Tax-Benefit Modeling and Accounting in the microWELT Continuous-Time Microsimulation Platform

Thomas Horvath, Barbara Kalar, Natasa Kump, Martin Spielauer, Philipp Warum

Discussion Note – last update 15-4-2024 (response to comments by Justin van de Ven added)

Abstract

This report outlines the tax-benefit modeling and accounting approaches applied in the dynamic microsimulation model microWELT.

The tax-benefit calculation follows a synthetic database imputation approach based on Euromod. This approach supports the comparative nature of the model, i.e. its applicability to many countries without the need to code tax-transfer systems in detail. Using the Euromod hypothetical household tool, we produce a synthetic database of ~250k families across five dimensions covering population heterogeneity by income and family characteristics and composition. In the simulation, families identify a matching database record, which is used to retrieve tax and social contribution rates, as well as a set of benefits.

The MicroWELT tax-benefit and accounting approach implements the National Transfer Accounting (NTA) framework at the individual level, incomes and benefits being organized by NTA categories and assigned to individuals. MicroWELT is a continuous-time model what is also reflected in the implementation of the accounting system which supports instantaneous updates whenever relevant characteristics such as employment status, income, or family compositions changes. In addition, the approach allows adjustments and alignments on a yearly basis. Individuals keep a memory of yearly totals supporting longitudinal calculations such as the present value of lifetime tax-benefit flows.

Keywords: Dynamic microsimulation, tax-benefit systems, NTA accounting

1. Introduction

MicroWELT is a dynamic microsimulation platform developed for the comparative study of the interactions between population aging, sociodemographic change, and welfare state regimes. This report outlines the tax-benefit modeling and accounting approaches applied in microWELT. The tax-benefit calculation follows a synthetic database imputation approach based on Euromod. While all relevant income and family characteristics entering the tax-benefit calculator are dynamically modeled in microWELT, tax-rates and net benefits are retrieved from matching synthetic database records. We create such databases for 8 European countries using the Euromod hypothetical household tool. Matches between dynamically simulated families and database records are deterministic, capturing the key dimensions of population heterogeneity essential for tax-benefit calculations.

Our approach aims at implementing the National Transfer Accounting (NTA) framework, incomes and benefits being organized by NTA categories and assigned to individuals. MicroWELT is a continuous-time model what is also reflected in the implementation of the accounting system which supports instantaneous updates whenever relevant characteristics such as employment status, income, or family compositions changes. In addition, the approach allows adjustments and alignments on a yearly basis. Individuals keep a memory of yearly totals supporting longitudinal calculations such as the present value of lifetime tax-benefit flows.

After outlining the main development goals and rationale of our approach, this report covers three aspects of microWELT tax-benefit modeling, namely (1) the individual and family characteristics and income components modeled within microWELT which provide the inputs of the tax-benefit calculator, (2) the synthetic database imputation approach for retrieving taxes and benefits, and (3) the continuous-time accounting framework developed for tracking and processing individual-level tax- and benefit outcomes.

2. Goals and Rationale

MicroWELT is designed for comparative analysis of the operations of welfare states in the context of population ageing. Concerning tax-benefit accounting, it aims at bringing the National Transfer Accounting (NTA) and National Time Transfer Accounting (NTTA) frameworks down to the individual level, allowing the study of the sustainability and redistributive operations of welfare state regimes across population groups and cohorts.

Given the high complexity and variety of tax-benefit codes across countries, a key development goal is the feasibility of implementation in a comparative setting. Rather than coding tax-benefit systems within the dynamic microsimulation model, we aim at drawing on existing static models, in particular on Euromod. A promising approach explored by Van de Ven et.al. (2022) is the dynamic simulation of taxes and welfare benefits by database imputation. While in their approach, simulated families search for the closest neighbor within an existing large-scale UK tax-benefit database, we modify this approach by creating a synthetic database of families constructed for covering all combinations of relevant characteristics in some key dimensions. In the context of comparative analysis, this approach addresses the following limitations, or has the following advantages, respectively:

- Sample size: overcoming the small sample size of the Euromod database in some countries
- (Non-)confidentiality of data
- Speed: Unique matches by defined combinations of categories allow direct addressing of data records avoiding slow database searches.
- Abstraction supporting comparative systems analysis by direct visualization of taxes and benefits alongside the dimensions applied for database creation.
- Direct implementation as multi-dimensional parameter tables within the Modgen / openM++ programming environment

The main disadvantage is the higher complexity of model validation, as results cannot be directly compared between the direct use of the tax-benefit calculator and the database imputation approach based on the original population sample. We find this disadvantage minor given the supposed overall needs of model alignments and adaptations stemming from another design goal, namely the consistency of tax-benefit with the NTA framework:

- Categorization of benefits by NTA categories which is not necessarily intrinsic to tax-benefit calculators: Education, Health, Old-age, Disability, Survivor, Unemployment, Childbirth related, Family
- Tax-benefit accounts summing up to National Accounts / NTA totals.
- Assignment of all incomes, taxes, and benefits to individuals rather than households.

A second set of goals and requirements stem from the continuous-time modeling approach followed in the microWELT model. This requires the implementation of an accounting framework which provides:

- Support of instantaneous updates whenever relevant characteristics such as employment status, income, or family compositions changes.
- Support for periodic (annual) adjustments and alignments.
- Storage of annual totals supporting longitudinal calculations such as the present value of life-time tax-benefit flows.

3. Modeled Inputs to the Tax-Benefit Calculator

Combining the longitudinal modeling of (1) individual employment, earning, and various components of income careers such as pensions and asset income with (2) tax-benefit calculations by database imputation requires a clear concept of what to be modeled within the dynamic microsimulation, and what to be retrieved from a tax-benefit database. We currently explore the feasibility of the following distinction:

Inputs to the tax-benefit database:

- Gross "labor-related" incomes comprised of earnings, and a collection of benefits which can be thought of earning replacements (typically contributory benefits for non-working periods implemented by social insurance systems): old-age and disability pensions, maternity and parental leave benefits, and unemployment benefits. These income components are to be modeled longitudinally and are a key input for the tax-benefit imputation.
- Asset incomes: we assume that asset income can be added in a second step, i.e. on top of the tax-benefit calculations presented here. Most assets are not relevant for most benefit calculations (e.g., home ownership, the most dominant and prevalent component of private wealth, is usually not affecting tax rates or benefits, except for housing related policies); also, asset incomes often follow their own tax rules (e.g. flat-rate tax on interest). Like labor-related incomes, these income components will be modeled longitudinally.
- Family-demographics: taxation and benefits often depend on family characteristics. Having a partner and spousal income can alter individual tax rates (due to income splitting rules; deductions for single earner; or splitting of tax allowances for dependent children), while the number and age composition of children are key inputs for various benefits.

We refer to "labor-related incomes" as either earnings or income from income replacement schemes that are generally contributory and expressed as a proportion of earnings. In contrast, benefits from social welfare schemes are handled by the tax-benefit calculator.

Outputs:

- The tax rates applicable to total gross "labor-related" incomes (separate for both spouses in the case of couples)
- Total net benefits by NTA category (Education, Health, Old-age, Disability, Survivor, Unemployment, Childbirth related, Family

3.1 Earnings

In the aggregate, earnings constitute the dominant source of household income, as displayed in Table 1 which contrasts total earnings to pensions, and benefits according to the Euromod data of 8 countries.

Table: Aggregate values of income categories in 2021, in Mio EUR (GBP in the UK)

	AT	DE	SI	ES	IT	FR	FI	UK
Earnings	14,138	146,912	1,955	40,803	52,219	80,530	8,266	79,936
Pensions	4,612	34,130	458	12,842	24,331	27,126	2,769	7,920
Means-tested benefits	573	3,545	51	1,861	2,131	7,097	455	6,731
Non-means-tested benefits	987	7,516	110	3,162	6,669	4,358	494	5,120

Source: EUROMOD output files. Earnings refer only to income received for active periods. Replacement incomes are included in pensions, means-tested and non-means-tested benefits.

Earnings are implemented in microWELT by means of Mincer equations. based on EU-SILC data.

3.2 Unemployment spells and benefits

Longitudinal work careers implemented in MicroWELT distinguish spells of employment, unemployment, out of labor force, and permanent retirement. Unemployment risks and transitions out of unemployment are – besides other characteristics such as age, education, number, and age of children – duration dependent and modeled in continuous time. The risks of entering as well as of leaving unemployment are typically highest in the first few months of the according spell. The modeling of unemployment is documented in a separate research note.

The eligibility conditions, the amount and the duration of unemployment benefit(s) vary widely across eight countries. Our goal was to find a stylized way to model unemployment benefits uniformly. Since the eligibility for unemployment benefits may last longer than a year,

EUROMOD is not the most appropriate source for unemployment benefits since the model covers only one year. Therefore, we estimated replacement rates using the OECD tax benefit calculator, which enables us to calculate the amount of unemployment benefits up to 5 years in monthly steps. We express the calculated gross unemployment benefit as a percentage of the average gross wage in a country, accounting for contribution period, age, and the presence of a partner and/or children in the family. Children are considered since top-ups are available for children in some countries. We found that rates sometimes differ for couples with 2 children (compared to singles) which therefore form a separate category, in this case replacement rates are calculated assuming that both partners earned the same level of income before one becomes unemployed. For each case, also the minimum and maximum unemployment benefit expressed as a share of the average wage is provided.

Parameters by month of unemployment are currently available for the following groups:

- No dependent children, age 40, contribution period 1 / 5 / 15 / 22 years
- No dependent children, age 55, contribution period 25 years
- One dependent child, age 40, contribution period 1 / 5 / 15 / 22 years
- Single with 2 (also applied if more than 2) dependent children, age 40, contribution period 1 / 5 / 15 / 22 years
- Couple with 2 (also applied if more than 2) dependent children, age 40, contribution period 1 / 5 / 15 / 22 years

The current set of options might be extended or modified in the process of model development and testing.

Parameters use OECD definitions for what is considered an unemployment benefit - for example, in Austria, benefits include the "Notstandshilfe", which follows unemployment insurance benefits.

3.3 Maternity and parental leave spells and benefits

This chapter – and the approaches – are currently being revised at IER. Specifically, a distinction between maternity leave and paternal leaves will be made, the latter to allow for simulations of increasing take-ups by fathers.

The Eligibility for maternity and parental benefits is linked to the event of giving birth. EUROMOD input or output databases for most countries include non-simulated parental benefits, which come directly from SILC data. The EUROMOD model for all countries contains the simulation of

maternity/parental benefits, but they are switched off by default since the results are not reliable due to the lack of data (the base for benefit calculation, number of months in receipt). However, the model allows to switch on the calculation of maternity/parental benefits. Maternity/parental benefit policies are originally (in EUROMOD) modelled in a way to calculate the benefit amount given the input data, including the age of the child and the most likely base for the benefit. We changed the countries' models to ensure that the benefit is calculated for the whole year (assuming the child was born at the beginning of the year) and that the base for the benefit calculation is employment income. We used the hypothetical household tool and run adapted models to calculate maternity/parental benefits for different levels of wages expressed as a percentage of the average wage (33%, 67%, 100%, 200%, 300%). Like the unemployment replacement incomes, we expressed maternity/parental benefits amount as a re-placement rate, the benefit level relative to previous earnings. Additionally, the minimum and maximum thresholds are expressed relative to the average wage and the duration of the benefits for each country is taken into account.

3.4 Public pensions

- Modeling retirement decisions for microWELT is just beginning
- Concrete work is currently being done at WIFO on modeling retirement decisions. In the case of microWELT, this is a revision of the current approach and is related to current work on improving the longitudinal consistency of work careers.
- Retirement is the subject of an Austrian case study (part of Sustainwell) comparing the results of the highly stylized microWELT approach with the detailed Austrian model microDEMS. Ongoing improvements are being made to microDEMS, which is currently being used to assess the labor force consequences of the current pension reform - a gradual increase in the retirement age for women from 60 to 65 (the age for men). This provides a test case for the case study; a working paper is available already on this topic: https://wifo.ac.at/jart/prj3/wifo/resources/person_dokument/person_dokument.jart?publikationsid=71474&mime_type=application/pdf

Pension systems vary widely from country to country. In addition, lifetime work histories and earnings affect the pension benefits individuals receive when they retire. Because pension legislation is very complex, some countries have dedicated pension microsimulation models whose main purpose is to simulate pensions. Since detailed modeling of pensions is beyond the scope of MicroWELT, the general idea is to estimate replacement rates for different career

lengths and income levels using available estimates (Pension Adequacy Report, OECD database) and available microdata.

3.5 Family demographics and characteristics

MicroWELT models nuclear families, i.e. singles or couples with or without dependent children. Family characteristics provide input parameters for tax-benefit imputations. As the number of possible family compositions is extremely large, we aim at finding classifications capturing heterogeneity in a meaningful and manageable way (see below).

4. The synthetic tax-benefit imputation database(s)

EUROMOD is a tax-benefit calculator based on EU-SILC data. EUROMOD has input and output data files. Input data files contain socio-demographic data of the population sample and disaggregated incomes, taxes, and benefits from EU-SILC data. The output files contain values of simulated benefits and taxes. As some benefits cannot be simulated due to lack of data, imputed values of these benefits are included in the output files.

Although the EUROMOD model provides a valuable tool for distributional analysis and for analyzing the effects of policy changes, it has some limitations for our purposes. The main limitation is the sample size of EUROMOD (EU-SILC). Therefore, the hypothetical household tool is used to model the simulated benefits in EUROMOD. The hypothetical household tool is a EUROMOD application that allows the creation of a synthetic population assuming numerous income levels and family composition options. However, it is important to note that the hypothetical household tool and the synthetic population can only be used to calculate benefits simulated in EUROMOD.

General idea (refinements are discussed below a/o are still to be decided):

- Creation of a synthetic database with a "large but manageable" number (100-250k) of potential nuclear families depicting the heterogeneity concerning family and income composition. Each simulated family retrieves EUROMOD-simulated social insurance rates, tax rates, and benefits by NTA category from a corresponding database record.
- Matching criteria are categorical variables alongside a set of dimensions (the following specification comprises 141k records):
    - Income categories expressed in 5% groups based on a reference (average or median) income. [100] This makes comparative systems analysis more

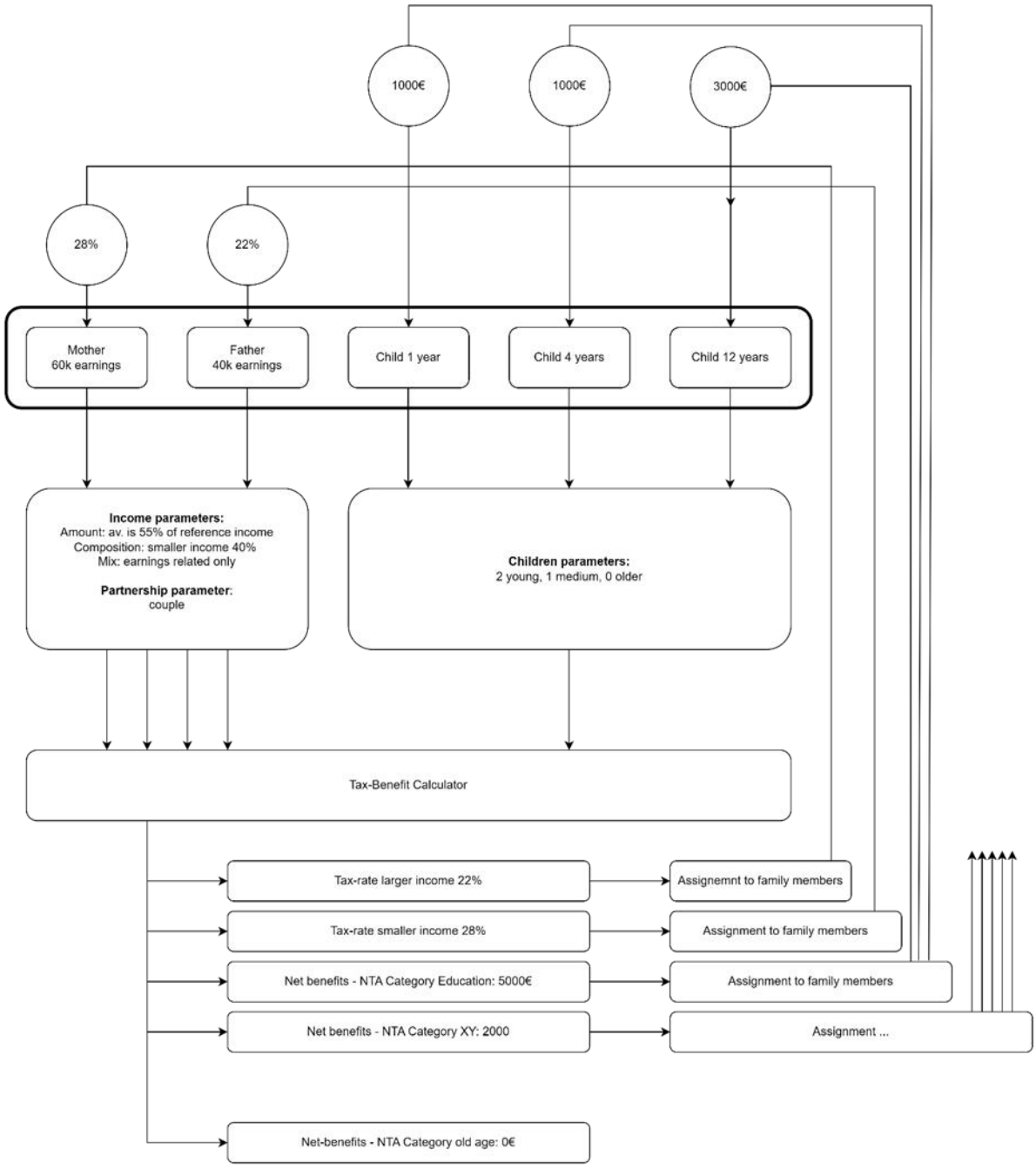
meaningful than nominal concepts and implicitly adapts income brackets over time relative to e.g., median earners.

  - Number of adults - single versus couple. [2]
  - Share of smaller income (couple) in 5% bands [11]
  - Number of children by age group, e.g., 0-3 children by 3 age groups [64]

- Continuous time updates (whenever any category changes)
- Filing of yearly "tax returns": at this point, adjustments can be made (allowing for alignments, revisions, amendments, etc.) and results are finalized and stored. Same at death.
- Modelled asset incomes will be considered when available. They will be used to correct or complement the results of the tax-benefit calculations of the first step.
- The database is used to retrieve tax rates (2 in the case of couples) and benefits by NTA category.
- The assignment of benefits to family members is treated in a next step and is informed by the NTA logic (e.g., child benefits are received by children).

The following Figure 1 illustrates the general approach.

Figure 1: General operation of tax-benefit imputations



Source: Own illustration of the general approach

The approach outlined above assumes that all gross "work-related" income can be summed up and converted into a single parameter that relates it to a reference income (e.g. median or average). While this is a reasonable simplifying assumption for most purposes, a not insignificant exception is pensions, which are typically treated differently with respect to social security contributions, which – not including pension insurance anymore - are much lower for retired persons. As social insurance contributions are usually applied on the individual level, and it is gross incomes after social insurance which constitute the tax bases, we envision a two- (or more) step solution, first determining social insurance contributions, and then taxes and benefits. Both steps can be implemented by separate synthetic imputation databases, while the first one (social insurance) operates on the individual level and - besides income brackets – differentiates by income type, i.e., earnings, pensions, unemployment benefits, leave benefits. We assume that at any given moment, only one of those types can be received, but – given the continuous time approach – the various types can be received in sequence in any given year.

Benefits will be categorized according to the NTA approach such as education, health, etc. Discussion and decisions are required in a range of areas:

- Treatment of non-simulated benefits in Euromod, especially if they depend on individual characteristics such as disability or providing care
- General consistency of concepts between Euromod and the NTA framework
- Mechanisms to assign benefits retrieved on the family level to individuals

Current discussion focuses on ways of a stepwise implementation – e.g., first determining individual income tax rates, and in a second step benefits on the family level.

For efficiency reasons (computational speed) we aim at creating synthetic databases alongside key categorical dimensions. Thus, each simulated entity can address a corresponding record directly, not requiring a database search of the closest neighbor.

5. The accounting framework

The continuous-time time framework of microWELT is reflected in the implementation of the accounting system which instantaneous updates whenever relevant characteristics such as employment status, income, or family compositions changes.

The implementation of the tax-benefit accounting system is under way; we currently implement a prototype based on sandbox data, here some specifications:

Modgen/openM++ support this approach in powerful ways, i.e., many updates are handled automatically and do not require the coding of update events. For example, the accumulating of a benefit over a year can be handled by a derived state:

```
double year_delta = current_spell_weighted_duration(year_spell, TRUE, current_rate);
```

The current_rate (such as a benefit) can change over the year, year_spell is re-initialised at each calendar year change, year_delta is the total benefit received so far in the current year, and can be used also within the year (e.g., when closing accounts at death)

At the end of each calendar year, or at death, a "tax return" is filed; at this point, adjustments can be made (allowing for alignments, revisions, amendments, etc.) and results are finalized and stored.

Accounts are implemented as a separate actor type (like persons); accounts are linked to a specific person and an observer actor (which is handling totals, alignments, etc.). This approach allows to switch all accounting off if not required.

The implementation of a generic prototype of the database-imputation approach is under way.

Synthetic tax-benefit databases are implemented as multi-dimensional parameter tables. Updates (new lookups) are typically initiated by the family head whenever a relevant characteristic of any family member (or the family composition itself) changes.

References

Van de Ven, Justin; Patryk Bronka, Matteo Richiardi† (2022). Dynamic Simulation of Taxes and Welfare Benefits by Database Imputation. https://www.iser.essex.ac.uk/wp-content/uploads/files/working-papers/cempa/cempa3-22.pdf

Benjamin Bittschi, Thomas Horvath, Helmut Mahringer, Christine Mayrhuber, Martin Spielauer, Philipp Warum (2024) Assessing the Labour Supply Effect of Harmonising Regular Retirement Age in Austria - WIFO Working Papers, 2024, (673), 28 Seiten - https://wifo.ac.at/jart/prj3/wifo/resources/person_dokument/person_dokument.jart?publikationsid=71474&mime_type=application/pdf