Modelling Fertility in a life course context: Some Issues

Vencatasawmy, Coomaren P.

working paper

Österreichisches Institut für Familienforschung
Austrian Institute for Family Studies

Nummer **16 – 2002**

Titel **„Modelling Fertility in a Life Course Context: Some Issues“**

Autor **Coomaren P. Vencatasawmy**

working papers have only received limited review

ÖIF, Gonzagagasse 19/8, A-1010 Wien
Tel. +43-1-535 14 54-0
Fax +43-1-535 14 55
url: http://www.oif.ac.at
email: team@oif.ac.at

P.b.b.: Verlagspostamt 1010 Wien; DVR: 0855561

ÖIF- Österreichisches Institut für Familienforschung
Austrian Institute for Family Studies

Coomaren P. Vencatasawmy

Modelling fertility in a life course context: some issues

Coomaren P. Vencatasawmy PhD

Senior Scientist at the Swedish University of Agricultural Sciences (SLU - Sveriges lantbruksuniversitet), Department of Forest Resource Management and Geomatics;

Contact: *coomaren.vencatasawmy@resgeom.slu.se*

Abstract

The understanding of fertility behaviours is an important task from a social and political point of view. Although there are a number of theories of fertility decline, not many of them are amenable to modelling. This reduces their usefulness in a policy context where better forecasts as well as better understanding of how policies interact with fertility decisions are needed. In this report we have examined some of these theories and extracted components that could be useful for modelling and simulation. Furthermore, we describe a number of modelling paradigms and argue about the benefits of simulation in studying social phenomena that are too complex to express in analytical forms. Finally, we examine two different statistical approaches for modelling longitudinal data in a life course context: survival analysis and longitudinal analysis. We show that although the longitudinal analysis is more amenable to simulation models, they are computer intensive to fit.

Content

1 Introduction

In all western countries, period total fertility rates are below, and often well below, replacement level, and population growth rates are close to zero. However, on closer examination one can notice geographical differences. In Italy, for example, the level of fertility dropped by half, from 2.4 to 1.2, within the span of two decades. On the other hand, in Sweden, the level of fertility fluctuated quite much but never went below 1.6 (Chenais, 1996). In fact, Sweden has one of the highest birth rates in Western Europe, although there is a shift in child bearing to later reproductive ages and that cohort fertility is almost constant (Walker, 1995).

A number of reasons for these conditions in Sweden have been identified. Hoem (1993) stressed the importance of policies in the recent increase in fertility in Sweden. He identified income compensation paid during parental leave, the flexibility of parental leaves, facilitation of women's entry in the labour force and the minimal cost of child bearing and child rearing as having important effects of the fertility trends in Sweden. These views are echoed by Chenais (1996) who compared western European countries and stressed the importance of social policies, which support families, on fertility rates. Walker (1995) argued that it is the combination of economic conditions and policy initiatives (expansion of child care system and parental benefit program) that has contributed to the observed pattern of period fertility rates in Sweden. The Swedish boom was followed by a bust that caused a lot of research regarding the separation of timing and quantum decisions of fertility; now it is dominantly believed, that policies in Sweden did effect fertility decisions mostly regarding timing, as policies created a "speed premium" for short birth intervals that increased TFR temporarily [Kohler, 1999].

During the 1970s and early 1980s, Sweden served as a model country for the Austrian government in its endeavour to modernise the Austrian society and the Austrian welfare state. Yet, since the mid-1960's Austria has experienced a drop in total fertility rate from 2.8 to what thought to be a stable value of 1.5 in the 80's but the decrease has continued into the 2000's, albeit at a much slower rate. Much of the fertility decline has been concentrated at parities two and three, which means that Austrians have had declining rates for third and higher-order births. In a country with a strong two-child norm, the third birth is the first birth that couples may decide not to have (Hoem et al., 2001).

The timing of the birth of first child relative to other life course events such as marriage has also changed considerably in the past twenty years, probably more so in some European countries than in others (Jensen, 1999). Given that an important function of marriage traditionally has been to secure men the right to the resource children represented in a society where they were an economic benefit, this function was lost as children were turned into an economic liability about hundred years ago (Caldwell, 1982).

The study of fertility rates and trends is, therefore, a complex process that is multifaceted. To study the dynamics of such a system, versatile tools are necessary. Microsimulation can be one such tool. Microsimulation enables the modelling of a phenomena on a micro level, which means that knowledge of individual behaviour, other actors and decision-making units can be integrated within the model and the consequences of many individuals' behaviour or responses to external influence can be explored. The central characteristic of the microanalytic approach is the identification and representation of individual actors in a system and the way in which their behaviour changes over time. These actors can include individuals, households, firms and so on. The shift of focus from sectors or aggregates to the individual decision-making units is the basis of all microsimulation work. In this way modelling is taken more of an educational tool rather than a predictive one.

One can think of microsimulation as encompassing a variety of specific methodological tools and techniques that are finding growing use in empirical social science applications (Wolf, 1999). Such models are useful from a policy point of view as they assess quite well the overall short-term effects of policy change. Dynamic simulated population can be used to forecast outcomes in domains such as fertility, and mobility, under different regimes and they can serve as a test bed for comparing competing theories about social processes in which verisimilitude to historical population distributions is a relevant criterion. Although aggregate variables are important in explaining fertility changes, variation in fertility is typically observed at the individual level as a function of physiological differences and attitudes and behaviours of women and their spouses (Smith 1989).

One prerequisite of microsimulation is the ability to model social behaviours at the micro level. In the last three decades a number of statistical methods have been used to model social behaviours. Longitudinal data analysis has been proposed to overcome many severe inferential limitations that cross-sectional analysis entails. A number of studies have examined the power of longitudinal analysis and longitudinal data are believed to be indispensable for the study of processes over the life courses and their relation to historical change. Longitudinal analysis shares many property of time series analysis but also overcome many of the their limitations, for example sample size requirements.

Event history analysis with its continuous-time observation design was proposed to augment the power of longitudinal analysis. It is believed that the observation of events provides an attractive alternative to the observation of states and that event-oriented observations design offers richer information.

Almost in parallel, multilevel analysis leads to research into the interaction between variables that describe the individuals and variables that describe the social groups. In multilevel analysis a hierarchical structure of the population is assumed. The multilevel analytic framework is very useful, but the national –level data sets to which it has been applied do not constitute optimal or even appropriate research designs as the units of analysis at the macro level are too large to sustain the assumption of homogeneity with respect to the institutional features deemed crucial in imparting variance in fertility and fertility-related behaviours (Smith, 1989).

2 Aims

These different approaches to modelling behaviour entails both different designs for collecting the relevant data and different modelling aspects that needs to be addressed. Many of these issues have already been discussed in the literature but not at the micro level nor in the context of simulation.

In this report, we aim to discuss in detail several of these issues in the context of developing a fertility module for a microsimulation model. Furthermore, in discussing these issues, we shall mention the suitability of these models for modelling causal relationships and for representing the modelling of agent-based behaviours.

The modelling of fertility: Theories and paradigms

Modelling approaches should be related to social theory- some assumptions of how human beings behave in a society and how they communicate and make decisions. Such assumptions can either be explicitly articulated or they may remain as unexamined presuppositions. In any case, these assumptions have implications for the type of modelling that is possible and the nature of any changes that may be made in consequence (Lane, 1999). For a long time 'social research' was used as shorthand for 'sociological research' but over the years this field has grown to encompass many disciplines that integrates social and economic analysis and at the same time encourage interdisciplinary research (Williams, 2000). A detailed survey of the different paradigms and philosophies of social sciences is given in Lane (1999)The modelling of fertility, however, is not a simple matter. Many models have been suggested and since some of the components of the models are not easily quantified and measured, it is difficult to verify them. It is often difficult to know precisely what a given theory asserts, or what it predicts or implies by strict logical inference (Burch 1995, 1997). Van de Kaa (1996) has also made a critical review of many theories of determinants of fertility and fertility decline. He mentioned that reviewing demographic transition theory as described by different observers, with different backgrounds and experience, as being very confusing. He added that many of these theories have elements of plausibility, which stand the proof of time and it is very difficult to choose between them. Hotz et al. (1997) have surveyed the intellectual development and empirical implications of the literature on the economics of fertility as it applies to fertility behavior in developed countries.

The overall result is a body of fertility theory that can leave even the careful reader hard-pressed to know precisely what is being explained, how key variables are defined, or precisely how variables are interrelated, functionally, causally or temporarily (Burch 1995). For example, educational attainment is an indicator of differences between individuals, and it is so on many dimensions. It is a measure of talent, income potential, social status or class. It should also measure individual autonomy, for one would expect more highly educated women to be more independent, of men in general, of their husbands in particular and perhaps also of general norms in society (Hoem et al., 2001). Educational attainment also visibly influences fertility indirectly. First, educational aspirations are a determinant of the age at which childbearing starts and there is a lot of evidence that age at first birth is important for the timing and extent of subsequent childbearing. Secondly, educational attainment influences labor force behavior, and for a woman an attachment to the labor force is likely to affect her childbearing decisions strongly, particularly for births beyond number two.

Furthermore, the field of demography lacks consensus on a unified, comprehensive theory of long-term aggregate decline in fertility (Burch 1997). It is recognised that there are heterogeneities present due to regional and personal differences. While some of these personal differences, such as educational attainment, are obvious many of them are not well known and the interactions between these personal attributes make it very difficult to study them separately.

The socio-economic perspectives assumes that individual fertility behaviour is affected by individual socio-economic variables and that both fertility change and fertility differentials within populations are functions of shifts in the marginal distributions of the socio-economic variables, induced by exogenous change in economic and social structures (Smith, 1989; Bongaarts and Watkins, 1996).

There are a number of theories that have attempted to relate aspects of the social and community structure to fertility. Easterlin (1975) assumes that couples make fertility decisions early in marriage in order to maximise life-time utility in which family size is defined as a function of income, subjective prices of children relative to other goods and tastes or preferences for children relative to other goods. An interesting feature of the Easterlin model is that it relates to supply as well as demand for children. Cain et al. (1979) recognize two major stratifications in Bangladesh: class relations (ownership of land and other resources) and gender stratification (gender inequity maintained by the institution of patriarchy). Caldwell (1982) theory of fertility and fertility decline recognizes the importance of institutional factors such as patriarchy and family structures and factors such as education but also stresses the importance of breath of education in the community rather than depth. Mason (1986) focuses on gender inequality.

Lesthaeghe and Surkyn (1988) attempted to reintegrate the sociology and economics of fertility by explicitly offering a model of joint economic and cultural dynamics. Their theory complement economic theories, and posit complicated interrelations between the cultural and economic spheres, with culture partly exogenous and partly endogenous with respect to economic change. They argued that ideational changes should be a significant part of economic optimisation.

Rosero-Bixby and Casterline (1993) put forward the hypothesis that diffusion effects act as a third type of causal agent of fertility transition above and beyond the traditionally studied demand and supply factors. They argued that substantial historical and contemporary evidence points to the existence of diffusion effects on the timing and pace of fertility transition. Diffusion is described as an endogenous feedback dynamic that occurs when adoption of birth control by some individuals influences the likelihood of adoption by others. They present three states through which married women of reproductive age go through: 'natural', in which they have no interest in controlling their fertility, 'latent', in which they want to control fertility but are not doing so for whatever reason, 'control' in which they are deliberately controlling fertility. This expressed in a dynamic deterministic model with constants controlling entry into the three states. Such models allows the simulation is closed and efficient way (Burch, 1999).

McDonald (2000) argued that the fertility transition from high to low levels has been associated mainly with improving gender equity within family-oriented social institutions, indeed most exclusively within the family itself. The fall in fertility is associated with women acquiring rights within the family that enable them to reduce the number of births to more desirable levels.

Walker (1995) used a simple neoclassical economic framework to examine the effects of Sweden's rapidly changing policy environment and economic conditions on fertility behavior. He found that female wage rate has significant negative effects on the timing and spacing of births. The estimated effect of parental benefits was also negative.

In Sweden Rephann (1999), found that working did not have an inhibiting influence on fertility decision, as it is true in other western country. A result, which he attributes to the way parental leave benefits are tied to working income. He also mentioned that there is evidence that the spacing between births and timing of births are influenced by the way the parental leave benefits are dispensed for women who elect to have a second child.

In correspondence to the various paradigms of fertility, a variety of variables are used in the modelling of fertility. Some good examples include Hoem et al (2001) who used the following groups of explanatory variables for modelling risk for third birthsSocial background (number of siblings, religiousness)

- Demographic outcomes (age, age at first and second birth, sex of first two children, whether the birth-union was her first or second union or of a higher order union and current civil status in any month of observation)

- Educational attainment and labour force participation (educational attainment of respondent and of her partner, her employment history, her current job status in any month of observation)
- Period factor (to reflect economic trends and the development of parental-leave policies).

Life course

The life course model arose out of the confluence of several major theoretical and empirical streams of research connecting social change, social structure, and individual action. It grew from research on how social system needs become articulated with individual goals through the connections between social structure and personality and how, in turn, individuals continuously try to change the large society. It also attempts to integrate the structural and dynamic approaches resulted into the life course paradigm. This paradigm has become the dominant one in demography.

Prior to the life course paradigm, social scientists generally either examined a snapshot view of the impact of social construct on the individual or took a dynamic approach that traced the story of lives over time. The former, called the structural approach, generally focused on either the interconnections of the macro social order (the whole society) or the micro level (small groups face to face interaction). The importance of the surrounding environment for the individual (ecological perspective) is stressed. In the latter, the stress was on special subgroups or individual actors. At the macro level, the dynamic approach was associated with conflict theories and at micro level it appeared in the life history tradition. This life history tradition was enhanced by theories of socialization and growth of self, resulting from social interaction and internalization but also by extended symbolic interaction theory that explained how new cultural and social movements arose. Both of these approaches are one sided.

Four key factors have been identified from the earlier works, which have an impact on life course of a person: historical and geographical location, social ties to others, personal control, and variation in timings (Elder, 1994). These are variants of the four-function model of the social system (latent pattern maintenance, integration, goal attainment and adaptation) as applied to the life course and social change. The insights that generated this four-faceted model of the life course have come as much from empirical research as from abstract theorizing.

The first factor (location in time and place) is a measure of cultural background that recognizes that individuals and social behavior is multi layered. Involving several different levels of the social and physical context. This can be understood that both the general and the unique aspects of individual location affect personal experience.

The second factor (linked lives) represents social integration. All levels of social action (cultural, institutional, social, psychological, and socio-biological) interact and mutually influence each other not only as parts of a whole but also as the result of contact with other persons who share similar experience. How well these different expectations, norms, or social institutions are integrated or internalized will vary. Some will show discontinuity and disruption, and others will show a smooth interweaving of individual attainments with social and cultural expectations; in any case, one expects to find differences among persons with different family backgrounds or experiences in work, education, or other domains.

The third factor (human agency) represents individual goal orientation. The motives of persons and groups to meet their own needs result in their actively making decisions and organizing their lives around the goals such as being economically secure, seeking satisfaction, and avoiding pain.

Finally, the fourth factor (timing of lives) represents strategic adaptation. To accomplish their ends, persons or groups both respond to the timing of external events and undertake actions and engage in events and behavior to use the resources available. How and when a person accumulates or deploys wealth or education, takes a job, or starts a family are examples of various possible strategies. The timing of life events can be understood as both passive and active adaptation for reaching individual or collective goals. For a more elaborate discussion on the emergence of the life course paradigm, see Giele and Elder (1998) and Rajulton (1999).

The life course paradigm moved research from single-factor explanations and introduced a combinatorial model with many variable. Life course refers to a sequence of socially defined events and roles that the individual enacts over time. It differs from the concept of the life cycle in allowing for many diverse events and roles that do not necessarily proceeds in a given sequence but that constitute the sum total of the person's actual experience over time. The life course concept also allows for the encoding of historical events and social interaction outside the person as well as the age-related biological and psychological states of the organism. Life course research looks for cohort effects that are due to the interaction of age and historical period. The resulting concept of life course has helped to operationalise the many meanings of timing and age. The sociology of age helps us to comprehend the many possible ways by which the life course of one person is linked to the fate of both age peers and the larger social order. A distinctive feature of life course studies today is the link between lives and societal changes.

The superiority of the life course idea lies in its flexibility and capacity to encompass many different types of cultural, social, and individual variation. Life course research looks for cohort effects that are due to the interaction of age and historical period. In the life course paradigm the effects of period (the distinctive historical and cultural events experienced by persons of a given age cohort), cohort (the socially shared experience of age peers) and age (the biological or developmental time since the birth of individuals) are represented as location in time, linked lives and human agency, respectively. Furthermore, the life course perspective introduces a fourth dimension, timing (i.e. the medium for integrating historical, social and individual activities). The model is about a system of action in which the individual group, and culture interact but with an added dimension of chronologically ordered through which all parts of the system are linked in a dynamic way.

Life-course analysis is oriented to the explication of historical effects, the process of status transmission, and behavioral change across the life span. These studies highlight the task of tracing the process by which certain conditions make a difference in cohort life patterns and the individual life course (Elder 1992). Introducing decision-making within a life-course framework makes explicit the existence of additional margins over which parents may choose to substitute their fertility, namely, childbearing at different ages over the life course. As such, changes in prices and income over the life course may result in changes in the timing of fertility demand, even if they do not cause lifetime fertility to change. The life-course context is also the appropriate setting within which to consider the consequences of the stochastic nature of human reproduction, including the choice of contraceptive practices and responses of parents to the realizations of this stochastic process.

Finally, the dynamic setting provides a more appropriate context within which to examine the relationships between women's labor supply, investment in human capital and childbearing decisions (Hotz et al., 1997). They propose a comprehensive model, in which the couple is assumed to make choices so as to maximize a well-defined set of preferences, subject to time and budget constraints, to technological constraints which govern the production and rearing of children and to constraints on the production of the woman's stock of human capital which determines the value of her time in the labor market at each age. The couple can make these decisions either in a certain (perfect foresight) or an uncertain setting, where the uncertainty they may face can arise either from the stochastic nature of the reproductive process or of the future income, prices or wage rates they may face.

Moen and Wethington (1992) compared the life course approach to the structural and the rational approach for modeling family adaptation strategies. Strategies, here, are broadly defined as the actions families devise for coping with, if not overcoming, the challenges of living, and for achieving their goals in the face of structural barriers. The structural approach emphasizes the ways that larger social structural forces constrain, and to some extent determine, the repertoire of adaptations available to individual families in a given society. The rational approach argues that family acts so as to maximize its household utility, that is, the well being of the family unit. They argued that the life-course models pace family and individual in a broader historical, social and cultural context of shifting opportunities and constraints, resources and demands, norms and expectations. Another advantage of the life course formulation, they argued, is the emphasis on the temporal the temporal nature of family strategies: their timing, duration and sequence within the family cycle and the life course of the individual. Finally, they mentioned that the life-course approach to family adaptive strategies emphasizes the importance of context.

The concept of family strategies and the broader paradigm of the life course are fairly important in modeling social changes. They can useful for modeling fertility, for example. Easterlin (1975) assumed that cohorts have higher or lower fertility rates (strategic decision) depending on their economic standing and prospects relative to that of earlier cohorts. Marriage 'choices' are similarly inferred from behavior of individuals in particular historical, social and economic contexts. Scholars draw on the family strategy construct as a sensitizing device to describe both macrolevel and micro level processes: large scale trends fertility, labour force, migration are seen as indicators of strategies, as are family level shifts in membership and role allocations within individual households (Moen and Wethington, 1992). Family strategies, in both their macro level manifestations (as behavioral patterns and demographic trends) and their micro level manifestations (as family and household decision making processes) become ways of talking about social continuities and social change.

Much of demographic research today adopts a life course perspective, frequently in combination with a historical perspective. The significance of the emergence of the life course perspectives is considered as important as the introduction of the cohort perspective in the study of demographic change. These two approaches are believed to complement each other. The cohort remains an important concept in the study of demographic change, but the significance of the life course concept has increased because of flexibility and individual pathways of life that produce greater intra-cohort variation. Life course perspective helps us to understand human behavior as it focuses on how individuals themselves perceive their lives; individuals experience events and organize their lives around these events Some events come by choice, while others come by chance (Willekens, 1999).

Some of the methods for modeling life course events are event-history analysis (Blossfeld and Rohwer, 1995; Ravanera et al., 1997), state space approaches (Fernando, 1999) and longitudinal analysis (Willekens, 1999). The state space approach mainly entails modeling the transition between states in discrete times. In the Markov model variant the occurrence of an event of interest depends directly on its preceding state, and only on it. In the semi-Markov variant changes in states depend on the state of origin. The occurrence of an event of interest depends on both the preceding and succeeding events, and on the length of duration between the two events. Both approaches can be computationally intensive without any closed analytical form. In this paper we shall compare the first and the third approaches.

3 Why modelling and simulation?

Computer-assisted modelling followed by simulation has become a useful method of inquiry in the sciences over the last quarter century. In some disciplinary areas, such as management sciences, economics and many physical sciences, the use of computer assisted modelling is a routine and highly accepted research method (Hanneman and Patrick, 1997). In demography the relative simplicity of the underlying process coupled with the relatively observed chaotic population dynamics make simulation an attractive proposition (Halpin, 1999). Simulation has helped to solve and understand a number of static and dynamic optimisation problems (Bonabeau and Theraulaz, 2000).

An important purpose of modelling, and simulation of the models produced, is to clarify concepts, rather than to replicate an observed transition. Researchers, concerned with both abstract theorizing and concrete prediction, have adopted variations of the methodology. The social sciences are over-rich in descriptive theories that have limited practical application (Lane, 1999). The process of modelling itself can produce valuable personal insight on the subject being modelled. The act of translating a theory into a simulation requires making everything explicit and quickly exposes internal inconsistencies and gaps. Mathematical models bind theory to precise formulations and by doing so accentuate logical inconsistencies (Rosero-Bixby and Casterline, 1993).

Simulations are also useful in the elaboration of macrosociological theories (Hanneman et al. 1995). Simulation provides a support for the debate on the different theories, and could potentially be useful for the resolution of alternative hypotheses, such as the discovery of the non-linear effect of education on fertility behaviours (Smith, 1989).

Computer simulation can be used to deal with problems that are analytically intractable within other modelling traditions. Simulation can offer a solution to circumvent the limitations imposed by intractable models. A simulation can be a "hypothesis-generating" exercise or as "thought experiments". Like formal models, a simulation model can isolate components and investigate them while keeping other details fixed. Empirical research has great difficulty separating the effects of institutions from political structure, for example, but simulation can circumvent that by exposing a variety of institutions to identical behavioural rules (Johnson 1999).

Computer simulation represents a third domain, complementing both natural language and mathematical/statistical analysis. Natural language is inadequate for the formulation of complex theories, especially when there is a system of multiple causations involved or where dynamics are important. Where possible, a mathematical formulation is preferable, but often the mathematics is intractable or outside the skill of the researcher, and then simulation provides a powerful alternative (Hanneman et al. 1995). One has to understand, however, that simulation does not replace theory or empirical inquiry but act as a bridge between the two (Hanneman and Patrick, 1997). Very often there are many levels of uncertainties and sensitivities that the user should be wary about (Kerr, 1997).

The modelling of causal relationships is an also important in simulation although it is a difficult scientific endeavour (Holland, 1986). There is a need to move beyond the listing of factors that significantly affect behavior and to pay attention to how these factors affect demographic behavior, i.e. the causal mechanism. It transcends disciplinary boundaries and much has been written about conceptualising and modelling causal relationships. Causal relationships are important in simulation as it allows the building of more robust models and also it allows a greater versatility of the models.

3.1 Approaches of modelling

One approach of modelling is systems dynamics, whose history is described in Lane (1999). The essence of system dynamics is that it models systems, in which systems consist of asset of quantities or stocks or levels that can grow or decline and in which the rates of growth and decline are affected by feedback loops from each stock itself and other stocks, in potentially complied ways (Halpin, 1999). This is very close to the microsimulation philosophy. Game theory is another major simulation tradition in social science, especially in economics, but also in sociology in association with the growth of rational choice theory and in recent years has developed into the complex and often more interesting direction of iterated games, in which the players may develop long-term strategies based on knowledge or assumption about the other participants (Davis, 1997).

Another approach uses the theory of complex systems (Goldspink, 2000; Holland, 1995), which includes cellular automata (CA) and agent-based simulation (O'Sullivan and Haklay, 2000). CA has been used in studying group segregation. (Hegselmann, 1996) while agent based modelling has used in many applications where dynamic evolutions of behaviours are followed in time (Troitzsch et al., 1996; Kirman and Zimmermann, 2001). Another approach that has the potential for providing a handle on the more complex emergence of social structures is the multilevel modelling approach (Möhring, 1996). One of the important features of the approach is that at all levels- individual, intermediate, and macro can bear on the development of the model, and thus, it offers the potential to improve both on the system dynamics style macro orientation and the bottom up micro orientation.

Modelling and simulation can be done at the micro level or at the macro level. The former, also called individual-based modelling, allows the more elaborate study of dynamics of behaviours. In the study of ecological populations, for example, there are two reasons for using individual based studies. On the one hand, it is important in the study in question to include the individual differences or heterogeneity. On the other, it suspected that what has been learned under state variable models would have to be revised if the discreteness, uniqueness, life cycles and variability of individuals were taken into accounts. Grimm (1999) refers to these two motivations as pragmatic and paradigmatic, respectively. A number of examples of studies falling under each motivation is given in Grimm (1999). Macro level modelling is more useful for forecasting purposes and this includes a large number of Economics models.

Microsimulation was introduced over four decades ago by Orcutt (1957) and has experienced somewhat of a revival in the social sciences over the past decade (Merz, 1991; Clarke, 1996; Isard et al., 1998). It has been used to study various social phenomena such as population growth and development, the effect of ageing and pension formulas on social insurance funding, and the effect of various tax regimes on fiscal budgeting. Microsimulation basically models changes at the most basic unit: the individual. It tries to address the variability inherent among individuals within a given society or country, while allowing for aggregation of the results at different scales.

Simulation has many applications in business processes. It can be used to reduce the risks and costs associated with business change projects, to drive strategic organizational analysis, to derive requirements and specifications for information systems (IS), or to support semi automated execution of processes (Paul et al. 1999).

Johnson (1999) gave a summary of the use of simulation in political sciences studies on the paradox of voting, congressional politics, the study of voting in presidential nomination, war games and iterative prisoner dilemma. A review of simulation requirements in studying technological innovation is given in Windrum (1999).

Drombroski (1993) used simulation to demonstrate that the practise of bride-wealth, usually understood purely in terms of solidaristic interfamilial exchange, may well have the important and unexpected consequence of reducing the risk of cattle herds dying out, thereby making economic the practise of small-scale husbandry.

4 Statistical issues

The ability to model behaviours depend very much on the data and the mathematical tools developed to fit them. Over the years, several statistical methods for modelling behaviours have been suggested, many of which are quite complex, which attempt to capture as much information about the process being modelled and the data. This section describes such new advances in modelling behaviours. It describes the different approach to data collection (cross-sectional and longitudinal) and describes the different approaches that can be used to model longitudinal data.

The advantages of longitudinal data analysis over cross-sectional studies are well known. Diggle et al. (1994) argued that even though it is possible to address the same scientific questions with a longitudinal or cross sectional study, the former is better suited for separating cohort and age effects. That is, longitudinal studies can distinguish changes over time within individuals (ageing effects) from differences among people in their baseline levels (cohort effects). Furthermore, in some studies, the period or calendar date can also be important.

Blossfeld and Rohwer (1995) listed twelve drawbacks of cross-sectional studies. They mentioned the problems of ambiguity when drawing inferences from such studies since an implicit or explicit assumption of statistical equilibrium is made. They argued also that there are only few situations in which directions of causality can be established on cross-sectional data. Connected with the latter is the drawback that these data cannot be used to discover the different strengths of reciprocal effects. Furthermore, social research is usually based on non-experimental designs, which are highly selective. This problem cannot be fully solved by longitudinal analysis, but it is even worse for cross sectional data. They also mention the importance of previous history, age and cohort effects, and historical settings as well as multiple events in modeling social processes and these are not easily addressed in cross-sectional analysis. Finally, they argued that cross-sectional analysis do not provide the opportunity to study contextual processes at different levels, duration dependence, variability in state dependencies and changes in outcomes.

The analysis of longitudinal data, however, requires the data to be collected based on a given, sometimes quite expensive or time-consuming designs. Longitudinal data can be collected either prospectively, following subjects in time, or retrospectively, by extracting multiple measurements on each person from historical records. The defining feature of a longitudinal study is repeated observations on individuals enabling direct study of change. One can also collect data retrospectively via life history studies that cover the whole life course of the interviewed individuals. This particular form of data is called history data. The strengths and drawbacks of these data collection design are well known and described in Blossfeld and Rohwer (1995).

In this document we compare two approaches for modeling fertility: survival analysis (Blossfeld and Rohwer, 1995; Klein and Moeschberger, 1997; Hosmer. and Lemeshow, 1999) and longitudinal analysis (Diggle et al., 1997). Both approaches fall under the umbrella of "Analysis of longitudinal data". While the former 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. Furthermore, the distributional assumptions under the different models are be different.

4.1 Survival analysis

Survival analysis is concerned with the analysis of time-to-event data. This type of data arises in diverse fields such as medicine, biology, economics (the analysis of which is called duration modelling) and demography (the analysis of which is called event history analysis). The analysis of survival experiments is complicated by issues of censoring, where an individual 's life length is known to occur only in a certain period of time, and truncation, where individuals enter the study only if they survive a sufficient length of time or individuals enter the study only if the event has occurred by a given date. Since there exists many good reference books for survival analysis, only the basic concepts are covered here. For a deeper understanding of survival analysis, the interested reader should consult the references listed in this report.

Let X be the time until some specified event. The event, in our case for example, would be getting pregnant. X is assumed to be a non-negative random variable from an homogeneous population. Four functions characterise the distribution of X, namely:

1 The survival function, which is the probability of an individual surviving beyond time x;

$$S(x) = \Pr(X > x) = \int_x^{\infty} f(t)dt \tag{1}$$

2 The hazard function which is the chance an individual of age x experiences the event in the next instant;

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x | X \ge x)}{\Delta x} \tag{2}$$

3 The probability density function which is the unconditional probability of the event occurring at time x;

$$\text{Pdf} = f(x) \tag{3}$$

4 The mean residual life at time x, which is the mean time to the event of interest, given the event has not occurred at x.

$$mrl(x) = \frac{\int_x^{\infty} (t-x) f(t)dt}{S(x)} \tag{4}$$

If one knows any one of the above functions, then one can uniquely determine the remaining three.

So far an homogeneous population has been assumed, but often when analysing survival data, one has to account for concomitant information (sometimes referred to as covariates). Populations that exhibit such heterogeneity are prevalent whenever the study involves a cohort study or an observational study. Consider an event time $X > 0$, and a vector $\boldsymbol{Z}=(Z_1, \ldots, Z_p)$ of explanatory variables associated with X. $\boldsymbol{Z}$ may include quantitative variables, qualitative variables, and/or time dependent variables.

There are two approaches to the modelling of covariate effects on survival. The first approach is analogous to the classical linear regression approach. In the second approach, which is more common, is to model the conditional hazard rate as a function of the covariates. Multiplicative or additive hazard rates models can be used.

Regression approach:

$Y=\ln(X)$ is modelled.

$$Y = \mu + \gamma' Z + \sigma W \qquad (5)$$

γ = the regression coefficients

W= Error distribution

Hazard rate models:

Multiplicative

$$h(x|z) = h_o(x)c(\beta' z) \qquad (6)$$

where *c* is a function and is usually taken as the exponential function.

Additive

$$h(x|z) = h_0(x) + \sum_{j=1}^{p} z_j(x)\beta_j(x) \qquad (7)$$

$h_0(x)$ is an arbitrary non-negative function.

Many of the design issues in modelling survival data, such as checking the validity of the model by examining the residuals, choosing the appropriate subset of explanatory variables and interpreting the model produced are described in Hosmer and Lemeshow (1999).

4.2 Longitudinal analysis

In longitudinal analysis the time dependence between the measurements is explicitly taken into consideration in the modelling (Fitzmaurice et al., 1993). As for the survival analysis, the covariates can be both time-stationary, that is, constant across occasions, and time varying. Since the references on binary longitudinal data analysis are patchy and highly technical, this statistical approach is described in more detail than the first one.

Since the response is binary (pregnant or not pregnant), the logit link is used. This raises several issues. For example, there is a lack of a discrete multivariate analogue to the multivariate Gaussian, which can be parameterised only in terms of mean and covariance parameters that vary independently in separable parameter spaces. In addition, with binary responses, the usual choices of non-linear models for the means make the parameters of random effects models difficult to interpret.

In marginal modelling, no attempt is made to examine the interrelationship within the same individual between outcomes at different points in time. A marginal model does not specify a full probability model for the outcome variables, considered jointly. To obtain appropriate estimates of the parameters, we need either to specify a full probability model for the available data, from which an appropriate likelihood function could be derived, or alternatively to decide on a reasonable estimation method (in the absence of a full model). If we attempt to specify a full probability model, we are faced with the problem of dependence or correlation between the values of y_{ij} on the same individual.

Using the maximum-likelihood method that assumes no correlation between the individuals, consistent estimates of the parameters are obtained, as long as the first order model (the model relating the explanatory variables to the response variable) is correct. In other words, with large samples, the point estimates should be close to the true population values. The estimated standard errors, however, will be too small, assuming the correlation between subjects is positive. A first approach to improving on the use of maximum likelihood is to stick to the point estimates but to 'fix up' the standard errors. One common approach is to use the information sandwich method, which is based on an asymptotic (large sample) approximation and which still provides good results in many situations (Carlin et al., 1999).

While the naïve maximum-likelihood estimates are consistent given only the specification of the first-order model, they are not as efficient as estimates from a method that more fully utilizes information on the data's structure, including dependencies over time. A method based on maximising the likelihood under a full probability model would be (asymptotically) the most efficient, but a useful method that does not require the specification of a full model yet still offers greater efficiency is the quasi-likelihood approach. In the context of repeated measurements, a generalization of quasi-likelihood that has become widely used is the generalised estimating equations (GEE).

Consider the simple logistic model where one is modelling a binary outcome variable, Y_{ij}, where i indexes subjects (i=1,…,n) and j indexes time (j=1,….,J):

$$\log it(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \sum_{k=0}^{K} \beta_k x_k \qquad (8)$$

where π_{ij} is an estimated probability, β_k are parameters to be estimated and x_k are covariates, with x_0=1. This is sometimes called a marginal model because it does not specify a full probability model for the outcome variables, Y_{ij}, considered jointly. If there was no correlation within individuals, one reasonably assume that

$$y_{ij} \sim Binomial(1, \pi_{ij}), \text{independently for all i, j}$$

and derive the likelihood function, up to a proportionality constant, as

$$\prod_i \prod_j \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{(1-y_{ij})} \qquad (9)$$

To find maximum likelihood estimates of β, it is simpler to work with the log-likelihood

$$l = \sum_i \sum_j \left(y_{ij} \log \pi_{ij} + (1-y_{ij}) \log(1-\pi_{ij})\right) \qquad (10)$$

Maximising this likelihood function amounts to solving simultaneous the K equations

$$U(\beta_k) = \frac{\partial l}{\partial \beta_k} = \sum_i \sum_j x_{ijk}(y_{ij} - \pi_{ij}) = 0 \qquad (11)$$

These equations are known as the likelihood score equations and can be written in vector form as

$$U(\boldsymbol{\beta}) = X^T(\mathbf{y} - \boldsymbol{\pi}) = 0 \qquad (12)$$

where **y** and **π** are vector forms of the data y_{ij} and parameters π_{ij}, respectively and X is the corresponding design matrix with number of rows equal to the length of the y vector and number of columns equal to the number of β parameters. The likelihood based score equations will give consistent estimates of β regardless of the full probability model (with its potential dependencies) as long as the first-order specification of the mean or prevalence (8) is correct. However, they are not efficient.

A method based on maximising the likelihood under a full probability model would be (asymptotically) the most efficient, but a useful method that does not require the specification of a full model yet still offers greater efficiency is the quasi-likelihood approach. Quasi-likelihood estimates are obtained in general by solving the equations

$$U^q(\boldsymbol{\beta}) = D^T V^{-1}(y - E(y)) = 0 \tag{13}$$

where the N×K matrix D contains the derivatives $\partial E(y_{ij})/\partial\beta_k$, and V=cov(y) is the covariance matrix of the y_{ij}'s, which contains known functions of $E(y_{ij})$ and possibly other unknown parameters. For the logistic regression specification, the estimating function (12) become

$$X^T A V^{-1}(y - \boldsymbol{\pi}) = 0 \tag{14}$$

In the context of repeated measurements (clustered or longitudinal) on I=1,…,n individuals, a generalisation of quasi-likelihood that has become widely used is the *generalized estimating equations* (GEE) method. The key of the GEE is the representation of V in the estimating equations (14) as a block diagonal matrix consisting of n submatrices V_i having the form

$$V_i = A_i^{1/2} R(\alpha) A_i^{1/2} \tag{15}$$

where A_i is the diagonal submatrix of A corresponding to individual i and $R(\alpha)$ is termed a 'working' correlation matrix and is a function of further unknown parameters α. The role of the working correlation matrix is to provide a guess at the true marginal covariance matrix $V(y)$ and it turns out that as long as the robust information-sandwich method is used for standard errors, the GEE method works well in large samples even if $R(\alpha)$ is misspecified. Incorporating a guess at the correlation structure increases the efficiency of estimation of β. Given this β, the GEE method is to construct standardised residuals for the y_{ij}'s from which an updated estimate of α is obtained. These iterations are repeated until convergence.

If the first and second moments are correct, the formal parallel between quasi-likelihood and likelihood can be used to show that large-sample standard errors for the GEE estimates, $\hat{\boldsymbol{\beta}}^{GEE}$, are provided by the quasi-information matrix

$$\text{cov}_M(\hat{\boldsymbol{\beta}}^{GEE}) = (D^T V^{-1} D)^{-1} \tag{16}$$

The subscript M stands for model-based, since the formula will no produce consistent estimates if the assumed form of V is wrong. In practise, estimates of D and V based on substituting the estimated values of β and α must be used. The information-sandwich idea can be used to 'robustify' these standard errors, leading to

$$\text{cov}_R(\hat{\boldsymbol{\beta}}^{GEE}) = (D^T V^{-1} D)^{-1} \sum_i \left(D_i^T V_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\pi}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\pi}}_i)^T V_i^{-1} D_i\right)\left(D^T V^{-1} D\right)^{-1} \tag{17}$$

which is a generalized version of the previous formula (16), which robustified the standard error of the independence ML estimate. In most cases, a given form for the correlation structure between the individuals, but as shown in Darlington and Farewell (1992) much can be gained by studying the dependence of the correlation structure on a set of time-independent explanatory variables.

5 Data

The data used for this analysis was obtained from a micro census, which each held four times a year. In each wave, a sample of about 60.000 persons is drawn. This makes 23.000 households, which is around 0.75% of the population.). Each provinces are sampled in equal amounts. In every wave 1/8th of the sample is replaced, so that every 2 years one have a completely new sample. The original purpose of this waves of micro census is to have some proxies to interpolate the general census held every ten years asked within the standard survey (respondents are LIABLE to answer) plus a battery of extra. In this particular survey, which was held on February 1996, is special as it uses an event history

6 Analysis

In the context of life course event, the analysis is based on the occurrence of a birth following a major event. In this preliminary analysis two major events were recognised: marriage and reaching the age of fifteen. Thus four sets of equations are produced for each analysis. The first equation determines the duration of the first birth after a marriage. One could, of course, also consider the event marriage after giving birth. Since this exercised is focused on comparing two statistical approaches at modelling life course events resulting in fertility, the latter is not considered in this report. Furthermore, the timing of events is a design issue that will be addressed in the future. The second equation models second birth after marriage. The third equation, considers, for those who are not married, how long it takes to give birth after reaching the age of fifteen. Finally, the fourth equation models second birth, for those who stay unmarried, after the age of 15.

In both analysis a number time independent variables are used. The same variables were used for all sets of equations, except for duration of first birth, which is only used for modelling the second births. These time invariant covariates could be considered as state variables describing the woman at the time of the major event. A list of variables, with the acronyms used and a brief description of the variables are given in table 1. Many of the variables are used as proxy for financial or material state of the woman.

Table 1: List of variables used in the analysis

Variable name	Description	Type
Agemar	Age at marriage	Continuous
Bdl	Province	Categorical 1 Burgenland 2 Kärnten 3 Niederösterreich 4 Oberösterreich 5 Salzburg 6 Steiermark 7 Tirol

		8 Vorarlberg 9 Wien
htype	No. of dwelling within the buildings	Categorical 1 no apartment 2 one apartment 3 two apartments 4 3-5 apartments 5 6-10 apartments 6 11-20 apartments 7 21+ appartments
Citizen	Citizenship	Categorical 1 Austria 2 Former Yugoslavia 3 Turkey 4 Other
School	Attending school	Categorical 0 yes 1 no
Labour	Labor market participation	Categorical 1 working 2 military/civil service 3 waiting period 4 just before a new job 5 searching for job 6 anything else
Prof	Professional status (main classes)	Categorical 1 self employed 2 assisting relative 3 employee 4 clerk 5 educated blue collar 6 blue collar in education 7 Employee in education 8 worker in education

		9 Not specified
Educ	Education level	Categorical 0 min. level not reached 1 min. level reached 2 technical apprenticeship 3 middle grade prof. School 4 high grade common school 5 high grade prof. school 6 high grade prof. school (1 year college) 7 Academic 8 University
Origin	Where lived at age 15	Categorical 1 Austria 2 Europe, USA, Canada, Oceania 3 Former Russian Block 4 Africa 5 South America 6 Asia
Municip	Municipality type	Categorical 0 rural 1 urban
Educf	Education of father	Categorical Same levels as educ
Educm	Education of mother	Categorical Same levels as educ
Dur1	Duration of first birth from major event	Continuous (months)

For the survival analysis, a parametric exponential regression is used. There were two reasons for this. First of all, based on analysis or residuals this model fitted best. Secondly, this particular model is used quite a lot in modeling social behaviors. Only the variables significant at the 90 % level are listed. All the analysis was done in Splus. For the longitudinal analysis, the library Oswald, was used.

The results for the durations in months for the first birth, following marriage, are listed in table 2. This model was based on 5488 observation points and missing values were excluded. The model fit resulted in a reduction in the log likelihood from –22660,5 to –22187,2. The Chi square statistics of the model was 946,51on 59 degrees of freedom, which is highly significantly at the 95 % level (p-value =0). At this stage, no interaction terms were included, because the model was too big for convergence to be achieved. Furthermore, one has to weigh very carefully the social significance of any interaction terms, before inserting them into the model.

Table 2: Parameters of the model of duration of first birth since marriage

Variable	Coefficient	Std. Err.	Z	P-value
Intercept	2,402	0,288	8,3	0
Agemar	0,062	0,005	13,4	0
Bdl (1)	0,097	0,035	2,8	0
Bdl (3)	0,026	0,011	2,3	0,02
Bdl (7)	0,021	0,006	3,5	0
Bdl (8)	0,024	0,007	3,6	0
Htype(2)	-0,049	0,020	-2,4	0,01
Htype(3)	0,111	0,022	5,0	0
Htype(4)	0,060	0,014	4,4	0
Htype(5)	0,023	0,012	1,9	0,05
Citizen(1)	-0,116	0,020	-1,9	0,06
School	-0,426	0,149	-2,9	0
Labour(5)	0,087	0,027	3.2	0
Prof(1)	-0,119	0,041	-2,8	0
Prof(2)	0,169	0,017	9,9	0
Prof(3)	0,068	0,017	4,1	0
Prof(4)	0,035	0,019	1,9	0,06
Prof(6)	-0,291	0,149	-2,0	0,05
Educ(4)	-0,030	0,014	-2,2	0,03
Educ(6)	-0,037	0,012	-3,0	0
Educ(7)	-0,019	0,011	-1,8	0,08
Origin(1)	0,158	0,056	2,8	0
Origin(2)	-0,075	0,035	-2,1	0,03
Municip	0,082	0,035	2,3	0,02
Educf(2)	0,087	0,024	3,6	0
Educm(3)	0,115	0,027	4,2	0

The results for the duration from marriage to second birth are given in table 3. This model was based on 6109 observations and resulted in a reduction in the log likelihood function from –26032,7 to –25285,4. The Chi square statistics of the model was 1494,6 on 61 degrees of freedom, which is highly significantly at the 95 % level (p-value =0). Note that the number of observations in this case has increased since the number of couple who got a child before marriage, married before the second birth.

Table 3: Duration from marriage to second birth for married women

Variable	Coefficient	Std. Err.	Z	P-value
Intercept	3,628	1,664	2,2	0,03
Dur1	0,011	0,000	25,1	0
Agemar	0,059	0,005	12,1	0
Bdl (3)	-0,070	0,011	-6,5	0
Bdl(4)	-0,030	0,012	-2,6	0,01
Bdl(6)	-0,021	0,007	-3,1	0
Bdl (7)	0,022	0,006	3,6	0
Htype(2)	0,048	0,020	2,4	0,01
Htype(4)	0,117	0,015	7,8	0
Htype(5)	0,046	0,013	3,6	0
Htype(6)	0,056	0,013	4,2	0
Citizen(2)	-0,133	0,053	-2,5	0,01
School	-0,369	0,178	-2,1	0,04
Labour(3)	0,128	0,061	2,1	0,03
Labour(4)	0,060	0,028	2,2	0,03
Prof(1)	-0,086	0,039	-2,2	0,03
Prof(2)	0,239	0,017	14,0	0
Prof(3)	0,136	0,018	7,4	0
Prof(4)	0,038	0,020	1,9	0,06
Educ(5)	-0,031	0,017	-1,9	0,06
Educ(6)	-0,048	0,012	-3,6	0
Origin(2)	0,077	0,038	2,0	0,04
Municip	0,216	0,038	5,7	0
Educm(5)	-0,072	0,028	-2,1	0,04

Table 4 shows the results of the model of duration from fifteen's birthday to first birth, for those who are not married. The model was parameterized on 4254 observations and the log likelihood value was reduced from –14444,1 to –13934,1. The Chi square statistics of the model was 1020,1 on 11 degrees of freedom, which is highly significantly at the 95 % level (p-value =0).

Table 4: Duration from fifteen's birthday to first birth, for unmarried women

Variable	Coefficient	Std. Err.	Z	P-value
Intercept	7,462	0,318	23,5	0
Bdl (1)	-0,194	0,056	-3,4	0
Bdl(4)	-0,045	0,015	-3,1	0

Bdl(5)	-0,027	0,011	-2,5	0,01
Bdl (7)	0,026	0,010	2,5	0,01
Bdl(8)	0,028	0,012	2,2	0,02
Htype(2)	-0,110	0,028	-3,8	0
Htype(5)	0,0329	0,017	1,9	0,05
Htype(6)	0,050	0,017	2,9	0
Citizen(1)	0,245	0,117	2,1	0,04
School	-1,520	0,13	-11,8	0
Labour(1)	0,290	0,072	3,9	0
Labour(2)	0,309	0,139	2,2	0,03
Labour(3)	0,152	0,066	2,3	0,02
Prof(2)	0,241	0,027	9,0	0
Prof(3)	0,122	0,028	4,4	0
Prof(4)	0,084	0,023	3,7	0
Prof(5)	-0,027	0,011	-2,5	0,01
Educ(0)	-1,215	0,251	-4,8	0
Educ(1)	-0,377	0,085	-4,4	0
Educ(2)	-0,143	0,045	-3,2	0
Educ(6)	-0,056	0,023	-2,4	0,02
Origin(2)	-0,131	0,077	-1,7	0,09
Educf(1)	0,133	0,045	2,941	0
Educf(2)	0,094	0,040	2,346	0,02
Educm(2)	0,117	0,041	2,9	0

Table 5shows the results of the model of duration from fifteen's birthday to second birth, for those who are not married. The model was parameterized on 902 observations and the log likelihood value was reduced from –3069,1 to –2908,4. The Chi-square statistics of the model was 321,4 on 58 degrees of freedom, which is highly significantly at the 95 % level (p-value =0).

Table 5: Duration from fifteen's birthday to second birth, for unmarried women

Variable	Coefficient	Std. Err.	Z	P-value
Intercept	4,700	0,643	7,3	0
Dur1	0,013	0,001	10,7	0
Htype(6)	0,076	0,038	2,0	0,05
Prof(2)	0,252	0,065	3,9	0
Prof(4)	0,141	0,057	2,5	0,01
Prof(6)	-0,439	0,255	-1,7	0,09

It was not possible to fit the longitudinal model to all the data sets due to computational problems. By reducing the data set significantly, one could fit such models but this required samples sizes of around a thousand, which was not enough to represent all provinces. One way to go around this is to fit separates models for each provinces but this will raise many statistical problems. First of all, given that the sample sizes vary quite much between the provinces, we will get estimates whose accuracy varied significantly between the provinces. Secondly, for parsimonious reasons, should try to reduce the number of equations to a minimum and this would entail finding robust statistical techniques for comparing the coefficients from the different equations. This is not an easy task. Furthermore, even though the sample sizes were reduced, in some data sets, no convergence was reached and a floating error message was produced.

Since only the models for second birth for unmarried women were able to converge, they are the only results that are given here. This dataset contained only 902 observations. The parameters of this model are given in table 6

Table 6 Parameters of the model of probability of second birth since fifteen's birthday for single women.

Variable	Coefficient	Std. Err.	Z	P-value
Intercept	-2,504	0,6056	4,1	0
Dur1	-0,007	0,001	-6,1	0
Htype(2)	-0,158	0,073	-2,2	0,03
Htype(5)	-0,084	0,046	-1,8	0,07
Htype(6)	-0,094	0,050	-1,9	0,06
School	0,999	0,415	2,4	0,01
Prof(2)	-0,360	0,072	-4,9	0
Prof(3)	-0,170	0,074	-2,3	0,01
Prof(4)	-0,194	0,067	-2,9	0
Origin(1)	0,6716	0,218	3,1	0
Educf(6)	0,192	0,113	1,7	0,09
Educm(3)	0,255	0,092	2,8	0

7 Discussion

One could argue that the fertility decision consists of two nested decisions: the decision to have a baby and the decision of when to have the baby, following specific events. Most probably, the factors that influence the two decisions are different or affect the decision differently. In this paper, we have assumed that the first decision has already been taken and we model the second decision. This is fairly reasonable given the two children norm in Austria (Hoem et al, 2001). Furthermore, dividing the data by using the number of children reduces the heterogeneity in the data and improves the fit of the statistical models and hence more robust models of behaviours are obtained. . This was also reflected in our data set since only 13% did not have any children and only 25% had only one child. The emphasis is thus on what determines the time couples wait to have their first and second child after getting married.

For the first child, the explanatory variables that had the highest significance are age at marriage, professional status and housing type. The variable citizenship was not significant. This was quite unexpected. This probably could be explained by the fact that once the decision to give birth is taken, the duration was not affected by cultural background. It is not easy to interpret the coefficients of the variables on their own, especially when the variables are not dichotomous. One approach is to look at the hazard function, which is the chance an individual of age x experiences the event in the next instant. The hazard function, for the exponential model, is given by

$$H(t,x,\boldsymbol{\beta}) = e^{-\sum \beta_i x_i} \tag{18}$$

For the duration of first birth, most of the variables were positive. One has to interpret the variables very carefully. For continuous variables such as age, a positive coefficient means that an increase in the variable would decrease the hazard function. Hence people, who get married late, usually postpone having a child. For categorical variable, one has to compare the variable to the reference variable. In all categorical variables the reference is taken as the last level of the variable as described in Table 1. Thus if one look at the provinces, one can observe that compared to Vienna, if all other variables remain constant, the probability of getting a child in the next instance increases. When a level is missing, that means that the coefficient for that level is not significantly different from that of the reference variable. For example, Kärnten have the same coefficient as Vienna. For dichotomous variables, such as school and municipality, a positive coefficient means that the non-zero level decreases the hazard function. As an example, living in an urban area will reduce the probability of having a child in the next instant. For school, with a negative coefficient, not being at school increases the hazard function and hence increases the probability of getting a birth in the next instant. People who live in buildings with only one apartment, most probably detached house, will have will a higher propensity to have a first child now than in the future. Austrian citizenship also decreases the time you wait for the first child. Searching for a job decreases the propensity to have a child now. A decrease in duration is observed for self-employed and blue colour workers, whereas assisting relative increases that duration. Compared to University degree holders, education levels 4,6, and 7 increases the probability of having a child in the next moment. Origin affects the duration also. For people with Austrian origin, the duration is increased whereas for western countries, the duration is smaller. Middle education of the parents also increases the duration.

For the second birth for a married woman, the most significant variables are time waited to have first child, age at marriage, and professional status. The longer the time waited the shorter the duration for the second child. This has noted in other studies. For age at marriage and professional level, the same conclusions as above apply. As for the provinces the sign of the coefficient for province Niederösterreich has changed, while that of Tirol has maintained its sign. Furthermore, the other two provinces that are significantly different from Vienna have changed. The same kind of changes can be observed in house types. Now all house types increase the duration, compared to the reference. In this model, only the education of the mother was significant.

For the first birth outside marriage, the most significant variables are being in school, professional status, and highest education reached. Note also that the significance of the education of the parents has increased. This could be explained by the fact that these two variables can be used as proxies for social classes the woman originates from. In other studies it has been shown that the behaviours of the children are influenced by those of the parents and thus giving birth outside marriage, which is subject to social norms, are affected by education of the parents. Professional status, another proxy for social status, was also significant. Thus being a worker or employee in education has a positive effect on the probability of getting pregnant in the next instant. This behaviour of having a child outside marriage is also geographical as now more provinces are different from Vienna.

As for the second birth outside marriage, this had, by far, the smallest sample size. The only three significant variables are duration of first birth from fifteenth birthday, housing type and professional status. Unfortunately, this is the only dataset that could be compared to the longitudinal data analysis. In comparing Table 5 and 6, one can help notice the large difference in the variables, which are significant in the two models. This is probably explained by the fact that since the longitudinal data analysis uses more explicitly the correlation between the individuals, better information is being used and a more elaborate model is obtained with the longitudinal data analysis that could otherwise be obtained from a small dataset using survival analysis techniques.

8 Summary and Conclusions

This report has examined some issues regarding the modelling of fertility in a life course context that are also intended to improve the fertility module of the ongoing FAMSIM+ microsimulation project. . Generally, we have argued that such models are very useful for investigating fertility behaviours and for comparing fertility models. We also mentioned that the support of fertility theories for building such models is rather sparse, although a number of variables were identified in this study that corresponded to earlier studies. Microsimulation could be an important tool to compare and evaluate the large number of fertility theories that already exist. But to able to do that, a number of issues need to be addressed.

Although the longitudinal data analysis were promising for the modelling of conditional and marginal behaviours of individuals, and they have fared very well in other studies, they were quite computationally intensive and required large memory for fitting datasets as large as those in this study Based on the only dataset that would converge in the longitudinal analysis, major differences have been noted between the models fitted by longitudinal analysis and survival analysis. We will have to investigate this more in the future.

For large datasets, the survival analysis approaches fared much better and were easier to fit. It is a bit more complicated to both validate and interpret such models but new developments such frailty models are promising for modelling individual differences. The models were fitted quite well and given their handling of time of events they are better suited for examining causal behaviours. Furthermore, they are also quite useful to model life course events and thus are very useful for agent based modelling to include causality and goal orientation into microsimulation models.

The inclusion of time dependent variables that characterise the socio-economic environment will be investigated. Frailty models that account for unobserved differences between individuals will also be investigated. Some investigations how best to fit the data by longitudinal methods will be investigated. The crux of the next report, however, will be the investigation of how to create dynamic groupings that could potentially increase both the fit and the simulation of the individual fertility behaviours. One of the aims of microsimulation is to model the heterogeneities inherent in groups of individuals. So far, most attempts at reducing such heterogeneity have consisted of arbitrarily dividing the sample into a number of static groups based on physical properties such as age, sex, educational levels and marital status. A behavioural model is developed for each group and this model is used only on individuals that are in that group at a given time. While this approach is successful in reducing some of the heterogeneities, it suffers from two major drawbacks. First of all it assumes that the groupings do not evolve over time and that attitudes and behaviours are static in time, restricted by a number of physical characteristics. Secondly, it assumes that behaviours can only be described by physical parameters. In recent years, the importance of goal orientation and life style as components of behaviours has been stressed in fields such as Sociology and Psychology, with applications in fields such as the segmentation of markets of customers. There already exist a number of statistical approaches for including such non-observable characteristics of individuals. The most well-known ones are factor analysis and mixture modelling. The inclusion of goals orientation and life style in microsimulation is rather new in microsimulation and can potentially allow better simulation of processes such as fertility behaviours and also allow further, more complex and elaborate analyses. One example of such analysis would be the way to understand how behaviours in general are affected by our choice of goals and orientations and how these in turn restrict our ability to adapt to and our inclination to adopt changes in our environment, whether economic or political. So one could analyse how incentives to increase fertility are affected by people from different goals in life.

References

Blossfeld, H. P. and Rohwer, G. 1995. Techniques of event history modelling: New approaches to causal analysis. Lawrence Erlbaum Associates, publishers, Mahwah, New Jersey, USA.

Bonabeau E, and Theraulaz G. 2000. Swarm smarts. Scientific American, **282**, 72-79.

Bongaarts, J. and Watkins, S.C. 1996. Social interactions and contemporary fertility transitions. Population and Development Review, **22**, 639-682.

Burch T.K. 1997. Fertility decline theories: Toward a synthetic computer model. Discussion paper**. No. 97-7.**

Burch, T.K. 1995. Icons, strawmen and precision: reflections on demographic theories of fertility decline. The Sociological Quarterly, **37**, 60-81.

Burch, T.K. 1999. Computer modeling of theory: explanation for the 21st century. Discussion paper **99-4**.

Cain. M. Khanam, S.R. and Nahar, S. 1979. Class, patriarchy, and women's work in Bangladesh. Population and Development Review, **5**, 405-438.

Caldwell, J. 1982. Theory of Fertility Decline. London: Academic Press.

Carlin, J.B., Wolfe, R., Coffey, C. and Patton, G.C. 1999. Analysis of binary outcomes in longitudinal studies using weighted estimating equations and discrete time survival methods: Prevalence and incidence of smoking in an adolescent cohort. Statistics in Medicine, **18**, 2655-2679.

Chenais, J. 1996. Fertility, family and social policy in contemporary Western Europe. Population and Development Review, **22**, 729-739.

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

Darlington, G.A. and Farewell, V.T. 1992. Binary longitudinal data analysis with correlation a function of explanatory variables. Biometrical Journal, **34**, 899-910.

Davis, M.D. 1997. Game Theory: A Nontechnical Introduction. Dover Publications. Mineola, USA.

Diggle, P.J., Liang, K.Y. and Zeger, S.L. 1994. Analysis of longitudinal data. Clarendon Press, Oxford, U.K.

Dombroski, K. 1993. Some considerations for the understanding of small herd dynamics in East-African arid zones: The long-term consequence of bride-wealth exchange networks. Human Ecology, **21**, 23-50.

Easterlin, R.A. 1975. An economic framework for fertility analysis. Studies in Family Planning, **6**, 53-63.

Elder, G.H. 1992. Models of the life course. Contemporary Sociology, **2**, 623-635.

Elder, G.H. 1994. Time, human agency and social change: Perspectives on the life course. Social Psychology Quarterly, **57**,4-15.

Fernando, R. 1999. Lifehist: Analysis of life histories, a state space approach. Paper presented at the workshop on longitudinal research in social science: A Canadian focus, Windermere Manor, London, Ontario, Canada October 25-27.

Fitzmaurice, G.M., Laird, N.M. and Ronitzky, A.G. 1993. Regression models for discrete longitudinal responses. Statistical Science, **8**, 284-309.

Giele, J.Z. and Elder, G.H. 1998. Methods of life course research: Qualitative and quantitative approaches. Sage publications. Thousand Oaks, California

Goldspink, C. 2000. Modelling social systems as complex: Towards a social simulation metamodel. Journal of Artificial Societies and Social Simulation, **3**. (http://www.soc.surrey.ac.uk/JASSS/3/2/1.html)

Grimm, V. 1999. Ten years of individual-based modeling in ecology: What have we learned and what could we learn in the future. Ecological Modelling, **115**, 129-148.

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

Halpin, B. 1999. Simulation in sociology. American Behavioral Scientist, **42**, 1488-1508.

Hanneman, R. and Patrick, S. 1997. On the uses of computer-assisted simulation modeling in the social sciences. Sociological Research online, **2**, http://www.socresonline.org-uk/socresonline/2/2/5.html.

Hanneman, R.A., Collins, R. and Mordt, G. 1995. Discovering theory dynamics by computer simulation: Experiments on state legitimacy and imperialist capitalism. Sociological Methodology, **25**, 1-46.

Hegselmann, R. 1996. Understanding social dynamics: The cellular automata approach. In Social Science Microsimulation K.G. Troitzsh, U. Mueller, G.N. Gilbert, and J.E Doran (eds.) 282-306, Berlin: Springer.

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

Hoem, J.M., Prskawtz, A. and Neyer, G. 2001. Autonomy or conservative adjustment? The effect of public policies and educational attainment on third births in Austria. MPIDR Working Paper **WP 2001-016**. Max Planck Institute for Demographic Research.

Holland, J.H. 1995. Hidden order: How adaptation builds complexity. Addison-Wesley, Reading, Mass.

Holland, P.W. 1986. Statistics and causal inference. Journal of the American Statistical Association, **81**, 945-960.

Hosmer, D.W. and Lemeshow, S. 1999. Applied Survival Analysis: Regression Modeling of Time to Event data. John Wiley and Sons, Inc. New York, USA.

Hotz, V.J., Klerman, J.A. and Willis, R.J. 1997. The economics of fertility in developed countries. In Handbook of population and family economics, Rozenweig M.R. and Stark O. (Eds). Elsevier Science.

Jensen, A. 1999. Children's families: Exploring the self-evident. Presented at the conference Research in Childhood, Sociology, Culture and History. Denmark.

Johnson, P.E. 1999. Simulation modeling in political science. American Behavioral Scientist, **42**, 1509-1530.

Kerr RA. 1997. Greenhouse forecasting still cloudy. Science, **276**, 1040-1042.

Kirman, A. and Zimmermann, J.B. 2001. Economics With Heterogeneous Interacting Agents. Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Berlin, Germany.

Klein, J.P and Mosechberger, M.L. 1997. Survival analysis: Techniques for censored and truncated data. Springer Verlag, New York, USA.

Lane, D.C. 1999. Social theory and system dynamic practise. European Journal of Operational Research, **113**, 501-527.

Lesthaeghe, R. and Surkyn, J. 1988. Cultural dynamics and economic theories of fertility change. Population and Development Review, **14**, 1-46.

Mason, K.O. 1986. The status of women: Conceptual and methodological issues in demographic studies. Sociological Forum, **1**, 284-300.

McDonald, P. 2000. Gender equity in theories of fertility transition. Population and Development Review, **26**, 427-439.

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

Moen P. and Wethington, E. 1992. The concept of family adaptive strategies. Annal Review of Sociology, **18**, 233-251.

Möhring, M. 1996. Social science multilevel simulation with MIMOSE. In Social Science Microsimulation K.G. Troitzsch, U. Mueller, G.N. Gilbert, and J.E. Doran (eds), 123-137. Berlin: Springer.

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

O'Sullivan, D. and Haklay, M. 2000. Agent-based models and individualism: is the world agent-based? Environment and Planning A, **32**, 1409-1425.

Paul, R.J., Giaglis, G.M. and Hlupic, V. 1999. Simulation of Business Processes. American Behavioral Scientist, **42**, 1551-1576.

Rajulton, F. 1999. Lifehist: Analysis of life histories, a state space approach. Paper presented at the workshop on longitudinal research in social science: A Canadian focus, Windermere manor, London, Ontario, Canada, October 25-27.

Ravanera, R.Z., Rajulton, F. and Burch, T.K. 1997. Early life transitions of Canadian women: A cohort analysis of timing, sequences and variations. Discussion paper, **No 97-3**.

Rephann, T.J. The fertility module of SVERIGE: Documentation **V 1.0**. Spatial Modeling Center, Kiruna, Sweden, http://www.smc.kiruna.se.

Rosero-Bixby,L. and Casterline, J. 1993. Modelling diffusion effects in fertility transition. Population Studies, **47**, 147-167.

Smith, H. L. 1989. Integrating theory and research on the institutional determinants of fertility. Demography, **26**, 2, 171-184.

.Troitzsch, K.G., Mueller, U., Gilbert, G.N. and Doran J.E. (eds) (1996). Social Science Microsimulation. Springer, Berlin

Van de Kaa, D.J. 1996. Anchored narratives: The story and findings of half a century of research into the determinants of fertility. Population Studies, **50**, 389-432.

Walker, J.R. 1995. The effect of public policies on recent Swedish fertility behavior. Journal of Population Economics, **8,** 223-251.

Willekens, F. 1999. Theoretical and technical orientations toward longitudinal research in the social sciences. Paper presented at the workshop on Longitudinal Research in Social Sciences: A Canadian focus, Windermere Manor, London, Ontario, Canada, October 25-27.

Williams, M. 2000. Social research- the emergence of a discipline? International Journal of Social Research Methodology, **3**, 157-166.

Windrum, P. 1999. Simulation models of technological innovation. American Behavioral Scientist, **42**, 1531-1550.

Wolf , D. 1999. The role of microsimulation in longitudinal data analysis. Presented at the workshop on Longitudinal Research in Social Sciences- A Canadian Focus held in London, Ontario, Canada 25-27.

Die bisher erschienenen WORKING PAPERS:

- Herbert Janig, Liselotte Wilk, Martina Beham, **"Puberty - A Challenge for Parents and Adolescent"**, Nr. 1/97 – in englischer Sprache
- G.Doblhammer, W. Lutz, A. Prskawetz, **"Covariates of First Marriage Dissolution derived from the Austrian FFS"**, Nr. 2/97 – in englischer Sprache
- Ulrike Zartler, **"Pubertät und ihre Bedeutung für Eltern und Kinder – ein Literaturüberblick"**, Nr. 3/97 – in deutscher Sprache
- Martina Beham, **"Förderliche und hemmende Faktoren zur Bewältigung der Entwicklungsaufgaben in Familien mit Pubertierenden"**, Nr. 4/97 – in deutscher Sprache
- Mag. Martina Beham, **"Elternbildung – Hilfe zur Selbsthilfe: Ergebnisse einer Pilotstudie"**, Nr. 5/97 – in deutscher Sprache
- Vera Nowak, Christiane Pfeiffer, **"Erwerbseinstieg von Frauen nach Geburt des jüngsten Kindes"**, Nr. 6/98 – in deutscher Sprache
- Veronika Gössweiner, Vera Nowak, **"Scheidungsfolgen – die langfristigen Auswirkungen von erlebter Scheidung auf die Lebensführung unter besonderer Berücksichtigung der ersten Lebensgemeinschaft"**, Nr. 7/98 – in deutscher Sprache
- Wolfgang Lutz, Vera Nowak, Stefan Ohmacht, Martin Oppitz, Christiane Pfeiffer, Helmuth Schattovits, Martin Schneider, Wilfried Schönbeck, Wolfgang Schulz, **"Rohbericht – Band 1 Ergebnisbericht zum Projekt"**, **"Entwicklung von Modellen eines Kinderbetreuungsschecks und Analyse der Auswirkungen"** **(Machbarkeitsstudie Kinderbetreuungsscheck)**, Nr 8/98 – in deutscher Sprache
- Vera Nowak, Rudolf Schipfer, **"Familien in Zahlen"**, Daten und Graphiken zur Familie in Österreich auf einen Blick, Nr. 9/98 – in deutscher Sprache
- Paloma Fernández de la Hoz, Johannes Pflegerl, **"MigrantInnen im Spital"**, Workshop für MitarbeiterInnen in Gesundheitseinrichtungen der Stadt Wien, Nr. 10/2001 – in deutscher Sprache
- Martin Spielauer, Norbert Neuwirth, **"Family Microsimulation"**, Nr. 11/2001 – in englischer Sprache
- Veronika Gössweiner, Christiane Pfeiffer, Rudolf Richter, **"Quality of Life and Social Quality"**, Nr. 12/2001 – in englischer Sprache
- Brigitte Cizek, Christiane Pfeiffer, **"HorseTalks",** Nr. 13/2001 – in deutscher Sprache
- Martin Spielauer, Franz Schwarz, Kurt Schmid, **„Education and the Importance of the First Educational Choice in the Context of the FAMSIM+ Family Microsimulation Model for Austria"**, Nr. 15/2002 – in englischer Sprache

Zu beziehen bei: Österreichisches Institut für Familienforschung (ÖIF)
Gonzagagasse 19/8, A-1010 Wien
Tel: +43-1-535 14 54-19
Fax: +43-1-535 14 55
E-Mail: edeltraud.puerk@oif.ac.at

ÖIF

Das Österreichische Institut für Familienforschung (ÖIF) ist ein unabhängiges, gemeinnütziges Institut zur interdisziplinären wissenschaftlichen und anwendungsbezogenen Erforschung und Darstellung der Vielfalt und Veränderungen familialer Lebenswelten aus Sicht von Kindern, Frauen und Männern.

Gedruckt mit Unterstützung des Bundesministeriums für soziale Sicherheit und Generationen sowie der Länder Burgenland, Niederösterreich, Oberösterreich, Salzburg, Tirol, Vorarlberg und Wien.