# A Microanalytical Simulation Model to Predict the Long-Term Evolution of Employment Biographies in Austria: The Demographics Module

Marcus Wurzer and Reinhold Hatzinger
Vienna University of Economics and Business

**Abstract:** The well-known problems of decreasing birth rates and population ageing represent a major challenge for the Austrian pension system. It is expected that the group of pensioners will grow steadily in the future, while the proportion of people that support them – the taxpayers – will shrink. In this regard, microsimulation provides a valuable tool to identify the impact of various policy measures. With microsimulation, it is not only possible to predict cross-sectional data (e.g., the distribution of age groups in 2050), but also to simulate lifecourses of people, providing longitudinal outcomes. The demographics module is the first in a series of modules that are part of a microsimulation prototype. This prototype is being developed in order to predict the long-term evolution of Employment Biographies in Austria.

**Keywords:** Microsimulation, Demography, Employment Biographies, Pension System, Population Ageing.

## 1 Introduction

The origins of microsimulation modelling date back to the 1950s, when Guy Orcutt, who can be seen as pioneer in this field, proposed that certain problems of macroeconomic models could be overcome by going back to the individual level and model interactions of the individuals (Orcutt, 1957). Microsimulation models usually contain a high number of variables and need a large number of individuals, making them very data-intensive. This can be seen as the main reason why microsimulation wasn't widely used during the decades after its introduction (Harding, 1996).

Since the mid-nineties, interest in microsimulation has steadily grown, which can be seen to have two main reasons:

- Computers that were powerful enough to perform microsimulations became widely available.

- The ongoing discussion about problems that may emerge when the population ages became more and more important. Questions that are posed in this subject area can be seen as a prototype application of microsimulation, which aroused the interest of politics in the method.

Traditionally, the whole microsimulation model is divided into different modules (see, e.g., SESIM (Flood, 2008), APPSIM (Keegan, 2007), DYNAMOD-2 (King, Bækgaard, and Robinson, 1999), etc.), each of them approaching a number of events that can happen to a person during its lifecourse. One of the most important modules is the demographics module, simulating death, birth, marriage, migration, etc. This module also constitutes the

first one of the microsimulation prototype that has already been finished. The following sections will cover the problem of the ageing society, a short introduction to microsimulation, features of the microsimulation prototype and the demographic module and the evaluation of some of the results, using projections of the Statistik Austria as reference (Statistik Austria, 2008).

Apart from the demographics module, at the moment three additional modules are planned:

- *Education and employment career*, treating schooling, academic studies, occupational status, unemployment, maternity leave, retirement, etc.
- *Income*, covering, e.g., income level, full-time/part-time income and unemployment compensation.
- *Pension level*.

# 2   The Ageing Society

## 2.1   Demographic Trends

As it is the case in many other OECD countries, the life expectancy in Austria has steadily increased during the past 50 years. Today, women as well as men can reckon with living approximately 12 additional years, compared to people in the 1960s. Figure 1 depicts this development.
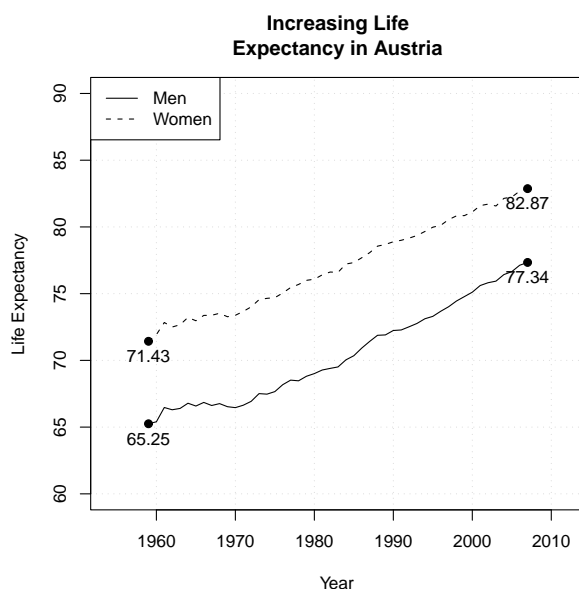


Figure 1: Increasing Life Expectancy at birth in Austria 1959–2007.

This trend is not likely to end very soon: Statistik Austria expects that in 2050, women will have a life expectancy at birth of 90 years and men of 86 years.

A second important demographic trend is the decrease in the Total Fertility Rate (TFR). The Total Fertility Rate – the average number of children a woman will give birth

to over her lifetime, assuming that the current age-specific fertility rates remain constant – sank from 2.78 at the beginning of the 1960s to approximately 1.5 in the 1980s, roughly staying at this level since then. In 2007, the TFR was 1.38, which is a decrease of over 50% during the past 50 years. To sustain the current population size, the TFR would have to be approximately 2.1 (not exactly 2.0 because of the influence of (childhood) mortality and the disproportionate sex ratios at birth). Total fertility rates since 1960 are presented in Figure 2.
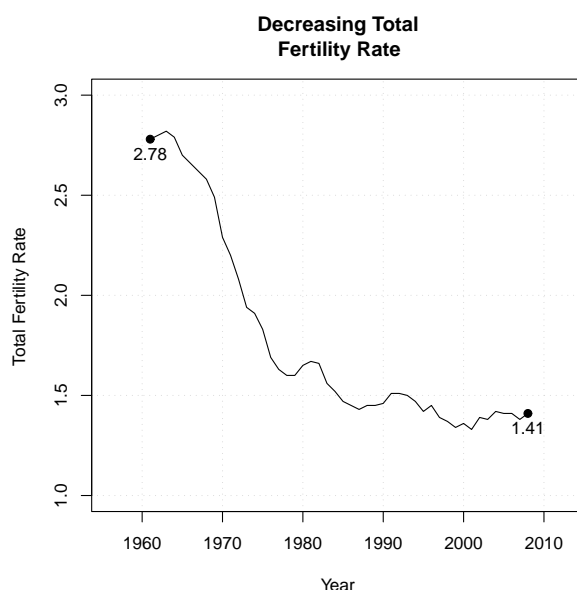


Figure 2: Decreasing Total Fertility Rate in Austria 1961–2008.

It is expected that the TFR will rise slightly in the next 20 years because of suspended births. (The mean age of women at childbirth rose from 28.0 years to 29.5 years between 1998 and 2008. Statistik Austria assumes that this is partly due to women who only suspended their births, but will catch up on them in the future.) However, the TFR will remain at a comparatively low level (1.5 in the year 2030).

Looking at the number of births, the decrease is also substantial, although not as dramatic as for the Total Fertility Rate. The tradeoff between falling fertility rates and the growing Austrian Population (mainly due to immigration) can be seen as reason for this. Since the 1960s, the number of births fell from over 130000 to 75000-80000 in the last few years—a 40% drop. This trend can be seen in Figure 3.

## 2.2   Consequences of the Demographic Trends

Due do increasing life expectancy and decreasing fertility rates, Austria faces the problem of an ageing society during the course of the next decades. Two major implications of this development are:

- The proportion of people working vs. people holding a pension will change: A smaller share of people in the labor force will have to pay for a larger share of pensioners.
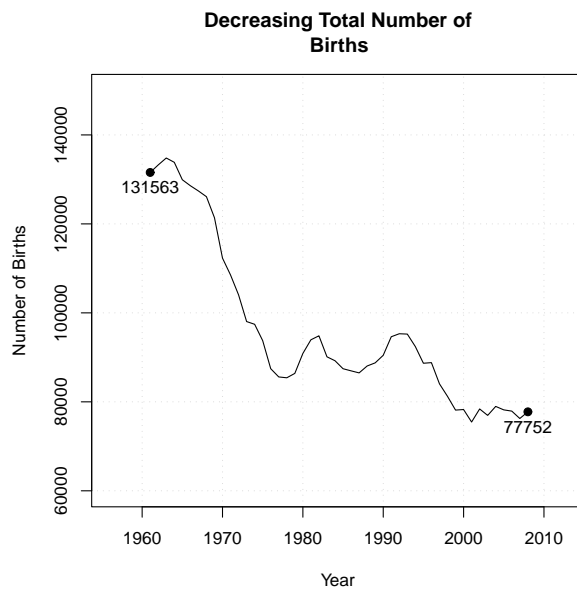
**Decreasing Total Number of Births**



Figure 3: Decrease in the Total number of births in Austria 1961–2008.

- Since the group of old people will not only increase relatively, but also in absolute numbers, it is expected that there will be an increased demand for health care and health services.

Policy makers are interested in the long-term financial viability of the retirement plan. Since the age structure of the population changes, it is questionable if the current retirement plan will still work out in the future. More likely, certain adaptations will have to made. Concerning this matter, there exists a Commission for the Pension Protection in Austria, giving advise to the Federal Ministry of Labour, Social Affairs and Consumer Protection (for further information, see `www.bmask.gv.at`). There exist several possibilities to compensate for the expected additional expenditure for future pensions, for example:

- Increase of the contribution rates
- Increase of the age of retirement
- Decrease of the federal share

The question of interest for policy makers now is: Which consequences will certain political measures have? Microsimulation models are an appropriate choice here: Due to their stochastic nature, it is easy to simulate different scenarios in order to get answers to such "What if..."–questions.

## 3 The Microsimulation Prototype

### 3.1 Using Microsimulation to Evaluate Policy Measures

Typically, not everyone is affected by certain policy measures in the same way. For example, if a new tax is introduced, the rates could be different for various sections of the

population and there may also exist persons who are not affected by the tax at all. This is because family composition, age levels, income thresholds, certain exceptional rules, etc. have to be be taken into account. To understand the full impact of a particular measure, one has to go back to the individual level - macro models are not feasible anymore.

Pension policy is a typical case for this. The amount of money a person gets when he or she retires is dependent on a number of events that happened during the life cycle of this person, e.g., phases of unemployment, disability, maternity leave, etc. Microsimulation can be utilised to model the future life profiles of people. Knowing the impact at the individual level, the impact on the aggregate (the population) can also be computed.

## 3.2    Functionality of Microsimulation Models

Microsimulation starts with drawing a simulation sample (also called base dataset) that is representative for a certain target population in a certain year. This sample consists of individuals that are characterised by a set of attributes. Starting at time $t$, the simulation sample is updated (in microsimulation terms: it is aged) which means that certain attributes may change their state between time $t$ and $t + 1$ according to certain transition probabilities - a person may die, marry, emigrate, give birth to a child, etc. This is called the *dynamic element*.

Having a model that is specified correctly, the simulation sample is also representative for time $t + 1$. Ageing continues until a certain point in time has been reached. The result of a single microsimulation run can be seen as one possible path into the future or as random sample of the events that could be observed in "reality". Microsimulation allows to gain insight into long-term processes (e.g., the number of deaths or births over time) as well as the estimation of the distribution of a certain attribute in the future population (e.g., age distribution of the Austrian population in 2055).

## 3.3    Features of the Microsimulation Prototype

The developed microsimulation model is characterised by the following features:

- *Population Model:* A sample of the whole (Austrian) population is used (in contrast to *Cohort Models* where a certain age group is simulated from birth to death).
- *Dynamic Model*: Each person has a certain number of characteristics. As the person is aged, these characteristics may change with probabilities corresponding to the population data (as opposed to *static models* that only evaluate the immediate effect of a certain measure and where variables that are not of primary interest, e.g. *family status*, are held constant).
- *Closed Model*: New persons can enter the model only due to birth or immigration. Interaction is only possible with persons included in the simulation sample (as opposed to *open models*, where, for example, spouses are generated out of nothing if a women is eligible to marry).
- *Discrete Time Model*: The person characteristics are updated in one-year steps (cross-sectional simulation). The order of the events that may change the characteristics has to be defined by the researcher in advance (as opposed to models using *continuous time* where survival functions are estimated and the events are

competing. This may seem favorable in theory, but data that would allow for the usage of continuous time are rarely available).

- *Modularity*: Division into several modules ensures the flexibility of the MSM. The model can easily be extended by the implementation of additional modules, changes in the existing code are possible without having to reconstruct the whole model. For each year, the single modules are processed in a fixed order as depicted in Figure 4.
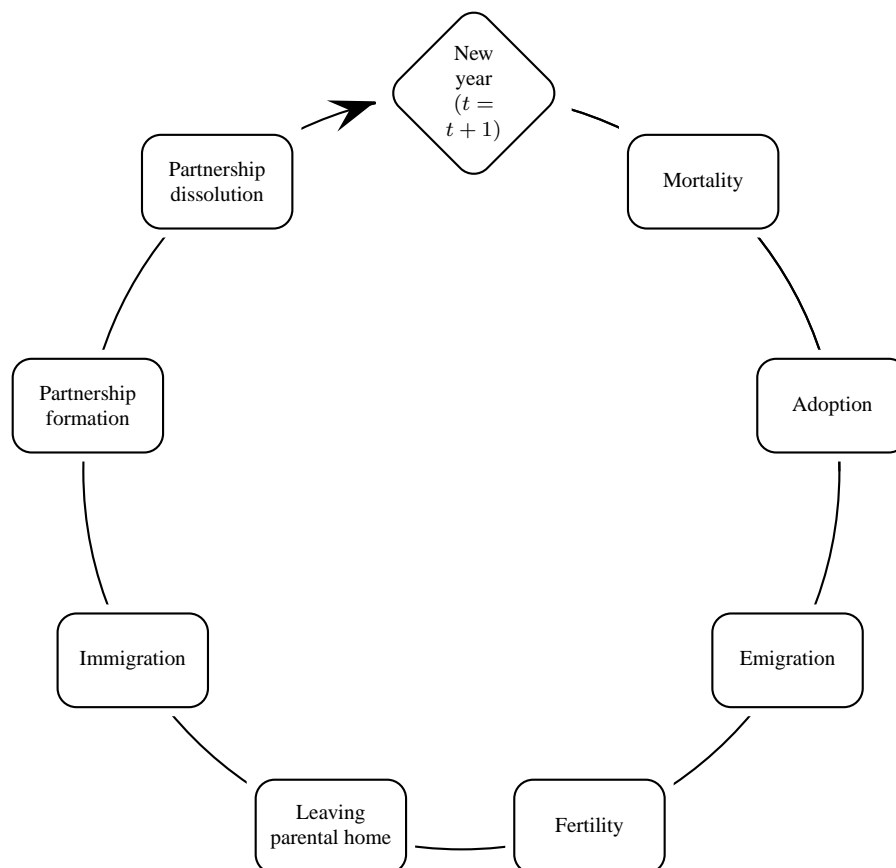
Figure 4: Order of the demographic submodules in the MSM.

## 3.4   Programming and Data Processing

The amount of data used by and produced by the MSM requires efficient data management. A database is utilised to store data for modelling, simulation data, simulation parameters, etc. The statistical programming language R (R Development Core Team, 2009) is employed for all types of data processing:

- Communication with the database
- Statistical modelling
- Running the simulation

### 3.5 Choosing Appropriate Data

In microsimulation, data is needed for three different purposes:
- Base Data for the simulation
- Data for Statistical modelling
- Validation Data

As the demographics module was the first one to be designed, the selection of the data was primarily based on the variables needed for this module. For the usage in MSMs, data has to fulfil two main criteria: It has to include a large number of cases which ensures that also smaller subgroups of a population are represented well and it has to cover as many of the important variables as possible to avoid or at least mitigate the need for variable imputation. Typically, there is a certain tradeoff between these requirements: Very large datasets like population registers normally only cover a few important variables, smaller datasets often provide more in-depth information – it is very unlikely that one single dataset is sufficient for all purposes of microsimulation. We decided to use Micro-Census data of the Statistik Austria, since it covers many of the variables needed, is of sufficient size and also meets two additional requirements: Disposability and Representativeness.

2004 was chosen as starting year for the simulation. The data of four quarters were pooled to get a simulation sample size of 76000 individuals. This corresponds to approximately 1% of the Austrian population, a number that is in line with many other microsimulation models (e.g., SESIM, APPSIM (Kelly, 2007), SAGE (Zaidi and Scott, 2001), DYNAMOD-2). For statistical modelling and validation issues, older Micro-Census data (2000-2003) are used. In some cases, probabilities for certain events do not have to be estimated using the Micro-Census data, but pre-assembled tables (e.g., life tables) can be used.

## 4 The Demographics Module

The demographics module consists of four *submodules* of different complexity, namely *Mortality*, *Fertility*, *Migration* and *Household*.

### 4.1 Mortality

Since only one event (death of a person) has to be modelled, *Mortality* is the simplest of the four submodules. Life Tables of Statistik Austria are used to assign probabilities of death, since estimation using Micro-Census data is not possible: If a person drops out early (it should be in the Micro-Census for 5 quarters), we do not know whether it has died, emigrated or moved from home.

The original life tables only differentiate mortality by gender and age. Based on an article of Klotz (2007), these tables have been expanded by the level of education. They also only contain probabilities of death up to the age of 95, but looking at the life table of 2007, it can be seen that almost 10% the women and over 4% the men are still alive at this age. Since these numbers seem too high to just eliminate persons from the dataset when they arrive at the age of 95, probabilities of death up to the age of 110 are modelled. A Weibull distribution was used for estimation, separately for men and women. The hazard

rate of the Weibull distribution should approximate the observed mortality rates in order to get reliable estimators for the mortality. The following two graphics show the comparison of mortality and hazard rates for men and women.
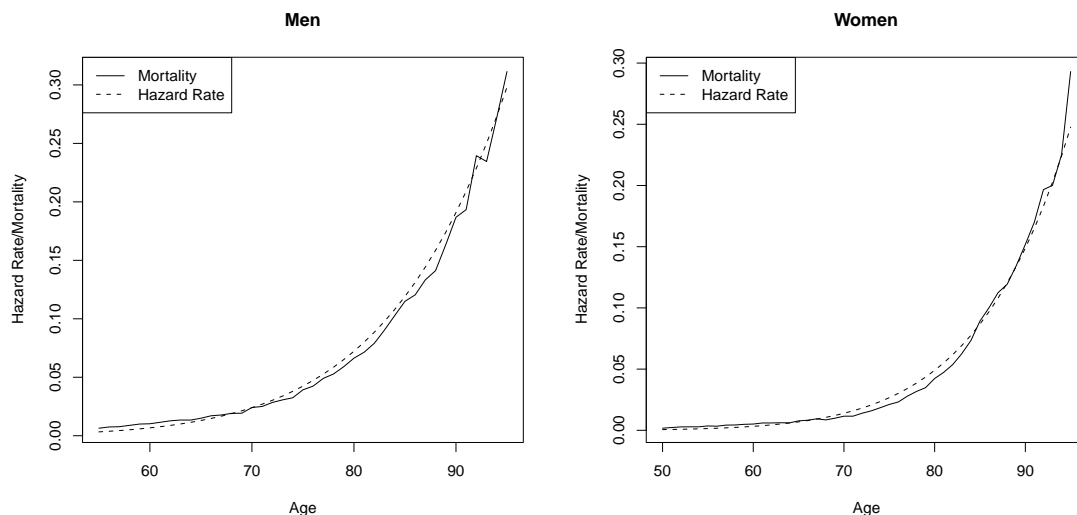


Figure 5: Comparison of observed mortality rates and hazard rates for men and women.

Since the probability of death is not monotonically increasing in the lower age band (infant mortality, child mortality, higher probabilities of death for males aged 18-25), modelling was only done for persons older than 50 - for persons below that age, the values given in the life tables of Statistik Austria are used. The approximation seems to be quite good, the estimated old-age mortality for men and women can be seen in Figure 6. Women have a lower mortality rate over the whole age band, but the differences get smaller for the very old. Note that the probabilities are set to 1 at the age of 110.

## 4.2  Fertility

*Mortality* effects every person in the dataset, *Fertility* is modelled only for women aged 16 to 45. Binary Logistic Regression (see, e.g., Hosmer and Lemeshow, 2000) is used to estimate the probabilities of the event *Birth of a child in the next year*. Two separate models are utilised for women cohabiting with a partner and single women (since there are possible predictors that may not make sense for some women: For example, the *employment status of the partner* can only be used for women actually having a partner). Additionally, tables of Statistik Austria are used to decide if there are multiple births and to determine the gender of the newborn children.

Table 1 shows the parameter estimates for women that are not cohabiting with a partner. The age of the woman, her participation in education and the number of children already present as well as the age of the youngest child have an influence on the probability of giving birth to a child next year. It can be seen that the relationship between age and birth of a child is not linear: The probability of birth rises up to a certain age (around
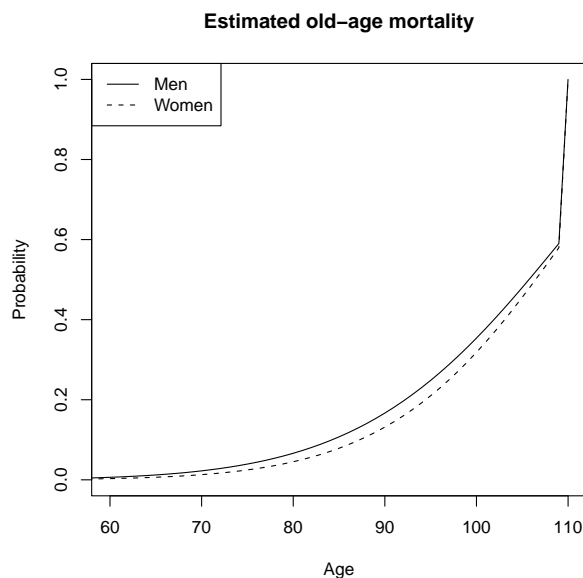
**Estimated old–age mortality**



Figure 6: Estimated old-age mortality.

30) and then goes back again. Therefore, a quadratic term is used to model this relation. For the categorical variables, the reference categories are italicised.

Table 1: Parameter estimates for the birth of a child in the next year (women not cohabiting with a partner).

|  | $\hat{\beta}$ | se($\hat{\beta}$) | z-value | p-value |
|---|---|---|---|---|
| (Intercept) | −13.189 | 2.173 | −6.068 | <0.001 |
| Age | 0.584 | 0.167 | 3.492 | <0.001 |
| Age squared | −0.011 | 0.003 | −3.744 | <0.001 |
| Participation in education no | 1.607 | 0.485 | 3.313 | 0.001 |
| *Participation in education yes* | 0 | — | — | — |
| One child (0-2) | 1.301 | 0.452 | 2.881 | 0.004 |
| Two children (3-9) | 0.685 | 0.459 | 1.493 | 0.135 |
| One child (10-19) | 1.178 | 0.514 | 2.291 | 0.022 |
| Two children (0-2) | 0.886 | 0.747 | 1.186 | 0.236 |
| Two children (3-19) | 0.291 | 0.633 | 0.460 | 0.646 |
| Three or more children | 1.274 | 0.638 | 1.998 | 0.046 |
| *No children* | 0 | — | — | — |

## 4.3 Migration

The future total *Migration* numbers are based on the projections of Statistik Austria for future migration: Because of the expiration of temporary arrangements for people coming from the eastern European countries that joined the EU in 2004 or 2007, respectively, it is

assumed that the immigration will slightly rise during the following years, up to 111.000 persons in 2011. After that it will go back, sinking to the current level of approximately 100.000 persons in 2020. It will then rise again until it reaches a level of 115.000 in 2035. This assumption is due to the falling population numbers of people in the working age in Austria, which cannot be compensated by a higher rate of labour participation alone. The emigration rates will remain at the current level, so that Austria will have a net immigration of 25.000 persons in the next years, rising up to 35.000 in the long run (Statistik Austria, 2008).

Migration in the microsimulation model always affects whole families: If a person is selected to migrate, the other members of the family also do. For emigration, people included in the simulation sample are selected randomly. At the moment, emigration just means that these people are removed from the dataset. It is planned to place them into a pool of emigrated person in the future, because they may still receive benefits from Austria, like pensions.

Immigrating families are clones of families already living in Austria. People of the simulation sample are selected for cloning at random until the specified number of immigrants for a certain year has been reached.

## 4.4　Household

The *Household* module is the most complex one, consisting of several different events:

- Formation of a cohabitation
- Dissolution of a cohabitation
- Marriage
- Divorce
- Choosing a partner (Partnership market)
- Child custody
- Leaving parental home

The four partnership formation and dissolution events are modelled using Binary Logistic Regression. Like the partnership market, they are female dominant which means that only women are eligible to search for a partner. The partner selection is influenced by a number of variables that depend on characteristics of the woman and of her potential spouse as well (like age difference, difference in education, etc.). Conditional Logistic Regression is utilised to estimate the relevant parameters. If a woman is selected for the formation of a cohabitation or marriage, 100 random suitors are assigned to her and each one of them may be selected with a certain probability, according to his characteristics. Monte Carlo Methods are used for the actual selection. This design is based on the work of SAGE, for details see Cheesbrough and Scott (2003).

In line with the Austrian numbers, in 85% of the cases the child custody is given to the mother. If only one of the parents is the biological one, this parent always takes custody of the child. All children aged 16 or older may leave their parental home. The estimation of the corresponding probabilities is based on a special programme of the Micro-Census of September 2001 (Statistik Austria, 2001). Leaving the parental home leads to the generation of a new household, consisting of the single person that left its home. New

households are also established if a couple chooses to cohabitate or to marry. These households consists of the couple and of children that may have been present before.

The following table summarises the methods utilised to get the transition probabilities for the different events of the demographics module:

Table 2: Modelling techniques and data sources for the events of the demographics module.

| Event | Method |
|---|---|
| Death | Life Tables of Statistik Austria (enhanced) |
| Birth | Logistic Regression |
| Immigration/Emigration | Tables of Statistik Austria's projections |
| Cohabitation formation/dissolution | Logistic Regression |
| Marriage/Divorce | Logistic Regression |
| Partnership market | Conditional Logistic Regression |
| Child custody | Table of Statistik Austria |
| Leaving parental home | Logistic Regression |

# 5 The Simulation Process

## 5.1 Alignment/Synchronisation

A MSM that only uses the transition rates that have been calculated utilizing the available data will not be capable of giving accurate estimates of future developments. There are two reasons for this:

- The data used for estimation captures conditions that may be valid only for a certain time period (e.g., fertility rates). This problem is normally solved by aligning/synchronising the model to match aggregate outcomes that are estimated using a macro model (see, e.g., Cassells, Harding, and Kelly, 2006). Official government projections can be used for this purpose, the changes of fertility and life expectancy as projected by the Statistik Austria can be seen as an example for this; in the ideal case, micro- and macro-model are developed simultaneously and experts of both fields work together.

- The data itself may have certain peculiarities that do not allow for the computation of estimates that are consistent with reality. As an example, the observed number of marriages in Austria in 2000 was about 39000. Using the Micro-Census Data and extrapolating to the total population, one would get a value much lower than that: ca. 26500. This can be seen as a result of a *lost to follow-up effect*: Because the Micro-Census is not person-, but household-centered, we have no information about persons who are not part of the initial household anymore (see Section 4.1). Events like marriage, formation of a partnership or birth of a child can be seen as a trigger for the establishment of a new household. In other words, a couple that has, e.g., the desire to marry has a certain likelihood to move from their parents' home in the same year the marriage happens, but if this sequence of events (leaving

> parental home → establishing a new household → marriage) occurs, the marriage
> – which was the source for leaving the parental home – is not captured in the data.
> Therefore, the real numbers of these events are underestimated; the estimates have
> to be adapted.

In the demographics module, alignment is utilised to adjust the number of births, deaths,
marriages and divorces.

## 5.2 Simulation Results

When using alignment, the results of the simulation and the projections from Statistik
Austria go together quite well. This can be seen in the following two graphics that com-
pare the projected total number of deaths and the projected population growth from 2005
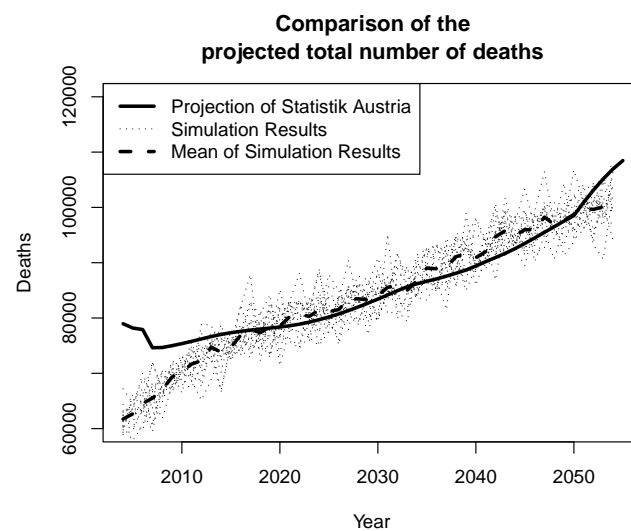to 2055.



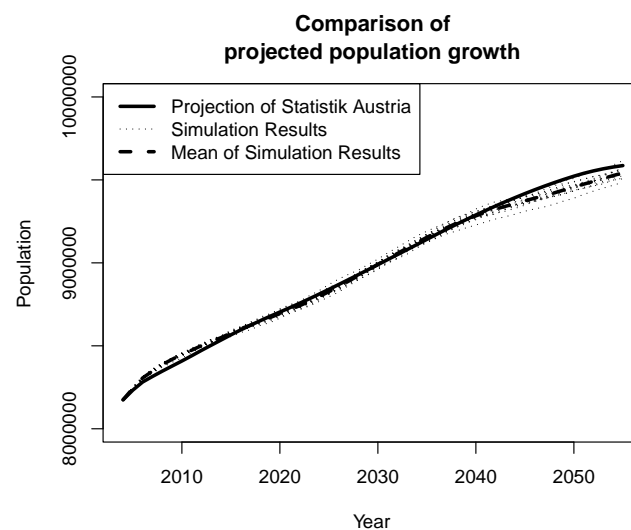Figure 7: Comparison of the projected total number of deaths.



Figure 8: Comparison of projected population growth.

The number of simulation runs was 20. The lack of fit in the first few years of the simulation originates from inconsistencies in the base data. Because of this, it makes sense to choose a starting point that lies a few years in the past so that the model has time to stabilise. A simulation run without alignment (constant mortality, constant fertility) resulted in an inverted U-shaped curve for the population numbers, with a shrinking population from ca. 2030 on (not depicted here).

# 6  Conclusion

In this paper, we presented first steps undertaken in the development of a microsimulation prototype that will be used to predict the long-term evolution of employment biographies in Austria: The design of the demographic module. Currently, education and employment career are modelled (Module 2). Education and employment variables serve as important predictors for a number of demographic events. Because of these dependencies, the quality of the prognoses of the demographics module cannot really be assessed at the moment. After module 2 has been completed, a reevaluation of the demographics module will be necessary.

## Acknowledgements

# References

Cassells, R., Harding, A., and Kelly, S. (2006). *Problems and prospects for dynamic microsimulation: A review and lessons for APPSIM* (Tech. Rep.). Canberra: National Centre for Social and Economic Modelling. (Discussion Paper no. 63 December 2006; Online: http://www.canberra.edu.au/centres/natsem/)

Cheesbrough, S., and Scott, A. (2003). *Simulating demographic events in the SAGE model* (Tech. Rep.). London School of Economics, UK: ESRC SAGE Research Group. (SAGE Technical note no. 4 December 2003; Online: http://www.lse.ac.uk/collections/SAGE/)

Flood, L. (2008). SESIM: A Swedish micro-simulation model.

Harding, A. (1996). *Microsimulation and Public Policy*. Amsterdam: North-Holland, Elsevier Science B.V.

Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Hoboken, NJ: Wiley.

Keegan, M. (2007). *Modelling the workers of tomorrow: the APPSIM dynamic microsimulation model* (Tech. Rep.). Canberra. (HILDA Survey Research Conference 2007 19-20 July 2007; Online: http://www.melbourneinstitute.com/conf/hildaconf2007/)

Kelly, S. (2007). *APPSIM: Selection of the main source data file for the base data* (Tech. Rep.). Canberra: National Centre for Social and Economic Modelling. (Working Paper No. 2 April 2007; Online: http://www.canberra.edu.au/centres/natsem/)

King, A., Bækgaard, H., and Robinson, M. (1999). *DYNAMOD-2: AN OVERVIEW* (Tech. Rep.). Canberra: National Centre for Social and Economic Modelling. (Technical Paper no. 19 December 1999; Online: http://www.canberra.edu.au/centres/natsem/)

Klotz, J. (2007). Soziale Unterschiede in der Sterblichkeit; Bildungsspezifische Sterbetafeln 2001/2002. In *Statistische Nachrichten April 2007*. Vienna: Statistik Austria.

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

R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org`

Statistik Austria. (2001). *Familienstrukturen und Familienbildung – Ergebnisse des Mikrozensus September 2001*. Vienna: Statistik Austria.

Statistik Austria. (2008). *Demographisches Jahrbuch 2007*. Vienna: Statistik Austria.

Zaidi, A., and Scott, A. (2001). *Base dataset for the SAGE model* (Tech. Rep.). London School of Economics, UK: ESRC SAGE Research Group. (SAGE Technical note no. 1 September 2001; Online: http://www.lse.ac.uk/collections/SAGE/)

Authors' address:

Marcus Wurzer and Reinhold Hatzinger
Department of Finance, Accounting and Statistics
Institute for Statistics and Mathematics
Augasse 2-6
1090 Vienna
E-mail: `marcus.wurzer@wu.ac.at`, `reinhold.hatzinger@wu.ac.at`