

**microWELT: A Dynamic
Microsimulation Model for the Study
of Welfare Transfer Flows in Ageing
Societies from a Comparative
Welfare State Perspective**

Martin Spielauer
Thomas Horvath
Marian Fink

microWELT: A Dynamic Microsimulation Model for the Study of Welfare Transfer Flows in Ageing Societies from a Comparative Welfare State Perspective

Martin Spielauer, Thomas Horvath, Marian Fink

WIFO Working Papers 609/2020
September 2020

Inhalt

This paper introduces the microWELT model. Starting from its objectives, we discuss design choices, the model architecture and key features. microWELT provides a demographic projection tool reproducing Eurostat population projections but adding details such as education, intergenerational transmission of education, fertility by education, partnership patterns, and mortality differentials by education. The model integrates transfer flows as captured by the National Transfer Account (NTA) and National Time Transfer Account (NTTA) accounting framework and calculates a set of indicators based on NTA literature. Individual accounts allow the study of transfers over the whole life cycle by cohorts and between generations.

E-Mail: martin.spielauer@wifo.ac.at, thomas.horvath@wifo.ac.at, marian.fink@wifo.ac.at

2020/246/W/7316

© 2020 Österreichisches Institut für Wirtschaftsforschung.

Medieninhaber (Verleger), Hersteller: Österreichisches Institut für Wirtschaftsforschung
1030 Wien, Arsenal, Objekt 20 | Tel. (43 1) 798 26 01-0 | <https://www.wifo.ac.at>
Verlags- und Herstellungsort: Wien

Die WIFO Working Papers beruhen nicht notwendigerweise auf einer abgestimmten Position des WIFO. Die Autorinnen und Autoren wurden über die Richtlinien der Österreichischen Agentur für wissenschaftliche Integrität (ÖAWI) zur Guten Wissenschaftlichen Praxis informiert, insbesondere bezüglich der Dokumentation aller Elemente, die für eine Nachvollziehbarkeit der Ergebnisse notwendig sind.

Kostenloser Download: <https://www.wifo.ac.at/wwa/pubid/66390>

microWELT

A Dynamic Microsimulation Model for the Study of Welfare Transfer Flows in Ageing Societies from a Comparative Welfare State Perspective

Martin Spielauer, Thomas Horvath, Marian Fink

Abstract

This paper introduces the microWELT model. Starting from its objectives, we discuss design choices, the model architecture and key features. microWELT provides a demographic projection tool reproducing Eurostat population projections but adding details such as education, intergenerational transmission of education, fertility by education, partnership patterns, and mortality differentials by education. The model integrates transfer flows as captured by the National Transfer Account (NTA) and National Time Transfer Account (NTTA) accounting framework and calculates a set of indicators based on NTA literature. Individual accounts allow the study of transfers over the whole life cycle by cohorts and between generations.

Keywords: Microsimulation, Welfare State, Demographic Change, National Transfer Accounts

JEL: E01, J11, P51, O57, C53

This paper is part of the WELTRANSIM project, a project of the Horizon 2020 Joint Program Initiative More Years, Better Live's second joint transnational call entitled Welfare, Wellbeing and Demographic Change: Understanding Welfare Models.

1. Introduction

This paper provides a non-technical introduction to the purpose and design of the microWELT model. microWELT is a dynamic microsimulation model developed at the Austrian Institute of Economic Research (WIFO) alongside the research program WELTRANSIM (Welfare Transfer Simulation) studying the distributional effects of four welfare state regimes represented by Austria, Finland, UK, and Spain. The WELTRANSIM project is an international collaboration funded by the Joint Programming Initiative "More Years, Better Lives". Project partners are the University of Barcelona, the Finnish Institute for Economic Research, and the Finnish Centre for Pensions.

Dynamic microsimulation can be perceived as experimenting with a virtual society of thousands – or millions – of individuals who are created in a computer and whose life courses evolve over time (Spielauer 2010). Individual actors attend school and make educational choices, form families, migrate, become parents, work, have earnings, pay taxes and receive benefits, make and receive monetary and time transfers within and between families, and eventually die. Dynamic microsimulation models come in a wide variety of model architectures and designs, reflecting differing priorities, data availability and restrictions, as well as current or past technical choices and limitations.

The paper starts with the concrete model requirements for the WELTRANSIM project. This is followed by a discussion of key architectural and design decisions – and reflections on what makes a good model –, identifying trade-offs between relevant approaches for microWELT. This provides the context for the description of the specific features of the microWELT model and its main components.

2. Requirements for the microWELT model

The purpose of microWELT is twofold: to provide a study and projection tool oriented at the aims of the WELTRANSIM project, and to provide an open-source modelling platform which can be adapted, extended, and refined for a broad field of new applications. The WELTRANSIM project studies the interactions between welfare state regimes, welfare transfers, and population ageing accounting for educational change, life expectancy, differentials by education, and changing family patterns. Development steps in the context of WELTRANSIM focus on socio-demographic behaviours combined with a highly abstract implementation of the key features of different welfare state regimes. These mechanisms are captured in National Transfer Accounts (NTAs) and National Time Transfer Accounts (NTTs) as well as in the distinct policy reform mechanisms responding to the pressures of population ageing. The study compares four European countries – Austria, the UK, Finland, and Spain – representing four welfare state regimes: Corporatist-Statist, Liberal, Social Democratic, and Familial / Mediterranean. Individual accounts allow for the longitudinal analysis of transfer flows between population groups and generations. From a platform perspective, we aim at developing microWELT in a

way that supports refinements for individual countries, implementing behaviours as well as social insurance and tax systems in more detail.

2.1 Requirements for socio-demographic projections

Given the comparative perspective of the project, we aim at identifying key socio-demographic characteristics and patterns of the four studied countries attributable to the respective welfare state regimes as well as the processes driving socio-demographic change. Studied processes with potential links to welfare state types include (1) the intergenerational transmission of education, (2) childlessness and fertility by education, (3) partnership behaviours and lone parenthood, (4) age at leaving home, and (5) mortality differentials by sex and education. Differences in these processes are captured in the model parameters. By means of microWELT projections, we identify the impact of these processes on the future population composition by age, sex, education, and family characteristics of the countries studied. At the same time, we consider it valuable to be able to reproduce existing Eurostat population projections at the aggregated level. This leads to the following list of key requirements:

- Fertility: while being able to reproduce given fertility projections, we aim at providing a realistic depiction of differences in fertility (distribution of family sizes, age pattern) by education with an emphasis on childlessness. Empirically, the distribution of family sizes in most European countries followed a trajectory from a high concentration of reproduction (high childlessness, fewer women having a high share of children; see, e.g. Spielauer 2005, Shkolnikov et al. 2007) at the beginning of the past century, to an equal distribution in the baby boom, back to an increasing concentration – especially higher educated women having increasing rates of childlessness. These patterns vary between welfare state regime and are of key relevance when studying transfer flows between population groups by education and family characteristics.
- Mortality: while reproducing Eurostat projections of mortality rates by age and sex, we aim at capturing life expectancy differences between education groups. The goal is to keep relative risks in mortality hazards as observed today while aligning overall results to the projected aggregated rates.
- Education: the application should be able to model educational attainment distinguishing three education levels, implement realistic patterns of school attendance by education outcome, and to account for the intergenerational transmission mechanisms in education, i.e. the influence of parents' education on education choices. The model should also support the (optional) alignment of educational outcomes by sex and birth cohort.
- International migration: the model should be able to reproduce Eurostat projections of migration flows by age and sex.

- Families: the model should be able to link persons to (nuclear) families. This involves the modelling of union formation and dissolution, re-partnering, leaving home, and the maintenance of links between persons over the simulation. To capture differences in welfare states on partnership patterns the model should be able to realistically depict the partnership status by education and the presence and age of children and match spouses by observed distributional patterns concerning age and education. From a longitudinal perspective, the model should allow distinguishing cohorts of parents from childless individuals and couples.

2.2 Requirements concerning the accounting framework

microWELT should implement an accounting framework based on the National Transfer Account (NTA) framework. NTA data are cross-sectional age profiles breaking down national accounting variables on consumption, income, saving, and transfers. Data distinguish between private and public consumption singling out education and health. Public transfers also distinguish pension benefits. NTA data by sex are available for more than 50 countries, including most European countries (agenta-project.eu). Alongside the WELTRANSIM Project, we further disaggregated NTA data by education and family type, these disaggregated NTA data constituting key parameters of the model.

- Implementation of individual accounts based on the National Transfer Account (NTA) framework. NTAs consist of a set of variables on private and public consumption by type, public transfers, private transfers within and between families, labour and asset income, and saving. Mechanisms should be provided to study transfer flows between population groups both in the cross-section as well as longitudinally accounting over the whole life cycle of cohorts.
- Implementation of time transfers based on the National Time Transfer Account (NTTA) framework. Time transfers are measured for various types of home production. Mechanisms should be provided to study transfer flows between population groups both in the cross-section as well as longitudinally accounting over the whole life cycle of cohorts.
- Ability to implement mechanisms of policy reforms addressing the sustainability of systems in the context of population ageing. While having to be highly stylised, given the limited number of monetary flows distinguished in the NTA framework, mechanisms should reflect typical ways corresponding to the different welfare models.

2.3 Requirements from a platform perspective

microWELT should provide a versatile platform allowing model refinements concerning tax-benefit and social insurance systems, welfare transfers, socio-economic life-courses (employment careers, earnings, retirement), wealth accumulation and inheritance, health, and

migration. We immediately found the microWELT modelling platform in high demand for other applications. In this context, so far, the following priorities were identified:

- Ability to easily include new individual characteristics such as place of origin or immigration status as an additional characteristic and explanatory factor in behavioural models.
- Ability to model pools of immigrants by geographical origin and immigration type supporting detailed migration scenarios. This also involves accounting for differences in emigration risks by origin and the support for modelling back migration.
- Ability to refine educational projections depicting the details in national education systems and the differences in school choices and success rates by detailed socio-demographic characteristics.
- The ability to add new modules, e.g. for the modelling of labour force participation, employment, health, and retirement decisions (and their interaction) as well as related policies.
- The ability to study processes and policies on a sub-annual scale, like unemployment spells and duration-dependent unemployment benefits. Other examples are family policies like maternity and parental leaves and health-related processes and policies like sick leaves and benefits.

3. Approaches

This section discusses the key architectural and design decisions and identifies trade-offs between relevant approaches for microWELT. This chapter is organised by key questions – and corresponding design options:

- How is the population created? Does the simulation start from a starting population file or is the population built synthetically?
- When does the simulation start? Does the model include a simulation of the past? Do all lives start at birth? What is the historical depth of the model?
- How does time evolve: continuously or in discrete steps?
- Are spouses matched within a closed population?
- In which order is the population simulated: case-based or time-based? How much communication between actors is required?
- What is modelled, what is assumed, and which processes require (or should include options for) alignment?

- In which order is the model developed? In what sequence, at which initial level of detail? Which development steps can be parallelised? Is the model designed and developed in a modular manner to enhance efficiency and flexibility?
- In which programming language is the model implemented?

These points give the context for discussing design choices for microWELT. In the following chapter, this discussion will be complemented by a structured list of criteria – a wish-list – frequently associated with a good dynamic microsimulation model.

3.1 How is the population created in the computer?

Basically, there are two methods of creating a population in a computer: (1) by reading a set of individual variables characterising each member of the population from a cross-sectional file, or (2) by creating the set of characteristics of each person of the population synthetically by modelling his or her life from birth.

The first option requires a starting population file. Most existing dynamic microsimulation models start from such a starting population. A frequent problem of this approach is that not all information required is available in a single existing data set or is not available at all from existing data. When variables are missing from the starting population, they must either be imputed or modelled. This can be challenging, especially with complex correlations between the variables.

The second option requires some distributional information of the population – e.g. how many people were born in each year by sex – and a set of models of individual life course careers based on historical information. The most prominent example of this approach was probably the Canadian LifePaths model (Spielauer et al. 2013) that created complex life histories from birth to death representative of the history of Canada's population. At birth, very few characteristics of a person are known – the year of birth, sex, province of birth –, all other characteristics are modelled over the life course. In consequence, all lives are fully synthetic and do not correspond to real individuals from a single micro-dataset. Another characteristic of this approach lies in the fact that it does not only model the future but also the past of each person. Benefits of the synthetic method include the avoidance of confidentiality issues and the consistency in variables across time. Another benefit is that there is no upper limit for the size for the modelled population, which can support robust modelling of rarer subpopulations (if the transitions are estimated at that level of detail). In a model with a starting population, the size of the starting population will influence the simulation results. Even though cloning is possible, without additional work, the characteristics of the starting population will remain constant, resulting in an incomplete depiction of the true diversity in the modelled population.

microWELT starts from a cross-sectional starting population file based on the Euromod (Sutherland and Figari, 2013) dataset. Reasons are the internationally standardised nature of this dataset, its completeness concerning required information, and the fact that it is used in the static Euromod tax-benefit model, thus supporting communication between the models. The latter feature is used in the WELTRANSIM project in a static ageing exercise, reweighting the Euromod database by weights generated by microWELT. The choice also allows the extendibility of microWELT as a rich set of additional variables relevant to tax-benefit calculation can be easily added to the starting population from the Euromod dataset. Also, the Euromod database is frequently updated.

3.2 When does the simulation start?

We can distinguish three important points in time:

- When are individual actors created?
- When does the projection of individual life courses start?
- At which point of time is the simulated population complete?

In synthetic models, all persons are created at the moment of their birth. In models based on a starting population file, there are basically two options: (1) at birth, using the starting population information of individual birth dates, or (2) at the point of time corresponding to the starting population file. In the latter case, persons are created with the age observed in the starting population. The choice of approach will depend on which past information of each actor has to be known in the simulation, how this information is obtained, and the accounting framework of the model. In the simplest case, all past information required is contained in the starting population file. In contrast, starting lives at birth allows the retrospective simulation and/or imputation of characteristics. Creating persons at birth also allows different approaches to be combined: some characteristics can be taken directly from the starting population, while other processes can be synthetically created imputing past histories consistent with observed cross-sectional outcomes.

When does the projection start? Or, what is a good starting point in time? When creating a population from a starting population, we must choose not only the best source of data but also its time. We can start with the most recent available data. Or use an earlier starting point, allowing some retrospective validation of the model. The Canadian DYNACAN (Morrison 1998) model was an example of selecting a historical starting population; the same holds true of the initial design of the US CORSIM model (SOA 1997), a choice which was later reversed. Similar to the choice between creating a synthetic population versus using a starting population file, going back in time comes at the cost of spending significant development time to reproduce the past. On the other hand, besides the opportunity for retrospective model validation, selecting

an older starting population has a second potential benefit connected to another important moment of time to be decided for a model: the point in time when the simulated population is complete in the sense that it is suitable for cohort analysis. This is because starting populations typically only contain information on persons who are alive (and resident) at that observation point. This impedes cohort analysis for cohorts born before the start of the simulation, as only survivors are observed. Also – when modelling a tax-benefit system – contributions and benefits will not add up to historical totals, as the representation of the underlying population is not complete.

In microWELT, the starting year of the simulation is 2010; thus, we create the starting population from the 2010 Euromod database. The choice of this date was taken as one of the main model parameters – NTA and NTA variables – are currently only available for this year. microWELT itself is implemented in a way supporting the easy change of the start year of the simulation and the according shift in period parameters¹. Individual actors are created at their birth; thus, all simulated lives start at birth and not 2010. This choice was taken in order to support the modelling of life-course careers from birth. In microWELT, this is used to back-impute full educational careers matching with the information on school attendance and attainment in 2010. Also, immigrants are created at birth; thus, we allow people living abroad before eventually entering the country. This choice was taken to provide full flexibility how immigrants are modelled, allowing the implementation of pools of immigrants, the inclusion of former residents currently living abroad and potentially back-migrating, and the modelling of immigrants from scratch, thus not being created from records of the starting population. While many of these features are not used in the microWELT project, the approach has turned out very beneficial already in other projects build on this platform (e.g. the modelling of the economic integration of immigrants in Austria, see Horvath et al. 2020). microWELT does not include people who died before the start of the simulation in 2010. For longitudinal analysis based on the full life-courses of birth cohorts, the consequence is that such analysis can be performed for birth cohorts 2010 and onwards.

3.3 How does time evolve in the simulation?

After having created a representation of the population in a computer – how does it move through time? There are two ways of modelling time: discrete and continuous. Discrete-time models move time in steps, like years or months. At each step, all characteristics of each actor are updated, actors are removed due to death or emigration, and new actors are added due to birth or immigration. This is often regarded as the "classical way" of modelling time, found in many pension models, such as the European family of MIDAS models in Belgium, Germany and

¹ Other projects built on this platform typically have other starting years.

Italy (Dekkers & Belloni 2009). Especially in the past when computing power was a serious bottleneck for microsimulation models, this was a natural approach, and databases holding the population starting file were readily available. In this tradition, dynamic microsimulation is essentially about periodically updating a database. Discrete-time models have several limitations and drawbacks, especially if the period between time steps is long (annual updates are typical):

- Some events might happen multiple times in a year. A person might have moved in and out some state (like employment) a number of times in a year or might have moved back to the initial state with all changes within the year going unnoticed. This state might have to be modelled explicitly as "combined events" with back-imputing of relevant information, such as time worked.
- When updating states in fixed time steps, information about the ordering of events is lost if not explicitly modelled. We might observe state changes to "married" and "mother" without knowing the order. We might have to model combined events, such as "became married followed by the birth of a child" and "birth of a child followed by marriage" if both events occurred. Given the multitude of events occurring in life – and in detailed life course simulations – this approach makes it difficult or impossible to model processes which influence each other in both directions, such as marriage influencing pregnancy, and pregnancy influencing marriage risks. Special care has to be taken if the last year of life is of importance, in order to account for incomplete years even if a person dies. Good examples are models of health costs, as the last year of life is typically the "most expensive" from a social insurance perspective.
- This same point applies whenever partial-year dynamics are important, for simulation or for reporting. For example, sub-annual episodes are typical for unemployment and the chance of finding new employment might be duration dependent; also benefit formulas might include the current state duration.

A way around these drawbacks is to model short discrete time steps (months, or days) making it unlikely that more than one event happens at a time. While it is still a clock that determines when changes are modelled, the clock ticks faster / rings more often, causing updates in smaller, more frequent steps. This approach was pioneered in the Australian DYNAMOD Model (King et al. 1999). While allowing the use of the same technical framework, model runs are slowed down. By shortening discrete time steps, models at some point can reach "pseudo continuity"².

² For a discussion from discrete time versus continuous time in the context of DYNAMOD see Galler 1997

Continuous-time models bypass the need for a clock by having events themselves move time. There is no clock event between other types of events. The next event that occurs in the simulation moves time to the time of that event. The statistical method associated with this approach is models of duration, such as piecewise constant hazard regression models. Given that all states stay as they are for any time segment short or long, events are competing. Whichever event happens first moves time forward. At any event, the waiting times for any modelled processes can be re-evaluated, as they might have been affected by the changes which occurred in the last event. So, for example, the birth of a child may increase the probability of marrying and decrease the probability of returning to school.

Continuous-time models are the natural approach from a life course perspective. This perspective has shaped how we study lives by splitting them into career domains (e.g. education, family, work, health, earnings, etc.), recognising the dependence and interactions between careers (e.g. the reconciliation of family and work careers), human agency (e.g. expressed how we order our life priorities over time), and the interaction between "linked lives", e.g. partnerships, families, or broader social networks. In consequence, the life course perspective has changed the way we collect data (collecting retrospective careers, linking observations of the same person in panels, or collecting administrative event data like health and treatment histories). In the context of dynamic microsimulation models, continuous-time models are frequently the first modelling choice of demographers and for health models.

While solving many problems associated with discrete-time models, continuous-time models require adequate data. On the other hand, given their increased flexibility (nothing prevents a modeller from implementing periodic time steps in a continuous-time model) the approach also accommodates hybrid models. In such models, some processes create events in continuous time – typical examples being demographic events like births and death – while other states (and typically most accounting) are updated at periodic events, like calendar year changes or birthdays.

microWELT is a continuous-time model which we see as the most natural approach from a life course perspective bypassing a series of problems associated with discrete-time models as outlined above. By using what we believe to be the most powerful microsimulation programming technology available today (Modgen/openM++; see below), we regard the higher computational complexity neglectable. In the context of WELTRANSIM, we see the biggest advantage of this approach in the extendibility of the modelling platform for applications requiring accurate accounting on a sub-annual level. The benefit is already demonstrated by various applications built on the microWELT platform, as documented at the project website <http://www.microWELT.eu>.

3.4 In which order is the population simulated

There are two ways of ordering the simulation of a population: (1) case-based and (2) time-based. In case-based models, cases are simulated one by one. In its simple variant, a case refers to a single person. Individual lives are simulated one by one. Case-based models are computationally very efficient, as simulations can be easily parallelised. Also, there are no population size restrictions, as only one case is kept in computer memory at a time.

The primary limitation of case-based models is that no communication or interaction between cases is possible. This has two main consequences. First, it means that spouses are not found in the existing population but are created as attributes of the simulated population. Secondly, not being able to communicate between cases impedes the alignment of outcomes, as population-wide accounting is only possible at the end of a given simulation. As actors cannot observe others outside their own case, population-wide measures or indicators are not available for modelling individual decisions, and it is also not possible to model a centralised actor like a government adjusting taxes and benefits according to periodic outcomes like tax revenue. Some of these latter limitations can be overcome through the use of sequential simulations, with the population-level outcomes of one simulation informing the micro-level outcomes of the next. However, this is not as efficient as direct communication within a single simulation.

In time-based models, all individuals alive in a given period of time are simulated simultaneously, allowing communication or interaction between all actors. This allows a wide range of simulation applications including the spread of diseases, the modelling of policy formulas that include mechanisms to adjust to total revenues and expenditures (e.g. to balance budgets), and models that include alignment routines adjusting aggregate outcomes to given targets. Spouses can also be found within the existing population, provided that the population is sufficiently large to support correlations in partner characteristics.

The price of such models can be reduced speed and more restrictions concerning the number of individuals in a simulation, the level of detail with which those individuals are represented, and the length of the projection horizon. This has the potential to be a larger issue in continuous time models, as actors must be kept in memory simultaneously and schedule their events in the same shared event queue. Whatever happens first to any of the simulated actors moves the time for the whole population as it might affect other actors. Specialised microsimulation programming languages like Modgen and OpenM++ implement various techniques (like the "just in time" approach) to manage these issues.

microWELT is a time-based interacting population model; thus, all actors are simulated simultaneously and can communicate at any moment of time. The choice is based on the requirement to allow the implementation of policy responses for balancing budgets over time, and the requirement to allow (optional) model alignments to external targets. We also link persons to families. The higher computational requirements of this approach mainly concern limitations

in the size of the simulated population. As microWELT is based on a starting population sample stemming from SILC (with typical country sample sizes of around 30,000 persons) we do not see this as a major limitation as larger simulations cannot add distributional detail beyond the starting population³.

3.5 Is the population open or closed?

A closed dynamic microsimulation model is a model in which, once the simulation has started, the only way a new person can enter the simulation is due to a simulated woman giving birth. This has three implications:

- Spouses must be found within the existing population
- When the model includes immigration, it must simulate not only the resident population but also the source of immigration.
- Closed population models require a time-based model architecture: this point was discussed in the previous section.

Open population models, in contrast, allow persons to be created and enter the simulation as required. In the case of immigration, this means that immigrants can be created at the moment of entering the country. Creating immigrants on demand can solve a typical problem of microsimulation models: finding a way to simulate a country without simulating the rest of the world. An open population model can avoid simulating people who have never entered the country and might never do so, but the problem then shifts to the question of how to model the characteristics of new immigrants. Typical solutions include cloning, with hosts being recent immigrants of the same age or people with another set of required characteristics. Depending on the degree of detail, the modelling of immigration can be a complex matter involving the problems of modelling family migration, back migration, and the modelling of various types of immigration (e.g. work migration, family migration, refugees) each associated with very different distributions of individual characteristics. This holds true both for open and closed models.

One of the key differences between open and closed models is the modelling of spouses. In closed models, spouses are found within the existing population. This requires a time-based modelling framework. In open models, spouses are attributes of the dominant population. Both case-based and time-based models can be open. For example, both the case-based LifePaths and the time-based SIMUL are open models and have spouses modelled as attributes. They are labelled "non-dominant actors" in LifePaths, or "ghosts" in SIMUL (Bissonnette et

³ Concerning Monte Carlo variability, we found simulation runs of initially 200,000 persons – created by sampling and cloning from the starting population – and running 5-12 replicates sufficiently large to eliminate Monte Carlo variability in most simulation results and easily manageable concerning computational time.

al. 2016a). Depending on the information required about the spouse's previous life, the creation of spouses with appropriate characteristics is typically a complex issue.

Comparing the closed population and the open population approaches of creating spousal unions, a typical trade-off is between the quality versus the consistency of matches. When modelling a closed population, we have the problem that we usually simulate only a sample of a population and not the whole population of a country. If the sample is too small, it becomes unlikely that high-quality matches can be found within the simulated sample. This is especially the case when modelling geographical detail.

On the other hand, open models face the possibility of inconsistent matches: without care, the set of modelled spouses might not be consistent with population totals. By their nature, closed models automatically achieve population-level consistency in spouse matching, as spouses have to be found in the population. So, if there are exactly 219,345 men married to women in the model, there will also be 219,345 women married to men. In an open model, the estimates may not be exactly symmetric. Relatedly, upon divorce or union-dissolution, any income or resources split or shared between spouses consistently add up in the simulation accounts in the closed model. In an open model, there will similarly be complete accounting consistency between a main (or dominant) actor and his or her non-dominant spouses. However, the accounting consistency within the population of main/dominant actors will not be perfect.

microWELT is closed concerning partner matching, i.e. spouses have to be found within the simulated population. As microWELT uses only age and education as a matching criterion and does not include geographical characteristics, this is not seen as a limitation. For other applications built on this platform (so far this has shown when modelling partner matching also based on place of origin, see Horvath et al. 2020), the problem of finding plausible matches can become more problematic and requires large size simulations. The microWELT simulation platform does not principally impede the implementation of an open population concerning partnerships but adding non-dominant actors as a basic model feature is beyond the scope of this project.

Concerning immigration, microWELT is an open model, the size of the immigrant population (and some characteristics as its age and sex distribution) being parameters. Immigrants characteristics in microWELT are acquired by cloning from an appropriate host population of current residents of matching age and sex. As an extendable modelling platform, microWELT also allows alternative approaches of modelling immigrants, including population pools and people currently living abroad contained in the starting population.

3.6 What is modelled, what is assumed?

Dynamic microsimulation models are used in various ways. To make projections at the aggregate level (Is the current system sustainable?), to make statements on distributional aspects (Is

the system adequate? Who are the winners and losers of a reform?) and to answer a broad range of hypothetical "what if" questions. The first use requires predictive power. The second, a good representation of the population heterogeneity, while projected aggregates might come from another source. The third, flexible and easy ways to express "what if" scenarios. How would an increase in life expectancy by five years affect public expenditures? To what extent would any such impacts be offset if fertility also increased to levels observed in the baby boom?

Aggregate prediction power, distributional detail, and easy scenario creation typically cannot be obtained simultaneously. Modelling distributional detail requires a larger number of variables. If adding variables (and the simulation of processes to update these variables in the simulation) requires estimation, additional variables add randomness. While too simple a model, missing key variables, might make wrong predictions as it is misspecified, adding variables beyond some point could compromise a model's predictive power (for a discussion see Van Imhoff and Post 1998). Furthermore, detailed variables are frequently only available in specialised surveys of limited sample size, thus subject to sampling errors which also increase randomness. To some extent, this problem can be addressed statistically, e.g. by proportional models. Such models have the advantage that aggregate outcomes and relative factors (distributions) can be estimated separately from different data sets, thereby combining the robustness of large data sets (like the Census or vital statistics) for overall outcomes with the wealth of variables offered by other sources such as surveys used to estimate relative factors. In many large-scale dynamic microsimulation models, some outcomes are aligned or calibrated towards aggregate numbers or external projections from other sources, at least in part in response to such issues.

What is explicitly modelled, what is assumed, and what is modelled but should allow for alignment or calibration to external scenarios are fundamental design choices. For example, many models do not try to make demographic forecasts but rather have mechanisms to reproduce official population projections. But even so, a model might want to add ways to account for relative mortality differentials by characteristics like education, marital status etc. Or it might want to be able to reproduce the age-specific fertility profiles as available from population projections but add mechanisms to realistically distribute births to women by a set of individual characteristics such as parity, time since last birth, current partnership and employment status. This also applies to economic projections. For example, models do not necessarily produce their own forecasts of future unemployment rates but might include mechanisms to account for individual differences in unemployment and re-employment risks into the model.

There are often trade-offs between specification and transparency. We might opt for a more transparent model even knowing that another, more complex model would improve predictions but make it hard to identify the underlining drivers of outcomes or impede scenario creation as it would require making changes to a multitude of complex equations by changing

parameters with little intuitive meaning. Ultimately, making choices about the appropriate level of complexity requires considerable judgement and should reflect the most critical needs of model users. Designing a model that aligns or calibrates to specific aggregate outcomes does not replace well-founded modelling work based on detailed microdata. Quite on the contrary, models should also be used to assess the projections of alternative modelling approaches and to challenge available macro projections. An implication is that alignment or calibration techniques must be able to be turned on and off.

microWELT does not aim at improving available demographic projections as published, e.g. by Eurostat on the aggregate level but adds distributional detail. In the case of fertility, this means, that microWELT reproduces age-specific fertility rates, but distributes babies realistically to potential mothers by characteristics like education and the number of previous children. Mortality is consistent with aggregate life tables (death rates by age, sex, and year), but relative mortality differences by education are applied. International migration is modelled applying directly published numbers, rates and age distributions of emigrants and immigrants. In the case of education, alignment to external targets is optional, as microWELT models the intergenerational transmission of education which – based on the scenario – can drive changes in the educational composition of the population either entirely or can be aligned to given outcomes. In the latter case, the individual differences in the likelihoods of outcomes by parents' education are respected. All processes of microWELT for which alignment is available are based on proportional models which express risks (in continuous time) or odds (for discrete-time decisions) by a base factor applying to all, and a set of relative factors applying to specific population groups. The logic behind this type of alignment is to adjust base factors in a way that – for a given population composition, and when applying individual relative factors – a target aggregated outcome is reached. In this way, relative differences between population groups are maintained. Proportional models are very convenient in microsimulation. From a modelling perspective, they have the advantage that aggregate outcomes and relative factors can be estimated separately from different data sets, thereby combining the robustness of large data sets (like the Census or vital statistics) for overall outcomes with the wealth of variables offered by other sources such as surveys used to estimate relative factors. Another advantage of proportional models is that they support very intuitive scenarios, e.g. trends which apply to all population groups, and changes in relative differences, like scenarios of closing gaps between groups. (For a discussion in the context of Statistic Canada's Demosim population projection model, see Caron-Malenfant & Coulombe 2015). When estimating overall outcomes and relative differences separately, base factors still must be found which, together with the relative factors, result in the target aggregate rates for a given population composition. In the context of parameter estimation, this step is performed outside of the model. Alignment routines in microWELT follow the same logic as when combining estimates of aggregate targets with estimates of relative factors to one proportional model but perform the search for base

factors within a simulation run. While the technical implementation of some alignment routines is complex (for detailed documentation see the implementation guide at the project website), the mechanism has a clear statistical interpretation.

3.7 In which order is a model developed?

The development order refers to the sequence in which project components and modules are created, and the timing of adding all necessary detail. A typical sequence is life-course driven, starting with demographics and socio-economic processes that take place earlier in the life-course, moving on module by module – ideally allowing for the parallel development of some components. Concerning detail, there are basically two choices: creating each module at the desired level of detail immediately or starting from a simple version and then refining the module stepwise.

The sequence in which modules are added largely defines the sequence of potential model applications. For example, a model whose development begins with the modelling of demographics and education could initially be used for educational projections or the analysis of education finance while other modules necessary for answering other questions still have to be added. Various existing multi-purpose models (including LifePaths in Canada, or SESIM in Sweden, see Flood et al. 2012) started as models for education finance. It is in the nature of retirement-income models that pensions, and other forms of retirement income, are received late in life, thus requiring detailed knowledge (and modelling) of the entire earlier life-course.

Modules do not necessarily have to be created with all the desired detail in one step. Model implementation can follow a top-down approach, aimed at modelling simple variants of all core processes and modules, creating a simple but "complete" model first, followed by a stepwise refinement by priority. Sometimes the microsimulation implementation of an available macro model is a good starting point, as – besides being simple to implement – all parameters are typically readily available. For example, the demographic core of a model can start with a microsimulation implementation of a cohort component model, adding more refinements (e.g. adding relative risks to baseline hazards) at a later step. Platforms built on this idea include DYSEM (Moore et al. 2017) and Dynamis-Pop (Spielauer & Dupriez 2019). Following this approach might also ease the parallelisation of model development; for example, accounting frameworks can be developed independently of the complexity or realism of other modules using simpler versions as dummies. Another benefit concerns scenarios: a model following the stepwise refinement approach also easily provides user options to run alternative models of the same processes, thereby assessing how model refinements change results. How does the modelling of life expectancy differentials by education impact social insurance systems?

While stepwise model creation is practical, there needs to be an initial overall design focused on the key analytical objectives. In addition, modularity is critical for efficient and flexible model

development and expansion. Modularity requires clean and agreed interfaces between modules, including precise definitions of variables that need to be passed between or accessed by different modules. Also, the chosen programming technology can either help or hinder modularity.

microWELT is implemented applying a top-down approach; thus, we start from simple base versions of all required processes and accounting routines which can be subsequently refined and extended. microWELT is highly modular; existing modules can either be complemented by alternative or refined versions leaving the choice between modules applied in a simulation to the user or can entirely replace the base version. From a platform perspective, this approach has been proven successful. As of today, applications based on the microWELT platform have introduced alternative models for education (e.g. depicting the detailed Austrian school system and making school careers dependable on factors like immigration status and place of origin, see Horvath et al. 2020, or changing the number of levels, see REF Bertelsmann), immigration (immigration pools and place of origin), and partner matching (adding place of origin). Also, new modules addressing health, health costs, labour force participation, and retirement have been developed and added.

3.8 In which programming language is the model implemented?

Dynamic microsimulation models can be implemented in three ways: using general-purpose programming languages like C++ or Python; using (and extending) statistical packages (e.g., R) in combination with other tools; and using specialised microsimulation packages like LIAM, JAS-mine or Modgen/OpenM++. All three approaches can be observed, and reinventing the wheel is still a common practice in dynamic microsimulation model implementation. Model design, especially the time concept to be used in the model, are important considerations in modelling technology decisions. Having opted for a continuous-time model and trying to avoid reinventing the wheel by using a well-tested and established programming technology, we opted for Modgen for implementing microWELT. To our knowledge, the only other product to date handling continuous time models within a comprehensive microsimulation technology (rather than combining a multitude of tools) is JAS-mine (Richiardi and Richardson 2017). Our choice for Modgen was also based on over a decade of experience in model development using this technology, and the existence of a family of other models following similar approaches. microWELT specifically re-uses and shares code with the DYNAMIS-POP modelling platform for socio-demographic and health applications in developing countries (Spielauer & Dupriez 2019) and DypenSI, a Slovenian Pension Microsimulation model (Kump et al. 2017).

Modgen is a generic microsimulation technology and language developed and maintained at Statistics Canada over more than two decades. A recent development is OpenM++, a platform-independent open-source implementation of the Modgen language, with (optional) extensions and capabilities. As models can be implemented in a cross-compatible way

(simultaneously compiled in both Modgen and OpenM++) in Windows, we do not treat them as different products in this description.

Modgen is generic microsimulation programming languages that support the creation, maintenance, and documentation of dynamic microsimulation models. Virtually all types of dynamic socio-economic and socio-demographic microsimulation models can be accommodated, from small and specialised to large and multi-purpose, in continuous or discrete time, with interacting or non-interacting populations. A key strength – compared to other modelling technologies – is the handling of continuous-time models. As a compiled language, Modgen/OpenM++ is also very fast, so large-scale models can be run on standard desktops as well as on networks. Like statistical packages, Modgen/OpenM++ does not require the use of dedicated programmers but is typically used directly by researchers/analysts. Enabling a researcher to cohesively design, implement and validate his or her model development in this manner creates substantial efficiencies. This is possible because Modgen/OpenM++ simplifies coding, hides underlying mechanisms, e.g. event queuing, and creates a stand-alone model-executable program with a complete visual interface and detailed model documentation. Tabulation is done on the fly in continuous time and includes a mechanism for estimating the Monte Carlo variation for any cell of any table. OpenM++ has additional support for uncertainty analysis and cloud computing.

4. What makes a good microsimulation model?

Design and architectural choices have to be informed by the goals of a model, the available data, and available staff and technical resources. In this section, we try to provide and structure a list of criteria – a wish-list – frequently associated with a good dynamic microsimulation model. Most of the criteria will apply to any type of model. Some are specifically important, hard to achieve, or easy to miss in dynamic microsimulation, and others – at least at first sight – go against the criteria often used for other types of models.

4.1 Clarity & transparency: clear model objectives

Clarity and transparency are design goals for any model claiming policy relevance. Clarity starts from clear objectives, the identification of the essential processes to be modelled, identification of options, and transparent decisions – as well as a developmental strategy – based on priorities. Transparency requires that all steps are traceable and documented. Such documentation includes (1) the documentation of the goals and requirements of the model – specifications – (2) the documentation of options and decisions – option analysis – and (3) the documentation of a development strategy. This documentation covers and reflects most architectural and design decisions outlined above and provides a base for project development and associated planning. Transparency also includes the thorough documentation of all statistical analysis and of the technical implementation of the model. It starts from (4) a documentation

plan. Various parts and aspects of documentation can be automated and/or integrated into the workflow of model estimation and implementation. For example, Modgen models create their own encyclopaedic documentation in the form of a hypertext file. Also, statistical packages (like R with R Markdown) support integrated documentation, and tools exist for automated website creation based on such documentation (e.g. the open-source tool Sphinx (www.sphinx-doc.org) initially developed for the documentation of Python; the latter has been used by the World Bank to document microsimulation models (Spielauer & Dupriez 2019). A practice which supports clarity and transparency is mechanical reproducibility: the ability to mechanically reproduce all estimated or tabulated input parameters. This principle ensures that all aspects of the model are explicitly specified. It also supports the goal of maintainability (see below), because there is always a method to reproduce the model and its parameters.

A clear presentation of the objectives, essential processes, design decisions and development strategy of microWELT is the very purpose of this paper. The microWELT project website contains thorough documentation of all statistical analysis and of the technical implementation of the model. Both the analysis code (Stata and R scripts) and the implementation code (Modgen) are self-documenting, allowing the automated creation of their web-documentation. Parameter generation is automated running the corresponding scripts and does not contain any manual reformatting copy-past steps ensuring a straight-forward mechanical reproducibility of all analysis and parameter generation. This also ensures that all aspects of the model are explicitly specified.

4.2 Feasibility: the model can be built with given resources

The potential to add more detail and sophistication to a model will always be a temptation to model developers – and historically, this temptation has resulted in many over-ambitious failures. Over-ambitious failures can have dramatic effects, putting a whole project in danger. This was experienced in Australia in the context of the DYNAMOD model, where an "overambitious development schedule resulted in faltering progress and culminated in a major break in the project when in late 1994/early 1995 the entire DYNAMOD team of researchers resigned" (Casells et al. 2006). Less dramatic – but common – are delays in model development, or the necessity to rethink and simplify models while in development. Many past and present large microsimulation models either never reached their initially planned size or expanded well beyond it; the struggle between ambition and feasibility is a common experience in microsimulation model development.

We address these issues by applying a top-down approach ensuring the termination of a complete model first and further developing the model by concrete demands in parallel to its specific applications within the WELTRANSIM project framework. We immediately found microWELT in high demand for other applications in various research projects which we addressed by

providing a highly modular platform rather than trying to incorporate all additional features into one single model.

4.3 The right degree of complexity

Models can reach a level of complexity where it is increasingly hard or even impossible to manage them. Decisions on the appropriate level of complexity are not trivial in microsimulation projects as one of the strengths of microsimulation is that it can handle detail as well as the complex interactions between processes – requirements for answering many of the questions that such models are built for. Some general principles relevant to many models cannot be directly applied to microsimulation. A prominent example is Occam's razor, which, when applied to modelling, states that given roughly equal predictive power, the simplest model is the best. But many microsimulation models are not built for predictive power alone (e.g. to make statements on the sustainability of a system), but also for distributional analysis (e.g. income adequacy) and thus require rich detail. Complexity has many dimensions. One dimension is the number and range of different modelling approaches, statistical methods, etc. used. From the perspective of model architecture, we must distinguish between the range of approaches an architecture supports (the fewer restrictions, the better) and the range of approaches used in the concrete application (and their consistency). An important dimension is the ease with which a model can be understood and communicated. Are the model and simulation results explainable, is it easy to understand the factors driving results? Model complexity impacts the sustainability of a model, impacting the ease with which it can be updated and maintained. The management of complexity is supported by modularity and thorough documentation, and by the judicious use of supporting tools and software.

microWELT is kept "simple on purpose", as it is used as an explorative tool on a highly abstract level – stylised welfare state regimes – emphasising the importance to understand which factors drive results. From the perspective of model architecture, we clearly distinguish between the range of approaches the architecture supports (which is broad as we envision microWELT as a flexible, versatile platform) and the range of approaches used in the concrete application which is kept narrow. microWELT focuses on future cohorts and does not reconstruct the past. microWELT is highly modular and fully documented and uses self-documentation of code and tools for an automated website generation.

4.4 Accessibility: the model can be used

Many design decisions will depend on the question of who is eventually intended to use the model. Accessibility depends on:

- Confidentiality: A model based on confidential data restricts both the range of potential users and the location of the model's use

- Required hardware – Required background and skills to understand and use the model, create scenarios, produce, tabulate and interpret the output.
- User Interface
- Documentation and user guides, available training resources
- The possibility to modify and extend the code, add modules, etc.
- Licensing – Pricing

microWELT is an open-source model, and all its components are freely available. Users need access rights to the Euromod SILC database – the starting populations of the model can be generated from this database using scripts provided. microWELT is a Windows application and runs on regular Windows PCs⁴. The model has a graphical user interface with its own help system. All model parameters can be edited within the graphical user interface. Simulation results are presented in an extensive set of output tables contained in the GUI. Tables include various views for not only displaying simulation results but also distributional measures like standard variations and the coefficient of variation of each single table cell for assessing Monte Carlo variation. microWELT also comes with a brief user guide accessible from the project website. The implementation of microWELT is fully documented in a step-by-step implementation guide which can serve as a training resource of microsimulation programming using Modgen/openM++. This allows developers to modify, refine and extend microWELT and/or use it as a platform for other applications.

4.5 Usefulness: the model answers the questions it is built for

A model is useful if it can answer the questions it is designed for with reasonable effort on the part of users. The usefulness of a model thus depends on the range of questions it can address, the quality of the answers, and how quickly and easily answers can be obtained, validated, understood and communicated. From a design perspective, the challenge lies in answering different types of questions or needs simultaneously, where each on its own would be best addressed with a different architectural approach.

Typical conflicting demands are the requirement to meet external aggregate targets (i.e. scenarios requiring alignment or calibration) and the need to produce accurate, reliable results for specific/small population groups (which requires simulating large population samples). Meeting external targets can be achieved by alignment and calibration. Alignment can address the need to meet external aggregate targets in a single model run but requires a time-based architecture which limits the size of a simulation, and, therefore, the ability to produce accurate, reliable results for rarer population subgroups. In contrast, a case-based model has

⁴ Recommended are 16-32GB RAM.

no size limitations, but meeting external targets requires multiple simulation runs, each modifying parameters until targets are met. This can be cumbersome or might be infeasible in practice.

In the long run, the usefulness of a model will also depend on how flexibly it can be adapted and extended to new questions.

Being build alongside a concrete multi-country comparative research project, the development of microWELT directly responds to project demands. To enhance practical usefulness, also post-processing of simulation results, including visualisations and scenario/country comparisons are supported by a set of automated tools.

microWELT implements mechanisms which make it easy to meet external aggregate targets if required. As a time-based model, alignment is automated within a single model run, thus not requiring tedious calibrations re-running the model based on iteratively modified parameters. Built as a flexible modelling platform, we aim at promoting its usefulness beyond its current application.

4.6 Reliability: the model can be trusted

The extent to which a model can be trusted depends on its specification, internal validity (being free of computer bugs), accurate parameterisation, how well results can be understood, and – empirically – how well simulations perform compared to reality.

Specification refers to the statistical modelling, the choice of data sources, selection of variables, and ways to model processes, i.e. items contributing to the prediction power of the model. But even a perfectly specified model can be unreliable if its code is not bug-free, it is too complex to be maintained, or if its parameters are not (or cannot be) accurately specified. Debugging can be a time-consuming process, and quite frequently bugs do not show up immediately. A successful strategy for avoiding and catching bugs early-on requires planning, organisation and discipline. Code must be tested at every stage of development. This process is supported by modern programming environments. In addition, bugs can be avoided by the use of well-tested specialised microsimulation programming languages which allow developers to re-use readily available components that have been well-tested in other applications.

Parameter uncertainty can be addressed by sensitivity analysis, i.e., a controlled way of assessing the impact of changing parameters on the simulation output. Modifying all parameters in a simultaneous way (re-running a model with parameters changed according to their statistical confidence intervals and correlations) can produce confidence-intervals of model output. For model validation (and a better understanding of underlying mechanisms driving results), studying the impact of changes in single parameters is essential. As sensitivity analysis is time-consuming, the use of (or investment in) technologies allowing automation is advised. By

its nature, sensitivity analysis cannot assess the accuracy of parameter estimates but can make statements on the robustness of simulation outputs, e.g. by specifying a band of parameter values for obtaining specific (narrow ranges of) results. Given historical computational constraints, sensitivity analysis and the study of parameter uncertainty was rather limited in the past but can be expected to become standard with the disappearance of computational bottlenecks.

In the case of microWELT, internal validity is supported in various ways. Being implemented in Modgen, it uses a well-tested specialised microsimulation programming language avoiding having to reinvent the wheel. The simulation engine and some modules share code that has been well-tested in other applications. In particular, being built on the experience with other microsimulation models, some modules were fed back into existing models replacing existing code by more efficient and/or flexible implementations which allowed testing microWELT components in other applications by comparing results. For example, the simulation engine handling the generation of a starting population of user-specified size by sampling and cloning from weighted records is shared with Dynamis-Pop (Spielauer & Dupriez 2019) and DypenSI (Kumpt et al. 2017).

Given the comparative perspective of the WELTRANSIM model, statistical analysis results and resulting parameters are explicitly studied and presented (1) by themselves for identifying welfare-regime specific socio-demographic patterns and patterns contained in NTA/NTTA age profiles and (2) by identifying their long-term consequences assessed by simulation projections. While the study of parameter uncertainty within the WELTRANSIM project is limited to sensitivity analysis, microWELT is built on a programming technology supporting the automation of multiple model runs modifying all parameters in a simultaneous way for producing confidence-intervals of model output.

4.7 Maintainability: the model can be updated and kept alive

Given the complexity of large-scale microsimulation models, updates and maintenance need careful planning; the question of how a model will be maintained and updated should be an integral aspect of the development plan for each of its components and modules.

Over time, new data become available, and the scope of available data might change, requiring flexibility concerning the use of new data for the modelling of processes. Models which begin from a starting population file will typically require periodic re-basements whenever a new wave of their key data-source becomes available. Planning model maintenance and update cycles in advance can save considerable resources. The ease with which a model can be maintained will depend on various factors including the model code (e.g. avoiding hard-coded dates and supporting the easy shift of time-ranges), the re-usability and adaptability of statistical analysis code, the degree of automation of parameter generation, the availability of

powerful tools for model validation, as well as efficient (self) documentation and version control. Updating individual modules will typically change projection results. To the extent that the model is supposed to meet given aggregate targets, this might require complex calibrations. In such cases, it might be more efficient to perform a full cycle of all module updates before re-calibrating the model. In the same way, external scenarios – like updates in Eurostat projections – typically come on a given schedule, which should inform a model's maintenance plan.

Potential updates of microWELT are supported by an easy mechanism of shifting the starting year – all related code is contained in a single context module – and the organisation of the parameter generation in self-documenting scripts which can be easily be adapted to more recent data sources.

4.8 Extendable: the model can grow and be refined

Given the variety of processes which must be modelled in most dynamic microsimulation applications, such models lend themselves to be developed into multi-purpose models. The range of applications that could be added to the model and the ease in which model extensions could be implemented should be considered when making model design choices. The range of applications will depend on the time concept used. For example, many applications in health and demography, as well as policy applications requiring the implementation of sub-annual processes (e.g. unemployment, maternity, hospital stays) will profit from a continuous-time model. Other applications might require a time-based approach, as they model communication and linkages between people (e.g. kinship networks, the spread of diseases, or social networks). Making a model useable and extendable beyond its prime focus also broadens the potential for collaboration and for a user community which supports model validation and debugging. Also, some extensions might feedback into the original application; a design which readily supports improvements and refinements will also allow the original application itself to adapt more easily to new policy demands. However, care should also be taken to decide whether extensions should become core parts of the model, and thus increase future maintenance, or be kept as one-off extensions.

microWELT is designed both as a specific application and as a flexible modelling platform. We have decided to make extensions which have been developed in other applications based on microWELT but not being required within the WELTRANSIM project not to become core parts of the model, keeping the future maintenance of microWELT simple.

4.9 Efficiency: no waste of time and resources

A still common practice in microsimulation is reinventing the wheel by implementing models from scratch using general-purpose programming languages. Using a dedicated microsimulation language allows focusing on the model itself, as many software components are provided

automatically, and programming is supported by many microsimulation-specific functions and concepts.

A considerable share of development time is consumed by data manipulation and statistical analysis for parameter estimation and generation. Well-organised and documented statistical code producing parameters and parameter files in their required format can avoid errors and waste of time, e.g. avoiding reformatting statistical output. Including visualisations of results in the statistical scripts (rather than producing it manually, e.g. in Excel) supports an automated workflow. The same holds true for analysis scripts for the post-processing, validation, and visualisation of simulation results and the comparison of scenarios. Maximum automation of the workflow from statistical analysis, parameterisation, model execution and post-processing makes results not only reproducible but also avoids many time-consuming, repetitive manual steps.

A common time-killer in dynamic microsimulation project development that can be avoided by automation is model documentation, i.e. by using self-documenting programming technologies and self-documentation systems of statistical analysis code.

As in all software projects, appropriate version control can avoid time-consuming and error-prone manual gate-keeping of code versions and support collaboration. Version control is a system that records changes to all relevant project files – together with the information regarding who performed which changes – over time. This allows going back to earlier versions of code or reverting selected files back to a previous state. It also allows branching project development and supports merging different branches back to a single version.

microWELT is implemented in Modgen, and we make extensive use of self-documenting systems, including website creation.

5. Characteristics and modules of the microWELT model

5.1 Data requirements

The starting population is generated from EUROMOD input data, and various parameters are estimated directly from this data set. Thus, it can directly correspond with EUROMOD (Sutherland and Figari, 2013) which is used for complementing microWELT for detailed distributional analysis within the studied population groups (by age, education and family) today, and to study the sustainability of the current system at future points. While microWELT does not simulate all necessary variables to run EUROMOD, its projections stemming from dynamic ageing are used for reweighting the EUROMOD database (static ageing). We use EUROMOD also to identify patterns of policy changes and mechanisms to balance budgets typical for the various welfare state regimes (Fink et al. 2020), mechanisms which – at a highly abstract level – feed back into microWELT. Another micro-data source used in the project are Labour Force Surveys. Demographic scenarios and corresponding mortality, fertility, and migration parameters come

from Eurostat population projections complemented by own estimations and projections found in the literature. Full documentation of the data steps regarding the modelling of socio-demographic processes – data sources, statistical models, current patterns and projection results from a comparative welfare state perspective – are available on the project website www.microWELT.eu. The model is designed to be portable to other countries and to a large extent parameter generation is automated.

The model allows the parallel parameterisation of NTA variables by three levels of disaggregation: by age, by age and sex, and by age, sex, education and family type. Data further disaggregated by education – and by family type – are calculated alongside the WELTRANSIM project for Spain, Austria, the UK, and Finland. For a growing set of countries, NTA data by education are also calculated and used alongside other research projects (e.g. Renteria et al., 2016 and Hammer 2015).

5.2 Fertility

microWELT uses age-specific fertility rates from Eurostat population projections. While reproducing these macro projections (which account for differences in total fertility and overall age patterns), microWELT aims at obtaining a more realistic distribution of family sizes by education. This is done by a fertility module which models individual birth risks by individual-level characteristics like education and parity. Empirically, the distribution of family sizes in most European countries followed a trajectory from high concentration at the beginning of the past century, to an equal distribution in the baby boom, back to an increasing concentration, especially higher educated women having increasing rates of childlessness (e.g. Spielauer 2005, Shkolnikov et al. 2007). These patterns vary between welfare state regimes, e.g. fewer women having a larger share of the children is typically associated with conservative regimes. We simulate birth events by using age-specific fertility rates to produce the desired number of births but assign the events to the woman with the shortest random waiting time. Currently, we concentrate on first birth patterns and levels of childlessness by education. The model uses current first birth rates by education calibrated to reproduce cohort childlessness by education as projected in the literature (e.g. Reher, & Requena 2019). After first birth rates are met, the remaining births are distributed to mothers by age-specific rates.

5.3 Mortality

The ability to quantify the impact of differential longevity by education on redistributive processes is one of the aims of the microWELT model. We assume that the relative mortality risks by education stay constant over time, while total mortality outcomes are aligned to reproduce Eurostat mortality projections. Model parameters are period life tables by sex and estimates of the current remaining life expectancy by education at age 25 and age 65 as published by OECD (Murtin et al. 2017). Relative mortality risks are calculated at the start of the simulation

and kept over time, while the age-specific baseline risks are aligned each year to reproduce the target mortality.

5.4 Education

microWELT distinguishes three education levels – low, medium, and high – corresponding to compulsory education, secondary education, and post-secondary education attainments. Highest education eventually obtained is decided at birth by selecting one of two modelling approaches – or by combining them: outcome-parameters for projected age and cohort-specific distributions, and parameters for the distributions by parent's education. When selecting the model simulating the intergenerational transmission of education, users can choose to align the aggregate outcomes to the external targets. In this case, the odds ratios between groups from the intergenerational transmission models are used to select students based on their parents' education, while the number of graduates is calculated from the outcome parameters. Accounting for parents' education is important when modelling NTAs by education, as the transfer a child receives (as well as education choices) depend on the education of parents. Besides the highest educational attainment, microWELT also models study patterns, i.e. school enrolment. Again, this is important from the NTA perspective, as we disaggregate NTAs by enrolment status.

5.5 Partnerships

microWELT models the female partnership status according to observed partnership patterns by age, education, and the presence and age of children. Appropriate partners are matched by age, education, and childlessness. The model assumes that *cet. par.* the probability of being in a partnership does not change over time; thus, all changes come from composition effects. Besides changes due to the death of a partner, updates are performed at a yearly basis to maintain cross-sectional consistency. Under the assumption of time-invariant patterns, the model is longitudinally consistent by education, childlessness and birth cohort – thus allows the calculation of consistent life-course measures by these groups – but (currently) does not model consistent individual life-courses within these groups. The partnership status is modelled for all women within the age range 15-80, no more union formation events are modelled thereafter when it is assumed the only union dissolution is due to widowhood.

Male partners are matched by age and education. Men destined to stay childless (a model parameter by education and cohort) avoid unions with mothers unless no other men are found; if in a union at a birth of a child, they pass on their "never father" status to another (childless) man of the same age and education to meet overall childlessness rates. Concerning age differences between partners, the model tries to fit observed distributions by age. Empirically, the spread increases with age. Part of this pattern arises from re-partnering, the distribution of age differences, thus differing for new partnerships compared to all observed partnerships. As the

former typically cannot be observed in data, at each partnership event, the current age distribution in the simulation is compared with a target distribution and partners are picked to best close the gaps between the two distributions.

5.6 Migration

The migration modules of microWELT allow reproducing aggregate projections of immigration by number, age distribution, and sex – and emigration rates by age and sex. The study of the effect of migration is not a priority of WELTRANSIM, but we added this module in order to allow reproducing Eurostat population projections. Immigrants are produced from scratch and clone their characteristics from persons of the same age and sex in the host population. Emigrants are removed from the simulation. From a platform perspective, microWELT does not restrict how migration is modelled. For example, the Austrian microDEMS model built on this platform implements migration pools by origin and status (see WP1 Brief 4). Migration can be switched off.

5.7 Family Links

microWELT models nuclear families. At birth, children are linked to their mother and – if present – their father. This link to biological parents is maintained within the simulation. As the partnership status is only updated on a yearly basis, we assume that if a single mother enters a partnership within the first year of her child, the partner is also the biological partner. Until leaving home, children also keep links to their current social parents: in the case of a union dissolution of parents, children choose with whom to live, and if this parent enters a new partnership, a link to this new partner is established. In the case of the death of a single social parent, children move back to a biological parent if available, or to grandparents. Children leave home when entering a partnership, becoming parent, or at age 18, if not enrolled in school. Students stay at home up to age 25, the probability for leaving home being modelled.

5.8 NTA Variables and Accounts

NTA data are cross-sectional age profiles breaking down national accounting variables on consumption, income, saving, and transfers. Data distinguish between private and public consumption singling out education and health. Public transfers also distinguish pension benefits. NTA data by sex are available for 50 countries, including most European countries: The AGENTA project produced NTA estimates using comparable European data. We further dis-aggregated NTA data by education and family type. microWELT allows the user to choose the aggregation level, allowing to compare simulation results based on aggregated profiles with simulations accounting for composition effects along the education and family dimensions. At their most dis-aggregated level, NTA variables in microWELT are parameterised by the following population groups:

- Children age 0-16 and students age 17-25: by highest parent's education

- Non-students age 17-60: by sex, education, presence of partner and children in the household
- Persons 60+ by sex, education, partnership status, and parental status/childlessness

NTA data provide a detailed picture, how resources are re-distributed by means of public and private transfers and asset re-allocation. microWELT implements a set of 19 NTA variables.

Consumption components:

- Private Consumption Education (CFE)
- Private Consumption Health (CFH)
- Private Consumption other than Education and Health (CFX)
- Public Consumption Education (CGE)
- Public Consumption Health (CGH)
- Public Consumption other than Education and Health (CGX)

Public transfers:

- Public Transfers Pensions, Inflows (TGSOAI)
- Public Transfers Other Cash Inflows (TGXCI)
- Public Transfers Other In-Kind Inflows (TGXII)
- Public Transfers Education Inflows (TGEI)
- Public Transfers Health Inflows (TGHI)
- Public Transfers Outflows (TGO)

Private transfers:

- Net Interhousehold Transfers (TFB)
- Net Intrahousehold Transfers (TFW)

Saving:

- Private Saving (SF)
- Public Saving (SG)

Income components:

- Labour Income (LY)
- Private Asset Income (YAF)
- Public Asset Income (YAG)

microWELT implements individual accounts for all variables as well as mechanisms to sum up over the life-course with and without applying discount factors. Accounts are updated in continuous time; thus, any change in characteristics affecting NTAs like a change of partnership status, leaving school, or children are captured immediately. NTA based indicators and accounting approaches are implemented in separate (and optional) modules, some directly corresponding to publications introducing the concepts.

5.9 NTA Indicators: Support Ratio, Impact Index

Based on NTA variables, a series of indicators was suggested in the literature. Two of these indicators – the Support Ratio and the Impact Index – are implemented in the base version of microWELT and others can be added. As of today, published projections of indicators rely on aggregate data (NTA by age and sex only). microWELT can be used to study the effect of accounting for the changing educational composition of the population and changes in family patterns adding realism to existing projections.

One of the most widely used indicators in NTA literature is the Support Ratio, i.e. the number of effective producers per effective consumer determined by the population age distribution and the age profiles of per capita consumption and labour income as observed today (Lee & Mason 2014). It refines the simple demographic dependency ratio by accounting for the age profiles of consumption and labour. As an index set to 1 in the base year, the measure shows the change in the relationship between available labour to the current level of consumption in the absence of economic growth or changes in the age profiles of consumption and labour.

The Impact Index proposed by Lee & Mason (2017) further accounts for changes in wages resulting from the changing labour supply in relation to capital due to demographic change. It measures the change in the relationship between consumption and the current consumption level. As in ageing societies – based on aggregated NTA data – labour becomes scarcer, wages increase, softening the consequence of ageing on future consumption. Calculation of the Impact Index requires a simple economic growth model based on a Cobb Douglas production function without productivity growth. Initial capital by age is calculated indirectly from the age shape of capital income, assuming an initial interest rate. The resulting capital endorsement by age is assumed to stay constant, with wages and interest rates adapting to population change (when assuming a closed economy; the Impact Index can also be calculated for open economies).

5.10 Longitudinal NTA accounting: NPV of transfers

A typical approach for studying the intergenerational dimension of transfers is by calculating their Net Present Value (NPV) for distinguished cohorts. Such an analysis was done by Lee et al. (2007), which we use as a template – further disaggregating by education and childlessness/parenthood. The approach requires an assumption of economic growth and on an

appropriate discount factor. NPV can be calculated with or without adjustments of taxes and benefits to balance budgets, thereby assessing the effect of population ageing. Balancing budgets requires assumptions on how this is done, e.g. selecting which benefits to decrease, and which taxes to increase by which amount. Lee et al. (2017) assume a very simple symmetric adjustment of transfers for balancing budgets each year; i.e., taxes are increased by the same extent as benefits are decreased. microWELT implements the same approach but allows going beyond these simplistic assumptions. Even in the simple case, microWELT allows adding (and testing) realism of this approach accounting for educational changes and exploring the NPV by education as well as how much different education groups are hit by the adjustments. microWELT also allows studying the effect of mortality differentials by education which might require a higher adjustment of public transfers (negatively affecting all education groups).

5.11 NTA Variables and Accounts

National time transfer accounts (NTAs) measure the provision and consumption of household production (which refers to services which could in principle also be provided by third persons). Captured are inter-household transfers distinguished in child-care and other services and intra-household transfers. As indicated by its name, time transfers measure household production in units of time. These times can be monetised by applying appropriate wage rates⁵. Like for NTAs, microWELT implements NTAs by age, sex, education and family type – and allows for both cross-sectional and longitudinal accounting.

5.12 Model Output

microWELT produces rich output supporting the detailed analysis of simulation results. Three types of output can be distinguished:

- A series of **tables** organised in several groups. The list of tables can be easily extended. Tables can be of any number of dimensions and are part of the user interface. All tables selected by the user to be part of the simulation output are automatically created with each simulation run. Tables have various views which can be accessed by right-clicking on them. Most important of these additional views complementing the output values is the output of the coefficient of variation of each table cell (which is calculated automatically when running various replicates of the model) allowing to assess the Monte-Carlo error of a simulation. All table output is stored together with the parameter tables and settings thus can be retrieved when opening a simulation scenario. Tables can be copy-pasted to Excel or exported all together as an Excel workbook. Excel output is

⁵ For details on measurement and pricing see Vargha (2017).

typically used for graphical analysis of results, including the comparison of alternative scenarios.

- A micro-data file with selected variables. This file is typically used for further statistical analysis using statistical software packages. The output of a micro-data file is optional, and users have control of the timing of the output which can be a single point in time or recurrent output events at user-defined time intervals. The output file comes in csv format and contains a header row with variable names.
- A tracking database of a sample of individual life course careers. The tracking database is a database of individual life courses which can be displayed by the BioBrowser tool, a software tool provided freely with Modgen. The main application of tracking individual life-courses is for model validation and debugging.

6. Summary and Additional Resources

This paper discussed the purpose and design of the microWELT model. microWELT is built to study the interactions between welfare state regimes, welfare transfers, and population ageing accounting for educational change, life expectancy differentials by education, and changing family patterns. microWELT simultaneously depicts and projects key socio-demographic characteristics of the studied populations, as well as transfer flows as captured by the NTA and NTTA accounting framework. The study compares four European countries – Austria, the UK, Finland, and Spain – representing four welfare state regimes: Corporatist-Statist, Liberal, Social Democratic, and Familial / Mediterranean. microWELT captures key features of these welfare state regimes expressed in socio-demographic patterns and processes, transfer flows, and the adaptation mechanisms addressing sustainability issues resulting from population ageing. Individual accounts allow for the longitudinal analysis of transfer flows between population groups and generations.

This paper is part of a series of related papers and other resources which together build comprehensive documentation and presentation of the research performed developing and using microWELT. All materials are available at the project website www.microWELT.eu. One of the objectives of microWELT is the provision of a modelling platform available for applications beyond the WELTRANSIM project. A collection of project descriptions and links to these projects is available on the project website. microWELT is an open-source project: the application, the model code, a step-by-step implementation guide as well as analysis scripts for parameter generation are available for download.

References

- Bissonnette, Luc, Boisclair, David, Clavet, Nicholas-James, Lacroix, Guy, Marchand, Steeve, Michaud, Pierre-Carl (2016a), SIMUL, A Demographic and Economic Microsimulation Model for Quebec. http://www.cedia.ca/sites/cedia.ca/files/dt8_simul_anglais_oct2016.pdf [retrieved 2020-09-17].
- Caron-Malenfant, Eric, Coulombe, Simon (2015), Demosim: An Overview of Methods and Data Sources. Statistics Canada Catalogue no. 91-621-X. <https://www150.statcan.gc.ca/n1/pub/91-621-x/91-621-x2015001-eng.pdf> [retrieved 2020-09-17].
- Cassells, Rebecca, Harding, Ann, Kelly, Simon (2006), Problems and Prospects for Dynamic Microsimulation: A review and lessons for APPSIM. NATSEM Working Paper Series 63, University of Canberra, National Centre for Social and Economic Modelling. <https://ideas.repec.org/p/cba/wpaper/63.html> [retrieved 2020-09-17].
- de Menten, Gaëtan, Dekkers, Gijs, Bryon, Geert, Liégeois, Philippe, O'Donoghue, Cathal (2014), "LIAM2: a New Open Source Development Tool for Discrete-Time Dynamic Microsimulation Models". In: Journal of Artificial Societies and Social Simulation 17 (3) 9 <http://jasss.soc.surrey.ac.uk/17/3/9.html>
- Dekkers, Gijs, Belloni, Michele (2009), Micro simulation, pension adequacy and the dynamic model MIDAS: an introduction. https://www.researchgate.net/publication/265069962_Micro_simulation_pension_adequacy_and_the_dynamic_model_MIDAS_an_introduction [retrieved 2020-09-17].
- Fink, Marian, Horvath, Thomas, Spielauer, Martin (2020), microDEMS – Ein dynamisches Mikrosimulationsmodell für Österreich. Illustration am Beispiel der Entwicklung der Erwerbsbeteiligung bis 2040 (microDEMS – A Dynamic Microsimulation Model for Austria. Illustration Using the Example of the Development of Labour Force Participation Until 2040). WIFO-Monatsberichte, Austrian Institute of Economic Research 2020, 93(1), 51-61 https://www.wifo.ac.at/publikationen/publikationssuche?detail-view=yes&publikation_id=62442 [retrieved 2020-09-17].
- Flood, Lennart, Jansson, Fredrik, Pettersson, Thomas, Pettersson, Tomas, Sundberg Olle, Westberg, Anna (2012), SESIM III - A Swedish dynamic micro simulation model. (Handbook) <http://www.sesim.org/Documents/Handbook%20of%20SESIM.pdf>
- Galler, H.P. (1997), Discrete-Time and Continuous-Time Approaches to Dynamic Microsimulation Reconsidered. Technical Paper 13. National Centre for Social and Economic Modelling (NATSEM), University of Canberra

<https://pdfs.semanticscholar.org/c243/ff74973c2069163f6902ea98c4bb6a4b5111.pdf>
[retrieved 2020-09-17].

Hammer, B. (2015), National Transfer Accounts by Education: Austria 2010. AGENTA Working Paper 2/2015.

Immervoll, Herwig, Lindström, Klas, Mustonen, Esko, Riihelä, Marja, Viitamäki, Heikki (2005), Static Data Ageing Techniques. Accounting for Population Changes in Tax-Benefit Microsimulation Models. EUROMOD Working Paper No. EM7/05, <https://www.euromod.ac.uk/sites/default/files/working-papers/em7-05.pdf> [retrieved 2020-09-17].

King, Anthony, Baekgaard, Hans, Robinson, Martin (1999), "DYNAMOD-2: An Overview". Technical Paper No. 19. National Centre for Social and Economic Modelling (NATSEM), University of Canberra. <https://www.natsem.canberra.edu.au/storage/tp19.pdf> [retrieved 2020-09-17].

Kump, Nataša, Majcen, Boris, Sambt, Jože, Lotric Dolinar, Aleša, Spielauer, Martin, Verbic, Miroslav, Spruk Rok (2017), Dinamici Mikrosimulacijski Pekojninski Model. Monograph. EkonomIERa, ISSN 2630-2896. Socialni razvoj. ISBN 978-961-6906-43-2.

Lee, Ronald, Mason, Andrew (2017), Some Economic Impacts of Changing Population Age Distributions - Capital, Labor and Transfers. Agenta Keynote, WP.

Lee, Ronald, McCarthy, David, Sefton, James, Sambt, Jože (2017), Full Generational Accounts: What Do We Give to the Next Generation?. Population & Development Review.

Mannion, Oliver, Lay-Yee, Roy, Wrapson, Wendy, Davis, Peter, Pearson, Janet (2012), JAMSIM: A microsimulation modelling policy tool. Journal of Artificial Societies and Social Simulation, 15 (1) 8. <http://jasss.soc.surrey.ac.uk/15/1/8.html> [retrieved 2020-09-17].

Moore, Kevin, Hicks, Chantal, Jones, Jenifer, Spielauer, Martin (2017), The DYSEM Microsimulation Modelling Platform. Statistics Canada, Analytical Studies 11-633-X No. 008, <https://www150.statcan.gc.ca/n1/en/pub/11-633-x/11-633-x2017008-eng.pdf?st=FUB-MXUwv> [retrieved 2020-09-17].

Morrison, Rick (1998), Overview of DYNACAN. A full-fledged Canadian actuarial stochastic model designed for the fiscal and policy analysis of social security schemes. https://www.actuaries.org/CTTEES_SOCSEC/Documents/dynacan.pdf [retrieved 2020-09-17].

Murtin, Fabrice, Mackenbach, Johan, Jasilionis, Domantas, d'Ercole, Marco Mira (2017), Inequalities in longevity by education in OECD countries: Insights from new OECD estimates. OECD Statistics Working Papers, 2017/02, OECD Publishing, Paris.

- Reher, David, Requena, Miguel (2019), Childlessness in Twentieth-Century Spain: A Cohort Analysis for Women Born 1920–1969. *European Journal of Population* volume 35, pages 133–160 (2019) <https://link.springer.com/content/pdf/10.1007/s10680-018-9471-7.pdf> [retrieved 2020-09-17].
- Rentería Elisenda, Mejía-Guevara, Iván, Patxot, Concepció, Souto, Guadalupe (2016), "The effect of education on the demographic dividend". In: *Population and Development Review*, 42, 4, 651-671.
- Richiardi, Matteo, Richardson, Ross (2017), "JAS-mine: A New Platform for Microsimulation and Agent-Based Modelling". In: *International Journal of Microsimulation*, 10(1): 106-134.
- Shkolnikov, Vladimir M., Andreev, Evgueni M., Houle, René, Vaupel, James W. (2007), "The Concentration of Reproduction in Cohorts of Women in Europe and the United States". In: *Population and Development Review* Vol. 33, No. 1 (Mar. 2007), pp. 67-99
- SOA (1997) Overview and discussion of CORSIM; in: Anderson, J.M. (2001), *Research Models for Retirement Policy Analysis, Report to the Society of Actuaries* https://www.soa.org/Files/Research/Projects/Chapter_5.pdf [retrieved 2020-09-17].
- Spielauer, Martin (2005), "Concentration of reproduction in Austria: general trends and differentials by educational attainment and urban-rural setting". In: *Vienna Yearbook of Population Research* 2005.
- Spielauer, Martin, Dupriez, Olivier (2019), *Dynamis-Pop, a Multi-Country Modular Socio-Demographic Microsimulation Model for Developing Countries*. Project Website: <http://www.dms-c.com/dynamis/index.html> [retrieved 2020-09-17].
- Spielauer, Martin (2010), "What is Social Science Microsimulation?". In: *Social Science Computer Review*, vol. 29, 1: pp. 9-20. <https://journals.sagepub.com/doi/10.1177/0894439310370085> [retrieved 2020-09-17].
- Spielauer, Martin, Hicks, Chantal, Gribble, Steve, Rowe, Geoff, Lin, Xiaofen, Moore, Kevin, Plager, Laurie, Nguyen, Huan (2013), *The LifePaths Microsimulation Model: An Overview*. Statistics Canada.
- Sutherland, Holly, Figari, Francesco. (2013), "EUROMOD: The European Union tax-benefit microsimulation model". In: *International Journal of Microsimulation*. 6. 4-26. 10.34196/ijm.00075. https://microsimulation.org/IJM/V6_1/2_IJM_6_1_Sutherland_Figari.pdf [retrieved 2020-09-17].
- Van Imhoff, Evert, Post, Wendy (1998), "Microsimulation methods for population projection". In: *Population*, 10(1), 97-136.

Vargha, Lili, Gál, Róbert Iván, Crosby-Nagy, Michelle O. (2017), Household production and consumption over the life cycle: National Time Transfer Accounts in 14 European countries. Demographic Research V36 A 32. <http://www.demographic-research.org/Volumes/Vol36/32/> [retrieved 2020-09-17].

Asghar, Zaidi, Rake, Katherine (2001), Dynamic Microsimulation Models: A Review and Some Lessons for SAGE. <https://pdfs.semanticscholar.org/7dab/b1bb07bb3885c7181e364f3f031b1280c00f.pdf> [retrieved 2020-09-17].