc.).

(2) $$\Pr\mathbf{ob}(\mathbf{ED_i} = 1) = \frac{\mathbf{e}^{\mathbf{ß'x}}}{1 + \mathbf{e}^{\mathbf{ß'x}}}$$

$$b'\mathbf{x}=\beta_0+\beta_1\mathbf{AGE}+\beta_2\mathbf{AGE2}+\beta_3\mathbf{SEX}+\beta_4\mathbf{IMM}+\beta_5\mathbf{NINC1}+\beta_6\mathbf{NINC2}+\beta_7\mathbf{NINC3}+\beta_8\mathbf{NINC4}$$
$$+\beta_9\mathbf{NINC5}+\beta_{10}\mathbf{MARRY}+\beta_{11}\mathbf{DIVSEP}+\beta_{12}\mathbf{CHILD5}+\beta_{12}\mathbf{EMP}$$

where:

β_0	constant
AGE	Age in years
AGE2	Age squared
SEX	gender (1=male, 0=female)
PATED	father's education level (PARED =1 if father completed Gymnasium)
MATED	mother's education level (MATED=1 if mother completed Gymnasium)
IMM	dummy variable indicating if individual is an immigrant
INC1	individual earnings (CARB) less than 600 (100s of SEK).-
INC2	individual earnings between 1000 and 1399 (1=Yes, 0=No)-.
INC3	individual earnings between 1400 and 1799
INC4	individual earnings between 1800 and 2199
INC5	individual earnings equal or above 2200
MARRY	dummy variable indicating if individual is married
DIVSEP	dummy variable indicating if individual is divorced, separated, or widowed.
EMP	dummy variable indicating if individual is employed
PCOLL	percentage of municipality (KOMMUNE) residents aged 25-65 with college/university education.

The specifications differ for each educational level and some explanation is offered here. Sex (SEX) and immigration status (IMM) is used in each specification, regardless of whether it is statistically significant or not. Marriage is used for educational levels where the participants are young adults (and therefore likely to be married), namely Gymnasium, Komvux, Högskola, and Forskarutbildning. Divorce and separation is used for Komvux, Högskola, and Forskarutbildning. Age (AGE, AGE2) is used as an explanatory variable for those levels which participants of various ages (i.e., Komvux, Högskola, and Forskarutbildning). Age is not used in Grundskola and Gymnasium because these equations are estimated using specific age cohorts. Earnings (INC1-INC5) and employment (EMP) are used only in estimating educational entry to Komvux, Högskola, and Forskarutbildning. These variables are not used in the persistence and graduating equations because educational participants are assumed to be not employed and hence would not have any earnings. In fact, the earnings and employment module is implemented only after the education module. Therefore, only those who complete or drop-out of education can become employed in any given year. Grundskola completion is the only level at which parental educational variables (i.e., PATED and MATED) are used as explanatory variables. These variables are used because they are easily obtained for the sample of 16 year olds. However, this information is much more difficult to extract for older individuals and altogether impossible to retrieve for anyone over the age of twenty-three.

The results are summarised in Appendix B, along with diagnostic statistics. All but a couple of coefficients have reasonable signs. Likelihood of completing Grundskola is positively associated with parental educational achievement and negatively associated with being an immigrant. Gymnasium completion is negatively associated with being married and being an immigrant. Entrance to Komvux is negatively associated with age and being a male. It is positively associated with being married, separated or divorced, and having low earnings. Komvux persistence is a quadratic function of age and negatively associated with being married. Högskola is a quadratic function of age and is positively associated with lower earnings, being employed, and the proportion of municipal residents with a college degree. It is negatively associated with being married and male. Högskola persistence is negatively associated with being male and age. Graduate school entry and persistence results are similar to those for högskola.

11. Alignment

The model should accurately predict two things: (1) the number of individuals enrolled in education at different levels and (2) the number of individuals who attain various levels of educational achievement. The former information is important because it has a bearing on the number of labour force participants. The latter is important because educational achievement appears as a variable in each of the microsimulation modules. CORSIM does not use educational alignment and this may have something to with the way the educational module is structured. Unlike the other modules, education is a compound event rather than a simple event. Mortality is a simple event. A person dies or they don't die in a given year. If too few people are predicted to die, the likelihoods of dying are boosted via alignment.

Educational participation, in contrast, depends on several events, entry and persistence through several years. Higher educational achievement is obtained only after entering and then persisting for several years. Alignment would be relatively straightforward if flows of pupils/students through the various years and levels of education were available. Unfortunately, they are not. In fact, only entrants and graduates for college/university and graduate school are available in a format that could be used for alignment (see Appendix D for this data). Data regarding grundskola and gymnasium entrants and graduates are available too, but they appear to sum both child and adult participants. In order to be used here, the adult participants would need to be separated from children in order to use separately for the Komvux equations. This information may be available from Skolverket.

The information in Appendix D allows partial alignment; that is to say alignment for equations (5), (6), and (7). Column (A), which listed högskola first-time entrants, is used to align the högskola entry equation (5). Column (B), showing högskola graduates, is used to align equation (6). It is used because information regarding student persistence is not available. Since the persistence may occur over three years, there is a need to make three alignments. However, since there is only one persistence equation, any corrective factor is assigned to the persistence equation by using an nth root (or in this case a cube root) of the ratio of actual to estimated graduates. Column (C), showing graduate school entrants, is used to align equation (7).

References

Bankston, Carl L. and Stephen J. Caldas. 1998. Family structure, schoolmates, and racial inequalities in school achievement. *Journal of Marriage and the Family* 60: 715-723.

Befolkningen: Sveriges Nationalatlas. 1993. Eds. Sture Öberg and Peter Springfeldt. Stockholm: SNA Publishing.

Caldwell, Stephen B. 1993. *CORSIM 2.0 Model Documentation*. Ithaca, NY: Cornell University.

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

Fischer, Peter A., Einar Holm, Gunnar Malmberg, and Thomas Straubhaar. 1998. *Why do people stay? The insider advantages approach.* Hamburg, Germany: Institute of Economic Policy Research.

Goetz, Stephan J. and David L. Debertin. 1993. Estimating County-Level Demand for Educational Attainment," *Socio-Economic Planning Sciences*, 27:25-34.

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

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.

M. Moulden, and M. G. Bradford. 1984. Influences on educational attainment: The importance of the local residential environment. *Environment and Planning* 16, 1: 49-66.

Megdal, S. B.. 1984. A model of local demand for education. *Journal of Urban Economics* 16, 1: 13-30.

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.

Nordhaug, Odd. 1990. Structural determinants of publicly subsidized adult education. *Adult Education Quarterly*. 40, 4: 197-206.

Statistiska centralbyrån. 1996. *Svensk Utbildningsnomenklatur (SUN): Del 1. Systematisk version*. (Statistics Sweden. *Swedish standard classification of education. Part 1. Numerical Order*). Stockholm: SCB.

Statistiska centralbyrån. 1997. *Statistisk Årsbok* (Statistics Sweden. *Statistical Yearbook.*). Stockholm: SCB.

Statistiska centralbyrån. 1998. *Universitet och högskolor Grundutbilding: Nybörjare, registerade och examina 1996/97* (Statistics Sweden. *Higher Education Undergraduate education: University entrants, enrolled and graduated students 1996/97*). Stockholm: SCB.

Smith, Eldon D. 1989. Reflections on human resources in the strategy of rural economic development. *Review of Regional Studies*. 19, 1:13-22.

Teachman, Jay D. and Kathleen Paasch. 1998. The family and educational aspirations. *Journal of Marriage and the Family* 60: 704-714.

Waslander, H.B. 1997. An education model for Dynacan: Some initial thoughts. December 1997.

APPENDIX A. CORSIM Equations

A.1 edu_hs_dropout: Dropout equation coefficients
| returns probability of person in 10th or 11th grade dropping out
| of school
| logistic regression
| Dimensions: race (0=white 1=black)
| sex (0=male 1=female)
|
| 1 highest parent education <12
| 2 highest parent education 12
| 3 highest parent education 13-15
| 4 highest parent education 16+
| 5 parents own home
| MISSING: living on own
|
| Source: HSB data

A.2 edu_hs_grad: Graduation equation coefficients
| returns probability of graduation for 12th-graders
| logistic regression
| Dimensions: race (0=white 1=black)
| sex (0=male 1=female)
| Variables: 0 constant
| 1 highest parent education <12
| 2 highest parent education 12
| 3 highest parent education 13-15
| 4 highest parent education 16+
| 5 Married (?)
| 6 Parent owns home
|
A.3 edu_co_enroll: College enrollment equation coefficients
| returns probability that HS graduate will enroll in college
| logistic regression
| Dimensions: race (0=white 1=black)
| sex (0=male 1=female)
| Variables: 0 constant
| 1 parents didn't finish high school
| 2 parents max edu == 12
| 3 parents max edu between 12 and 16
| 4 respondent has children
| 5 respondent is married
| 6 parents own home
|

A.4 edu_co_1st: completion of first year of college
| returns probability of person enrolled in college completing
| first year
| logistic regression
| Dimensions: race (0=white 1=black)
| sex (0=male 1=female)
| Variables: 0 constant
| 1 parents didn't finish high school
| 2 max parent education == 12
| 3 max parent education between 12 and 16
| 4 respondent has children
| 5 parents own home
| Source: HSB data

A.5 edu_co_2nd: completion of second year of college
| returns probability of person in college completing second year
| logistic regression
| Dimensions: race (0=white 1=black)
| sex (0=male 1=female)
| Variables: 0 constant
| 1 parents didn't finish high school
| 2 max parent education == 12
| 3 max parent education between 12 and 16
| 4 respondent married
| 5 respondent has children
| 6 parents own home
|
A.6 edu_co_3rd: completion of third year of college
| returns probability of person in college completing third year
| logistic regression
| Dimensions: race (0=white 1=black)
| sex (0=male 1=female)
| Variables: 0 constant
| 1 parents didn't finish high school
| 2 max parent education == 12
| 3 max parent education between 12 and 16
| 4 respondent married
| 5 respondent has children
| 6 parents own home
|
A.7 edu_co_grad: college graduation equations
| returns probability that seniors in college will graduate
| logistic regression
| Variables: 0 intercept
| 1 black
| 2 age
| 3 agesq
| 4 kids
| 5 father completed high school
| 6 father attended some college, but didn't graduate
| 7 father graduated from college
| 8 mother completed high school
| 9 mother attended some college, but didn't graduate

A.8 edu_co4_enroll: graduate school enrollment
| returns probability of college senior enrolling in grad school
| logistic regression
| Variables: 0 intercept
| 1 female
| 2 age (years)
| 3 age squared
| 4 graduated from college (COMP16)
| 5 mother completed high school
| 6 mother completed some college, but didn't graduate
| 7 mother graduated from college

A.9 edu_grad1_enroll: second year grad school enrollment
| returns probability that first year grad students sign on
| for year 2.
| logistic regression
| Variables: 0 constant
| 1 graduated from college

APPENDIX B. SVERIGE Equations

(1) Completion of Grundskola

The LOGISTIC Procedure

Data Set: WORK.REGDAT1
Response Variable: COMPGRAD
Response Levels: 2
Number of Observations: 4620
Link Function: Logit

Response Profile

Ordered Value	COMPGRAD	Count
1	1	4494
2	0	126

WARNING: 39 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	1158.203	1103.908	.
SC	1164.641	1136.098	.
-2 LOG L	1156.203	1093.908	62.295 with 4 DF (p=0.0001)
Score	.	.	76.442 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	3.0320	0.1790	286.9264	0.0001	.	.
SEX	1	-0.1882	0.1838	1.0483	0.3059	-0.051881	0.828
MATED	1	0.9764	0.1909	26.1615	0.0001	0.247021	2.655
PATED	1	0.4180	0.1944	4.6211	0.0316	0.114814	1.519
IMM	1	-1.0487	0.2495	17.6679	0.0001	-0.134691	0.350

Association of Predicted Probabilities and Observed Responses

Concordant	= 62.6%	Somers' D	= 0.354
Discordant	= 27.2%	Gamma	= 0.394
Tied	= 10.2%	Tau-a	= 0.019
(566244 pairs)		c	= 0.677

(2) Completion of Gymnasium

The LOGISTIC Procedure

Data Set: WORK.REGDAT4
Response Variable: HIGHGRAD
Response Levels: 2
Number of Observations: 3452
Link Function: Logit

Response Profile

Ordered Value	HIGHGRAD	Count
1	1	3155
2	0	297

WARNING: 2250 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	2026.749	2006.102	.
SC	2032.896	2030.689	.
-2 LOG L	2024.749	1998.102	26.647 with 3 DF (p=0.0001)
Score	.	.	35.284 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	2.5223	0.0909	770.2108	0.0001	.	.
SEX	1	-0.1473	0.1223	1.4505	0.2284	-0.040617	0.863
MARRY	1	-1.4408	0.7295	3.9008	0.0483	-0.040514	0.237
IMM	1	-0.9785	0.1978	24.4683	0.0001	-0.122137	0.376

Association of Predicted Probabilities and Observed Responses

Concordant = 33.5%	Somers' D = 0.090
Discordant = 24.5%	Gamma = 0.156
Tied = 42.0%	Tau-a = 0.014
(937035 pairs)	c = 0.545

(3) Entry to Komvux

The LOGISTIC Procedure

Data Set: WORK.REGDAT5
Response Variable: FIRSTKOM
Response Levels: 2
Number of Observations: 56446
Link Function: Logit

Response Profile

Ordered Value	FIRSTKOM	Count
1	1	674
2	0	55772

WARNING: 26130 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	7310.607	6043.108	.
SC	7319.548	6159.341	.
-2 LOG L	7308.607	6017.108	1291.499 with 12 DF (p=0.0001)
Score	.	.	1601.643 with 12 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-1.9946	0.5521	13.0517	0.0003	.	.
AGE	1	-0.0583	0.0268	4.7466	0.0294	-0.401542	0.943
AGE2	1	-0.00067	0.000348	3.6669	0.0555	-0.406960	0.999
SEX	1	-0.7745	0.0868	79.5816	0.0001	-0.212487	0.461
MARRY	1	0.3892	0.1019	14.5830	0.0001	0.107114	1.476
DIVSEP	1	0.9151	0.1358	45.4174	0.0001	0.182196	2.497
INC1	1	1.1531	0.2554	20.3850	0.0001	0.313227	3.168
INC2	1	0.4521	0.2697	2.8112	0.0936	0.075860	1.572
INC3	1	0.0117	0.2741	0.0018	0.9660	0.002111	1.012
INC4	1	-0.3426	0.2815	1.4809	0.2236	-0.069810	0.710
INC5	1	-0.0483	0.2922	0.0273	0.8688	-0.008632	0.953
EMP	1	0.5859	0.0987	35.2290	0.0001	0.150752	1.797
IMM	1	0.0510	0.1010	0.2546	0.6138	0.010144	1.052

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Concordant	= 83.6%	Somers' D	= 0.703
Discordant	= 13.4%	Gamma	= 0.724
Tied	= 3.0%	Tau-a	= 0.017
(37590328 pairs)		c	= 0.851

(4) Persistence through Komvux

The LOGISTIC Procedure

Data Set: WORK.REGDAT6
Response Variable: SECKOM
Response Levels: 2
Number of Observations: 674
Link Function: Logit

Response Profile

Ordered Value	SECKOM	Count
1	1	306
2	0	368

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	930.651	927.858	.
SC	935.164	959.451	.
-2 LOG L	928.651	913.858	14.793 with 6 DF (p=0.0219)
Score	.	.	14.274 with 6 DF (p=0.0267)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-2.1736	1.1043	3.8741	0.0490	.	.
AGE	1	0.1418	0.0653	4.7106	0.0300	0.794567	1.152
AGE2	1	-0.00199	0.00087	5.2080	0.0225	-0.825721	0.998
SEX	1	-0.2699	0.1678	2.5872	0.1077	-0.071176	0.763
MARRY	1	-0.4918	0.2147	5.2485	0.0220	-0.129820	0.612
DIVSEP	1	-0.1680	0.2759	0.3709	0.5425	-0.032676	0.845
IMM	1	-0.1504	0.2003	0.5639	0.4527	-0.033755	0.860

Association of Predicted Probabilities and Observed Responses

Concordant	= 56.6%	Somers' D	= 0.150
Discordant	= 41.6%	Gamma	= 0.153
Tied	= 1.8%	Tau-a	= 0.075
(112608 pairs)		c	= 0.575

(5) Entry to College/University

The LOGISTIC Procedure

Data Set: WORK.REGDAT7
Response Variable: FIRSTED
Response Levels: 2
Number of Observations: 94250
Link Function: Logit

Response Profile

Ordered Value	FIRSTED	Count
1	1	4207
2	0	90043

WARNING: 53282 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	34386.173	26047.039	.
SC	34395.627	26179.391	.
-2 LOG L	34384.173	26019.039	8365.134 with 13 DF (p=0.0001)
Score	.	.	9919.890 with 13 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	0.6001	0.2529	5.6325	0.0176	.	.
AGE	1	-0.2274	0.0123	342.3015	0.0001	-1.582313	0.797
AGE2	1	0.00178	0.000162	119.6363	0.0001	1.003210	1.002
SEX	1	-0.4771	0.0360	176.0567	0.0001	-0.131459	0.621
MARRY	1	-0.2052	0.0526	15.2184	0.0001	-0.056576	0.814
DIVSEP	1	0.0629	0.0930	0.4576	0.4987	0.010821	1.065
INC1	1	1.8322	0.1132	262.0466	0.0001	0.430675	6.248
INC2	1	0.7748	0.1193	42.1441	0.0001	0.133856	2.170
INC3	1	0.2739	0.1199	5.2169	0.0224	0.056712	1.315
INC4	1	0.0378	0.1204	0.0984	0.7538	0.008523	1.038
INC5	1	-0.1042	0.1348	0.5979	0.4394	-0.020161	0.901
EMP	1	0.4206	0.0486	74.9164	0.0001	0.075999	1.523
IMM	1	-0.0111	0.0663	0.0282	0.8667	-0.001712	0.989
PCOLL	1	0.0469	0.00277	286.5271	0.0001	0.143571	1.048

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Concordant	= 84.3%	Somers' D	= 0.697
Discordant	= 14.6%	Gamma	= 0.705
Tied	= 1.2%	Tau-a	= 0.059
(378810901 pairs)		c	= 0.849

(6) Persistence through College/University

The LOGISTIC Procedure

Data Set: WORK.REGDAT8
Response Variable: SECED
Response Levels: 2
Number of Observations: 4215
Link Function: Logit

Response Profile

Ordered Value	SECED	Count
1	1	2379
2	0	1836

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	5775.084	5697.412	.
SC	5781.430	5741.837	.
-2 LOG L	5773.084	5683.412	89.672 with 6 DF (p=0.0001)
Score	.	.	89.597 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	1.5455	0.4275	13.0667	0.0003	.	.
AGE	1	-0.0513	0.0272	3.5668	0.0589	-0.231652	0.950
AGE2	1	0.000285	0.000389	0.5363	0.4640	0.087308	1.000
SEX	1	-0.2685	0.0645	17.3117	0.0001	-0.073128	0.764
MARRY	1	-0.0839	0.1073	0.6124	0.4339	-0.017859	0.919
DIVSEP	1	0.0972	0.1712	0.3223	0.5702	0.011202	1.102
IMM	1	-0.0592	0.1219	0.2357	0.6273	-0.008467	0.943

Association of Predicted Probabilities and Observed Responses

Concordant = 56.2%	Somers' D = 0.160
Discordant = 40.2%	Gamma = 0.166
Tied = 3.6%	Tau-a = 0.079
(4367844 pairs)	c = 0.580

(7) Entry to Graduate/Professional School

The LOGISTIC Procedure

Data Set: WORK.REGDAT7
Response Variable: GRADED
Response Levels: 2
Number of Observations: 16953
Link Function: Logit

Response Profile

Ordered Value	GRADED	Count
1	1	615
2	0	16338

WARNING: 10202 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	5288.806	5054.363	.
SC	5296.544	5154.960	.
-2 LOG L	5286.806	5028.363	258.443 with 12 DF (p=0.0001)
Score	.	.	256.572 with 12 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-4.4091	0.8013	30.2775	0.0001	.	.
AGE	1	0.0894	0.0375	5.6918	0.0170	0.500479	1.094
AGE2	1	-0.00145	0.000436	10.9942	0.0009	-0.724863	0.999
SEX	1	-0.7172	0.0946	57.4416	0.0001	-0.197681	0.488
MARRY	1	-0.3681	0.1009	13.3034	0.0003	-0.095718	0.692
DIVSEP	1	-0.1241	0.1506	0.6791	0.4099	-0.022409	0.883
INC1	1	0.8655	0.1741	24.7107	0.0001	0.137215	2.376
INC2	1	0.9060	0.1791	25.5848	0.0001	0.092077	2.475
INC3	1	0.4304	0.1674	6.6136	0.0101	0.056221	1.538
INC4	1	0.3354	0.1343	6.2420	0.0125	0.059094	1.399
INC5	1	0.3581	0.1141	9.8415	0.0017	0.078793	1.431
EMP	1	0.3474	0.2214	2.4614	0.1167	0.043986	1.415
IMM	1	0.0338	0.1508	0.0502	0.8226	0.005122	1.034

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Concordant	= 67.1%	Somers' D	= 0.367
Discordant	= 30.4%	Gamma	= 0.376
Tied	= 2.5%	Tau-a	= 0.026
(10047870 pairs)		c	= 0.683

(8) Persistence through Graduate/Professional School

The LOGISTIC Procedure

Data Set: WORK.REGDAT8
Response Variable: SECGRAD
Response Levels: 2
Number of Observations: 615
Link Function: Logit

Response Profile

Ordered Value	SECGRAD	Count
1	1	232
2	0	383

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	817.114	817.806	.
SC	821.536	844.335	.
-2 LOG L	815.114	805.806	9.309 with 5 DF (p=0.0974)
Score	.	.	9.364 with 5 DF (p=0.0954)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	0.1679	0.3942	0.1814	0.6701	.	.
AGE	1	-0.0210	0.0104	4.1114	0.0426	-0.109985	0.979
SEX	1	0.0957	0.1831	0.2731	0.6013	0.024498	1.100
MARRY	1	0.1846	0.2144	0.7414	0.3892	0.050570	1.203
DIVSEP	1	0.0662	0.3130	0.0447	0.8326	0.012349	1.068
IMM	1	0.6249	0.2951	4.4854	0.0342	0.096764	1.868

Association of Predicted Probabilities and Observed Responses

Concordant	= 56.3%	Somers' D	= 0.139
Discordant	= 42.4%	Gamma	= 0.141
Tied	= 1.3%	Tau-a	= 0.066
(88856 pairs)		c	= 0.570

Appendix C. Probabilities of obtaining field for högskola graduates

	Female	Male
(1) Engineering and technical	0.109991	0.443008
(2) Business and social sciences	0.255016	0.295940
(3) Allied health	0.217741	0.080107
(4) Education	0.344129	0.122002
(5) Arts and humanities	0.073124	0.058943

Source: SCB (1997)

Appendix D. Partial alignment data, 1985-1995

	(A) COLLEGE ENTRY	(B) COLLEGE GRADS	(C) GRAD ENTRY
1985	72404	31120	2293
1986	75092	30759	2254
1987	73759	30634	2565
1988	78535	30521	2452
1989	76537	29729	2469
1990	82818	28985	2494
1991	92655	31649	2708
1992	99589	33165	3465
1993	61784	33482	2697
1994	62899	33761	2960
1995	66336	33523	3349

Source: SCB (1987-1997) and SCB (1998).