

Macro and Micro modelling

A draft paper for further discussion by the Social Security Subcommittee of the AAE

The Pension Adequacy Report and the Pension Chapter of the Ageing Report engage in evidence-based guidance of pension systems of Europe. This endeavour uses country-by-country and cross-country analysis based on a number of indicators. Several indicators are derived from projected cash flows of pension systems. However, the underlying methodologies show a diverse picture. One reason might be that macro modelling suits better for some indicators while for others micro models can be derived from the definition. It would be a reasonable expectation to suppose that the results for a country from the two sources are consistent in the sense that a statistical relationship can be established between the macro and micro models. For example, we may suppose such a relation exists between the Aggregate Replacement Rate and the Individual Replacement Rate. We accept that they are different but would like to have a reasonable explanation and estimate for that difference.

In this paper we propose the use of the Markov chain framework¹ and the Expectation Maximization and Hidden Markov Model method to provide a consistent approach to pension systems modelling at different levels of detail. We reason that this model is applicable to all levels of aggregation and therefore could be used from the onset of the pension projection exercise, including the planning of data and modelling methodology, in this way establishing the relation between the results of micro and macro modelling.

Methodology behind the indicators: Macro and micro modelling

Some indicators are derived from aggregate projections while others are defined at individual or household level. The simple and straightforward method of modelling the first category is using macro models, with micro models for the second. The microsimulation methodology was developed originally around household statistics.

But adopting the other solution is possible for both categories; there are no accepted rules. In such applications additional considerations have to be included in the original model and the interpretation of the results of the calculations is highly dependent on the methodology. The representative member approach can be used for calculating individual measures from macro models. Here the definition and interpretation of the representative members matters. In case of micro simulation applied to macro we have to derive the total amounts from a sample population and compare them to aggregates from macro statistics. The question here is the methodological consistency between the development of the total of the individual indicators and the aggregate values in a future state.

¹ We restrict our cases to discrete-time Markov chains defined in every year, and with random variables of, in fact, finite natural numbers. This also applies to the Hidden Markov Models. In certain cases the HMM will be supposed to be a Hidden semi-Markov Model.

An EU project assessed the use of microsimulation pension models in the Member States in 2009.² While a lot of development has been made since then,³ some observations of the first report might be useful for reference. The report classified the pension system projection models as standard⁴ (cohort or typical agent) [macro] or microsimulation models.

Standard [macro]: Cohort models use cross-sectional [present and past] information on labour market and social security data of cohorts or other social groups. In standard EU reporting gender also differentiated. In theory, these models are also capable for further disaggregation by labour-market attributes and demographic attributes like marital status, education. Most cohort models are partial equilibrium models. The elaboration of the social groups and economic model largely depends on the available data and calculation complexity. However, there are also in use special overlapping generation general equilibrium cohort models, covering the economic behaviour of households, firms and pension funds.

Typical agent models study theoretical career or life paths of individuals, which regarded as typical from the perspective of the analysis.⁵ This method is best applicable for longitudinal examination of future policy changes. Transitions between the examined career paths are not accounted for, and therefore the outcome largely depend on [the method of] the definition of the typical agents. Parity with exogenous aggregates is required only in the case of pension system projections where the system is described by a set of typical agents.

Microsimulation models simulate changes in a sample of a total population. The aim is to achieve manageable trade-off between data and calculation difficulty and the methodology, assumptions and parameters. Still, microsimulation models have the objective of capturing population, policy, behavioural and temporal complexity at individual, household or organisation (elementary unit) level.

Static microsimulation models focus on two states of a system by simulating the effects of constant factors. *Dynamic microsimulation* models take into account development in time and hence able to model changing parameters and assumptions of a system. Pension projection is an example where, development in time and, for example, changing life expectancy must be taken into account. There are two approaches to introducing ageing in a dynamic microsimulation model. In its simpler form cross-sectional model values are updated and re-weighted by an adjustment algorithm to add-up to the exogenous aggregate data in projected states of the system. In a fully dynamic model the individual life histories are built up using dynamic parameters. This is a more complex task and checking and securing parity to the aggregate values still have to be done.

The Member States reported the following approached to pension modelling in 2009:

- Only standard (cohort, typical agent) model but no micro-simulation model
 - Import model from international or foreign agency customised to special needs (6)
 - Model developed in-house
 - No steps towards microsimulation (6)
 - Steps towards microsimulation (4) – these countries must have completed the development
- Utilisation of microsimulation model
 - Microsimulation model as complement
 - Modelling work contracted out (2)
 - Modelling work in-house (5)

² Róbert I. Gál, András Horváth, Gábor Orbán and Gijs Dekkers: PENMICRO - Monitoring pension developments through micro socioeconomic instruments based on individual data sources: feasibility study, Final Report for The European Commission DG EMPL E4 Unit, Brussels, TARKI Social Research Institute, Hungary, 2009

Note: This introduction to the report includes the critical observations of the author.

³ Gijs Dekkers, Raphaël Desmet, Ádám Rézmovits, Olle Sundberg, Krisztián Tóth: On using dynamic microsimulation models to assess the consequences of the AWG projections: Simulation results for Belgium, Sweden and Hungary, 2015

⁴ In this paper we keep the terms macro model and macro modelling.

⁵ An example is the Theoretical Replacement Rate (TRR) analysis in the EC Pension Adequacy Report.

- Microsimulation as primary modelling instrument (3)

The report finds increased data availability requirement for micro than standard/macro models. They identified five ways to improve micro datasets for microsimulation models, starting from the minimal requirements to the ideal cases. These are the following (countries falling in the respective classes are in parentheses; in case of more models by country, the name of the model is also added):

- Using social insurance administrative data only (Hungary, Sweden)
- Matching social insurance administrative data with tax records and/or census data (Spain, Netherlands, Austria, Finland)
- Using large household surveys (various countries)
- Matching social insurance administrative data with household surveys (France, Luxemburg, United Kingdom)
- Matching numerous administrative datasets (Belgium, Denmark, Sweden)

Data availability is improving by developing e-government in most countries. The 2021 Census Directive allows for using government databases to improve census data.

Note, that better data availability and set of assumptions used for microsimulation may also be used for improving the detailedness of macro models.

Micro simulation modelling with macro alignment

Modelling – in the context of our interest – is used to describe states of a system in time by formulas (functions or algorithms) operating on data consistently structured with the states of the system and capturing all relevant aspects of the system and representing the relevant relationships between the different elements of the system.

The results of macro models are usually accepted as accurate because the methods are simple and easily verifiable and there are close relationships with well-established statistical data systems. However, the methodology conceals important information. Such analysis is made possible by starting the modelling from a micro level. Microsimulation must have its role in policy analysis, but the results have to be interpreted according to its nature. We can only get outcomes which are delimited by the assumptions and input parameters, even if the result has been randomized. For this reason when setting up the statistical model Bayesian statistics would be better used to avoid ecological fallacy when extending the results of the sample or the typical agent to the total population. In the specific case of modelling individual decisions there is still another theoretical challenge. Again, the results of a parametrized model (decision tree or other) can be enveloped, and still can only model observations of past behaviour. The problem is that the classical economic model of rational behaviour of the well-informed consumer does not hold in financial economics, and therefore future events cannot be necessarily deduced from past experience. In pensions this phenomenon is discussed by *Peter Diamond and Nicholas Barr*⁶.

Accepting the difference between the results of micro and macro models leads to the consequence that the longer the projection period is the larger the difference becomes. Eventually this led to the proliferation of the so-called adjustment methods, that is some microsimulation models build adjustment algorithms to macro constraints into the process.

⁶ Barr, Nicholas and Diamond, Peter (2009) Reforming pensions: principles, analytical errors and policy directions, International social security review, 62 (2). pp. 5-29. ISSN 0020-871X DOI: 10.1111/j.1468-246X.2009.01327.x This version: <http://eprints.lse.ac.uk/25099/>

*Li and O'Donoghue (2014)*⁷ define alignment as "a process of constraining [microsimulation] model output to conform more closely to externally derived macro-data ('targets')." They suggest two broad categories for alignment:

- parameter alignment, whereby the distribution function is changed by adjustment of its parameters; and
- ex-post alignment, whereby alignment is performed on the basis of unadjusted predictions or interim output from a simulation.

They categorize the algorithms as the following:

- Multiplicative Scaling
- Sidewalk Shuffle, Sidewalk Hybrid and their derivatives
- Central Limit Theorem Approach
- Sort by predicted probability (SBP),
- Sort by the difference between predicted probability and random number (SBD), and
- Sort by the difference between logistic adjusted predicted probability and random number (SBDL)

The first three are defining or re-defining the outcome distribution, and might have connection with the EM model in this paper. In this case the only question is how the connected macro model has been defined.

However, the second group of three algorithms sort the outcomes and keep only the highest probability cases for later use, and doing so cut the tail of the distributions off. This is against the basic axioms of probability theory, and might be one of the reasons of the said difference, simply by general reasoning.

But why do we have to accept the difference, in the first place, without trying to give an analytical estimation? The point is that I feel uncomfortable with the solution that we accept the difference between the macro and micro results and apply a posterior adjustment, however sophisticated it might be.

In this paper we propose a framework which aims to provide not just parameter alignment, but a common approach to data and modelling methodology from the onset of the pension projection exercise, in this way bringing the results of micro and macro modelling closer.

Authors, comparing micro and macro simulation models, point out the complementarity of the two approaches. In regard to the basic assumptions and data it is easy to see or require comparability when complementary is obvious. But the applied methodologies may diverge and lead to non-comparable results.

The macro simulation models provide aggregate results closely related to the CGE (Computable General Equilibrium) models and are regarded as robust solutions. They aim for conclusions about the total population (aggregate values) and may use representative units described by average values and total numbers.

Standard actuarial pension modelling is based on estimating future cash flows of contribution and pension payments, and the underlying assets and liabilities. Part of the data is regarded as external parameters and assumptions, e.g. future employment rates are regarded as external economic parameters, and assumptions are made about trends of future changes. Stochastic elements are usually

⁷ Jinjing Li, Cathal O'Donoghue: Evaluating Binary Alignment Methods in Microsimulation Models, *Journal of Artificial Societies and Social Simulation* 17 (1) 15 (2014) <<http://jasss.soc.surrey.ac.uk/17/1/15.html>>

limited to demographic development and the occurrence or timing of events. In several cases conditionality is also supposed or standardized. If a funded element is part of the system, investment returns might also be randomized. Other elements, such as distribution of contribution payments, career patterns, and redistribution provisions of the pension formula, are hard to analyse by macro pension modelling. There might be other scheme rules, like optional choices, which difficult to model.

Microsimulation models build up the results from elementary units of the system characterized with multiple variables, changing their states at random and/or in a behavioural manner during the observation period. The results arise from calculations using the descriptor variables of the units. Here the social elements of the life-path are assessed to produce sample household and career developments, and the pension status descriptors are consequential.

In comparing the aggregate values of a macro model and the totals of a micro model applied to the same problem – even if using the same set of assumptions, data and parameters –, without adjustment the results will be different if the methodologies are different. In general, macro models need less assumptions. Microsimulation models require many assumptions, and this may in itself reduce the robustness of the results.

The above considerations raise some questions. Are macro model results really adequately exact without taking into account the distribution and optionality of the observed units? Could it be improved by utilising more of the detailed data and assumptions necessary to micro modelling? And do micro models adequately reproduce the underlying distributions of the elementary units when they are aggregated for macro results? Could a method be used iteratively improve the assumptions and parameters? The answer depends to a large extent on the specific models, or model specifications.

Intuitively, the models have to be “compatible” (i.e. use *mutatis mutandis* the same methodology and data, differing just in its depth and details), and each model has to be “consistent and consequent” in their methodologies to produce comparable results. Then, in theory one should be able to assess and give an estimation of the expected differences between the two models.

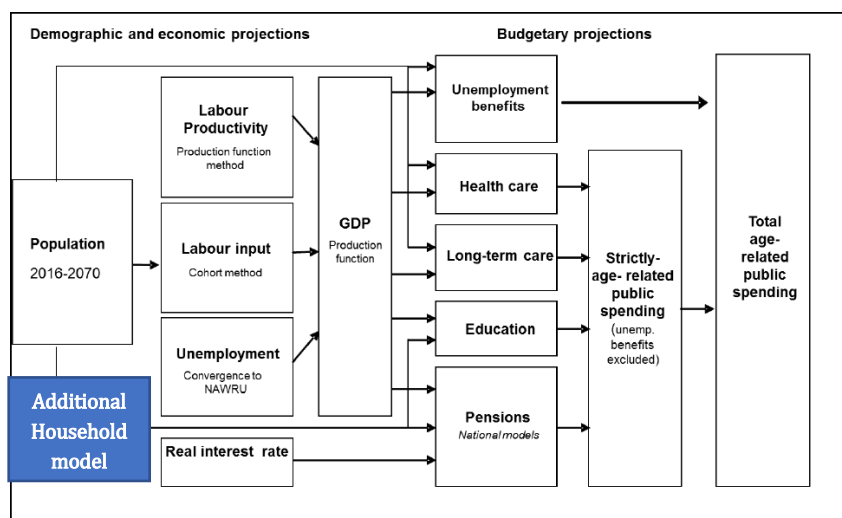
Data and modelling considerations

Macro modelling requires cross-sectional decomposition of the age groups of the population according to insured status: active worker, unemployed or pensioner or other beneficiary. The financial status of active workers may differ according to full- or part-time employment. Other specific forms of employment or tax status may also influence the benefit accrual. Differentiation of unemployment and other credit periods should also be recorded. All this influences the distribution of the contribution payments in a calendar year and the benefits later on. We may also be interested in different beneficiary statuses with their respective benefits. Benefits may also depend on the constitution of the household.

The total changes of the groups are subject to macro limits from observation and projections using the definitions from the rules of the pension system. The number of individuals in the subgroups changes with new entrants and those leaving to another group or dying. In fact, in establishing policy decisions we are not interested in individual cases, but only to the extent that they contribute to the exact and precise results. We are interested in the number of members and their average income in the groups. Additionally, other descriptors, with well-defined averages wherever it is necessary, might be of interest. However, the aim is to define the groups so that the average should be representative and adequate for the purpose of the cash-flow analysis. From the modelling point of view a key factor is the definition of the details of how the total population is broken down according to insured status and socioeconomic subgroups. This is a cross-sectional categorisation. Having solved that, we may examine the transitions of the number of members of the groups from year t to $t+1$ using cross-sectional and longitudinal relationships.

Pension adequacy decisions are made at household level. Family relationships influence supplementary pension benefits and, in some jurisdictions, basic or minimum pensions. Individual labour market decisions and other behavioural matters are also decided at family level. Modelling decision making may have a simple model in general, but has drawbacks at individual level in pensions. The reason is non-predictable economic decisions in the context of second-best analysis (imperfect information, non-rational behaviour and incomplete markets/contracts).⁸ This issue is in fact similar to modelling discrete choices in CGE models, for example household and labour answers to macro shocks.⁹

Household statistics are available at micro level as representative samples and also at an aggregate level in less granular totals.



The EC macro calculation models with additional household/family module

The EC semi-aggregate macro projection models include GDP labour and employment modules, similar to the classical ILO model. The models could and/or should be extended with a macro household module projecting macro household data like the total numbers of households, marriages and divorces, earners and children by family, single or two parent families, etc; that is the data relevant from a pension insurance/rights accrual and benefits perspective. This way the assumptions of the actuarial model would include macro level family/household estimations which could link economic and demographic assumptions.

In most countries detailed historical information is used for the determination of pension rights individually in annual pension statements and at retirement. That is a posterior longitudinal career path record. In any specific year the original cross-sectional data is usually generated in fragmented administrative databases of employers, tax, pension and labour administrations and other social institutions. When insured status changes the transition details are generally not recorded in the database. In most social security pension administrations, contribution and benefit payment transactions are recorded, but the information used for and included in pension statements and the resulting award of pension for the insured person may not be recorded. Even these records do not include some important demographic factors, like education and family relationships at the time of transitions during the career.

Pension system projections are made to analyse payments made at the events depending on retaining membership (still working or still a beneficiary) or changes of status in (transitions between) demographic and socioeconomic/pension insurance status groups. Cross-sectional information on

⁸ Nicolas Barr

⁹ John Cockburn, Luc Savard, Luca Tiberti: Macro-Micro Models (2014)

present status is derived from administrative statistical data. Information on changing status – transitions – might be obtained from historical records or administrative procedures using past records. Because of the large scale, missing data and separated databases we cannot suppose that all data is available and the consistency of the information must be established. A special case of missing data is where information might be available but contradictory to data from other sources or estimations.

Securing data consistency should preferably apply the same methodology as the projection.

Modelling pension systems

We use the macro projections of economic, labour market, population and household variables given as assumptions.

In the following discussion we go from macro to micro for simplicity of discussion. This is not necessarily to suggest a top-down approach. We are attaching financial descriptors to all observation units – including social and family categories – so as to support the objective of the cash-flow projection exercise at all levels of observations.

Modelling pension insurance status groups

We examine the social security pension system insured status groups as parts of the total population. Pension insurance status groups are modelling the total population as the sum of semi-aggregate population components.¹⁰ Population component cohort models follow specific demographic rules and model assumptions that ascertain the conditions of the stochastic process. In case of pension modelling the rules of the pension system apply as the filtration of the process. In most cases we are not simply interested in the number of members of a group, but rather in the pension insurance and related socio-economic status and its financial outcomes described by the filtration.

The models specify the number in a population G as the distribution of the members among the status groups $G^{(k)}$, $k=1, \dots, N$ at time t , $t = t_0, t_1, \dots, t_{max}$. Pension insurance status groups are categories of active, retired, etc. states with reasonable definitions. The states of the system are the consequence of the changing number of members of the component groups. The change in the total number is the result of the increase in respect of new members and the decrease by withdrawals from the status of membership in $G^{(k)}$. In pension insurance the insured status changes for demographic reasons and according to scheme rules and regulations. The $G^{(k)}$, $k=1, \dots, N$ status groups are described and distinguished by M descriptor variables of interest in pension insurance, like the number and measures or averages of salary/contribution or credit, accrued rights or benefits, age and sex structure and marital relations.

More formally let $G(t) = [G^{(k)}(t)]^T \in \mathbb{N}^N$ a *discrete-time Markov chain* modelling a pension system in calendar year t where we denote by $G^{(k)}(t)$ the insured status groups in calendar year t , by $g^{(k)}(t) \in G^{(k)}(t)$ a representative member of the status group k , $k = 1, \dots, N$, and N is the number of pension insurance groups describing the population. The sequence of numbers of the groups of population might be projected directly with an adequate level of accuracy for certain studies, without explicitly

¹⁰ Wherever this cannot be ensured special mapping algorithms apply (partition). Overlapping categories occur in case of similarities of family statuses, salary bands, insured spouses' benefits, etc. Special group cases will also apply to births and deaths. New entrants and withdrawals change the total number of the population.

relying on the transition matrix parameter of the Markov chain.¹¹ At micro level $g(t) \in \mathbb{R}^M$ with M descriptors is also subject to a Markov chain where *state #k* of g is defined as $g(t) \in G^{(k)}(t)$, following an actuarial multi state pension model.

We have defined the population component model as *the states of the Markov chain* $G(t)$ by $G^{(k)}(t)$, $k = 1, \dots, N$, at $t, t = t_0, \dots, t_{max}$: $G^{(k)}(t)$ with the turnover rules as filtration. From time t to $t+1$ $G(t)$ changes because of (i) demographic changes of births and deaths and (ii) transitions between insured status groups, with withdrawal from one status meaning a new entrant in another. As for the turnover of the membership, let us denote $G^{(i,j)} = \{g \in G(t): g \in G^{(i)}(t) \text{ and } g \in G^{(j)}(t+1)\}$, that is the group of members changing *status #i* \rightarrow *status #j* while $t \rightarrow t+1$, and $G_{i,j} = \|G(t)^{(i,j)}\|$ denotes the number members g of population G moving from *state #i* to *state #j* in time t . The process is parametrized with $P^G = P^G(t) \in \mathbb{R}^{N \times N}$, $P_{i,j}^G = P_{i,j}^G(t) = p(G(t) = j | G(t-1) = i)$.¹² We may have information about changing descriptors of individual members and/or individual statuses as $t \rightarrow t+1$, and/or observation or projection of the total number of members in G , *state* $G^{(k)}$ in t and $t+1$.

In a pension system the number of transitions between the insured status groups may change significantly, leading to large but scarcely populated transition matrices. However, most cases are similar and/or irrelevant, and for the relevant cases the observed turnover matrices give realistic estimations. We define the transition matrix of the semi-aggregate Markov chain by the Expectation Maximisation (EM) method, which is applicable to missing data and large/complex cases.¹³ With additional information on the sequence or transitions the algorithm can be modified.¹⁴

The EM has the following parameters:

- (1) The initial probability distribution over the N possible hidden states G : $\pi = [\pi_1, \dots, \pi_N]^T$, where $\pi_i = p(G(1) = i)$ is estimated from statistics, etc.
- (2) An individual state transition probability matrix $P \in \mathbb{R}^{N \times N}$ that specifies the probability of transitioning from *state i* to *state j*: $P_{i,j} = p(G(t) = j | G(t-1) = i)$.

We use the EM method for the maximum likelihood estimation to find the best fit state sequence $G(t)$ and transition matrix P^G .

¹¹ Although transition probabilities from state $t-1$ to t of G are limited in number and the required accuracy in reality.

¹² Note that in general discussion and in some cases $P_{i,j}(t)$ should be required. In this paper we use constant transition intensities for simplicity of discussion.

¹³ Maya R. Gupta and Yihua Chen: Theory and Use of the EM Algorithm; Foundations and Trends in Signal Processing Vol. 4, No. 3 (2010) 223–296 DOI: 10.1561/20000000034 and

Sherlaw-Johnson, Chris, Steve Gallivan, and Jim Burrridge: Estimating a Markov Transition Matrix from Observational Data, The Journal of the Operational Research Society 46, no. 3 (1995);

¹⁴ The Expectation–Maximization (EM) algorithm is an iterative method to find the maximum likelihood estimator of a model parameter when the data might be incomplete or has unobserved (hidden) latent variables. The iteration has two steps: creation of the expectation function from the latest parameter and finding the new parameter by optimizing the expectation over the parameter (arg max). The complexity of the algorithm depends on the missing or the available additional information. In simpler cases it is shown that the maximum likelihood estimation can be solved by multi-proportional adjustment.

Semi-aggregate family-household model

The micro level decisions in relation to spending or saving, employment, marriage and children which influence the pension system are made at family or household level.¹⁵ While we still do not aim to model individual behavioural decisions, our objective is to avoid obvious contradictions at model level.

We may have information on total cross-sectional data on households, but to a lesser extent on transitions. Household categories could be classified according to cohabiting generations, number of parents, earners, children, income. We define the transition matrix of the semi-aggregate Markov chain by the Expectation Maximisation method.

Let the Markov chain be $F = F(t) = [F^{(k)}(t)]^T$, $k=1, \dots, U$, where U is the number of family/household or other relevant socio-economic status groups. The total population, also $F = G$, $g = \{g: 1, \dots, I\}$ is distributed over the U status groups. In a similar way to the insured status model, the states of the system are the consequence of the changing number of members of the component groups. The process is parametrized with $P^F \in \mathbb{R}^{U \times U}$, $P_{i,j}^F = p(F(t) = j | F(t-1) = i)$. We may have information about changing descriptors of individual members and/or individual statuses as $t \rightarrow t+1$, and/or observation or projection of the total number of members in F , state $F^{(k)}$ in t and $t+1$.

The EM has the following parameters:

(1) The initial probability distribution over the N possible hidden states F : $\pi = [\pi_1, \dots, \pi_N]^T$, where $\pi_i = p(F(1) = i)$ is estimated from statistics, etc.

(2) An individual transition probability matrix $P \in \mathbb{R}^{N \times N}$ that specifies the probability of transitioning from *hidden state* i to *state* j : $P_{i,j} = p(F(t) = j | F(t-1) = i)$.

For the EM the complete set of parameters to estimate F is $\theta = \{\pi, P^F\}$. Our task is to find the best state sequence $F(t)$ and transition matrix P^F !

Typical Career Paths models and modelling income inequality

In the Typical Career Paths model we define the component cohort population model according to the usual demographic components and predefined career models. We may construct theoretical careers or identify career paths from observations from tax returns or pension administration data.

The theoretical replacement rate (TRR) methodology of the Pension Adequacy Report defines the base case with the average worker, and the cases of high- and low-income workers as 66% and 100% (increasing to 200% during the career) of the average. Other model cases include careers with breaks because of childcare/family reasons, sickness or unemployment. The EC recommends preparing the calculations for private/public sector and for self-employed.¹⁶ This uniform definition is well applicable for cross-country comparisons. However, this is a theoretical approach, not aiming to add up to the values of the macro model. Other career paths might be defined based on clustering historical career data. In such exercises the explanatory variables could be the sector (public, private, self-employed,

¹⁵ Macro-Micro Models, microsimulation techniques within a computable general equilibrium (CGE) by John Cockburn, Luc Savard, Luca Tiberti: Macro-Micro Models, Universite de Sherbrooke, GREDI Working Paper 15-08

¹⁶ The 2018 Pension Adequacy Report: current and future income adequacy in old age in the EU – Joint Report prepared by the Social Protection Committee (SPC) and the European Commission (DG EMPL) – ANNEX 1. Methodological background for the theoretical replacement rates European Commission; Directorate-General for Employment, Social Affairs and Inclusion and the Social Protection Committee © European Union, 2018 ISBN 978-92-79-85657-0 doi:10.2767/406275 KE-01-18-457-EN-N European Commission

farmer), career type (owner, manager, employee, etc.), and fragmentation of career because of unemployment, maternity, child care, other family care, etc. Jumps between career paths also have to be supposed, for example because of unemployment and voluntary or involuntary self-employment.

The above states use individual income positions, but we are interested in the groups of individuals who are in a similar position. This is even more obvious in the case of the following indicators.

In Gini models and income percentiles of the population models we define the component cohort population model according to income percentiles or other wealth/poverty status. The income categories of the AROP/AROPE measures include all household income and not just salaries and pensions. We may focus on salaries and pension and use it as a proxy for household income.

Note that we still cannot model individual behavioural decisions, but only estimate the outcomes. We use observations from the semi-aggregate insured status and/or socio-economic groups Markov chain to estimate the parameters of the Hidden Markov Models (HMM)¹⁷ of career path and/or income groups by the Expectation Maximisation method for HMMs.

Now the Hidden Markov chain is $H = H(t) = [H^{(k)}(t)]^T, k=1, \dots, V$, and V is the number of the career paths or income groups in the examination. We suppose that the total population $G, G = \{g: 1, \dots, I\}$ is distributed over the V groups. The descriptors of $g \in H$ are different from the case of G and might be hidden. The hidden process is parametrized with $P^H \in \mathbb{R}^{V \times V}, P_{i,j}^V = p(H(t) = j | H(t-1) = i)$. We may have information about changing descriptors of individual members and/or individual statuses as $t \rightarrow t+1$, and/or observation or projections of the total number of members in H , state $H^{(k)}$ in t and $t+1$.

In the HMM we have a *sequence of T observations* from the total population, described by the semi-aggregate model of G and/or F , which sequence we denote as $g = [g_1 g_2 \dots g_j \dots g_T]$. In an HMM the observations are in random relationship with the hidden states from the perspective of the subject matter of the exercise, e.g. in our case average income of the observed group can be a linkage. Let the related hidden sequence of H denoted by $h = [h_1 h_2 \dots h_j \dots h_T]$.¹⁸

The HMM is parametrized with $b_i(k)$: the conditional probability distribution of observation G given the hidden state i : $p(G_j = g | H_j = i) = p(g | b_i)$, so that $p(G_j = g^{(k)} | H_j = i) = p(g^{(k)} | b_i)$

The HMM algorithm has the following parameters:

- (1) The initial probability distribution over the N possible hidden states H : $\pi = [\pi_1, \dots, \pi_N]^T$, where $\pi_i = p(G(1) = i)$ is estimated from statistics, etc.
- (2) A hidden-state individual transition probability matrix $P^H \in \mathbb{R}^{V \times V}$ that specifies the probability of transitioning from hidden state i to state j : $P_{i,j}^H = p(H(t) = j | H(t-1) = i)$.
- (3) The probability distribution of observations G : $p(g_j = h_k)$ given *hidden state i* ; is parameterized with parameter set $b_i(k)$ so that $p(G_j = g | H_j = i) = p(g^{(k)} | b_{i,k})$.

¹⁷ Discrete-time, finite natural number random variable Hidden Markov Models or Hidden semi-Markov Models

¹⁸ In practice the complete data set should be defined so that, given x , it is relatively easy to maximize $p(x | \theta)$ with respect to the parameter θ . Theoretically, the complete data X must satisfy the Markov relationship $\theta \rightarrow X \rightarrow Y$ with respect to the parameter θ and the observed data Y , that is, it must be that $p(y | x, \theta) = p(y | x)$. In the special case when the random variable of the observed data Y is a function of X , that is $Y = R(X)$, then $X \rightarrow Y$ is a deterministic function, and thus the Markov relationship holds by definition.

Maya R. Gupta and Yihua Chen: *Theory and Use of the EM Algorithm; Foundations and Trends in Signal Processing Vol. 4, No. 3 (2010) 223–296 DOI: 10.1561/20000000034*

Now for the HMM the complete set of parameters to estimate H is $\theta = \{\pi, P, b\}$, where $b = \{b_{i,k}\}_{i=1,k=1}^{V,N}$. The HMM algorithm estimates the three parameters $\{\pi, P, b\}$ by EM sequentially.

For example, a problem statement can be the following: we observe average wages in the insured status or household model and suppose it is an outcome of a hidden Typical Career Path model. We would like to find the best hidden state sequence $H(t)$ and hidden transition matrix P^H !

Micro models

Microsimulation models typically simulate behavioural processes such as demographic (e.g. marriage), labour market (e.g. unemployment) and income characteristics (e.g. wage).¹⁹ Methods use statistical estimates of parameters and simulation techniques to generate the new populations. The simulation models may include transition of individuals from one state to another (e.g. unemployed to employed) where a probability of making such a transition is assigned to each individual and/or household, given their underlying characteristics, described by the variables of the database of individuals.²⁰ Dynamic models build up complete synthetic life histories for each individual in the dataset, including data on mortality, labour market status, retirement age, income, savings, etc. The simulation of transitions might be modelled maximum likelihood methods, where higher probabilities will lead to more likely transitions.

While other microsimulation methods may exist²¹, the above general description might be consistent with the standard actuarial multi-state model defined as a Markov chain. Then again, we can use the EM Hidden Markov Model (HMM) to estimate the parameters and data points – if and when it becomes necessary – and we can apply the same method and dataset at all levels aggregation.

Conclusions

Several models are used to derive macro and micro indicators to analyse social security pension systems. Some indicators can be easily calculated from macro projection models, while others may require micro simulation. Usually the results correspond to each other, for example ARR and TRR move in line as a tendency in a country. But if a different database and calculation model was used, a relationship cannot be proven. Furthermore, if a microsimulation model is used together with macro projection, the results have to be adjusted. A wide range of so-called alignment methods have been developed and their theoretical foundations examined. The application of Markov chains, maximum likelihood estimates and optimization methods for such problems have been proposed in different papers.

A general framework of social security pension system projections can be established by using the Markov chain approach. At macro level the solution is straightforward and in fact that is already widely used based on the ILO and the Ageing Report macro model. However, this model should be extended by a macro household module for the purpose of supporting more detailed calculations. More granular projections can be performed at semi-aggregate level using the Expectation Maximization model. In case of the pension system the model could be applied to the Markov chain of insured status groups. A similar semi-aggregate model could also be set up for appropriately defined family or household categories. Finally, the semi-aggregate models can be used to derive Hidden Markov Models for studies

¹⁹ Li and O'Donoghue (2014)

²⁰ The methods use only a manageable size sample of the total population.

²¹ Other models, for example, simulate behavioural decisions relying on ex post statistical observations of past outcomes.

like typical career path or income inequality examinations. In this set-up the semi-aggregate model serves for observations and we would like to learn the hidden career/income categories. The hypothesis is that this framework can also be developed for other hidden or latent random variables. The relationship between these models is established and the connection between the states of the system at different levels of aggregation is secured by the underlying pension system rules.

The Markov chain is widely used in multi-state actuarial models and in several other disciplines. There is the possibility to adopt their results in our practice. The solution described in this paper is one of the possible approaches for social security pension system projections. This is to illustrate the point that a consistent modelling solution could be used from the beginning of the pension projection exercise to design the data collection, modelling and finally the calculation of the indicators.

References

[...]