Modelling fertility and mortality in the APPSIM dynamic microsimulation model

Sophie Pennec
Institut National d'Etudes Démographiques (INED), 133 boulevard Davout, 75980 Paris cedex 20 (France) ; pennec@ined.fr
National Centre for Social and Economic Modelling (NATSEM), University of Canberra, 2601 ACT (Australia); sophie.pennec@natsem.canberra.edu.au

Bruce Bacon
National Centre for Social and Economic Modelling (NATSEM), University of Canberra, 2601 ACT (Australia); bruce.bacon@natsem.canberra.edu.au

Paper prepared for the 1st conference of the International Microsimulation Association,

Vienna, 20-22 August 2007

Modelling fertility and mortality in the APPSIM dynamic microsimulation model[1]

Sophie Pennec
Institut National d'Etudes Démographiques (INED), 133 boulevard Davout, 75980 Paris cedex 20 (France) ; pennec@ined.fr
National Centre for Social and Economic Modelling (NATSEM), University of Canberra, 2601 ACT (Australia); sophie.pennec@natsem.canberra.edu.au

Bruce Bacon
National Centre for Social and Economic Modelling (NATSEM), University of Canberra, 2601 ACT (Australia); bruce.bacon@natsem.canberra.edu.au

ABSTRACT: This paper first presents the new NATSEM dynamic microsimulation model APPSIM. It reviews the latest international practice when modelling demographics - and in particular fertility and mortality - in dynamic microsimulation models. The next section describes the data available in Australia for estimating fertility and the final section describes the method used and envisaged to estimate the demographic events in APPSIM and the method to link micro and macro variables in the model.

---

[1] The authors would like to acknowledge comment and input from Ann Harding, Mandy Yap, Simon Kelly (NATSEM) and Laurent Toulemon (INED). They would also like to gratefully acknowledge the funding and support provided by the Australian Research Council (under grant LP0562493), and by the 13 research partners to the grant: Treasury; Communications, Information Technology and the Arts; Employment and Workplace Relations; Health and Ageing; Education, Science and Training; Finance and Administration; Families, Community Services and Indigenous Affairs; Industry, Tourism and Resources; Immigration and Citizenship; Prime Minister and Cabinet; the Productivity Commission; Centrelink; and the Australian Bureau of Statistics. The authors would also like to acknowledge the two international partner investigators on this grant, Professors Jane Falkingham and Maria Evandrou of the University of Southampton

1 INTRODUCTION

This paper is associated with the development of the Australian Population and Policy Simulation Model (APPSIM). The APPSIM dynamic population microsimulation model is being developed at NATSEM and will be used by Commonwealth Government policy makers and other analysts to assess the social and fiscal policy implications of Australia's ageing population. The particular focus of this paper is how to most effectively model fertility and mortality within APPSIM.

Fertility has been a major topic of interest for demographers for some decades. One reason is the baby boom that occurred after WWII that increased the population from the end of the 40s to the beginning of the 60s. This bulge in the population was not regarded as a problem until recently. However, in the near future, this large population cohort will reach retirement age, resulting in slower economic growth and increasing the burden of caring for them (in terms of income through retirement pensions or in terms of health and the need for formal care and carers). As a result of these pressures, many have questioned the sustainability of current welfare schemes in some developed countries (e.g. see the studies in Harding and Gupta (2007)). The baby boomers are also responsible for initiating the 'baby bust' — that is, the low fertility period experienced during the past few decades.

Nowadays, with the legalisation of and the wide access to contraception, the number of children a person has reflects mainly their choice. Of course, for some persons, this choice is constrained by biological factors (sterility), by social factors (no partners during reproductive life) or even by economic ones. This choice results from the interaction of a range of different variables — such as family context, number of siblings, socio-economic status of parents that affect the education level of children; partnership status, and the socio-economic status of the couple.

When modelling fertility, one must pay attention not only to the number of children (quantum of fertility) but also to the time schedule of fertility (tempo). The effect of fertility upon such factors as careers and loss of wages (between a woman that stops working to raise all her children and a woman that remains in the labour force) can differ greatly. Some parents will move to or choose a part time job. To have a good representation of the spacing between births is also essential — particularly as some policy entitlements depend on the number of children aged under a certain age or when studying kinship and the likely number of children available for caring for elderly parents. A 65 year old retiree who is not disabled can more easily care for his or her disabled parents than a 55 year old person still in labour force or a 80 year old person who is also disabled.

2 APPSIM

This section gives an overview of the microsimulation model APPSIM. Following a comprehensive review of other successful dynamic population microsimulation models (Cassells and al, 2006), the choice made for this new model is to have a cross-sectional, dynamic, annual discrete transitions model.

2.1 Dynamic

As APPSIM is aiming at simulating the Australian population and policy up to 2050, the choice of a dynamic model is quite straightforward; this means that all the events have to be projected in the future and assumptions on both the trends and the individual behaviour must be determined.

Dynamic models make their population evolve over time. This ageing process is done according to the nature of the events, using therefore different techniques (deterministic, stochastic). Some events such as ageing are deterministic, as each member of the sample, is getting older at the same fixed rate (e.g. year or month according to the step size chosen for the model). Most of the other events are determined stochastically, in that they may or may not occur and are based on estimated probabilities. A draw determines for each persons if she or he will face the event according to a probability of the event (this probability is estimated for each person depending on different characteristics, age, education, income, health status, parity…). Another stochastic method for simulating events is to implement a market in which its outcome is stochastically determined. Some behavioural features may be introduced in these markets. Markets methods are often used to model labour force participation and matrimonial matching.

2.2 Cross-sectional

Among dynamic microsimulation models, one can distinguish between cross-sectional and cohort models. Cohort models build life histories of individuals, belonging to one or a quite small number of cohorts, whereas cross-sectional models are based on usually a representative sample of the whole population and show how this will evolve through time (see Harding and Gupta, 2007b, p. 6 for more details).

APPSIM is a cross-sectional model. Cross-sectional models are more day to day policy oriented. They give, for a representative sample of the overall population, the cost and saving of any policy schemes and how any changes impact on the income or level of life of the different groups of individuals in an accounting period.

2.3 Continuous event probabilities vs. discrete transition

Dynamic microsimulation models can use different features to model the events. One is continuous time methods that assumes that events occurs in continuous time or in pseudo continuous time and rely on the use mainly on survival analysis and model a random waiting time until the next event (or until the next occurrence of the event if the event can arise more than once). This waiting time is drawn from a distribution duration between the current state and the event (Imhoff & Post, 1998; Galler, 1997).

On the other hand, discrete time methods consider for each period of time and according to the current status, the probability that the event under consideration, will occur during the next period of time.

Although pseudo-continuous time methods can be seen as theoretically a more desirable way to reflect the demographic lifecourse, and is more elegant and more computationally efficient as it computes a waiting time until the next event instead of computing a draw at each period of time, APPSIM will use a discrete time approach.

This choice is mainly based on NATSEM's earlier experiences with DYNAMOD and on the fact that discrete time is easier to update and align with aggregates and over the time as pointed out by Kelly and King (2001). "On balance, considerable research needs to be done to ensure that the survival function advantages of efficiency and elegance are not outweighed by the disadvantages of alignment difficulty, longitudinal data requirements and transparency".

This choice for a discrete time approach does not mean that we will not use the advantage of longitudinal data whenever possible. As pointed out by Vencatasawny (2002, p. 2) "longitudinal data are believed to be indispensable for the study of processes over the life courses and their relation to historical change" and "while survival analysis assumes continuous time and models the duration an individual stays in a particular event, for example being non-pregnant, the latter assumes discrete time and models the risk of getting pregnant, for example at different points in time".

2.4 Closed population

In a closed model, all individuals that face any event belong to the population either because they belong to the starting population, or because they enter during the simulation by birth or immigration. In an open model, on the contrary, individuals are created when necessary (see Rowe and Gribble, 2007). This latter option is simpler to implement but may end up with a distorted structure of the population and therefore seems to have found less favour with government policy makers. APPSIM uses a closed population as it needs to be fully representative of the population and its structure over time and to be aligned to demographic targets.

2.5 Level of simulation

One advantage of microsimulation is that simulation can be performed at different levels - different levels for the individuals, different levels for the events. Individuals can be gathered as members of the same family/household and therefore the probabilities of facing an event can be applied to each individual or to the whole group (e.g. migration can be estimated at the individual level or at the family level). Different levels for the events, facing an event for an individual is based on only his/her individual characteristics or it can take into account some other kind of variables (characteristics of some other member of his/her family; macro aggregates like growth rate, unemployment rate.

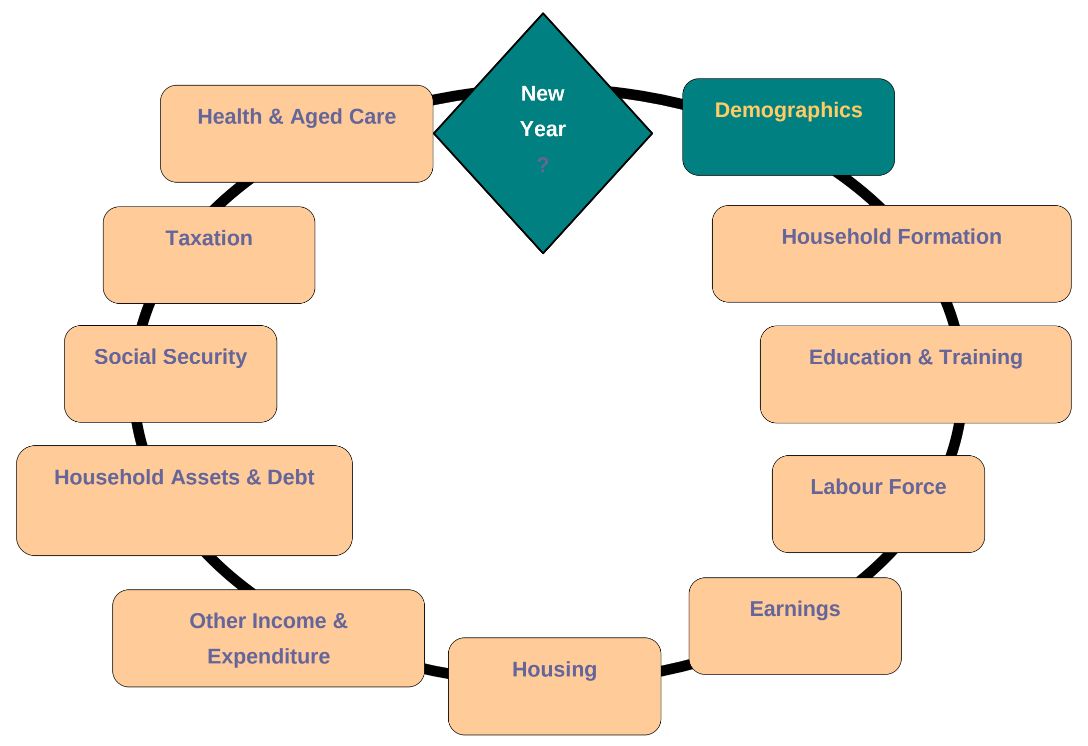
2.6 Starting population

After reviewing potential datasets, it has been decided that the starting population for APPSIM will be based on the 1 per cent Census Sample, drawn from the 2001 Census (Kelly, 2007). It consists of a sample of 1 per cent of private dwellings, with their associated family and person records, and a 1 per cent sample of persons from all non-private dwellings together with a record from the non-private dwelling (ABS, 2001) – that is 75,451 dwellings, 79,320 families and 188,013 persons. The 1 per cent census has the advantage of gathering information on both private and non-private dwellings, and has a big sample size. The main drawback is that the number of variables is limited and imputations or data matching must be performed to be able to gather all the information needed for the simulations.

2.7 Order of the modules

During each period of time, each individual goes through all the different groups of events. This includes the set of demographic events (including household formation), education, labour force participation, earnings, housing, and financial events such as unearned income , household expenditure, taxation, and finally welfare (with an emphasis in APPSIM on health care and aged care).

Figure 1 **The different modules in APPSIM**



As you can see, demographics are the first group of events. Because of the model structure, that means that we assume that demographic events can affect the other events that happen in the same year — but the later events, such as labour force status and earnings, can affect demographic events only in the following year. For example, this means that having a baby in year t can affect the probability of labour force participation in that year — but being in the labour force in year t can only affect the probability of having a baby in year t+1.

3 DEMOGRAPHICS IN DYNAMIC MICROSIMULATION MODELS

Demographics are very often the backbone of any dynamic microsimulation models. They are also usually the first step of the simulation. Microsimulation models rely on individuals, so if the number of persons and their characteristics are not well estimated, not only the results for population aggregates will be wrong but the results for income, social transfers, and the effects of policy will not be reliable either. Even in models that are not population-focused, this module is important.

Microsimulation modelling has been used for a long time by demographers to investigate fertility. Since its definition by Guy Orcutt (Orcutt, 1961)), whose first model included an extensive and detailed demographic module, a number of biological and demographic models have been built in the USA (Sheps et al., 1973), in Sweden (Hyrenius and Adolfsson, 1964; Hyrenius et al., 1966), and in France (Léridon, 1977). The biological fertility models are women-based and reconstitute the monthly fecundity and fertility processes during the reproductive life. Recently, Leridon (2004) used such a model to investigate the effect of the postponement of childbearing age on the total number of children — and, in particular, on the number of children that people will not have due to age-acquired sterility that prevents them having the number of children they would like to have and its effect on the total fertility rate. The demographic models (Hyrenius et al., 1966), Schofield (Smith, 1987); and more recently using the Family and Fertility Surveys, (Spielauer and Vencatasawmy, 2003) also investigate fertility but using different data, age and parity-specific fertility rates instead of using conception rates and then following the outcome of each conception. They investigate the fertility process both from the quantum and tempo point of view.

A related use, and often one of the main interests of demographers using microsimulation, is that it allows the study of kinship. Whether they live in the same household or not, ties between members of a family can be kept and the size and characteristics of families can be investigated (Hammel et al., 1981) , (Imhoff and Post, 1997), (Le Bras, 1973), (Pennec, 1997). With the ageing population and the potential threat to the care of the disabled elderly, microsimulation is the perfect tool to estimate the future population of potential carers. As the ties between the different members of the family are kept, the number of carers can be computed — not only their number, but also their characteristics (age, sex, being in the labour force or not, being retired or not, being disabled themselves or not …) (Keefe et al., 2004; Duée et al., 2005; Tomassini and Wolf, 2000).

This is why many models use a set of covariates to try to create a better distribution and include some time and birth spacing variables. This is, of course, critical in the case of the demographic models, as it is their aim to mimic fertility both in quantum and in tempo. But this is often not the only goal: more policy oriented models often also try to have a good representation of fertility and of family structures.

The following section outlines how demographics have been modelled in different dynamic microsimulation models. For some of them, demography is only a means to achieve a plausible, total number — and, maybe, the age and sex structure of the population. In that case, the modelling is therefore quite simple. But most models include quite detailed demographic modelling, because they need to also track the characteristics of the population as a whole and, very often, household and the family characteristics. For example, the family is often a decision-making unit in terms of economic behaviour and such expenses as housing. Welfare benefits, taxation schemes and some other variables are also often family-based.

Tables 1 and 2 give some examples of how births and deaths have been modelled in models like APPSIM all around the world. It shows the different modelling approaches and the covariates that came through the imagination of the modellers. This is a good starting point when investigating approaches for new modelling, as some "regularities in the covariates can be seen".

In summary, Table 1 suggests the modelling of fertility is mainly based on demographic data, such as age of the mother, marital status, duration since beginning of the union/marriage, number of previous children, and duration since previous birth. But, whenever possible, data on socio-economic status are also included — either by education level or by income level. A third set of variables are more dedicated to the modelling of the individual's current situation, such as being in full time education; being in the labour force, or having a very young child.

Where fertility is concerned, different benchmarks must be hit — some period ones (like total number of births and total fertility rates); and some cohort ones (like the distribution of women by number of children and completed fertility rates). Not only the total number of children and their distribution, but the age at childbearing and the duration between two births are of importance and must be included in the modelling (as they can impact on the labour force participation of the mother or change the receipt of benefits that vary according to the number or age of children, or are family related). Childlessness is another important phenomenon that has to be taken into account, as it has a strong impact of some of the topics related to ageing.

Mortality is a one time event and therefore is easier to model than fertility. Age and sex mortality rates are the basic indicators used in the models and, whenever possible, these rates also include health, family and socio-economic variables. Health variables are mainly related to disability, while socio-economic status can include higher education level or being in the labour force or occupation (and, for those retired, previous occupation).

Some models use survival analysis, but most of them use transition rates modelled by regression techniques (hazard functions or logistic regression equations). We will not go further into the comparison of models here as our aim when gathering the covariates used in the different models is not to compare the models. Obviously, the modelling of each event and the level of detail used depends on one side on the aim of the model; some objectives and policies may be more or less related to demographic behaviour. On the other side, the availability of data severely constrains modellers, to different degrees in different countries. As Harding points out: 'dynamic models are only as good as the data upon which they are based' (1993, p. 29).

Table 1 **Fertility covariates in selected dynamic microsimulation models**

Model	**Births**
Camsim	set of parity progression ratios, time between marriage and first birth. Only married women can bear a first child but the birth interval distribution from marriage to first birth permits pre-marital conceptions (Smith, 1987)
CORSIM	Age, birth(t-1), birth(t-2), duration of current marriage, earnings, family income, homeowner status, marital status, parity, schooling status, work status (F/T, P/T) , have child, marital status, race, work status
Demogen	age, marital status, parity
Destinie	parity 1: duration since end of education, age , education parity 2 (if same union as parity 1): duration since previous birth, education, age parity 3 and over (if same union as parity 2): duration since previous birth, education, age, parity parity 2+ (if not same union as previous birth): duration since new union, age, parity (Duée, 2005)
DYNACAN	Age, marital status, education level, employment status, parity
Dynamite	age, marital status, parity
DYNAMOD	model pregnancy and the outcome of the pregnancy (still born, live birth). 3 distinct kinds of births: premarital, 1st marital and subsequent marital (marital=de jure and de facto) equations based on women's education participation, educational qualifications, employment status, age. Marital births are also based on marital status, employment status of the husband, duration of the relationship, parity, duration since last birth (Abello et al., 2002)
DYNASIM III	Seven equation parity progression model; varies based on marital status; predictors include age, marriage duration, time since last birth; uses vital rates after age 39 (age-race-parity-specific probabilities); sex of newborn is assigned by race; (Favreault and Smith, 2004a)
Famsim	Age, Marital status, duration in marital status, in education, duration of schooling, in work, duration work, trend, duration since last birth (birth parity 2+), parity (birth parity 4+) (Spielauer and Vencatasawmy, 2003)

Harding	age, sex and parity (only for married women) (Harding, 1993)
Irish dynamic microsimulation model	Age, parity, education level, father's education level, marital status
Italian cohort model	age of mother, parity
Japanese dynamic model	age, marital status, parity
kinsim	age, marital status, parity (Imhoff and Post, 1997)
Biological models	age, marital status, time since last pregnancy according to outcome, sterility (Hyrenius and Adolfsson, 1964; Léridon, 1977)
LIFEMOD	marital status, age, parity progression ratio
LifePaths	age, marital status, parity, time [education, birth cohort, age at marriage, province of birth; the decision to have a child is modelled and then the different steps towards the birth, each birth occurs nine months after conception, 2) a spell of infertility lasting three months follows each birth and 3) a new fertile spell begins immediately after that. As a side benefit, the careful attention given to the timing of these events makes possible straightforward models of, for example, maternity leaves from employment and marriage following pregnancy.](Statistics Canada, 2006)
Melbourne	age of the mother, marital status, number of dependants
Microhus	hourly wage rate, disposable income, education, years of work experience, duration since previous child, other socio-economic variables
MOSART	age, parity, age of youngest child (Fredriksen, 2003)
Nedymas	age, year of birth, marital status, parity
PRISM	marital status, age, parity, previous employment status
SAGE	women living with a partner: age, marital status, duration in marital status, parity and age of youngest child women not living with a partner: age, participation to full education, parity and age of youngest child multiple births: age (Scott, 2003)
SESIM	Only women who move from her parents can give birth, age 18-49 First birth: age, marital status, pensionable income (quartiles), indicator for market work, highest education. Following birth: parity (up to 3, 4 is the maximum number of children modelled), age, marital status, pensionable income (quartiles), indicator for market work, highest education, age of youngest child (Flood et al., 2005)
Sfb3	Age, marital status, duration of marriage and parity

SOCSIM	monthly rates/ age, sex, marital status, waiting times between events (Hammel et al., 1990)
SVERIGE	marital status, family earnings, education level, working status (working or not), number of live children already born, whether a birth occurred last year, whether a birth occurred the year before the last, the country of birth of the mother, number of years that the mother has spent in Sweden (Holm et al.)
Swedish cohort model	age

Source: (O' Donoghue, 2001) and authors mentioned in the table

Table 2 **Mortality covariates in selected dynamic microsimulation models**

Model	**Mortality Covariates**
Anac Model	age, sex
Belgian Dynamic model	age, sex
Camsim	age, sex (Smith, 1987)
CORSIM	Age, birth place (US or other), education, employment status, family income, marital status, sex, race, marital status
Demogen	age, sex + health module
Destinie	age, age at end of education, sex, disability (with disability module)
DYNACAN	gender, age, education, marital status, employment status, disability status, time, region
Dynamite	age, sex
DYNAMOD	age, sex, disability
DYNASIM II	Married women 45-64: age, race, sex, marital status, education, number of children others: age, race, sex, marital status, education
DYNASIM III	Three equations; time trend from vital statistics 1982-97; includes socio-economic differentials; separate process for the disabled based on age, sex and disability duration derived from Zayatz (1999) (Favreault and Smith, 2004a)
Famsim	age, sex, marital status (Spielauer and Vencatasawmy, 2003)
Harding	age, sex, education status (Harding, 1993)
Irish dynamic microsimulation model	Age, occupational group, gender, in education

Italian cohort model	age, sex
Japanese dynamic model	age, sex
kinsim	age, sex (Imhoff and Post, 1997)
Biological models	age, sex (Hyrenius and Adolfsson, 1964; Léridon, 1977)
LIFEMOD	age, sex
LifePaths	age, sex, [for one study and for elderly people, disability status, living in nursing home or not]
Microhus	age, sex
Midas	age, ethnicity, gender
MINT	age, time, ethnicity, education, marital status, permanent income, disabled
MOSART	age, sex, disability, marital status, educational attainment (Fredriksen, 2003)
Nedymas	age, sex, marital status
Pensim/2	age, gender
PRISM	disability status, age, sex, years of disability
SAGE	age, sex, social class (Scott, 2003)
SESIM	age 0 to 29 : sex and age age 30 to 64: sex, age, indicator for early retirement, pensionable income (quintile), marital status age 64+: sex, age, indicator for early retirement at 64 years of age, marital status, highest level of education. (Flood et al., 2005)
Sfb3	Age, sex, family status
SVERIGE	Age, sex, marital status, family earnings, education level, working status(Holm et al.)
Swedish cohort model	age, gender

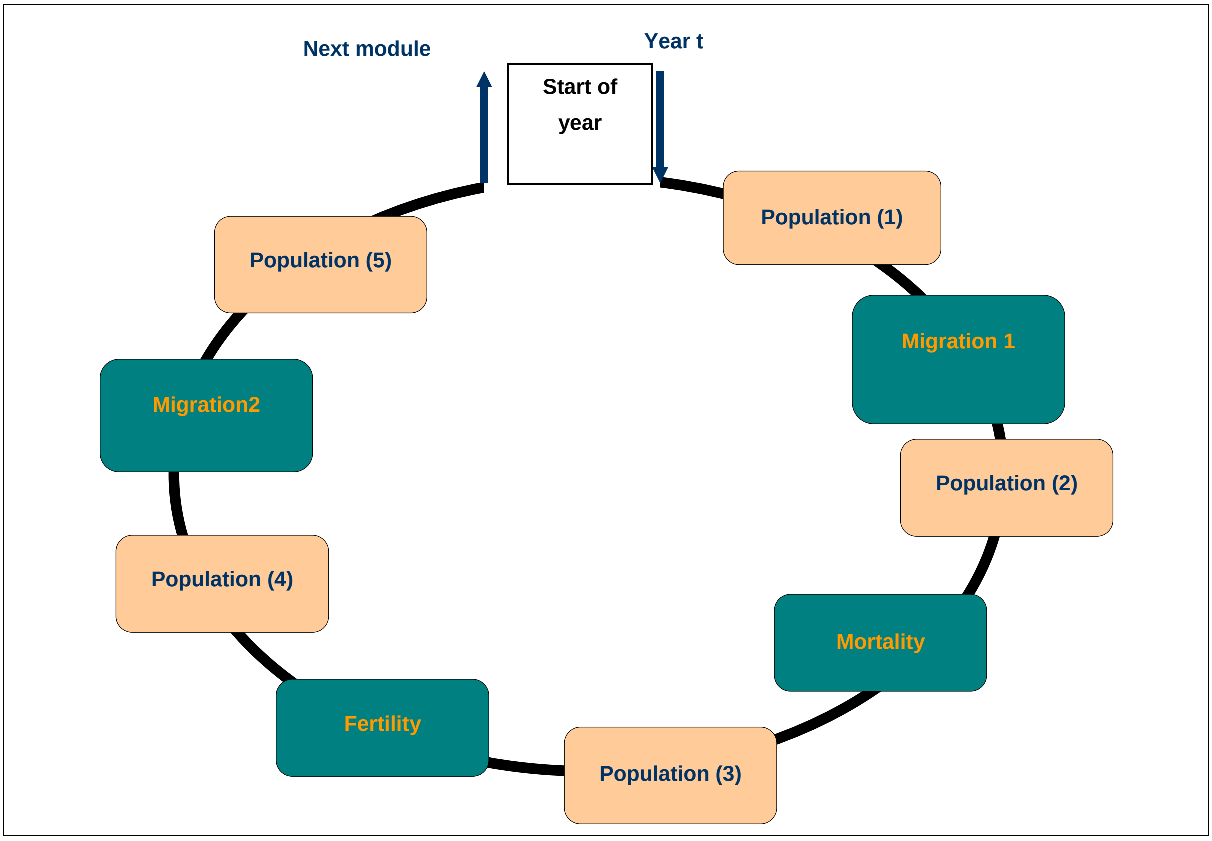
Source: (O' Donoghue, 2001) and authors mentioned in the table.

4 DEMOGRAPHIC EVENTS IN APPSIM

4.1 Order of the events in the demographic module

After an overview of the different events included in the demographic module, we will focus on the data and estimates for fertility and mortality.

Figure 2 **The events and their order in the demographic module**



The proposed order of events in the fertility, mortality and migration modules is summarised in Figure 2. In dynamic microsimulation models such as APPSIM, we apply probabilities to determine whether each event will happen or not — for example, every year we will test whether a person will stay alive or will die (or will emigrate or have a child).

The probability that an event arises is estimated using the most detailed data possible and whenever possible using longitudinal data — as they allow us to determine transitions between different states. For example, the probability of divorce is more related to the number of years spent in marriage than to age.

The order of the variables is not neutral, as if we model first mortality, a person selected to die cannot emigrate nor have a child later on. The order of the modules must thus reflect a logical interaction pattern and the probabilities of each event must be conditional probabilities taking into account the implication of the order.

In microsimulation models, each event is modelled in a more or less dependent way. While projections run using a cohort-component method use the average population and not the population at the beginning of the projection step for events such as fertility (so as to be able to take into account the fact that more than one event can arise during the year), in microsimulation, each event arises independently. If death is the first event then, for example, any women selected to die will not be able to give birth. We impose a causality in the model that has to be taken into account in the estimate — at least when the probability that both events arise during the same period of time is not negligible.

The order proposed here is first, immigration, then mortality, fertility and emigration. With this order, we can take into account:

Immigrants are alive when they arrive but can die and have children as soon as they arrive on the Australian soil;

Mortality before fertility: this assumption reduces the number of orphans at birth but, in a country like Australia, it is fortunately a quite rare event to have the death of a woman in the year of the birth of her child. These two events are estimated independently — that is, the risk of mortality is not dependant upon whether a person will have a baby. Therefore, this is a less risky assumption in many respects, as estimating births after mortality will reduce orphanhood in the 1st year of life but the error will not be great and we assume this error to be better than the reverse. The other reason for using this order is that the data used to determine the fertility estimates are based on surviving women.

Emigrants are alive at the moment of their departure but could have had a child just prior to emigration.

This order presents a drawback related to the fertility and mortality of migrants. If we assume like all projections that immigrants and emigrants stays on average half the projection period in the country, death and fertility rates of migrants are half those who are staying the whole year in the country. While it is possible to have a variable distinguishing immigrants and applying half the rates to them, it is not possible to do the same for emigrants because we don't know who they are when fertility and mortality are performed. This can lead to a small increase of mortality and fertility.

It is possible to avoid this problem by using the following method[2]. Let's consider that half the immigrants will arrive at the beginning of the period and half of them will arrive at the end of the projection period. Those arrived at the beginning of the projection period will face the whole year the events death and births while those arriving at the end of the period will not face any event. The same way for emigrants, we consider that half of them will leave the country at the beginning of the projection period and therefore will not face any other events in the country. The other half of the emigrants will leave at the end of the period and will face the other events during the whole year.

This shortcut is possible with the assumption that they stay on average half of the period; instead of all the immigrants and emigrants facing the events with half the probabilities because they stay half the year in the population, we will have half the immigrants and emigrants with the full rate.

4.2 Mortality

No data are available within Australia according to socio-economic level, disability status, occupation foreign vs. Australian born etc at individual level - only at SLAs (statistical local area).

The first version of mortality within the model is therefore very simple as it uses mainly the probability of dying by age and sex, determined by ABS for their population projections. We are still investigating how we can include some differentials in the model.

---

[2] The authors would like to thank Laurent Toulemon for this very pertinent suggestion.

The ABS data are up to age 100+, but we wanted to increase the upper age-group, given the likely increase in the number living past 100 years. Based on the work by Thatcher-Kannisto, we have mortality rates from 100 to 105+. We have used this distribution by age to determine the distribution by age over 99 from 2005 to 2050 (Pennec and Bacon, 2007).

4.3 Fertility

Data on fertility available in Australia.

Vital statistics

Births and confinements are registered by different bodies – the different State Registrars and the Australian Institute of Health and Welfare through the registers of midwifes. These two series present discrepancies, as a time lag exists between confinements and registration. The time difference differs from state to state and according to year. In 2003, the Perinatal Data Collection reported the occurrence of 255,100 live births in Australia, 1.6 per cent more than the 251,200 births registered in the same year (ABS, 2004); (McDonald, 2005).

The contents of the birth registration forms differ by state but, nevertheless, most of the variables are the same in all the states. One difference lies in the parity of the child; in some states, the parity of the child within the overall number of children of the woman and within the current relationship is available (Queensland, South Australia, Western Australia and Tasmania) whereas, in the other states, only the latter is available.

Specific surveys

The first possible source, one that was used for the DYNAMOD model, is the ANU survey “Negotiating the life course”. It is a panel of 2 247 persons designed to study the interaction between family life and labour force participation.

The second and more recent survey is the Household, Income, and Labour Dynamics in Australia survey (HILDA). HILDA is a broad social and economic survey. As it has a longitudinal design, most questions are repeated each year. In addition, each year a special topic is covered — such as in wave 1 the family background, in wave 2 the household wealth, and in wave 3 retirement and plans for retirement. Private health insurance and youth are covered in wave 4. The panel began in 2001 with a national sample of Australian households occupying private dwellings of 6 872 households and 13,969 individuals. Members of the original survey in 2001 have been traced and interviewed annually, along with members of their new households.

Census

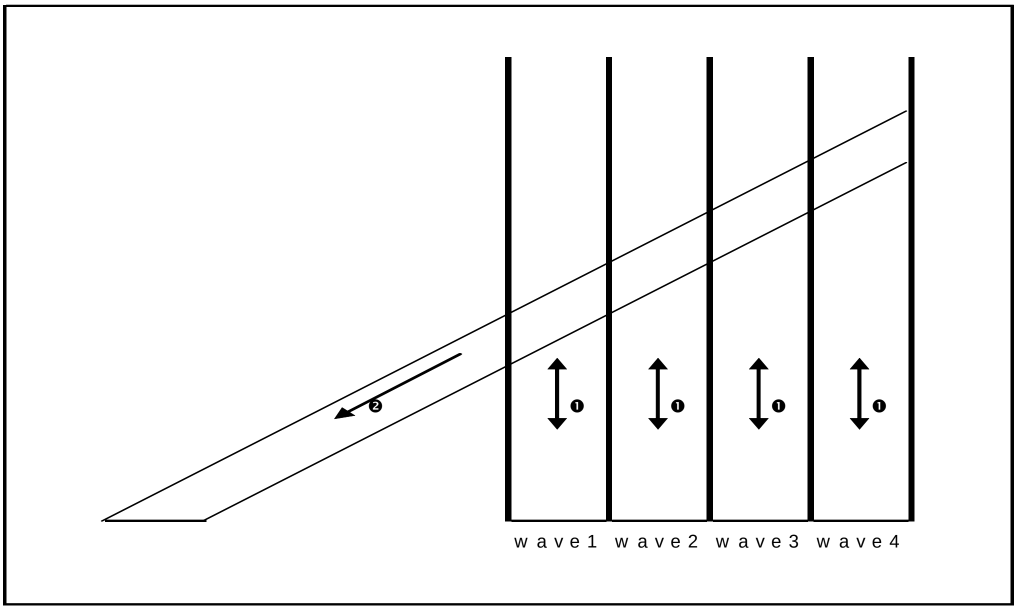
The quinquennial census is another useful source of data for fertility analysis. For some years, some specific questions on the total number of children ever born were added to the usual question of the dependent children living in the household. Some interactions of demographic variables with housing, economic, and labour force characteristics can be studied using the census data. A major advantage of the census, like vital statistics, is that it is an exhaustive source — but the number of variables is limited, as is the level of detail for each topic.

Vital statistics and census data will be used mainly to benchmark the APPSIM estimates, as our estimates need to rely more on micro-level data able to report interactions between variables. To determine the fertility estimates, we chose to use the HILDA panel data. The reason for this choice is that it contains both longitudinal historic data and cross-sectional data and its sample size is bigger than the ANU survey. One set of benchmarks to check the summed micro fertility estimates against is the ABS population projections.

Fertility in HILDA

For each responding person, for each year, a set of questions are asked about the children currently living in the household and those not living in the household — and another set are asked to elicit the likelihood of more children arriving in the near future. According to the place of residence of the child, questions on the involvement of the parent not living with the child if applicable are registered; child care and financial aspects are major topics investigated with this questionnaire. Fertility can be studied according to different approaches: a retrospective longitudinal approach (2) that looks at the past fertility of the different women and a cross-sectional approach (1) that looks at fertility at a point of time (figure 1).

Figure 1. Possible approaches with the HILDA survey



With the retrospective approach, for each woman, we can reconstitute her fertility history and, by aggregating the individual records by birth cohort, we can see the trends that affected the fertility level of earlier decades and can have a better insight into the future (as we can compare the behaviour of different cohorts at the same age and deduce from that what is likely to happen in the future).

On the other hand, with the cross-sectional approach, we have the current fertility patterns — and the design of HILDA is such that the key questions about economic characteristics and labour force participation are only available at the date of the interview, not for many years earlier as with the retrospective perspective.

Fertility estimates in APPSIM

As already mentioned, it is important that the projections performed with APPSIM give coherent results, at both cross-sectional and longitudinal levels. Our estimates must be as close as possible each year to the right total fertility rate, so that we will have the right number of births — but we also have to pay attention to the completed total fertility rate (that is, the fertility at birth cohort level). These two figures are both averages — but another important aspect of fertility modelling at an individual level is to obtain a good distribution of the number of children by woman. One aim of APPSIM is to study ageing, and in ageing we can also include disability and need of care, formal or informal. We know that informal care is provided mainly by the spouse and the children, so it is crucial to have a good representation of the distribution of children and, in particular, childless persons. Childless persons and those with no disability-free spouse or children rely almost entirely on formal care. The timing of childbearing can also be of importance, for labour force participation of women, for education cost, and for caring when old age of the parents arises.

The fertility estimation process

To estimate fertility, we follow an approach similar to the SAGE model as we have a similar dataset (Scott, 2003).

From its fertility module, HILDA allows us to use variables from both retrospective variables and current status or status last year as predictors of whether the woman will have given birth by the time of the next annual interview. HILDA combines both a retrospective approach at first interview and a panel approach.

From HILDA, a person-year file was created using records from 2001 to 2004. To generate a sufficient sample size these pooled records were then treated as independent. Table 1 gives the number of births registered in each wave. The eligible sample was restricted to records for women aged between age 15 and 49. Women who had a child that died have been dropped from the analysis because we don't know the date of birth of this child and therefore some of the intervals between births are incorrect and we cannot know which ones[3]. Cross sectional weights provided with the HILDA data were applied to the pooled records to adjust for unequal probabilities of response.

---

[3] This limit will be able to be removed in the future, as in a future wave of HILDA questions on children who have died will be asked.

Table 1: Number of births by parity in the different waves of HILDA.

	Number of 1st births	Number of 2nd births	Number of 3rd births	Number of 4th and over births	All births
Wave1	88936 (92)	90262 (85)	38902(45)	19493(28)	237593(250)
Wave2	88194(68)	86100(76)	30499(34)	20199(25)	224992(203)
Wave3	134581(95)	110548(86)	34482(34)	24817(18)	304428(233)
Wave4	107136(98)	79453(61)	36477(31)	25716(21)	248782(211)
All 4 waves	418847	366363	140361	90225	

Note: The figures within parentheses are the unweighted figures.

Due to the retrospective recollection of the data, births are registered after they occur, whatever the approach chosen. This is obvious for retrospective longitudinal approaches as we reconstitute the fertility biography of the women interviewed in 2001 for instance, but it is also the case even for the panel and cross-sectional approaches. In wave x, the births we observed are those that have occurred since the previous wave – e.g. the children aged 0 in wave 3 are the babies born between the 30th June 2002 (wave 2) and the 30th June 2003 (wave 3)[4]. Therefore, to determine the probability of having a child, we need to use the situation of the woman at the 30th June previous to her childbearing – i.e. for the mother of the baby born between wave 2 and wave 3, we have to use the value of the variables at wave 2. This means that, for the babies born in wave 1 (2001), we have to reconstitute the variables as if there will be a wave "0" [5].

---

[4] Some corrections to the file have been performed in order to create the age and number of children at the 30th June previous to the interview rather than at the date of the interview.

[5] For the first cut of the estimates, we assume that the value of variables education, labour force status and marital status remain the same in wave 0 as their values in wave 1. This is a strong assumption, in particular for young persons, but it will be corrected in the future.

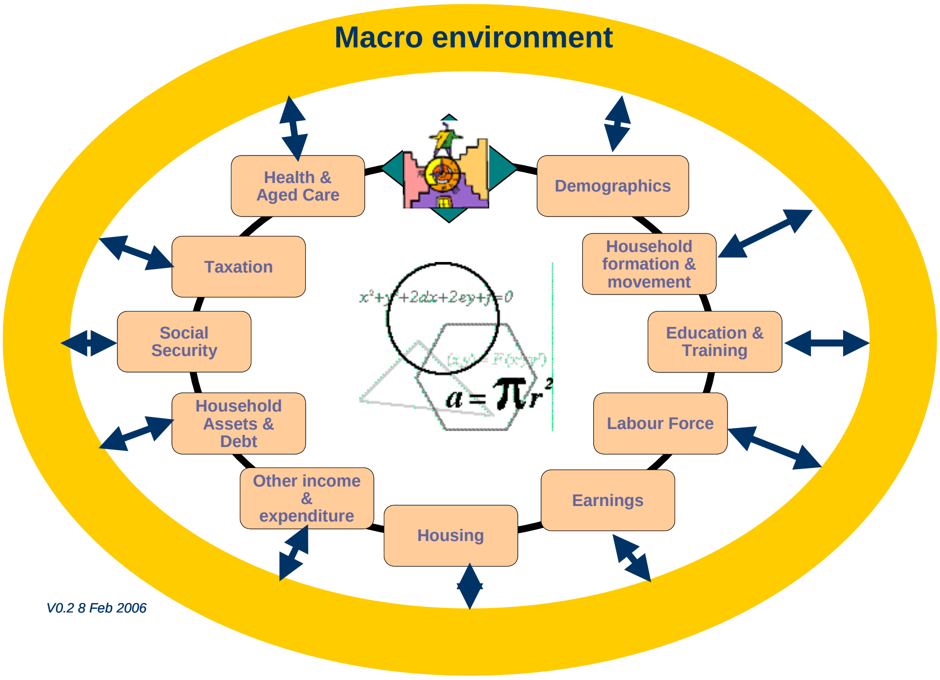
The first version of the fertility estimates developed here are mainly demography based — that is, mainly demographic variables are used along with quite simple education and labour force status variables. A future model could include more detailed demographic variables (such as duration since union) and some socio-economic context variables (like education and labour force status of the spouse, level of income, effect of being in a new relationship, duration of cohabitation with partner, etc). This first version will also rely only on a cross-sectional approach, using limited duration variables. A second version could try to capture more of the long-term trends that cannot be replicated with a 4 year long panel.

One constraint for the first version is to use in the estimation variables that are available in the 1 per cent household census sample or relatively easy to impute. (The 1 per cent census sample is the core dataset chosen as starting population of the microsimulation model APPSIM.)

To produce the probabilities of having a child during the following year, a set a logistic regressions have been performed. These regressions have been used to check the significance of the variables and to determine the parameters of the different covariates. A different equation is estimated by parity — and, for the probabilities of having a first or a second baby, a second level of stratification of whether the women is living with a partner or not has been introduced. The variables included in the estimation are demographic variables (age, age2, current legal marital status, current living condition –with a partner or not -, and age of the mother at 1st birth, duration since previous birth for parity above 1), level of education, labour force participation.

Calibration/Alignment

One of the issues that has received much attention during the past decade is the need to align summed micro estimates to benchmark data sources that are regarded as reliable within a country (such as official population projections) (Anderson, 2001). This topic is more fully developed in another paper (Bacon and Pennec, 2007). We note here the aggregates and benchmarks that we should be able to 'hit' with the fertility estimates in the APPSIM model



The number of births ought to align with the aggregate number of births based on fertility rates. In theory, it is desirable that the model outcomes also align with external benchmarks by age of the parents, number of births by the mother (first child, second child, etc), marital status of the parents, and aggregate number of multiple births (sets of twins, etc).

The fertility rate should reflect historical rates where available and the projected rates (general fertility rates and age-specific rates) should be a user-defined input parameter.

Different sources of information will be used in order to check and calibrate our estimates. Among these are the vital statistics, some surveys like the ABS survey on family characteristics and the ABS population projections.

One proposal for this micro-macro linkage when demographics are concerned is to link a cohort component model to the microsimulation model. Our idea is not only to align the microsimulation results to the "official results" but to built a cohort component model that take into account certain specificities of microsimulation like (1) the events are determined according to the age at the beginning of the simulation step that could not be equal to the age at the event; (2) rates are not applied to the mean population but to the population as it is when this event is calculated.

This approach leads to a real consistency of the macro and micro levels; gives a better consistency to the user when he/she wants to use some other scenario e.g even if the user wants to increase only the fertility, it has effects on the number of deaths so this method recalculates all events involved.

4.4 Implementation in APPSIM

Each women aged 15 to 49, will enter in the fertility loop. Her probability of having a child each year is related to a set of covariates — the number of children she has ever had (parity), her age, education level, labour force participation status, migration status, and for parity 2 and 3, her age at the first birth.

This implementation can be drawn as follows:

Once a women is set to have a birth within the year, some characteristics of the outcome of the confinement have to determined, such as whether the women gives birth to one or more children. Next, some characteristics of the newborn have to be determined; their birth cohort is deduced from the year the birth arises, but we need to determine the sex of the newborn and also whether the newborn will survive to his or her first birthday.

Multiple births

Once a woman has been selected to give birth in the coming year, we will determine if this confinement leads to multiple births or not. The probability of a woman giving birth to more than one baby is based on the average rate of multiple births observed in Australian in 2000-2005. The maximum number of multiple live births that can be allocated is triplets. As in other countries, women aged 30 and over are more likely to have multiple births than younger women. This is related to the ageing of childbearing and the more frequent use of medically procreative fecundation at these ages. On average 1.7 per cent of confinements lead to more than one child. Women aged 30 and over are 1.5 times more likely to have twins or triplets than younger women. In the model, for those having more than one child, 97.9 per cent will have twins and 2.1 per cent triplets

Sex of the infant

The allocation of the sex of a child follows the sex ratio at birth in Australia of around 0.512 (105 boys for 100 girls) male to female babies. This ratio will be replicated in APPSIM, with this choice being influenced by the fact that APPSIM is a national model which doesn't incorporate the State level. (If State and Territory were included within APPSIM, then the sex ratio should be state-related, as there are differences according to states in the sex ratio.)

Infant mortality

The probability of death between the birth and the end of the projection period (here the financial year) is given in the projected ABS population projection by sex. The death rate is 4.92 per cent for boys and 4.59 per cent for girls in 2004.

5 CONCLUSIONS

This paper has provided an introduction to the likely overall structure of the Australian APPSIM microsimulation model and the modelling of fertility and mortality within that model. Further information on the development of APPSIM will be progressively documented and placed in the Working Paper Series at www.natsem.canberra.edu.au. As this paper has made clear, we face considerable challenges in modelling these processes at the required degree of detail, due to data deficiencies – a dilemma faced by most dynamic microsimulation modellers!

REFERENCES

Abello Annie, Kelly Simon, King Anthony. 2002. *Demographic projections with Dynamod-2*. Technical Paper No 21, NATSEM, University of Canberra (available from www.natsem.canberra .edu.au) Canberra, Natsem, 21, 50 p.

ABS. 2004. Births Australia.

Anderson, JM. 2001. *Models for Retirement Policy Analysis*, Report to the Society of Actuaries, USA, (available from www.soa.org.ccm).

Bacon Bruce, Pennec Sophie .2007. Calibration and alignment in APPSIM, mimeo, NATSEM.

Duée Michel, Cyril Rebillard, Sophie Pennec. 2005. Les personnes dépendantes en France: Evolution et Prise en charge. IUSSP website (www.iussp.org) International Union for the scientific study of the population. XXV International Population conference., Tours (France), 18-24 juillet 2005.

Duée Michel. 2005. *La modélisation des comportements démographiques dans le modèle de microsimulation DESTINIE*. Paris, G 2005 / 15, 47 p.

Favreault Melissa, Smith Karen. 2004a. *A Primer of the Dynamic Simulation of Income Model (Dynasim3)*. The Urban Institute, 22 p.

Flood Lennart, Jansson Fredrik, Pettersson Thomas, Sundberg Olle, Westerberg Anna. 2005. "SESIM III - a Swedish dynamic microsimulation model." Sweden: Ministry of Finance.

Fredriksen Dennis. 2003. The MOSART model - a short technical documentation. In: Norway Statistics *Paper presented at the International Microsimulation Conference on Population, Ageing and Health: Modelling our Future, 7-12 December*. Canberra.

Galler, H. 1997. 'Discrete Time and Continuous Time Approaches to Dynamic Microsimulation Reconsidered, Technical Paper No 13, NATSEM, University of Canberra (available from www.natsem.canberra .edu.au)

Hammel Eugene A., Wachter Kenneth W., McDaniel C. K. 1981. The kin of the aged in 2000 A.D. In: Morgan James, Oppenheimer Valerie, Kiesler Sara New York, Academic Press, vol. 2 -Social Change, p. 11-39.

Hammel Eugene A, Mason Carl, Wachter Kenneth W. 1990. *SOCSIM II. A sociodemographic Microsimulation Program Rev. 1.0. Operating manuel*. 29, 76-(5) p

Harding Ann. 1993. *Lifetime Income Distribution and Redistribution: Applications of Microsimulation Model*. Amsterdam, North-Holland.

Harding Ann , Gupta Anil (eds). 2007a. *Modelling our future: Population Ageing, Social Security and Taxation*. Amsterdam, North-Holland.

Harding, A and Gupta, A, 2007b, 'Introduction and Overview', in Harding, A and Gupta, A. (eds), pp. 1-29.

Holm Einar, Holme Kirsten, Makila kalle, Mattson-kaupi Mona, Mortvik gunnel. *The Sverige microsimulation model - content, validation and example applications*. Kulturgeografiska Institutionen/SMC - university of Umea, 54 p.

Hyrenius Hannes, Adolfsson Ingemar. 1964. *A fertility simulation model*. Göteborg, [Distributor: Almqvist & Wiksell, 31 p. p.

Hyrenius Hannes, Adolfsson Ingemar, Holmberg Ingvar. 1966. *Demographic models*. Göteborg, v. p.

Imhoff Evert Van, Post Wendy. 1997. Methodes de micro-simulation pour des projections de population. *Population (French Edition),* 52 (4, Nouvelles approaches methodologiques en sciences sociales), p. 889-932.

Keefe Janice, Légaré Jacques, Carrière Yves. 2004. *Projecting the future availability of informal support and assessing its impact on home care services, Part I: Demographic projections.* Health Canada. Halifax: Mount Saint Vincent University.

Le Bras Hervé. 1973. Parents, grand-parents, bisaïeux. *Population,* 28 (1), p. 9-38.

Léridon Henri. 1977. *Human fertility: the basic components*. Chicago, University of Chicago Press, 202 p.

Leridon Henri. 2004. "Can ART compensate for the natural decline in fertility with age? A model assessment", *Human Reproduction* 19(7), pp.: 1548-1553, July .

McDonald Peter. 2005. Has the fertility rate stopped falling. *People and Place,* 13 (3), p. 1-5.

O' Donoghue Cathal. 2001. Dynamic Microsimulation: A Methodological Survey. *Brazilian journal of economics,* 4, c 77 p.

Orcutt Guy H. 1961. *Microanalysis of socioeconomic systems; a simulation study*. New York, Harper, xviii, 425 p. p.

Pennec Sophie. 1997. Four-Generation Families in France. *Population an English Selection,* 9, p. 75-100.

Pennec Sophie, Bacon Bruce .2007. APPSIM - Modelling mortality and fertility, Working Paper No 5, NATSEM, University of Canberra (forthcoming), 53 pages.

Pennec Sophie, Bacon Bruce .2007. APPSIM - Imputing mortality Proposal for a first approach to extend ABS population projection after age 100 for APPSIM ,NATSEM, University of Canberra, mimeo.

Scott Anne. 2003. *Implementation of demographic transitions in the SAGE Model*. London, 5, 54 p.

Sheps Mindel C., Menken Jane A., Radick Annette P. 1973. *Mathematical models of conception and birth.* Chicago, University of Chicago Press, xxiii, 428 p. p.

Smith James D. 1987. The computer simulation of kin sets and kin counts. In: Bongaarts John, Burch Thomas K., Wachter Kenneth W. *Family Demography: Methods and their applications*. Oxford, Clarendon Press.

Spielauer Martin, Vencatasawmy Coomaren P. 2003. FAMSIM: dynamic microsimulation of life course interactions between education, work, partnership formation and birth in Austria, Belgium, Italy, Spain and Sweden. In: *Vienna yearbook of population research*. p. 143-164.

Statistics Canada. 2006. The Lifepaths Microsimulation Model: An Overview. Dernière modification.

Rowe, G. and Gribble, S., 2007, 'Lifepaths Model', in Gupta, A and Harding, A (eds), *Modelling Our Future: Population Ageing, Health and Aged Care,* Elsevier B. V., Amsterdam, pp 449-452

Ruggles Steven. 1987. *Prolonged connections: The rise of the extended family in nineteenth-century England and America*. Madison; London, The University of Wisconsin Press, 283 p.

Tomassini Cecilia, Wolf Douglas A. 2000. Shrinking Kin Networks in Italy Due to Sustained Low Fertility. *European Journal of Population,* 16 (4), p. 353-372.