Deloitte.

Feasibility study on the alignment of inputs and outputs of the MoLSA's microsimulation model of the pension system with external projections

Prepared by: Deloitte Advisory s.r.o.

Prepared for: Ministry of Labor and Social Affairs (project VS/2018/0380)

Project VS/2018/0380 "Development of microsimulation tools for social insurance projection (DEMTOP)" has been funded with support from the European Commission. This feasibility study reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Project VS/2018/0380 Development of microsimulation tools for social insurance projection (DEMTOP) and this document has received financial support from the European Union Programme for Employment and Social Innovation "EaSI" (2014-2020). For further information please consult: http://ec.europa.eu/social/easi The information contained in this feasibility study does not necessarily reflect the official position of the European Commission.

April 2020

Content

List of Tables

1 Summary

This study describes methods that can be used to adjust the calculation of retirement benefits used in the NEMO pension insurance microsimulation model to correspond as much as possible to the projections of retirement benefits generated by other entities, thus bridging the differences arising from the use of varying assumptions and modeling principles. The aim is not to refine the current calculation in the NEMO model, but to modify it so that the results may compare to, and reasonably complement, those of other projections.

At the beginning, we describe the methodology of the Ageing Working Group (AWG) and Czech Fiscal Council projections, which the representatives of MoLSA consider important for their future work. We also explain how the economic scenario is generated in the current NEMO model.

The more variables we want to control, the more complicated the calibration is; therefore, we discuss below the conditions under which two projections can be considered aligned. We find that for many variables, consistency can be achieved by simply adopting certain inputs and assumptions. Among the remaining factors, we identify assessment base, employment and unemployment as the most important. We recommend calibrating these three values using more advanced methods.

On the basis of extensive research of literature and foreign microsimulation models, we have described seven possible calibration methods in detail. Each of them is evaluated in general terms and in terms of implementation into the NEMO model.

From these options we have selected two methods that we recommend integrating into the MoLSA microsimulation model, one for the calibration of employment and unemployment and one for the calibration of income. We further describe the general calibration procedure and the specifics to take into account when aligning the model with the AWG projection or the Czech Fiscal Council calculations.

In conclusion, we present a timetable for the implementation of selected methods into the current NEMO model.

2 Initial Situation

In this chapter we describe the purpose of the study, the reasons why MoLSA wants to proceed with the calibrations, and the requirements it lays down for them. This sets the starting points for all subsequent chapters.

2.1 Purpose of the Study

The purpose of the study is to analyze the possible ways in which MoLSA can align its calculations with the external projection specified and to serve as a basis for possible later implementation of the selected calibration method. Its main goal is to provide MoLSA primarily with an overview of the possible solutions and their advantages and challenges and to recommend the most suitable approach. However, a part of this study which is no less important is the discussion of other possible options.

2.2 Reasons for Calibration

In addition to the calculations performed by MoLSA in its current NEMO model, discussions on state pension insurance are often based on other projections of the development of the Czech pension system. These alternative projections differ from MoLSA results in the data and assumptions used, in the methodology chosen and, ultimately, also in the nature of the results. The most important differences result from the fact that they are not typically microsimulation models but cohort models. Therefore, they lack detailed information on the probability distributions of the resulting variables across the population which are provided by microsimulation models.

Therefore, the purpose of the calibration is to add to the external projections this missing information, i.e., to determine the probability distributions of the resulting variables that would be simulated by the external projection if the projection were generated by a microsimulation model based on assumptions similar to those used in this external projection.

2.3 Theoretical Requirements for Calibration

In order to meet the calibration objectives described in the previous section, at least the following conditions must be met:

- The result of the calibrated run of the MoLSA model must match the values of the external projections for the important variables. The specification of the level of precision of the match we require and which variables we consider important is provided in chapter 3.1.
- The calibration should preserve as much as possible the probability distribution of the important variables and their dependencies. Only then the microdata added to external projections will be reliable and suitable for further analysis.

In addition to these two conditions, the methods will be evaluated, at the request of MoLSA, according to the following criteria:

- effectiveness,
- difficulty of implementation,
- financial demand/cost of the implementation, and
- user comprehensibility.

2.3.1 The Calibration Does Not Change the Model Functionality

The description of methods in this document is based on the assumption that during the calibration, the aim is the model results approaching the external projection through appropriate adjustments and reconciliation of model inputs and probabilities and the occurrences of the modeled events, provided that the formulas used for modeling interactions and amounts, i.e., in particular the formulas for determining entitlement to a particular transition, entitlement to individual benefits, or formulas for determining benefits, will remain unchanged.

In particular, this document does not discuss the possible approaches to reconcile the results using approaches such as:

- introduction of a scaling factor in the formula determining the amount of pension, which would result in this formula deviating from the formula valid in current legislation;
- adjustments to functionality (e.g., formulas for determining the benefits) so that the functionality approaches or matches the formulas in the external model.

2.4 Description of Important External Projections

Although a larger number of projections of the pension system or parts thereof can be found, discussions with MoLSA representatives have revealed two projections that are particularly relevant to their needs:

- Projection of the Ageing Working Group for the needs of the European Commission, and
- Projection of the Czech Fiscal Council.

Therefore, when assessing the suitability of calibration methods, we will pay particular attention to how well the MoLSA calculations approach these two projections specifically.

2.4.1 Ageing Working Group (AWG)

The main result of the Ageing Working Group projection (Ministry of Finance, 2017) is the projection of pension expenditure as a percentage of GDP divided by type of pension.

The basic input to this calculation is two external projections:

- Demographic projection of EUROSTAT, and
- Projection of the labor market of CSM (divided into cohorts), including, for example, projections of the average salary, average duration of pension insurance, degree of economic activity, rate of employment, personal assessment base, and the like.

Total expenditure on a particular type of pension is calculated as the sum of the number of pension benefits awarded in the past multiplied by the average amount of the previous year's benefit and the valorization index and the number of newly awarded pension benefits multiplied by the average amount of the pension benefit in the current year. Calculations are performed separately for each cohort represented by year of birth and sex.

The average amount of newly awarded benefit is determined by weighing the number of newly awarded pension benefits, determined on the basis of the personal assessment base and the period of participation in the insurance, by the number of recipients of benefits.

An important input for determining the number of newly awarded pension benefits is the matrix of the number of new pensions (divided by type of pension) for given combination of personal assessment base and insurance period.

The projection assumes that the effective retirement age will always be below the legal threshold in the future. When a person reaches this effective age, he or she always retires early (disregarding the fact that waiting until the statutory retirement limit could be economically more advantageous for a person).

2.4.2 Office of the Czech Fiscal Council

The main result of the projection of the Office of the Czech Fiscal Council (Hlaváček, et al., 2019) is determination of the average amount and total amount of pension benefits paid, divided by the type of pension. The total amount of expenditure on specific pension benefits is calculated on the basis of number of pensions and the average amount of pension (distinguishing between the amount of the newly awarded benefit and the amount of the benefit awarded in the past).

The starting point of this projection is several versions of the demographic Population Projection 2018 – 2100 from the Czech Statistical Office (divided into cohorts).

The number of old-age pensions is divided by sex and is obtained from the rate of retirement and the size of the population concerned. Retirement rate is the ratio of the number of people of a given age who receive old-age pension to the total population of a given age (excluding the disabled). In addition, the retirement rate for women is unified for the old-age pension – it is assumed that each woman raises two children. In case of disability pensions, pension rates are differentiated according to the degree of disability. For survivor pensions paid individually, a constant proportion of pensioners in the population is considered; survivor pensions that are paid in parallel with old-age or disability pensions are also modeled on the basis of the retirement rate.

The amount of newly granted old-age pensions is determined as a percentage of the average income. This percentage is determined separately for women and men and its development during projection is part of the model's assumptions (i.e., it is not counted in the projection itself). The pension is then valorized annually by a percentage dependent on the calendar year.

The number of terminated old-age pensions is determined on the basis of mortality tables; the number of new old-age pensions is then determined to keep the current state in line with the rate of retirement. The amount of terminated old-age pensions in the simulation is set at 95% of average old-age pensions.

For the projection of disability pensions, a constant ratio between the average disability pension of a certain degree and the average old-age pension is always assumed.

The average amount of survivor pensions is determined as the average of the average old-age pensions for the last three years multiplied by the respective entered coefficient according to the type of pension.

The costs for a given pension benefit are calculated on the basis of known average amount of a particular type of pension and the number of pensioners.

The labor market forecast can be derived from other assumptions: the estimated number of employed individuals in the population can be calculated on the basis of the total income of the pension system and the average wage.

2.5 Creating the Economic Scenario in the NEMO model

Now that we know what projections the calibration will want to approach, we will look at the other side of the task and describe the inputs and assumptions that govern the projections in the NEMO model. This will subsequently make it easier for us to assess how essential the different properties of the described calibration methods are in the specific case of the model.

The term ‘economic scenario’ refers to the results of the projection of the pension system as a whole: the number of pensioners and the amount of money spent on their pensions, the number of working individuals and pension insurance payments collected from them, the number of persons who neither receive pension nor pay respective premiums. This scenario is the result of individual projections of individuals in the microsimulation model. The relevant part thereof is the life path of each person or, in other words, knowledge of the period when the person studied, worked, received unemployment benefits, received a pension, or was in another life situation relevant for the calculation of the pension insurance.

The life paths of individuals are randomly generated in the NEMO model using transition probabilities and events that may result in individual transitions. The parameters defining these processes are in many cases dependent on other variables (e.g., age, sex, education, income, etc.) or indirectly interconnected. This chapter describes the basic principles and effects that lead to transitions between the statuses.

2.5.1 Basic Principle

All calculations in the model are performed at the level of the model point which represents one individual from the population of the Czech Republic. Given the fact that in addition to the life path of the individual, some cash flows modeled also depend on their family, the model point includes the principal individual and several secondary persons (partner and children). However, the cash flows themselves are calculated only for the principal individual, because for each secondary person there is a model point where he or she is the principal individual. There are two basic variables which define the random processes described:

- STOCH_EVENTS:

In this variable, all the calculations associated with the generation of random events and their parameters are performed. The calculations are performed on a monthly basis and are applied to each person at the model point.

In the modeling process, the fulfilment of the requirements listed in table events_req.fac is first tested for each event. If any of the required conditions is not fulfilled, the event will have a zero probability.

If all conditions for the given event are met, the probability of occurrence is uploaded (see probabilities in table events.fac). A random number is generated and its value determines if the event will occur.

- STOCH_MOVE:
    - In this variable, all the calculations of the transitions between the working statuses of an individual are performed. The calculations are performed on a monthly basis for each living person.
    - It includes several methods of leaving a status:
        - Matrix method:
            - First, possible new statuses are identified based on the transition matrix;
            - Subsequently, based on a generated random number, a decision is made whether a transition to one of the possible statuses occurs, or whether the status remains the same.
        - Duration method:
            - At the moment of entering the status of this type, the status duration is randomly generated based on the defined distribution;
            - Subsequently, a test is performed every month whether the time in the given status has reached the duration. Then a return to the previous status occurs.
        - Combined method:
            - It combines the above-mentioned two approaches.

2.5.2 Modeled Phenomena

The modeled events can be divided into the following interconnected areas:

- Events;
- Career paths – they capture the economic (in)activity of an individual throughout his or her entire life;
- Family relations – they reflect the marital status of the individual and the number of children born and raised; and
- Cashflows consisting of:
    - the individual’s income;
    - payments to the pension system; and
    - payments of benefits from the pension system.

The following chapters analyze in detail the individual events and statuses modeled with emphasis on their assumptions and inputs affecting the respective probabilities of transitions.

2.5.3 Events

All important events (see table *events.fac* in the NEMO model) are randomly modeled on a monthly basis using defined probabilities. State variables are constructed based on these events (see table *status_vars.fac*); the career paths and family relations are linked with these status variables in the model. The list of events with their assumptions, parameters affecting the probabilities of their occurrence and the related state variables are given in the following table. The basic status variables (age and sex) on which most of the probabilities depend are not explicitly listed in the table.

Event	Related status/variable	Conditions of occurrence	Variables affecting the probability of occurrence
Birth	Living		
Death	Living		
Completion of sudies	Student		
Occurrence of disability	Disability	Legislative conditions	
Change of the degree of disability	Degree of Disability		
Cessation of disability	Disability		
Marriage	Married		Numerical rank of a marriage
Divorce/Widowing	Married		Numerical rank of a marriage
Birth of a child	Care of a child		Number of children already born
Termination of care of a child	Student		
Start and end of care of a family member	Care of a family member		Number of years since the last child was born
Retirement	Old-age pensioner	Legislative requirements	Number of years from/to entitlement to a regular pension
Emigration	In the pension system		
Change of the possibility of concurrent work and receiving an old-age pension	Choice of the possibility of concurrent work and receiving an old-age pension	Status of old-age pensioner	Type of old-age pension (regular, late, early))
Change of salary	Salary		

Table 1: List of events captured in the NEMO model and their characteristics

2.5.4 Career Paths

A career path for a model point is a sequence of statuses of economic activity and inactivity (see table *statuses.fac*). Change of the status triggers the events occurrence. In some cases, statuses are divided into sub-statuses (see table *sub_statuses.fac*). A list of statuses and sub-statuses of the career path is presented in the following table. The probabilities of transitions between them depend on age and sex.

State	Sub-state
Employed	Healthy
	Sick
Unemployed	Registered with the employment office and receiving unemployment benefits
	Registered with the employment office without unemployment benefits
	Other cases

Inactive individual	Registered with the employment office and receiving unemployment benefits
	Registered with the employment office without unemployment benefits
	Not registered with the employment office
Person outside the pension system	Emigrant
	Member of armed forces
Deceased	

Table 2: List of states and sub-states of an individual in the NEMO model

The reasons for inactivity are also modeled for inactive persons (see table *inact_periods.fac*). The modeled reasons for inactivity are as follows:

- Taking care of a child;
- Disability pensioner of degree 1-3;
- Old-age pensioner;
- Student;
- Child; and
- Taking care of family.

2.5.5 Family Relations

Events leading to changes in family relations (e.g., marriages, divorces, and remarriages) are randomly modeled based on assumptions regarding marriage rates and divorce rates. One person is always assigned the same partner in the model, whose parameters are defined within the creation of model points. The principal individual has the partner assigned already at the beginning of the projection, even if he or she is single or divorced. Similarly, the principal individual's partner remains the same also in the event of remarriage after divorce or widowhood.

Another important event is the birth of a child. This is modeled on the basis of fertility rates of a woman in the pair. Career paths of children are modeled in a simplified way, the calculation takes place in variable STOCH_CHILDREN using 4 statuses:

- Dependent child;
- Independent child (e.g., earning income or older than 26 years);
- Orphan (here a distinction is made between a half-orphan with the principal individual deceased, a half-orphan with the secondary person deceased and a double orphan);
- Deceased.

Parameters affecting the probabilities of transition between the statuses mentioned in this chapter are determined only by external statistics (e.g., fertility rate, divorce rate, number of orphans, etc..) and basic status variables such as sex and age.

2.5.6 Cash Flows

The following cash flows are modeled:

- Gross monthly salary determined on the basis of:
    - Wage inflation;
    - Career growth; and
    - Economic status.
- Payments to the pension system:

    - According to legislative rules;
    - Derived from the individual’s income.
- Old-age pensions:
    - According to legislative rules;
    - Derived from the individual's income history, years in work and, for women, the number of children raised.
- Disability pensions:
    - According to legislative rules;
    - The admissible status is disability pensioner.
- Widow’s pensions:
    - According to legislative rules;
    - The admissible status for the entitlement is married.
- Orphan's pension:
    - According to legislative conditions.

Size of the majority of modeled cash flows is therefore determined, given the knowledge of several status variables, based on legislative conditions. Only in case of salary, the calculation takes into account additional random effects.

2.5.7 Salary Modeling

The starting salary of an individual after completion of studies is a value which serves as an input parameter of the model point. In case that an individual works before completion of studies, his or her work is considered a short-term termporary job (typically with lower salary), and after graduation their salary jumps to a value from the model point. Salary increase occurs once a year and comprises three components:

- Wage inflation (average wage growth in the economy);
- Career growth:
    - Two versions of modeling:
        - Stochastically based on distribution of salary; or
        - Deterministically from average age-dependent growths.
    - Controlled via variable USE_STOCH_SAL_GTH in the table *global*;
    - Possibility to add dependence on the amount of salary via variable USE_SAL_DEP_SAL_GTH from the global table.
- Salary decrease due to inactivity and unemployment:
    - Dependent on age.

In order for the average salary growth in the projection to really correspond to wage inflation, a so-called residual wage inflation calibration is being introduced to correct the deviations, if any, from this average.

For more information on this mechanism see chapter **Chyba! Nenalezen zdroj odkazů.**.

2.5.8 Summary

The following assumptions are essential for the economic scenario:

- Demographic assumptions, i.e., mortality, birth rate, disability tables
- Assumptions for modeling life paths, i.e., transition probabilities between different statuses
- Parameters determining the individual's income, i.e., wage inflation and prerequisite for career development.

Many of these variables are determined as a result of the interaction between individual transition probabilities. It is therefore not easy to calibrate them separately. The most important of these are the number of employed and unemployed individuals, and the occurrences also concern, for example, the number of individuals on sick leave.

3 Calibration Suitability Criteria

While in the previous chapter we described the general situation in which the calibration takes place, here we set out the specific criteria by which we will select and evaluate the calibration methods in the following sections.

3.1 Criteria of Consistency of Two Projections

When assessing the consistency of two projections, a balance needs to be found between two effects. On one hand, we want to consider as many variables as possible, because no variable can guarantee consistency individually. For example, two projections, which coincidentally have the same expenditure on the pension system but have quite different population size, are certainly not considered to be reconciled and therefore no micro-data can be assumed/adopted between them.

On the other hand, the calibration becomes significantly more complicated with each addition of an extra variable. With a new variable, not only will the targets monitored be added and the calibration will therefore have to be finer, but at the same time, we will not be able to use the calibrated variable to achieve the desired values for the other target variables. For example, if only the total amount of old-age pensions paid is specified as a calibration target, it can be controlled by adjusting the number of pensioners, salaries or insured periods in the history of individuals. If we wanted to calibrate all these four variables, we could get into a situation where the number of pensioners, salaries and insured periods can be calibrated, but the total pensions paid still do not match. Then it will be necessary to adjust the other variables such as the length and type of the different substitute periods or the education level achieved, which have a less direct influence on the amount of the pension and which also affect the other variables we want to calibrate.

We will partially solve this problem by dividing the calibration into two parts (see also chapter 4.1). In the first part, we only take the inputs that can be directly supplied to the projection in the NEMO model. This alone will help us achieve a part of the calibration targets. In the second part, we will use selected additional methods for a smaller number of key variables.

3.1.1 Variables Relevant for Calibration

Now, for the most important variables from the external pension projections, we will assess how relevant they are for the calibration result. Then we will select from them a final list of variables that we want to monitor within the calibration.

Volume of pensions paid

Since this is the main result of both the NEMO model and both external projections, it is necessary to monitor this variable. At the same time, however, its value is directly determined by other variables, such as the number of pensioners, the average wage and, in particular, the lifelong career paths of individuals. There is therefore no point in calibrating directly the volume of pensions paid as such – by doing this, we would separate the amount of pensions from the variables on which they depend and lose the opportunity to explain the results using other parts of the projection. This decision corresponds to the requirement laid down in chapter 2.3.1.

It is therefore better to focus on the variables that determine the amount of pensions. Once these are completed, the volume of pensions paid should also match the external projection. If this is not the case, other variables that have not been previously included in the calibration may need to be addressed; or it is possible that the calibration target is not actually achievable at all (this may occur, in particular, because in the microsimulation pension, projections of pensions are very closely linked to the individual's work history and hence the past payments to the system, whereas in macroeconomic models, individual variables are projected more independently, and thus with a greater degree of freedom in the development of individual variables).

Average amount of pensions paid

Similarly to the volume of pensions paid, here, too, it is more appropriate to calibrate the variables that determine the amount of the pension paid.

Volume of payments into the system (pension insurance payments)

In the NEMO model, this is another key output for assessing the system's balance, so it is necessary to monitor this variable. However, the amount of payments in the model is directly determined by statutory pension insurance rates, individuals' salaries and their employment. Direct calibration of this variable is therefore not possible, and it is necessary to calibrate salaries and employment.

Total size of the population

This assumption is crucial for any projection. It is also based on a small number of independent inputs: the initial population is determined by the database of model points and the future development is determined by the assumption regarding new individuals (who are also entered among the model points) and mortality tables. Calibration of the population size should therefore always be performed as the first step of all. We expect that in usual situations, it will be possible to achieve a state where both projections reasonably correspond to each other in terms of the number of people in all age cohorts for all calendar years.

Number of pensioners

At the time when the total population size has been calibrated, the number of old-age pensioners depends only on the probability of retirement and on the age at which people normally achieve the required insurance period.

The AWG projection assumes that people will always retire once they reach the effective retirement age. This is usually lower than the statutory retirement age in these projections. Theoretically, this effect could be fully transferred to the NEMO model by setting the probability of early retirement for the relevant age and calendar year combination to be set to 1 (table *retirement.fac*). This would result in a full balance of the number of old-age pensioners, under assumption that the periods of insurance correspond in both projections. However, we would lose the plasticity of the NEMO simulation, which is an essential advantage. Therefore, we will adjust the probability of retirement so that the average retirement age is shifted to the value that we see in the external projection, but at the same time, the distribution in surrounding ages remains as unchanged as possible. This method is described in more detail in chapter **Chyba! Nenalezen zdroj odkazů.**

The projection of the Czech Fiscal Council determines the number of new old-age pensioners based on the population size, mortality rates, and retirement rates (divided by cohorts). These values can be used to calculate how many new old-age pensioners are assumed by this projection for each calendar year. As soon as we find out what part of the cohort has met the legal conditions for old-age retirement, these numbers of new pensioners can be replicated by appropriately adjusting inputs to Prophet. For more details see again chapter **Chyba! Nenalezen zdroj odkazů.**

In both external projections, the number of disability pensioners depends on disability rates divided by age and gender. For AWG projection, we use data about newly approved retirements sorted by type of retirement, age and gender. Therefore, the probabilities of origin of disability can be easily calculated as a ratio of new disabled persons and the origin population without already existing disability and old-aged pensioners (this information is also included in AWG projection). We will determine the probability of termination of disability conversely, so the total number of disabled persons after application of mortality rate and terminated disabilities is matched.

However, the projections of Czech Fiscal Council do not feature the number of newly incepted disability pensions, so we need to find another way, further specified in chapter 4.1.

Widow's pensions are significantly less relevant than old-age pensions. The advantage is that their number depends, apart from variables that must be calibrated also the purpose of old-age pensions (in particular mortality and salary), only on the marriage rate and the method of assigning partners, which in turn do not translate into any other substantial outputs of the model. Calibration of widow's pensions can therefore be included in analyses for which this type of benefit is relevant.

Orphans pensions are not very significant in terms of the number or amount of benefits paid. In addition, their number depends, among other things, on the birth rate which also determines the number of parental leaves, which is generally a more important parameter. Attempting to calibrate the number of orphans' pensions could hinder the calibration of the number of parental leaves without generating adequate added value in itself. We therefore recommend not to include them in the calibration at all.

Number of people standing out of the retirement system

The next group of individuals are the ones outside the retirement system. In particular, these are members of armed forces whose pensions are calculated differently from pensions of the rest of the population. Although, sometimes their primary data is missing (they do in case of both used external projections), it can be useful to compute them from other available data and further calibrate them as so-called residual group (see chapter 4.2).

Average salary in the population

The average salary is a key input into any projection of the pension system, as each person's pension depends directly on his or her previous income. It is present in the AWG projection as well as in the calculations of the Czech Fiscal Council. Therefore, reconciling the average salary should always be part of the calibration.

At the same time, it is a value which depends, in the NEMO model, not only on the starting values and the wage inflation, but also on the number of unemployed individuals and their characteristics. It is therefore indirectly affected by values such as probability of finding a job or probability of taking a parental leave. Therefore, we expect the use of more sophisticated methods to calibrate this variable.

Number of patients

The number of patients does not occur in any of the known external projections. In the NEMO model, it depends on the income of the pension system - the person does not contribute to the system of incapacity for work. However, since there are typically not many patients, this effect is not significant. Therefore, we will not calibrate the number of patients or we will use it only for fine tuning at the very end of the calibration.

Number of employed individuals

The number of employed individuals is one of the basic characteristics of the labor market. The alternation of periods of employment, unemployment and inactivity fundamentally affects the amount of the individual's pension because the pension is directly proportional to the length of insurance which is largely achieved by employment. Without earning a sufficient insurance period, no pension entitlement arises. Therefore, it is always appropriate to consider calibration of the number of employed individuals. In the next chapters, we will focus on their specific implementation and its steps as well.

In the NEMO model, the number of employed individuals depends primarily on transition probabilities: the probability that an employed individual will lose job and that an unemployed or inactive individual will find a job. Other essential assumptions include mortality and disability rates, length of studies for students, or the number of people taking parental leave. Moreover, we differentiate sub-status healthy and sick. This complexity means that calibration cannot be performed by simply adjusting the input parameters.

While in the AWG projection, the number of employees is available divided by cohort, the Czech Fiscal Council does not model it explicitly, and only the total number of employed individuals in the population can be determined. Therefore, in their projections this variable cannot be used as a criterion of approach of projections.

Neither one of the projections differentiates employment and self-employment. Therefore, we will consider both these groups together in all applications.

Number of self-employed persons

As we have already mentioned, neither AWG projection nor the Czech Fiscal Council includes self-employed projections, so their numbers need to be combined with numbers of employed persons.

In the NEMO model, the behavior of self-employed persons differs from employees, particularly noticing that the self-employed have lower incomes. Thus, if a model for a cohort assumes a different ratio of self-employed to employees than is implicitly surmised by an external study, this may result in an average salary mismatch. However, as we have no way of knowing whether the discrepancy was due to differences in the numbers of self-employed or in the salary projection, we will choose a simpler method and straighten this difference solely by salary calibration.

Therefore, we will not calibrate the number of self-employed per se.

Insurance period

The amount of the old-age pension directly depends on the length of insurance and a sufficient period of insurance is necessary for the entitlement to the (old-age or disability) pension to arise at all. It is therefore one of the most important values in any pension system model. Its amount is derived from the time spent by the person in employment and the substitute insurance periods he or she can obtain – unemployment and parental leave are particularly important in this context. Thanks to this, the length of the insurance can be skipped in the calibration and one can focus directly on the calibration inputs: employment, unemployment and time spent on parental leave. A similar approach is used by the projection of the Czech Fiscal Council where the length of insurance does not explicitly appear at all and the amount of pension is calculated directly as a percentage of the average salary, the amount of which is part of the model inputs. This percentage mainly includes the assumption for the average length of insurance.

It is therefore recommended not to include the insurance duration in the calibration and to focus on the variables on which it is based.

Number of unemployed individuals

The distinction between the unemployed (i.e., those seeking employment with the help of the employment office) and other non-working individuals is important in the pension model, in particular because the unemployed individuals receive, under certain circumstances, a substitute period of insurance. The way of modeling unemployment thus significantly affects the level of pensions in the projection. It is therefore appropriate to consider the calibration of the unemployed individuals.

Nevertheless, in terms of importance, it is a second-rank variable. Unemployment brings a substitute period of insurance only for a limited time (only as long as the person is entitled to an unemployment benefit and, on top of this period, for no more than three years in a lifetime), and so its impact on the pension is limited. At the same time, even if we do not calibrate unemployment, there are not so many ways to classify individuals incorrectly in terms of unemployment decisions (meaning inconsistently with the projection we want to approach). Assuming that both the total population and the number of employed have been calibrated, the remaining persons can be, in particular, unemployed, on parental leave or inactive. During parental leave, the parent also obtains a substitute insurance period, so any shortcomings in the numbers of the unemployed can be compensated for by the length of parental leave (although, of course, there is no guarantee that this compensation will always occur). So there is reason to believe that if we do not calibrate the number of unemployed individuals, the projection of MoLSA calibrated in other ways and the external projections we want to approach will still be quite close in normal circumstances.

Theoretically, the calibration of number of unemployed persons might be used for example to influence period of influence. However, we strongly do not recommend such approach. Due to non linear relation of these two variables (both in terms of time and implementation) it would be very difficult to try to reach the calibration goal this way.

Therefore, we recommend monitoring the differences in the unemployment rate between the two projections and implementing calibration only when we notice a very significant difference.

The calculations of the Czech Fiscal Council do not include the number of unemployed individuals, so it cannot be used for calibration.

Average duration of unemployment

The average duration of unemployment is closely related to the number of unemployed individuals, but it is still a figure that can influence the model's results: if the average length of unemployment is short and unemployment is distributed among a large number of people, virtually everyone will have a sufficient insurance period and everyone will have the period of unemployment counted as a substitute insurance period. Conversely, if the average unemployment period is long, a group of long-term unemployed individuals with shorter insurance periods will arise, many of whom may not even become entitled to the pension. It is sensible to consider calibration of such data. In normal circumstances, however, such extreme cases should not occur, so the calibration of this variable should not affect the results too much.

In the NEMO model, transition to and from unemployment is governed by the probability of transition. These transition probabilities depend on each person's age and sex, so the overall average depends on the number of people in the population. This may change during the projection. Calculating the average length of unemployment would therefore not be straightforward. Given that the calibration of this variable will have a rather minor effect, we recommend that you do not perform it in the first step and only perform it if the pension system expenditure differs significantly from the external projection even after the other variables have been calibrated.

This information, however, cannot be read from the projections of AWG or of the Czech Fiscal Council.

Number of individuals on parental leave

Like unemployment, parental leave is a condition from which a substitute period of insurance may be granted to a person who is not working. The number of persons on parental leave therefore does not reach the importance of the number of employed individuals and should be calibrated only in the second rank, but it is nevertheless a variable that may have a noticeable effect on the model results, its calibration cannot therefore be disregarded at once.

The numbers of parental leaves depend on two figures: the birth rate which determines their beginnings and the average duration thereof. The birth rate can be calibrated with great accuracy by merely taking the inputs, so there should be no fundamental differences in the number of parental leaves in normal circumstances. The duration of parental leave is based on the probability that the mother will find a job or go into unemployment status. These are part of the NEMO model inputs and are entered according to age. Therefore, in the context of taking the inputs, these probabilities can be adjusted so that the average duration of parental leave matches the external projections.

Neither the AWG projection nor the projection of the Czech Fiscal Council, however, operate with the average duration of parental leave. Therefore, in case of these two external projections, the number of persons on parental leave cannot be calibrated. However, if there is ever a calibration against a projection that has the necessary inputs, we recommend adjusting the probabilities of exiting parental leave included in the NEMO model based on this input.

Number of persons providing another care

The provision of care for close associates influences affects the amount of the pension paid, but it occurs only rarely. The external AWG projections and the projection of the Czech Fiscal Council do not include this information into their simulations, so there is no need to include this variable in the MoLSA calibration.

Number of inactive individuals

Inactive individuals are complementary to all other statuses for our purposes - in this group, we include individuals who are not employed or unemployed, do not study, do not receive a pension and are not on parental leave. These may be, for example, individuals who live on savings, who have remained in the household after parental leave, or who take care of a sick family member. Since this group is defined by the absence of another status, we will not calibrate it and instead, will focus on other statuses where we can use specific transition probabilities and conditions of occurrence for the calibration. Once all other statuses have been reconciled between the two projections, the number of inactive individuals will also match.

Number of students

Information on ongoing studies of an individual is relevant in the NEMO model during the derivation of model points where it is used as a basis to determine the expected future highest educational level achieved. It is less relevant during the projection; it is in particular manifested by the fact that students receive lower salaries and slightly different events may occur for them. Otherwise the student's behavior does not differ much from a non-student – he or she can become employed, unemployed and inactive, and their transitions between these statuses are with similar probabilities.

There is also a method in the NEMO model where studies play a more prominent role: when chosen, the working student will remain at work until he or she completes his or her studies. The method is currently not used. If MoLSA analysts choose to work with this method in the future, the impact on student employment will need to be considered: students who entered the model as working students cannot be excluded from this status by calibration. All the more reason to adjust the transition probabilities of the other members of the respective cohorts.

Neither the AWG projection nor the projection of the Czech Fiscal Council provide the number of students. Moreover, as working students do not have a major influence on the modeling of pensions, we recommend focusing on other variables when calibrating.

Marriages

Modeling of married couples is essential in the NEMO model mainly because of the prediction of widow and orphan pensions. However, old-age pensions, which are only marginally influenced by marital status, are mainly important for calibration. Therefore, we will not calibrate any of the prerequisites for modeling marital status (especially marriage rate, divorce rate and the number of married people in the initial population).

Overlapping of variables

Some people may be in multiple statuses at the same time, for example an old-age pensioner can work at the same time. The number of such persons can play a role even if both variables have been calibrated separately – for example, it may present a difference if the old-age pension within the group of 70 year old working individuals is received by everyone (maximum overlap) or no one (no overlap).

However, the concurrence of old-age pension and employment is the only important overlap that can occur. The statuses of unemployed, on parental leave and inactive are mutually exclusive, and they are not permitted to overlap with employment or with old-age pension. The concurrence of first or second degree disability pension with employment is not excluded; however, disability pensions are not frequent and are distributed across all age categories, so any mismatch in overlapping modeling will not be so manifest.

As a result, the overlap rate is evident once we know the numbers of people in all the major statuses that occur in a given cohort. For example, if a certain group is supposed to have a total of 100,000 people and 80,000 of those are pensioners, 10,000 inactive individuals and 20,000 employed individuals, then some people must necessarily be in multiple categories (because the sum of numbers in individual statuses exceeds the number of people in the population), and since the only possible concurrence is that of old-age pension and employment, we will calculate immediately that there must be 10,000 working pensioners.

There is no need to calibrate the number of people in multiple statuses.

3.1.2 Variables Recommended for Calibration

Developing on considerations in the previous chapter, we can see that calibration can be done in two steps.

In the preparatory phase, we expect that it is possible to calibrate the following by taking the inputs and input assumptions correctly:

- Initial population;
- Birth rate;
- Mortality;
- Disability rate (and hence the number of disability pensioners);
- Number of old-age pensioners (here we will not achieve an exact match, but at least an approximation).

In the second phase, we recommend focusing on the following using more sophisticated methods than simply taking over the inputs:

- Average salary in the population;
- Number of employed individuals.

By successfully calibrating these values, we can achieve an approximation also in some other variables without having to explicitly deal with them:

- Calibration of the number of employed individuals automatically entails proximity in the number of unemployed individuals and in the period of insurance;
- Match in a population projection shows side effects in similarity in the number of parental leaves;
- Approximation of the population, salaries, numbers of pensioners and periods of insurance should also help approximate the total income and expenditure of the pension system.

However, full match of the system revenue or expenditure modeled may not be achievable because the internal functioning of the micro-simulation model and cohort macroeconomic models is different (in particular because the microsimulation model is based on a firm link between the whole work history of an individual and his or her pension and also takes into account the distribution, not only the income averages). In certain cases, depending on the methodology of the external projection compared, calibration, for example, of salaries and employment in the first years of the projection may even cause greater pension variations in the subsequent years.

Another point of risk may present itself in setting the ratio between unemployment, duration of parental leave and inactivity, for which no requirement is laid down by the above-described calibration (in this case, inactivity includes all periods without pension insurance, including, but not limited to, studies). As a result, there may be a difference in the insured periods and subsequently also in the amounts of pensions between two projections even after calibration (parental leave is always insured, inactivity is never insured, and unemployment is insured only for a certain period of time). If this happens, the calibration can be refined by including unemployment.

3.2 Calibration Achievability Criteria

Some of the variables we mentioned in the previous chapter are very complex in terms of data they depend on. For example, a person can only retire after having worked for an appropriate period of time, so the number of pensioners depends directly on the person's work history (including employment, sick leave, parental leave, etc.). Because of this complexity, calibration is generally a very difficult task in the context of the NEMO model. In order to maintain the internal consistency of the model, it may be necessary to change all the components of the person's life path in addition to the calibrated variable itself.

In some cases, calibration may not be possible at all. This happens when the external projection target values lie outside the values that the NEMO model can achieve within the implemented rules. Examples include a hypothetical situation in which an external source would forecast very low average salaries, a high number of old-age pensioners, and very high old-age pensions: in addition to salary, the pension depends only on the time worked. A very high pension can be achieved by working significant number of extra years, but such a solution would be inconsistent with the requirement for a high number of pensioners. Such a hypothetical external projection is therefore not internally consistent with the rules in the NEMO model and cannot be achieved by means of calibration.

The range of achievable values can be determined by running the model on extreme assumptions selected to always maximize or minimize one of the states monitored. This test does not need to be performed before each calibration, as there are many runs and in most real cases, the calibration target should not be too far from the model results. However, it can be a useful method if the suspicion of unreachable targets arises for other reasons, such as a situation where the calibration has failed.

Below we will describe how to determine the range of achievable values for each calibration target.

Demographic assumptions

The initial population should not need to be calibrated because this is a known figure. Theoretically, however, it is possible to enter any person into the model and thus achieve any values.

Similarly, any values can be entered for mortality and birth rates, differentiated by age cohort, sex and calendar age, thus achieving any development in the number and age structure of the population.

Number of pensioners

An old-age pensioner may be an individual who has acquired sufficient insurance and substitute insurance period during his or her life. Retirement after the fulfillment of entitlements then depends on his or her decision and in the model, it is driven by the probability dependent on distance from the statutory retirement age. If we set the probability of retirement to zero, we can even achieve zero number of new pensioners. The number of existing pensioners develops with mortality. Increasing the number of pensioners is significantly more difficult. We get the maximum number of new pensioners by setting the probability of retirement to 1 (i.e., everyone retires early as soon as they meet the legal requirements), and by ensuring that the legal conditions are met as quickly as possible. For this purpose, we set the probability of leaving employment to zero, set the probability of transition from unemployment or inactivity to employment to 1, and set the probability of taking a parental leave to zero. This setting will have a drastic impact on the younger generations, but as we do not rewrite history, older individuals will be significantly less affected. Thus, the achievability limits for the number of pensioners in older cohorts may not be very far from the original results of the uncalibrated model.

Theoretically, the number of disability pensioners can be manipulated arbitrarily by changes in disability rates. Any number is thus achievable in both directions. In addition, the reduction in disability pensions is due either to mortality rates that are consistent with a healthy population or to the loss of disability. If we zeroed the probability of the occurrence of new cases of invalidity and set the probability of invalidity to 1, we would achieve zero number of new disabled. However, such a scenario is unrealistic in practice. In order to obtain a more realistic estimate of the permissible values, we will at least maintain the initial numbers of invalidity by setting zero probabilities for the occurrence of disability and its termination. Then the number of disabled pensioners will naturally decrease according to mortality rates. Theoretically, it is also not difficult to modify the model so that the invalids die according to their own mortality, which can then be manipulated to achieve a lower number of disability pensions. It should be noted, however, that the invalidity limit calculated in this way is not strict and, if exceeded by external projection, it is not necessarily an error.

If it is required to apply a mortality table or a disability table that is obviously incompatible with reality when assessing the achievability of the number of old-age and of disability pensioners,, it is appropriate to search in the external projection methodology instead and verify that the same values are being compared – for example, it would be possible that the external projection of the disability pensioners intentionally models only a part.

Average salary

As mentioned earlier, the average salary depends not only on the initial values and input macroeconomic assumptions, but also on the number of employed individuals and their characteristics.

Theoretically, of course, any average salary can be achieved by simply setting salary inflation to an extremely high level or to a value very close to zero. However, such a result is not very informative, so we will now describe an approach where we avoid extreme interventions in the salary inflation.

We will take advantage of the fact that the average salary may vary considerably for different population groups. We will therefore attempt to adjust the inputs in such a manner that low-income individuals leave the job.

- We set the probability of retirement to 1 (i.e., they everyone retires as soon as they are entitled to it) and we will prohibit the concurrence of receiving pension with gainful activity. This will exclude not only working old-age pensioners, but in particular working disability pensioners who typically have lower monthly salaries.
- We let working students leave the job immediately after the start of the screening and set the probability of finding a job while studying to 0. This will exclude working students; the students will enter the labor market only after completion of their studies with their full starting salary.
- We will set the salary decrease in the event of one-year unemployment/inactivity to 0%, thus ensuring that the starting salary of a newly employed individual who has come from an unemployed or inactive status will not be affected by his or her previous work history.
- We leave the residual inflation with the values used for common calculations.

After this transformation, only full-time employees who have had their annual salary increased on a regular basis or at least not reduced will remain in the population.

The lowest average salary can be achieved by the following settings:

- We will set the probability of retirement so that everyone works five extra years. The percentage of career salary growth is often negative for the elderly.
- We will leave the probability of overlap of working and receiving pension according to the settings for the common calculations of MoLSA. As a result, working disability pensioners, who generally have significantly lower salaries, will remain among working individuals.
- We set zero probability of leaving the job without an event for the working students (however, we will not interfere with leaving the job as a result of an event). Working students receive lower income compared to the earnings they receive after graduation.
- We will leave career growth and wage inflation unchanged (even the assumption of a constant salary in the course of the entire life is not very realistic and it is not clear how to determine the lowest meaningful year-on-year increase).
- We leave the residual inflation with the values used for common calculations.

Thus, with the model values set in this manner, we maximize the number of working individuals from low-income groups. This ultimately leads to the minimum average salary.

Number of employed individuals

The achievable values of the number of employed individuals can be examined from two points of view. If we are interested in purely mathematical extremes based solely on model properties and disregard the feasibility of our set values, we can achieve both full employment (i.e., all individuals in the population work, including pensioners, students or mothers immediately after birth), and zero employment (i.e., nobody works). This finding serves primarily as a warning that input parameter settings which are not well thought out can lead to completely unrealistic values. As the next task, we will try to find the limits of achievable results with input values consistent with reality.

Important factors affecting the number of employed individuals are not only the probability of transitions between working statuses, but also the probability of retirement or the probability of taking a parental leave.

To achieve full employment, we set the probabilities of transition from the inactive / unemployed status to the employed status to 1 and, conversely, set the probability of loss of employment to zero. Likewise, we set the probability of working concurrently with (old-age and disability) pension as equal to 1. Another equally important factor is the care of a family member. In order to achieve a 100% workforce, we set the probability of transition from care of a child or care of family to employment as equal to 1, irrespective of the time since the last child was born. On the other hand, we set the probability of the individual taking a parental leave when a child is born to zero. A more radical solution can also be implemented, namely setting the probability of the birth of a child to zero, thereby cancelling the women's departure to maternity leave and subsequently parental leave. If the population modeling is determined in this way, the development of the number of employed individuals will depend only on the birth rate, mortality and migration balance.

In order to achieve the minimum number of employed individuals, we substitute the above parameters with parameters of the opposite value.

In order to achieve a "logical" maximum for the number of employed individuals, we recommend not to interfere with the probability of occurrence of events (e.g., birth of a child) and only set the probability of transition from unemployment to employment to 1 and also set the probability that a student will be employed after completing education (transition linked to an event) to 1. We will not allow retirement until the person has worked five extra years. On the other hand, the probabilities of transitions from inactivity to employment will remain unaffected, in particular, the probabilities of returning to employment after the birth of a child and the probability of old-age pensioners working will be preserved. This setting thus describes a very high demand for work, but at the same time respects life situations in which starting an employment is unlikely.

Finding the opposite limit is more difficult, as it is not clear how low the demand for work can be in reality. Therefore, we will maintain the original probability settings for all transitions between working statuses and will focus on groups with generally lower employment, i.e., students, pensioners and mothers on parental leave. For students, we set the probability of finding a job during their studies without an event to 0 and the probability of leaving a job during their studies without an event to 1 (preserving the effect of the events). For pensioners, we set the probability of early retirement to 1 and prohibit overlap with gainful activity. Similarly, for women taking care of a child, we will allow them to start working again only after the maximum parental leave period has elapsed. Of course, this approach does not give us the lowest conceivable values – for example, there are many cohorts of men around the age of 40 for whom we have not made any adjustments because they do not include members of any of the three groups adjusted. It is therefore an indicative threshold; if the value according to the external projection is to be lower, it may not immediately mean that it is unachievable, but caution is advisable.

Number of unemployed individuals

Similarly to employment, in the number of unemployed individuals it is necessary to distinguish between limits achievable purely mathematically from unrealistic inputs and limits the model should actually keep within.

We get the lower theoretical limit of the number of unemployed individuals under the same conditions as those for the full employment in the previous section. The upper limit will not be exactly 100%, as some inactive individuals cannot switch to unemployment (e.g., old-age pensioners) or such transition is limited (e.g., students). Again, however, we are able to get very close to full unemployment.

The upper (lower) logical limit of the number of unemployed individuals can be achieved by means of the following adjustments:

- We set the probabilities of finding a job after completion of studies to 1 (to 0).
- We set the age of retirement to be five years after reaching statutory retirement age. Some older people will lose their opportunity to retire and a certain percentage of them will remain unemployed (for the lower limit, everyone retires early when they are entitled to it and thus becomes inactive).
- We change the probabilities of transitions from the admissible types of inactivity to unemployment to 1 (to 0). This adjustment can be made for disability pensioners and for individuals in the state of caring of a child, but not for students, old-age pensioners and mothers on parental leave.

3.3 Values Implied by External Projections

The variable that we would like to calibrate does not have to be projected in the external source at all. At the same time, however, this may be a relevant assumption because of which the volume of pensions in the external projection and the MoLSA projection will differ, even if all other important variables are calibrated. This is especially a problem of the projections of the Czech Fiscal Council where the explicit information on employment and unemployment is missing. We will not deal with the AWG projection in this respect because all variables required in chapter 3.1.2 are included in it, and more detailed calibrations are not recommended.

The projection of the Czech Fiscal Council can at least be used to find out the total number of unemployed individuals because the model's assumptions include the total income of the pension system in the calendar year and the average wage. However, there is no division by cohorts. And it is entirely impossible to get the number of unemployed individuals, because there is no distinction between the unemployed and the inactive in the projection.

We have not been able to get more implicit data. Although, for example, the number of new pensioners is available in the projection, it is not known how their number is divided by cohorts or how the number of pensioners is influenced by the time worked and how willing people are to retire early or work extra years.

In general, the calibration options cannot be largely extended.

3.4 General Requirements for Calibrations

3.4.1 List of Requirements

According to (O'Donoghue, et al., 2014), each calibration method should meet four basic requirements:

1. Projection approaches required values. This is the primary goal of any calibration.

2. Preserves the relationship between explanatory and dependent variables. For example, if the amount of the pension is clearly determined by a person's work history, the calibration should not break that relationship by approaching the pension target value, but leaves the history of the individuals unchanged.
3. Preserves the shape of all probability distributions. For example, if in the real population the employment rate of 30-year-old men is higher than the employment rate of 60-year-olds, the calibration should maintain that difference.
4. It is sufficiently computationally efficient. This criterion is important for MoLSA as without calibrations, the full run of the model takes approximately 23 hours and the complete preparation of the inputs takes about five days.

Please note that fulfilment of these conditions will ensure fulfilment of the theoretical requirements described in chapter 2.3.

Stephensen (Stephensen, 2016) extends the list of calibration requirements and claims that a good calibration method shall meet the following criteria:

1. Approach (in the expected value) the target values;
2. Preserves the original shape of probability distributions;
3. Preserves zero probabilities;
4. It is formulated symmetrically;
5. It is able to calibrate multinomial events, i.e., probability decisions with more than two possible results;
6. It is particularly effective when probabilities determined by the logit function are calibrated;
7. It is computationally efficient; and
8. It is simple in terms of implementation.

The first two points correspond to points 1, 2 and 3 postulated by O'Donogue and Li. The third point emphasizes that setting a probability equal to zero is a very strong claim – meaning that an event is completely impossible. If under no circumstances a certain event can occur prior to calibration, it should not occur as a result of the calibration.

Symmetry means that if a given probability decision has several realizations, it does not matter which of these realizations we will calibrate. For example, if we calibrate the number of deaths, we should arrive at the same number of deaths and survivors as we did with the calibration of the number of survivors. In chapter 4 we will see that, although this condition is natural, it is not fulfilled for some commonly used calibration methods.

Although much of the common probability decisions contained in microsimulation models are binary (e.g., the decision between survival and death), in some cases, multiple target statuses may be acceptable – for example, one can imagine a model in which an individual, after having completed his or her studies, can enter the status of employment, unemployment, or inactivity. In this case, an advantage is if the existing calibration method can be applied to these transitions without much intervention. However, a large number of calibration methods allow only binary decision making.

Logit functions are widely used in practice and especially for probability calculation. Therefore, when evaluating calibration methods, for example in terms of maintaining the form of distribution or computation difficulty, it may sometimes be advisable to consider this case specifically. From the MoLSA point of view, it is not a particularly relevant point because the probabilities used in the NEMO model mostly use empirical distribution instead of logistic regression.

The last two criteria are essential for the ability to deploy and use the method in practice.

In this study we will evaluate according to Stephenson's criteria. In doing so, we will consider that point 5 and especially 6 are not particularly relevant for the needs of the NEMO model.

3.4.2 Checking Compliance with the Requirements

In addition to laying down requirements for calibration methods, Li and O'Donoghue also propose ways to assess the degree of compliance with these requirements for different calibration methods and evaluate some of the selected methods according to them (O'Donoghue, et al., 2014). In the following chapters we will refer to their results for those combinations of criteria and methods that the authors included in their study. Other cases will be evaluated on the basis of our own experience and considerations, or very indicative calculations. Therefore, we will not use the methods mentioned in this chapter because the actual implementation of the calibration methods described goes beyond the scope of this study. However, the recipient of the study is advised to perform these tests if the decision on the calibration method cannot be made using other criteria.

Specifically, Li and O'Donoghue recommend monitoring four variables.

Target deviation index

This variable assesses the degree of fulfillment of the condition point 1 set out above. It compares the target number of events with the actual number according to the following formula

$$TDI = \frac{T - S}{N},$$

where *TDI* is the deviation index, *T* is the number of actual events, *S* number of events simulated by the model, and *N* the total number of individuals (or other units) for whom the event might have occurred. This index can only be applied to binary variables (i.e., variables that can only have two statuses, for example death, unemployment, childlessness). In addition, for MoLSA purposes, it is necessary to select a variable for which the number of events can also be determined from the results of the external projection – for example, the number of pensioners is included in these projections, but the number of sick individuals is not.

Distribution derivation index

This variable assesses the degree of fulfillment of the condition point 2 of the Stephenson requirements for calibrations. It is based on the assumption that if the calibration perfectly maintained the probability distributions, the expected values of any two groups of persons would remain in the same ratio (for example, if the average salary of men before calibration was 20% higher than that of women, it shoiuld be 20% higher also after calibration).

Such situation will only occur if the only adjustment made during the calibration is to multiply all values by a common parameter. First, we define the parameter *R* as the ratio of the mean value of the measured variable after and before calibration. Then we divide the population into *n* groups and calculate the following for each group i:

- expected value of the measured variable before calibration and we denote it O_i, and
- expected value of the measured variable after calibration and we denote it S_i.

In addition, if we state that N_i is the number of individuals in group i and N the total number of individuals in the population, we can use the following formula for the calculation:

$$DDI = \sum_{i=1}^{n} \frac{N_i}{N} (S_i - RO_i)^2.$$

DDI is out testing value. The term RO_i expresses the expected value that group i should have if the distribution were perfectly preserved. Therefore, the more significant breach of the distribution for the given group, the greater the parenthesis. Fraction $\frac{N_i}{N}$ subsequently weighs this value by the size of the group – if there is a large deviation for a group that is insignificant in size, it may still be a good calibration. Zero value of DDI therefore indicates perfect match of the distribution before and after the calibration, and the higher DDI, the more distant the new distribution is from the original one.

The values of the index depend on the original expected value of the calibrated variable and the number of groups, so it makes no sense to compare the calibrations of different variables with each other (e.g., it cannot be said that any value below any threshold is acceptable), but only different methods of calibration of the same variable to the same target value.

The selection of the groups for which the DDI is calculated can be customized using the calibration method. Generally, we recommend using two basic sections. In the first one, our groups will correspond to common cohorts, i.e., we will divide the individuals in the model by year of birth and sex. This will verify that the dependencies on these two basic parameters have been preserved. As a second section, we then divide each cohort according to the size of the calibrated parameter into, for example, 100 levels and bring together people of the same level from all cohorts into one group (i.e., for example, when calibrating salary, we would include in the first group 1% of the highest earners from each group). This will verify that there is no bias of the probability distributions within the cohorts.

It should be noted that this method is made for calibration during the model run, and not for calibration performed solely by adjusting the input parameters. We expect that some input values will always be adjusted for calibrations performed by MoLSA (e.g., mortality tables will be taken over) which may or may not be followed by calibration during the run of the model. Values of the original distribution before calibration during model run will be then substituted for O_i, but after the adjustment of the input parameters. These are therefore not the results of the MoLSA’s own projection.

If the calibration is performed solely by changing the input parameters, there is no point in using the distribution derivation index. There may be differences from the distribution observed in MoLSA's own projection, but in this case the difference will be a natural and desired result of the change in the parameters.

Run time in seconds

This value assesses point 7 of Stephenson's calibration requirements. When comparing two calibration methods, the method that makes all adjustments in a shorter time with the same inputs and the same calibration targets is more suitable in this respect. Since the effectiveness of some methods may considerably depend on the inputs, we recommend performing this test for several different initial settings, varying by the number of model points and the calibration targets in particular.

General fit measure

This method specifies more fully the degree of fulfilment of item 1 out of the calibration requirements. It indicates the number of events that occurred in the calibrated model, but not in the target statistics to which we want to approximate the model by the calibration, and, on the other hand, the number of events that did not occur in the model, but did occur in the target statistics.

For the purposes of MoLSA, it would be possible to examine the results of the model against reality. However, this is not part of the objectives set for this study. This measure does not say anything about the approximation to the external projections because only the aggregate values are available in the external projections concerned, and information on the occurrence of individual events is missing.

In addition to these numerical methods, we recommend that you perform a general test of meaningfulness after each calibration, whether it is performed by interfering with the model run or by adjusting the input parameters. The user should list the results of the main variables for some important population groups and check to what extent they meet the expectations. If the observed deviation from MoLSA’s own projections is too drastic, it will be necessary to consider whether the resulting microdata after calibration is sufficient for its purpose or whether the calibration needs to be adjusted. Therefore, the exact form of this test, i.e., the group of people examined and the tolerance allowed, may vary depending on the purpose of the calibrated model.

4 Analysis of the Relevant Calibration Methods

This chapter constitutes the main body of the study. We will describe several possible approaches to calibration, including their main features, information on whether they are used in existing models or discussed in the literature, and considerations for their implementation in the current NEMO model.

The following calibration methods will be described:

1) Calibration by residual population
2) Calibration by iterative model runs
3) Refinement of average values
4) Multiplicative scaling
5) Sidewalk method
6) Alignment by sorting
7) Bi-proportional scaling .

The methods can generally be divided into two categories depending on where the calibration takes place. The first option is to calibrate solely by adjusting the input parameters; the second option is to make adjustments during the run itself based on random events realized so far. The key disadvantage of the second category is that these methods necessarily distort the probability distribution of individual variables compared to the distribution based on the population characteristics as described by the input parameters of the model. Nevertheless, the practical application of these methods prevails over the methods of the first group to such extent that O'Donoghue and Li (O'Donoghue, et al., 2014) only work with the second group in their study on calibration methods. The key advantage of these methods is their relative straightforwardness and, in case of the more sophisticated ones, a guaranteed result. Calibration of input parameters, on the other hand, can be a very complex process involving several preparatory runs of the model and uncertainty as to whether a satisfactory solution will eventually be found.

A list of calibration methods, which were referred to above, follows. The list starts with methods based on calibration of input parameters (the first group from the previous paragraph). Before we address the calibration methods as such, however, we will describe the preparatory steps that can help achieving initial approximation of two projections without using any sophisticated methods.

The description of some methods is based on Deloitte experience in creating real microsimulation models: we have previously created microsimulation models of state pensions for clients in four countries. In some cases, we are not authorized to disclose the name of the client in the study; for this reason, the method description will not include identification of the country and model name.

4.1 Preparation for Calibration

Chapter 3.1 suggests that calibration should consist of two stages: preparation phase, where both projections are approximated in variables that can be aligned by simply taking over the inputs and that, simultaneously, affect all other values, and the phase of calibration procedure, which uses more complex methods described later in this study. In this chapter, we will describe how to take over the inputs so that the most significant approximation occurs during preparation. We expect all external projections to use very similar inputs and therefore the procedure for them will be very similar. That is why we will focus on the example of the AWG and the Czech Fiscal Council projections.

Mortality and disability rates

Each population projection draws from a certain mortality table. We want to adopt this table. Even if we are unable to acquire it directly from the authors of the external projection, it can be easily calculated once we know the number of persons for each cohort in each calendar year. The mortality tables used for both the AWG projection (taken over from EUROSTAT) and the Czech Fiscal Council projection (taken over from the Czech Statistical Office) are publicly available.

In addition, we will focus on cohort disability rates, which are the starting point for the number of disability pensions in both external projections and are calculated as a ratio of disability pensioners to the total population. Obtaining relevant data for the calculation is not a problem; data on the number of disability pensioners and the size of population at a given age either serve as input to the projection (AWG) or are publicly available (EUROSTAT, CSO, CSSA). Thus, we will align the initial numbers of pensioners receiving disability pension.

The probability of a new disability is represented by the *MORB_RATE* variable in the NEMO model and can easily be calculated from the AWG projection inputs as a ratio of the number of new disability pensions to the population reduced by disability pensioners already existing in the given age group. We expect that the number of new pensioners aggregated with the number of last year's pensioners who did not die will exceed the number of pensioners predicted by the projection in the current year. Based on the difference, we will determine the probability of disability termination.

In case of the Czech Fiscal Council projection, we do not have data on the number of new pensioners receiving disability pension; we only have their total number. As a result, we will maintain the same probability of disability termination as in the normal runs of the NEMO model and determine the probability of incidence so as to match the total numbers of disability pensioners in both projections.

The probability of change in the disability degree will be calculated in a similar way.

Birth rate

The birth rate assumption enters the NEMO model projection in two ways. On the one hand, the newborns whose life path is then subject to modeling enter the calculation each year. These persons are included in the database of model points and, in the preparation of model points, their parameters, including their number, are also determined; therefore, for an approximation to an external projection, the relevant assumption has to be inserted in the model point creation procedures. In addition, the probability of a child being born is part of the projection itself. This is also based on the birth rate projection, which is used as one of the input tables in the model. All data including birth rate can therefore easily be taken over.

Initial population

We assume that the initial numbers of employees and pensioners (of all types) and the total initial population will correspond between the two projections, as the current state of these variables is known from the CSO and CSSA data. If there are significant differences, it is possible to randomly select some persons from the inputs of the NEMO model and remove or duplicate them (the selection should be based on age and education). However, the utility of a projection using such modified data is questionable - in such case, we recommend, first of all, to examine the grounds for the difference in the initial data and the extent to which their existence will affect further calculations.

Bigger differences may arise in the numbers of parental leaves, students, the unemployed and inactive, as detailed statistics are not available for these data and the distribution of persons among these groups may not be clear. In such case, we will again randomly select the appropriate number of persons and reassign them to another status, so that the representation of persons in each status corresponds to the external projection. Based on an expert decision, we can formulate several assumptions for this reassignment; we may, e.g., decide that newly created students should not have previous work experience. Concerning younger people especially, working with persons with no history in the NEMO model points should be sufficient for the task of reassigning people among these groups.

However, the only such state contained in the AWG projection or the Czech Fiscal Council projection is the number of the unemployed in AWG's work. Therefore, adjusting inputs in this way should rarely be needed.

Number of pensioners

The AWG projection assumes that each person will retire on the day he or shereaches the so-called effective retirement age. However, we want to adjust the average retirement age to best match the mentioned effective retirement age, but, at the same time, to maintain the number of persons who decide for early, or postponed retirement.

We will start out with the results of the NEMO model before calibration (e.g. the results regularly reported by the MoLSA). From the number of new pensioners in individual years we will calculate the average retirement age for individual cohorts (each cohort is assigned a unique SPCODE also for the basic run in order to be able to sort the results by cohort). Furthermore, for each cohort and each calendar year, we will determine the rate of retirement observed in the model (as the ratio between the number of new old-age pensioners and the initial number of non-retired persons) and compare it to the retirement probabilities entered to the model. The observed probabilities will be lower because the probability entered in the model does not apply to all persons not yet receiving an old-age pension, but only to those who have already reached the required insurance period. There, we will calculate the size of the population to which the probabilities of retirement were applied – i.e. the size of the population eligible for retirement.

If we are permitted to manipulate the statutory retirement age, we will change the retirement age in the *retirement.fac* table by the number of years equal to the difference in the average actual age of retirement between the model and the AWG projection. We then set the probabilities of early retirement and postponed retirement so that after their application to the population that has met the statutory requirements in the given calendar year, we receive the same numbers of persons who retire at an age different from their retirement age as those observed in the basic run.

If the user wishes to maintain the retirement age because the value is strictly defined by law, we can omit the above step and only work with the probabilities of retirement, at the likely cost of the need to extend the table of probabilities of retirement to include other possible differences between pensionable age and the actual retirement age.

We will take similar steps in relation to the Czech Fiscal Council projection. We will set the statutory retirement ages in accordance with the applicable legislation, only we will not consider the number of children of each woman and will (consistently with the projection) use the value applicable to a woman with two children according to the law. Similarly to the AWG projection, we will determine the number of persons who have reached eligibility to retire in the NEMO model (depending on the calendar year and the year of birth), and set the probabilities of early retirement, regular retirement, or postponed retirement so as to get the numbers of new pensioners projected by Czech Fiscal Council.

Please note that in this manner, we will only achieve approximate alignment. The probability of retirement interacts with employment, as the change in employment will also change the periods of insurance of individual persons, and hence the moment of eligibility to retire. Thus, even if we reach a perfect alignment in the first step, we will break it again as soon as we calibrate employment. Therefore, we will repeat this adjustment once again after having calibrated employment and unemployment in order to improve the alignment.

4.2 Calibration by Residual Population

4.2.1 Method Description

In some cases, the population captured by the model can be divided into several groups for which information with varying level of details is available. In a situation where there is reliable data on certain subpopulations and the overall result needs to be adjusted, it is achievable by calibrating the persons from groups with greater degree of uncertainty.

This method was used by Deloitte to calibrate mortality in a microsimulation model of one of their clients' retirement plans. This model considers the beneficiaries of various types of benefits separately and differentiates some of their input assumptions, including mortality. At the same time, there is a group of persons who have never been included in administrative records because they are participants of other pension schemes (e.g. members of armed forces). The mortality of these persons was determined so that its aggregate with the mortality of other groups corresponds to the mortality of the total population derived from external statistical data.

In its simplest form, this method is very straightforward. If, for example, x_i is the value of a variable in the group i (e.g. mortality at a certain age) and, together, we have n groups, of which we know the value of the variable for all but the last, and if N_i are the numbers of persons in each group, N the total number of persons and x the total value of the variable for the whole population (e.g. total mortality), then it holds

$$Nx = \sum_{i=1}^{n} N_i x_i$$

and the value for the last group can be calculated based on the following formula

$$x_n = \frac{Nx - \sum_{i=1}^{n-1} N_i x_i}{N_n}.$$

This procedure has two drawbacks. First, it can only be used with independent variables (i.e. only those for which there is no formula based on other variables in the model), and only with variables where it makes sense to divide the population into several groups whose members will all have the same value of calibrated variable (which is a reasonable assumption for e.g. mortality, but a considerable simplification in case of a salary). Furthermore, this procedure per se does not guarantee that the calculated value will be realistic. If the values for other groups have already been derived with some degree of uncertainty, the individual deviations may add up and the last value may significantly deflect from expectation. Therefore, it is appropriate to set in advance the limits within which the resulting values should fluctuate, either in absolute terms or relative to other mortalities (we can e.g. assume that, at most ages, men will have higher mortality rates than women, and calibration should therefore not decrease the mortality of men below that of women). If the calibration renders a value outside the specified limits, we will need to go back to one of the previous steps (i.e. deriving inputs for other groups or obtaining the input data as such), because our data and set assumptions may not be consistent.

A variation of this method can also be used to calibrate variables that should have a slightly different value for each individual (such as salary). First, we will calibrate the expected value of the variable for the complementary group according to the formula provided above and then we will differentiate this value to obtain the value for each individual using some other procedure. Some variation of the refinement of average values described in Chapter 3 may be appropriate.

So far, we have only considered a situation where there is a single complementary group; in its basic form, the method cannot differentiate a larger number of complementary groups. We can, however, consider a variant where for each complementary group we will determine the default value of the variable of interest, in addition to the permissible limit, and establish an optimization algorithm that will try to find the values of the variable for all complementary groups so that

- Each value found is within the specified limits;
- The x_i values in the complementary groups are approaching the initial values (measured, for example, by the squared deviation);
- The formula is $Nx = \sum_{i=1}^{n} N_i x_i$.

For example, Excel Solver is fully sufficient for such optimization. We need to remember that the result of this variant of the calibration method will strongly depend on provided input – the limits and the initial values.

4.2.2 General Evaluation

In order to be able to better discuss the advantages and disadvantages of this calibration method, we need to distinguish between two modes of use. In the first one, we want to calculate the value of the variable in only one group. In the second mode, we need to calculate the values for multiple groups.

Single group calibration

If we know the values of all groups but one, this method provides a straightforward means of calculation of the last value. The calculation is simple, takes almost no time and does not require any inputs other than the sizes of all groups and the known values of the reviewed variable.

The drawback of this method is the possible error accumulation. The values that we now consider known have probably been derived by statistical methods and may therefore deviate from the real value. In the calculation of the last value, these deviations may sum up to a considerable amount. It is therefore necessary to examine the logical outcome of this method.

The advantage of this method is that it accurately maintains symmetry. If, instead of the value of the last group, we calculate its complement to 1, we follow the following formula for the calculation, which is obviously equivalent to the calculation of the original value.

$$1 - x_n = \frac{N(1-x) - \sum_{i=1}^{n-1} N_i(1-x_i)}{N_n} = \frac{N_n - Nx - \sum_{i=1}^{n-1} N_i x_i}{N_n} = 1 - \frac{Nx - \sum_{i=1}^{n-1} N_i x_i}{N_n}$$

Multiple group calibration

In the second case, assuming unknown values in multiple groups, the situation becomes more complex from mathematical point of view. It leads to a non-linear programming task which, in general, may not have a solution, or may lead to a solution that is not unique. In addition, the calculation under this method may be rather complicated. The drawback is that the time complexity cannot generally be determined in advance.

In this case, too, the deviation of values from the known groups accumulates.

Another possible disadvantage of this method is the need to define both the initial values of the calibrated variable for each group and the limits that the calibrated values must not exceed. In certain cases, these limits may be clearly defined (for example, we can certainly assume that male mortality is higher than female mortality), but the method will be very sensitive to changes in all these parameters. If we wrongly determine solution feasibility limits, we may not find the solution, even if it exists. Or, the calculated value will not make any logical sense.

The advantage of this method is that if the individual groups are independent and their relationship in limit constraints is at most linear, the calibration using the residual population is symmetrical. Unfortunately, in case of more complex constraints, symmetry is not maintained. For example, the following constraint maintains symmetry: $x_i \leq 5$ or $x_i + 3x_j \geq 6$. However, non-linear (i.e. square, exponential, etc.) constraints break symmetry.

Shared properties

In Chapter 4.3, we provide two indices used to measure the quality of the calibration methods; unfortunately, neither of them make much sense reviewing for the calibration by residual group.

To use the target deviation index (TDI), we would need to know the number of actual events in the entire population and compare it with the number of simulated events. However, the MoLSA does not possess information on the actual number of events in the groups in which we calculate the variable values.

It is important to note that this method does not at all take into account explanatory variables other than the group in which each individual is included.

The logical limits that we determine (whether for one or more calculated values) ensure, in most cases, that zero value is maintained at zero probability. Nevertheless, specific cases may arise where this feature is violated. Imagine, for example, that we model the third pillar of the pension system with both the amount contributed by certain groups and the total amount of contributions to the third pillar. This method would then attribute the same proportion of the missing amount of contributions to all those for whom the information is missing. That would probably be a mistake, as we have reasons to believe that many of these individuals did not contribute at all, rather than believing that each of them contributed the same small amount.

Summary of strength and weaknesses

One of the main advantages of this method is that, for the determination of a single value, the calculation is very fast and usually maintains symmetry.

The main disadvantages are the consideration of only one explanatory variable, namely the group of the individual. If we want to calibrate several groups at once, we need to obtain more information and initial values, the method may not find a solution and, in case of more complicated constraints, it may run slowly.

4.2.3 Suitability for Application in NEMO

Distributing population into groups based on a particular attribute can be done in the NEMO model. At the beginning of the projection, these groups are defined; then, after simulating life paths, the model will assign an individual to the corresponding group. After simulating the paths of the last individual, the total simulated population is distributed to predefined groups for which the values of important variables are known.

Then the following considerations need to be discussed:

- the number of constructed groups,

- selection of (a) group(s) to represent a complementary group for the application of the described calibration method; and
- the limits of feasible values.

It should be noted that it is not possible to classify one individual "twice", i.e. to define separate sets of groups at the beginning of the projection. For example, it is not possible to initially define groups according to sex (male, female) and separately according to age groups, and expect the model to include an individual once in a group according to sex and once in another group according to age. If we need to classify the population into *n* age groups according to sex, we need to create *2*n* groups to which individuals will be assigned at the beginning of the projection, i.e. one group for each sex-age combination.

The upper limit for the number of groups should also be given consideration. In the NEMO model, the maximum allowance for categorization is 10,000 groups (upper limit of the SP code), which is not optimal. More groups lead to a significant slowdown in the model run, which lasts at least one day anyway. If we wanted to divide the population by sex, such operation would not burden the model too much. Concerning age groups, we would have to discuss the optimum number not only in terms of the complexity of the model run, but also in terms of reliable interpretation of the results; allocation of individuals to e.g. five-year age groups is less informative than allocation to actual age groups, but in terms of the number of groups, this categorization is more acceptable for the model while still providing sufficiently detailed information about the reviewed variable and its distribution (even in the sex-age combination).

The core of the calibration method is the selection of the group that will represent the complementary group and the value that will be calibrated to the calibration target by applying the method. In general, it should be the group with the highest degree of uncertainty; e.g. in modeling the amount of salary by age groups, this could be the group of working students aged 15-20, for which we usually have a low number of records. Another option for selection of the complementary group is to choose the one that most differs from the external projection values. The choice of the complementary group is not always easy or straightforward and requires expert judgment supported by sufficient analysis. Then, the complementary group would be calibrated to the desired target by a simple calculation using the above formula, while the necessary input values representing the values of total population would be imported as constants from the table. By comparing the calculated and the desired value of the complementary group, the calibration constant would be calculated and stored as a new value, in the table of constants that is imported on input. Then the model would run again, and at the end of the projection, the complementary group would be multiplied by the appropriate calibration constant. This approach to calibrating values would also be useful as an alternative to the average salary correction method used so far, which is represented by the “residual component” in the NEMO model.

After selecting the complementary group, it is important to verify that the calculated value falls within the interval of feasible values, which was determined in advance. This verification can be easily implemented into the NEMO model - the limits of the feasible values are read by the model upon entry as constants from the table.

The calibration method using the residual component can also be used to calibrate input variables in the event that rather unreliable or no data is available for a group/category. This approach was used in a model of Polish state pensions, where input mortality was calibrated for specific population groups. The advantage of this application is avoiding to run the model twice.

Finally, the simulated population can also be categorized using a time-varying variable (e.g. the age group, the current working state). In such case, the possible transitions between groups have to additionally be defined in the model using so-called extended formulas. After each simulation, the states in each group (entry/leaving of individuals) need to be manually recorded and, at the end of the whole projection, the final table can be printed. For calibration purposes, however, there is no point in further exploring this option.

4.2.4 Final Evaluation

The method supplies variable values for population groups for which less reliable or no data is available. Therefore, it only makes sense to use it in such cases. It only works at an aggregate level (e.g. death probability of persons in a group, average salary in a group). It can be used to prepare assumptions prior to the first run of any model, or to carry out the calibration based on the results of the preparatory run.

4.3 Calibration by Iterative Model Runs

4.3.1 Method Description

The basic idea behind this method is that the user will run the model with default inputs, adjust the inputs based on the results, and repeat this process until the desired approach to the target values is achieved. The main challenge of the method is the efficiency of the process.

The natural solution is to perform calibration on a simplified model that will only cover those relationships that are subject to calibration. For example, this process is used to calibrate the labor market in the government pension model of one of Deloitte‘s clients.

Case study – use of the method in Deloitte's client model

Each person in this model belongs to one of many possible statuses. Distribution into statuses is very fine, each of the natural main statuses – e.g., employed, unemployed, inactive, farmer or student – has several variants; there are 55 statuses in total. Transitions occur based on a transition matrix (i.e. a set of probabilities that a student will get a job, an employee will resume studies, an inactive person will take parental leave, etc.). The employment and unemployment rates for individual years and age cohorts are used as the calibration target. Another requirement for calibration is to maintain the entire transition matrix proximate to the state prior to calibration (i.e. to change individual transition probabilities as little as possible).

The initial stage of calibration is a preparatory run in Prophet. On the basis of this run, the initial transition matrix is determined by simply dividing the number of persons in the respective statuses at the beginning and at the end of the given period. We have a total of 55 × 55 transitions.

This matrix will be adjusted during calibration to approach the calibration targets. However, we will not include all its components in the process, as some transitions are driven by more complex mechanisms than the transition matrix can sufficiently cover:

- Some transitions in the model are the result of deterministic development and their occurrence is not affected by randomness (at the time of occurrence). An example might be the transition between sick and healthy employee in the MoLSA model, as the duration of the sickness is determined randomly at sickness commencement while the termination of illness is no longer subject to randomness. There is no meaning in calibrating such cases because they behave differently in the main and in the complementary model.
- Some transitions occur very rarely because they affect only a small proportion of the population or because their probability is very low. The number of such cases cannot be stable in the model and, consequently, its calibration makes no sense.

We will not change the parameters affecting transitions between these two groups within calibration. Nevertheless, for the calibration of other transitions, their existence must be acknowledged; for this purpose, the probability taken from the Prophet preparatory run will be used and will not be further adjusted.

The calibration itself takes place in a complementary model in MS Excel. The initial numbers of persons in each status, divided by sex and age cohort, are used as inputs. The model then includes a transition matrix for each combination of calendar year, age cohort, and sex. The input values of these matrices are taken from the Prophet preparatory run. Based on these inputs, the model projects the development of the number of persons in each status and compares the squared deviation of the employment and unemployment rate from the calibration target for each year. The model thus does not include calculation of the amount of benefits or verification of statutory requirements.

The file is finally optimized using Excel Solver. Non-linear GRG (generalized reduced gradient) is used as the optimization algorithm; it is a standard optimization method using (like many other optimization algorithms) partial derivations of the minimized function. The result is a set of transition matrices dependent on the cohort, calendar year and sex. Importing these matrices into the Prophet model and running the model subsequently will yield calibrated results.

Other specifics of the method

In general, it is not guaranteed that the calibrated run will actually meet the calibration target - transitions that have been excluded from calibration may affect the results differently compared to the simplified method that was entered in the MS Excel model. The results should therefore be thoroughly tested. It is necessary to check both the actual deviation of the calibrated variables from the calibration targets and the impact of calibration on the other variables.

If satisfactory calibration has not been accomplished, the entire process can be repeated. It depends on the expert judgment of calibration analysts whether to use, as the new input for the optimization algorithm, the model results after the first calibration or the results of a completely new preparatory run (two identical preparatory runs always give the same result, but it is possible to change the order of persons in the list of model points, thereby intentionally disrupting control over the realization of random variables in the simulation). The decision will mainly depend on how much the calibration target was approached and how much distortion in the distribution of other variables was caused by the calibration so far.

The difficulty of the task greatly depends on the complexity of the calibrated process and the number of calibration targets. In the Deloitte project described above, only two target variables were monitored, namely the employment rate and unemployment rate; still, it was not trivial to achieve satisfactory calibration. Furthermore, the method relies on the existence of a sufficient number of transitions that can be calibrated, i.e. for which the number of persons can be expressed as a percentage of the initial population and is spared the necessity of meeting statutory requirements.

Since calibration is performed outside Prophet, the calibration model can generally be created in a number of application tools. We consider MS Excel to be a reasonable compromise among the ease of use, the ability to replicate important parts of the model in Prophet and the availability of optimization procedures. While e.g. R would certainly allow for a tool that would replicate Prophet more precisely or offer more optimization algorithms, its creation would probably require far more work with uncertain added value.

An auxiliary model may not be necessary to calibrate certain specific microsimulation models, but in general cases it cannot be avoided. If one wants to use the main model for calibration, it should run the calculation in a sufficiently short time, as it will have to be run many times during the optimization. Another requirement is that the main model can be subject to an automated optimization procedure, i.e. there must be a tool capable of running the model, analyzing its results, adjusting the model's assumptions based on the results and repeating the process. Generally, such a tool is not easy to implement, unless the model has been created in a language with easily accessible optimization libraries (such as R).

An important question of this method is the degree of simplification that will be introduced in the auxiliary model. In the case described above, the model was simplified to the greatest extent possible - it was limited to labor market forecasting and calibrated only transitions with a simple form. The advantage of such significant simplification is the relatively easy implementation of the auxiliary model and its comprehensibility. If the auxiliary model were to be more complex, we would have to make additional assumptions - for example, if we wanted to calibrate the transition of employees to retirement, we would have to determine the percentage of employees with enough time worked to become eligible, ensure consistency of this assumption with the other assumptions used in the auxiliary model and consider it in the interpretation of results. We therefore recommend to always start with the simplest possible variant and introduce more complex functionalities to the auxiliary model only in exceptional cases.

4.3.2 General Evaluation

The major disadvantage of this method is that it constitutes a purely empirical exercise. There is no guarantee that the calibration will achieve the objective, and success or failure can only be evaluated after the model has been run in Prophet with the calibrated inputs. For this reason, the method needs to be tested for functional accuracy, and each individual calibration has to be thoroughly validated.

There is some hope though, as the client for whom this method was originally created by Deloitte, dealt with a task very similar to the task currently pursued by the MoLSA, i.e. calibration of a model in Prophet with basic structure similar to that of the MoLSA model (shifting time in monthly steps, having each individual transition between certain states based on a transition matrix, while verifying which states are acceptable in the given situation; all other calculations are based on current and historical states). In both cases, there are transitions that cannot be calibrated using this method because they depend on other legislative conditions (e.g. a person can only retire after having worked for a certain number of years). The main difference is that the MoLSA model contains a significantly lower number of statuses. We can therefore hope that the difference between the situation of the MoLSA and the original application of this method is not such that the calibration could not give satisfactory results.

Potential issue of this approach is its stability. The algorithm will generally only find a local minimum of the problem, so admittedly, a small change in the calibration input parameters will lead to different results, which can yield a different projection when run in Prophet. The difference between alternative projections will not be significant (since we always verify that the results of a calibrated run do not differ much from the original ones), it may, nevertheless, hinder the interpretation of results and diminish confidence of other users in the calibrated projections.

In addition to the computation demand, another disadvantage leaks in the inability to maintain symmetry of the task. The optimization algorithm is generally unable to work with symmetry.

Another problem may arise in maintaining zero transition probabilities. The condition for maintaining them must be explicitly included in the optimization algorithms, which further increases the complexity of the whole system.

It should be explicitly mentioned that since this method only modifies the input values, it fully maintains relationships between the explanatory variables and their responses, thus well maintaining the consistency of inputs and results.

4.3.3 Suitability for Application in NEMO

As the calibration itself takes place in an auxiliary external file, a significant part of the implementation issues is eliminated. However, certain complications cannot be completely avoided.

The transition probability matrices that represent the input for calibration are obtained in the model by initially defining the groups for which the model will record quantities during the run, and, at the end, recording quantity relating to December of the given year for each group, i.e. after completion of the run, the model will create a table of quantities at year-end, with predefined groups in rows and calendar years of projection in columns. These groups will be a combination of the year of birth of the individual, their sex and all possible transitions between statuses (e.g. transitions: employed-childcare, employed-student, student-student, student-unemployed and so on).

Then the final transition matrices for calendar year x are obtained by dividing the number of persons in the respective states in year x-1 by the number of persons in year x. The transition matrices thus prepared enter the auxiliary model, in which the values are calibrated.

The result of the calibration is again a set of transition matrices. Their import to the NEMO model and re-run will yield results that are already calibrated. When subsequently converting the calibrated transition matrices back to Prophet, we need to deal with the transformation of annual probabilities to monthly probabilities, but also to consider that the model works with two types of transition probabilities – those associated with the event that causes the change of state and those that are independent of the event. There are also instances in the model where some transition probabilities for a given individual are set to zero until the expiry of the predetermined period in which the individual has to remain in the given state (e.g., a sickness). The ways to deal with these complications are described in detail in Chapter 4.8.3. However, this is a non-trivial task.

4.3.4 Final Evaluation

This method requires a complex external tool, does not guarantee finding a solution, and even if a solution is found, it will not be generally stable. In addition, the computation time itself may be quite high. The method of calibration by iterative model runs is suitable for calibrating transition probabilities, but cannot be used for calibrating salaries. Its main advantage is the ability to simultaneously calibrate a large number of transition states. If we calibrate one variable only, the use of a simpler method is preferable.

4.4 Refinement of Average Values

4.4.1 Method Description

This method is again based on adjustment of the model's input parameters. Its primary purpose is to introduce individual variances into population in which, without the use of this method, all persons would have the same value of the adjusted variable. It does not necessarily involve a calibration in the sense of a shift of the expected value. A prerequisite for this procedure is the knowledge of the resulting probability distribution of the variable we want to adjust (or at least the existence of an a priori assumption on the distribution).

This method was used by Deloitte to calibrate salaries in a microsimulation state pension model for one of their clients.

If the variable that is being adjusted does not depend on the individual's history, the procedure is simple. It is sufficient to simulate a random value from the desired distribution for each person (this distribution is known in advance and has the correct expected value). Such straightforward manner can be used to determine, for example, the initial salaries of persons entering the labor market. Although these depend on attained education and the age of commencement of work, once we classify people into groups according to these variables, the specific value of their salary is completely independent (and thus independently simulated from the distribution relevant to that group).

A more complicated case arises if the adjusted variable is subject to other criteria - for example, when modeling the development of salaries, we should take into account the given person's previous year salary and general prospects on the labor market, in addition to maintaining the calibration target for the expected value. Here we can distinguish between several options according to complexity. We will assume that a percentile of the relevant distribution will be assigned to each individual. Let us recall, for example, if a person has a 60% percentile, this means that 60% of persons from the reviewed population have reached a lower value for that variable.

In the first option, we will exclusively compare the distribution of the explanatory and dependent variables (e.g., the salary in year R and the salary in year $R+1$) and assign each person such percentile of distribution of the dependent variable that corresponds to their percentile of the explanatory variable. We defined both distributions a priori when determining calibration targets. Succession of persons will therefore be retained - if person A had a salary higher than person B in year R, he or she will have a higher salary in year $R+1$ as well. We will use this method if we want to achieve a credible development of individuals (e.g. a person with a high salary in year R will have a high salary in year $R+1$), but we do not require any randomness in the development. The initial position of a person may be random, if simulated from the first distribution, but all other positions are already fixed by that initial position.

This approach is therefore as follows: on the basis of a person's position within a certain group in terms of the first variable (e.g. salary in year R), we will determine the position of the same person within the same group in terms of a new variable (e.g. salary in year $R+1$). This is particularly useful if we monitor a certain determinate group over the entire period of projection - for example, if we are interested in the development of salaries of men with university education born in 1980. Nevertheless, the group does not necessarily have to be completely impermeable – persons who have found a new job may transition here or, on the other hand, persons who have died or who have e.g. been granted a disability pension may transition out of the group. Important is that the position of a person within the group remains the key parameter.

Complications arise when we do not want to simulate some the new entrants to the group from the unchanged resulting distribution. An example could be a situation where someone returns to work after prolonged unemployment. If we neglected the effect of unemployment, we would assign such person the same percentile of salary they had before losing their job, or we would simulate a whole new value from the distribution that would not be modified in any way. We could take into account the unemployment by assigning the person a slightly lower percentile or by simulating from a distribution with higher probabilities of lower salaries compared to the original distribution. This would, however, distort the overall distribution of salaries in the population. The only solution is to make adjustments also to the persons who remain in the group, thus balancing the distortion. In our example, we would increase some persons' salaries by one-time jumps. This adjustment must also be subject to random distribution; the development of the variable will no longer be deterministic, even for the persons in the group.

Thus, we come to considerations as to how to introduce random development of the career path to the method. The option to simulate each variable for each person at random, irrespective of the previous results, is not suitable, as there would be unjustified jumps in individuals' life paths. (For example, a high salary in year R, low salary in year R +1and high salary again in year $R + 2$). In addition, the effect of unemployment referred to above would not be sorted out. Thus, we want to introduce randomness such that we can control it somehow. There are surely many ways to do this. We propose the following.

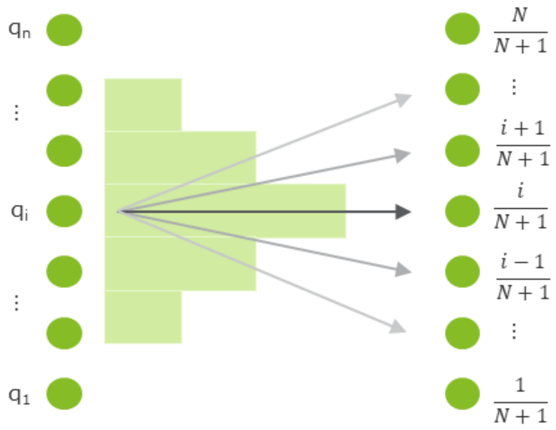
Introducing randomness – option one

Our inspiration for this method was the pattern used to allocate partners to persons in the NEMO model: an "ideal" partner for a person is identified according to certain criteria, and then the existence of such partner is verified in the model point database. If no such partner exists, another partner, the most similar to the ideal partner, is assigned instead. The method assumes that the database of possible partners is predetermined. In the NEMO model, one partner can be assigned to several persons; we will not use this simplification here.

Similarly, in this calibration method, we will in advance determine a list of permissible percentiles of the target probability distribution to which a person can transition. Since the percentiles of any distribution are always uniformly distributed, we will choose the target percentiles q_i as follows: if the total number of persons (and hence percentiles) is N, then $q_i = \frac{i}{N+1}$. Thus, the distance of two neighboring percentiles is always equal to $\frac{1}{N+1}$ and the distance of the first percentile from zero and of the last percentile from one is $\frac{1}{N+1}$ as well. We do not include zero or one among the permissible percentiles, because some distributions may converge to infinity in these percentiles.

We then assign one of these percentiles of the resulting distribution each person based on his or her initial percentile. Let us assume first that the initial value of the variable (e.g. the salary in year R, if we want to calibrate salary in year $R + 1$) meets the condition that each person is in one of the percentiles $q_i = \frac{i}{N+1}$ and none of the percentiles is repeated within the initial distribution. Let us define a priori a discrete random variable X that will determine the probability of a person staying in the same percentile, moving up a percentile, etc. The choice depends on the user's expert judgment; a possible candidate is e.g. uniform distribution with the same probability of transition to all percentiles. A discretized version of the normal distribution (truncated for the highest and lowest percentiles that do not have enough neighbors on one side) may be another suitable option for X. We will, step-by-step, work with each person and simulate the value from the distribution of X that will determine his or her transition to the next state. We will reassign the person to the percentile that is determined by the initial percentile and the value of X, while recording which percentiles of the target distribution have already been used. If we get a percentile for a particular person that has already been occupied, we will want to assign the nearest free percentile instead. Again, we have several options to choose this replacement. The simplest option is to solely rely on their order: if percentile q_i is taken, we will first try to move the person to percentile q_{i+1}, if it is already taken, then q_{i-1}, q_{i+2}, q_{i-2}, etc. In other words, we will minimize the value $|q_I - q_V|$, where q_I is the "ideal" percentile, i.e. one that resulted from the distribution of X, and q_V denotes available percentiles that a person can move to. This option is especially useful if X is an even distribution with odd number of percentiles, where the middle one corresponds to the person's percentile in the initial distribution (that is, the distribution that serves as a basis for calibration).

A more complex option is to take into account the shape of the X distribution. This, on the other hand, is reasonable to use if X is significantly non-uniform, i.e. if the probabilities of the two neighboring target percentiles differ considerably. An example may be a discretized normal distribution with a low standard deviation. The probability of transition to neighboring percentiles using discretized normal distribution is shown in the figure below. From the percentile $q_i = \frac{i}{N+1}$, we are most likely to move to percentile $\frac{i}{N+1}$, or less likely to percentile $\frac{i+1}{N+1}$ and $\frac{i-1}{N+1}$ and so on; for sufficiently distant percentiles, the probability is zero.



Probability of transition to neighboring percentiles

In this case, we confirm for all possible values of X, whether the person can move to the percentile determined by the realization of X and its initial percentile. We exclude unfeasible values (i.e., those that are already occupied) from the distribution and will choose only from the remaining values for which we keep the probability proportion. This will ensure that a percentile that has not yet been occupied is selected. We will only use the search for the closest available percentile as described for the previous case if transition to none of the percentiles that can be simulated on the basis of X is feasible. We will then start the algorithm from the values which would come out for the maximum and minimum realization of X.

So far, we have assumed that within the initial distribution, each person is in one of the percentiles $q_i = \frac{i}{N+1}$. If this is not the case, because the values are based on real observations or because we have simulated them in a different way, we take the default values as an empirical distribution. Percentiles of empirical distribution are always the same distance from each other; our only problem is that the highest value has percentile 1 and the lowest one has percentile 0. Therefore, to calculate the percentiles of the empirical distribution, in this case we extend the set of observations by two auxiliary cases, one extremely high and one extremely low. Percentile 0 and 1 will then acquire these auxiliary cases, and to real observations we assign the percentiles $q_i = \frac{i}{N+1}$ (for specific individuals in the model, i is still considered from the range of values 1 to N) that meet the requirements stated above.

If the population size changes between determining the initial and target percentile (e.g., new people will enter the labor market in the new period), we first calculate the initial percentile for each person, which it would have had if the initial population had the same count as the target population, by simple linear scaling and rounding. For example, let us suppose that we calibrate the salary at time $R+1$ when the size of the population is N_{R+1}, on the basis of the salary at time R when the size of the population was N_R. Then the ith richest person at time R was, in terms of salary distribution, in percentile $\frac{i}{N_R+1}$. The order of the new percentile in the new distribution will then equal

$$i\frac{N_{R+1}+1}{N_R+1},$$

rounded to the nearest integer multiple of fraction $\frac{1}{N_{R+1}}$.

Introducing randomness – version two

So far, we have assumed that we know the target distribution for modeling the transition from one percentile to another. Let us now focus on the case where, in addition, we assume the distribution of the behavior of the transition itself (i.e., by how much the target percentile will increase or decrease).

Let us choose a random variable X on which we will rely in the transitions between the percentiles. We require that X have the following properties:

- Expected value of X is zero;
- The range of X is any subset of the interval $\left[\frac{-1}{2}, \frac{1}{2}\right]$;

- X is symmetric, i.e., for any $0 < a < b < \frac{1}{2}$, it holds that $P[X \in (a,b)] = P[X \in (-b,-a)]$.

We will denote the realization of random variable X by a small letter x.

Let U be the variable which determines the starting distribution, i.e., the covariate. The relevant starting percentile for person i will be denoted by u_i (e.g., percentile of salary in year R). Similarly, let V be the target random variable and v_i its percentiles, i.e., values we wish to achieve by the calibration (in this case, percentile of salary in year $R+1$). When determining value v_i, we would like to use the formula $u_i + X$.

However, value $u_i + X$ cannot be directly used without other modifications because for some realizations of variable X, the sum $u_i + X$ will not lie in the interval $[0,1]$. Therefore, we create a new random variable X_i (this variable will be different for each person) and set $v_i = u_i + X_i$. We define it as follows:

- If the sum $u_i + X$ lies in the interval $[0.1]$, we do not make any adjustment, i.e., we put $X_i = X$;
- If the sum $u_i + X$ is negative, we set X_i such that $u_i + X = -(u_i + X_i)$, i.e., we put $X_i = -2u_i - X$;
- And finally, if the sum $u_i + X$ exceeds 1, we set X_i such that $u_i + X - 1 = -(u_i + X_i - 1)$, i.e., $X_i = -2u_i - X + 2$.

We created variable X_i so that the results are uniformly distributed; we will justify in due course that this really holds. However, the fact that the sums $u_i + X_i$ so defined lie in the interval $[0,1]$ will be shown immediately.

Let us recall that variable X belongs to interval $\left[\frac{-1}{2}, \frac{1}{2}\right]$. In addition, the values of the percentiles u_i are always in the interval $[0,1]$. Therefore, the highest possible value of the sum $u_i + X$ is $\frac{3}{2}$ and the lowest is $\frac{-1}{2}$.

If $u_i + X$ is negative, this expression will assume value greater than $\frac{-1}{2}$. Therefore, the expression $u_i + X_i = -(u_i + X)$ will certainly be in the interval $[0,1]$.

We will use a similar argumentation also when $u_i + X$ is greater than 1. This expression clearly has values between 1 and $\frac{3}{2}$, and therefore $u_i + X_i = -(u_i + X) + 2$ again belongs to interval $[0,1]$.

As mentioned above, we define the calibrated variable, shortly written, as follows:

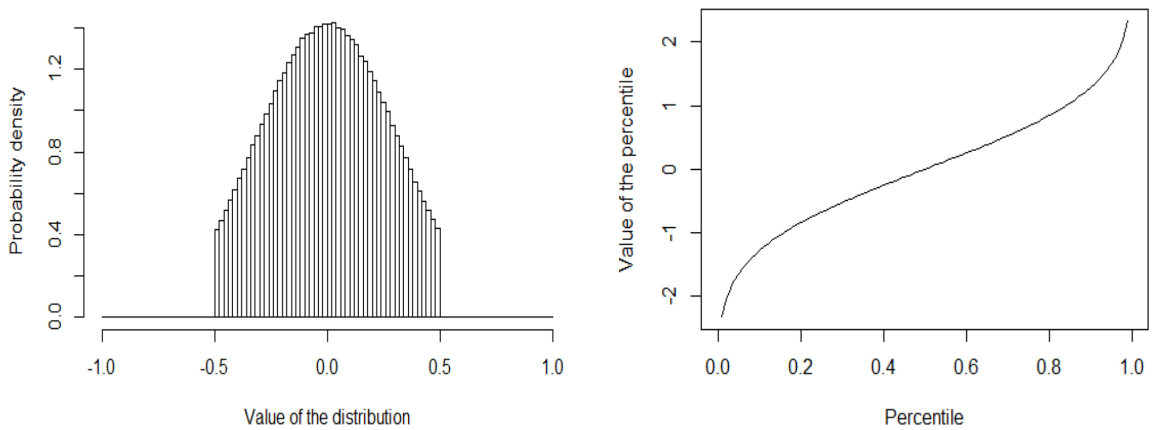
- $v_i = u_i + X_i = u_i + X$, if $0 \leq u_i + X \leq 1$,
- $v_i = u_i + X_i = -u_i - X$, if $u_i + X < 0$ and
- $v_i = u_i + X_i = X_i = 2 - u_i - X$, if $u_i + X > 1$.

The random variable that we want to calibrate has a set $[0,1]$ for these options; each individual can therefore be assigned a percentile from the target distribution. Each person can transition to a range of different states, and there is generally no guarantee that the expected value of the shift will be zero for each individual. However, this is a necessary consequence of our requirements: the richest person can only get to a worse position in terms of percentiles, and if we want him not to remain the richest throughout the projection for sure, the expected value of the change in his or her percentile must be negative.

At this point we could stop and distribute the resulting values of the individual according to their respective values v_i into N quantiles step by step just like we did in the previous version of the method. But that would be an unnecessary extra step. It turns out that the values v_i are uniformly distributed over the interval $[0,1]$, i.e., if we choose any subinterval $[a,b]$, the probability that a person (of whom we have no initial information) has a calibrated variable in this interval equals $b - a$. Equivalently, every two intervals of the same length contained in the interval $[0,1]$ have the same probability that the resulting variable will belong to them (if we don't know anything about the initial state of the person). Therefore, values v_i can be used as a result of the calibration without any further adjustments.

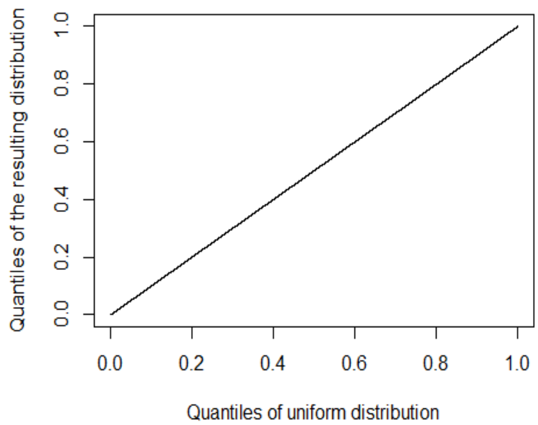
We will now illustrate this procedure with an example in which we substitute X with a distribution that satisfies the above conditions (symmetry, range of values from minus to plus one-half, and zero expected value), we will simulate the values of the uniform distribution over the interval $[0,1]$ and realization from the distribution of X, and check whether the result is in fact uniformly distributed. This example was created in the R software and its code is provided in Appendix B.1.

We use truncated normal distribution with values in the interval $\left[\frac{-1}{2}, \frac{1}{2}\right]$ with expected value 0, unit standard deviation, with the following percentile function and probability density:



Histogram of truncated normal distribution (left), percentile function of truncated normal distribution (right)

It is therefore a symmetric distribution that strongly prefers values around 0. However, if we compare the percentiles of the resultant distribution to the percentiles of the uniform distribution in the QQ-plot, we find that they match very well. (If two distributions perfectly match, all their quantiles, especially the percentiles, match as well. The QQ-plot compares the quantiles of the two distributions and, when perfectly matched, displays the direct proportion function $y = x$. The curve in the chart below approaches direct proportion.)



QQ-plot of distribution created from the normal distribution

We will use a formal statistical test to verify the uniformity of the resulting representation. We specifically choose the Kolmogorov-Smirnov test which is based on the comparison of (empirical) distribution functions. As the null hypothesis, it chooses that the values v_i come from uniform distribution. The resulting p-value is about 0.8 which means that the null hypothesis is resolutely not rejected by the test. The results can therefore be interpreted as a confirmation of the uniformity of the distribution.

The script that accompanies the task contains tests for two other choices of the X distribution, all of which have lead to the same conclusion.

4.4.2 General Evaluation

In this case, too, we will deal with properties for the first and second version of this method separately.

Version without randomness

First, let us comment on the version that (apart from the initial distribution) does not include randomness at all.

Due to the deterministic definition of the vast majority of values, this version is highly computationally effective.

Since this version accurately maintains the order of individuals, it does not matter whether we rank them in ascending or descending order; the selected order will be retained also after the calibration. In other words, this method respects the symmetry of the task.

However, its main disadvantage is that it may not be sufficient to describe the events our model is attempting to capture. For example, in modeling salary levels in the individual years, we probably do not want to determine the development of salaries deterministically, but we would also like to take into account the role of randomness.

The first version with randomness

The first random version a priori does not maintain symmetry. However, it is possible to change some details during the implementation (when percentile q_i is occupied, choose as the first option the percentile q_{i-1} instead of q_{i+1}), so that a symmetric task generates corresponding results.

Computation complexity of this version is significantly higher. While the other two versions are linear in terms of complexity, here the complexity can be up to quadratic. Since the MoLSA counts with large volume data, this should be taken into account when considering the method.

The second version with randomness

Although this version of the refinement method is the most complex in terms of thought, it is still computationally efficient. That is because for every individual in each step (imagine a salary change between two time periods), we only generate the value of random variable X which will be subsequently simply transformed by the defined formula and added to the individual's previous value. Thus, it is still a linear problem, but with a higher multiplicative constant.

The advantage of this version is that it can maintain symmetry. This is true because we require X to be a symmetric random variable. We will gradually analyze all three cases in turn according to how we can get the value of the target quantile v_i. Let P be the probability of an event.

- Let $v_i = u_i + X \epsilon (0,1)$. Then also $(1 - u_i) - X \epsilon (0,1)$. In addition:

  $$\mathrm{P}((1-u_i) + X = 1 - v_i) = \mathrm{P}((1-u_i) - X = 1 - v_i) = \mathrm{P}(1 - u_i - X = 1 - (u_i + X)) = 1.$$

  The first equality results from the symmetry of X. Therefore, we have shown that the probability of transition from one percentile to another is equal to the probability of transition between the respective complementary percentiles. We will proceed similarly in the next two cases.

- Let $u_i + X < 0$ and $v_i = -u_i - X$. Then $(1 - u_i) - X > 1$. Therefore, from percentile $(1 - u_i)$ we would switch to percentile $2 - u_i - X$.

  $$\mathrm{P}(2 - (1 - u_i - X) = 1 - v_i) = \mathrm{P}(1 + u_i + X = 1 + u_i + X) = 1$$

- Finally, let $u_i + X > 1$ and $v_i = 2 - u_i - X$. Then $(1 - u_i) - X < 0$, the appropriate percentile to which we would switch from $1 - u_i$ therefore has the expression $-(1 - u_i) - X$. Then

  $$\mathrm{P}(-(1-u_i) - X = 1 - v_i) = \mathrm{P}(-(1-u_i) + X = 1 - v_i) = \mathrm{P}\big(-1+u_i + X = 1 - (2 - u_i - X)\big)$$
  $$= \mathrm{P}(-1 + u_i + X = -1 + u_i + X) = 1.$$

We have shown that with a unit probability, this method preserves symmetry.

Shared qualities

Since these methods are based on accurate knowledge of the target distribution, they offer a great advantage in achieving the calibration targets.

If the covariate of the observed variable is the individual's history (as it is, for example, in our case of the amount of income that is directly influenced by past income), the relationship between the covariate and the response variable is preserved. This method is unable to take into account any other covariates.

The expected value for calibration is determined in the first two cases by the expected value of the defined distribution. In the latter case, it is again equal to the expected value of the target distribution, which, however, depends on the choice of the random variable X and also on the initial distribution.

Of course, once we accept that transitions between percentiles are controlled randomly, it is possible that individual moves from any percentile to any other percentile. Therefore, with very little probability, the poorest individual can become the richest. Thus, none of the transitions, when using this method, have a probability exactly equal to zero.

Summary of Strengths and Weaknesses

The main advantage of this calibration is that it safely approaches the calibration target. Moreover, the first and third versions are computationally fast and easy to implement. In addition, if the development of an individual (or, more precisely, the relevant observed variable) is influenced by his or her history, this dependence is fully taken into account during the calibration.

As for the disadvantages, it is necessary to mention that the simplest version (without randomness) may not model the situation well enough because of its lack of consideration of the stochastic development of the individuals. The other versions do not preserve zero probabilities.

For the second version, it is necessary to mention also the quadratic time complexity, which can cause problems for large data.

4.4.3 Suitability for Application in NEMO

A key point in the application of the method of refinement of average values is the knowledge of the target distribution of the calibrated variable. It is therefore necessary to enter the probability distributions into the model from which individuals will be assigned values. The easiest way to do this is to always specify the distribution type (e.g., exponential) and its parameters, and introduce a formula into the code that calculates the relevant quantile of the given distribution. An alternative would be to enter the entire distribution in the input table, but it would have to be very extensive, because every individual in the model gets their unique quantile and the table would have to contain them all. Therefore, we do not recommend this solution.

The knowledge of the monthly target distribution would be ideal for the model, but such data is not available. This data requirement can be avoided by setting additional assumptions, such as assuming that the salary will only be increased once a year in the simulation. For a one-year or a five-year timeframe, the distribution of the calibrated variable can be easily acquired, for example, from external statistics of the Czech Statistical Office or from the databases of MoLSA.

In all versions of this method, it is necessary to always know about the person at what percentile he or she is. The easiest way to do this is to introduce a new variable in the input database, which, depending on the method, can be stable or may change every year.

In the version where we want to verify which percentiles remain free, it is necessary to store the results of all individuals allowing the model access when simulating the following individuals. This can be done by introducing an extended type of variable. Such an implementation would not be entirely trivial, and more importantly, the variable would have to contain information on the occupancy of all quantiles in all projection years. Working with such a large array would place considerable demands on the memory and computing capacity of the computer.

So we can see that the version that is suitable for implementation in Prophet is the first version (the one that holds a stable quantile for all persons) or the third version (where the quantile is changed by random distribution regardless of the results of other persons). The version where people can only move to free quantiles is computationally demanding and cannot be recommended.

4.4.4 Final Evaluation

The strength of the method lies in its ability to assign individual persons a position within a predetermined probability distribution of the calibrated variable, consistently with the position of the person in the probability distribution of another variable. Therefore, the method can be used to calibrate salaries; on the contrary, using it for probability calibration would be complicated. Implementation in Prophet would be feasible but not entirely trivial. In general, therefore, it makes sense to use the method primarily for the calibration of salaries in a situation where their probability distributions given by external projections change from year to year.

4.5 Multiplicative Scaling

4.5.1 Method Description

This method can be used at least in certain versions both during the parameter calibration and during the main model run. It consists in multiplying the values for all persons by a certain coefficient. This can be either identical for all individuals or individualized, for example, according to the value of the variable before calibration.

A similar approach is already used in the current NEMO model for calibrating salary growth. A person's salary is influenced by the modeled career development. However, the overall salary growth of all persons in the population should correspond to the average wage growth in the economy. Therefore, the so-called residual wage inflation is introduced into the model as another multiplicative parameter, which will correct the deviation, if any, caused by career growth. Residual wage inflation is an input for the model, so it is necessary to run the model first without it, compare the salary growth calculated by the model with the expected average wage growth in the economy, and calculate the residual inflation on the basis of this comparison.

If the multiplicative factor is common to all persons, it is not difficult to find it: one preparatory run takes place, the value of the monitored variable is compared with the target value of the calibration, and the factor we are looking for is equal to the proportion thereof. If the model inserts a multiplier read directly from the inputs into the formula for the calibrated variable and the interpretation of this multiplier allows it, it will be possible to multiply this input directly by the obtained factor. This is possible, for example, for a monthly salary, which is typically calculated as the previous salary multiplied by salary inflation – the calibration can be done directly by changing the inflation. In other cases, it may be preferable to upload the calibration factor as a separated input and multiply the variable by it directly in the model only after all the other adjustments. This procedure is typically used when the calibrated variable is not controlled by a suitable multiplicative parameter. An example could be the calibration of salaries in the current NEMO model: this can be done neither by changing the wage inflation (since the aim is to approximate the overall salary growth in the model to the wage inflation entered in the input), nor by adjusting the way the salaries increase with advancing careers (because in one year, different people go through different points of their careers), and it is therefore necessary to introduce a third coefficient, namely the residual salary.

However, multiplication by one common factor is not very appropriate when we want to adjust probabilities (e.g., probability of finding a job), since it is not excluded that some of the adjusted probabilities will exceed 100%. This constraint can be eliminated by introducing a more complex version of this method, but only at the cost of disrupting the original probability distribution, since the probabilities of different amounts must be adjusted in different ways.

The following procedure may be an example. If we want to increase the expected value of the observed binary variable by calibration, we increase the probability of each individual event by the portion of the distance between the initial probability value and 1, by whichever portion of this distance we want to increase the average probability. Formally written:

- Let $E[X_0]$ be the initial expected value of the calibrated binary variable with values 0 and 1, $E[X_1]$ the expected value after calibration, and $E[X_1] > E[X_0]$.
- Let $p_0 = \frac{E[X_0]}{N}$ and $p_1 = \frac{E[X_1]}{N}$ be the average probabilities before and after the calibration, i.e., if all individuals had such probabilities, we would achieve the values $E[X_0]$ and $E[X_1]$ for the observed variable. The letter N here denotes the number of individuals in the population.
- Let $k = 1 - \frac{1-p_1}{1-p_0}$ be the said part of the distance between p_0 and 1 that we have exceeded by the shifting to p_1. Now $0 \le k \le 1$ and $p_0 + k(1 - p_0) = p_1$.
- If $p_{0,i}$ is the initial probability of event i, then we get the new probability as

$$p_{1,i} = p_{0,i} + \left(1 - p_{0,i}\right) \times k.$$

- Since $E[X_0]$ is the initial expected value, $\sum_{i=1}^{N} p_{0,i} = E[X_0]$ holds, and from here, after substituting for k
$$\sum_{i=1}^{N} p_{1,i} = \sum_{i=1}^{N} \left(p_{0,i} + \left(1 - p_{0,i}\right) \times k\right) = \sum_{i=1}^{N} p_{0,i} + \left(1 - \frac{1 - p_1}{1 - p_0}\right) \sum_{i=1}^{N} \left(1 - p_{0,i}\right) =$$
$$= E[X_0] + \frac{p_1 - p_0}{1 - p_0}(N - E[X_0]) = E[X_0] + \frac{p_1 - p_0}{1 - p_0} N(1 - p_0) = E[X_0] + E[X_1] - E[X_0] = E[X_1],$$
i.e., the algorithm always adjusts the expected value as desired.

If the goal is to reduce the expected value, i.e., $E[X_1] < E[X_0]$, the procedure is greatly simplified. This is because the ratio of the expected values $l = \frac{E[X_1]}{E[X_0]} = \frac{p_1}{p_0}$ is less than 1. Therefore, if we multiply a number between 0 and 1 by it, we will still stay in that interval.

We will therefore calibrate the probabilities by multiplying the original values by the parameter l, i.e., we will calculate using the following formula:
$$p_{1,i} = p_{0,i} \times l.$$
Maintaining the expected value is very simple in this case:
$$\sum_{i=1}^{N} p_{1,i} = \sum_{i=1}^{N} p_{0,i} \times l = l \sum_{i=1}^{N} p_{0,i} = \frac{E[X_1]}{E[X_0]} E[X_0] = E[X_1].$$

4.5.1 General Evaluation

As Stephensen notes (Stephensen, 2016), this method is not symmetric. Symmetry makes sense when we calibrate transition probabilities. In case we calibrate, for example, the level of salaries, the issue of symmetry does not have a meaningful interpretation.

We illustrate the asymmetry of the method on a short example of transitions between statuses. If we have two persons with a probability of death $p_{0,1} = 0.2$ and $p_{0,2} = 0.4$, the expected value of the number of deaths will be $E[X_0] = p_{0,1} + p_{0,2} = 0.6$. If we want to increase this expected value to 0.8 using multiplicative scaling (i.e., $E[X_1] = 0.8$), we have two options – to calibrate either the number of deaths, or the number of survivals.

First, we use the simple method where each probability is multiplied by the same coefficient. In the first case, we multiply both probabilities of death by a factor determined as the ratio of the two expected values: $k = \frac{0{,}8}{0{,}6} = \frac{4}{3}$ and then $p_{1,1} = k \times p_{0,1} = \frac{4}{15}$, $p_{1,2} = k \times p_{0,2} = \frac{8}{15}$. Since neither of our adjusted probabilities exceeds 100%, the multiplication by one common factor is correct. Next, we calculate that the probabilities of survival will be $1 - p_{1,1} = \frac{11}{15}$ and $1 - p_{1,2} = \frac{7}{15}$. Conversely, if we calibrate survival probabilities, we want to reduce the expected value of the number of survivors from 1.4 to 1.2. Therefore, we multiply both probabilities of survival by the coefficient $l = \frac{1{,}2}{1{,}4} = \frac{6}{7}$ and we get the survival probabilities as $1 - p_{1,1} = \left(1 - p_{0,1}\right) \times l = \frac{24}{35}$, $1 - p_{1,2} = \left(1 - p_{0,2}\right) \times l = \frac{18}{35}$, or probabilities of death respectively as $p_{1,1} = \frac{11}{35}$, $p_{1,2} = \frac{17}{35}$, which are different results.

Let us look on the same case, at how the values change if we use the second calculation method, which prevents any probability from exceeding 100%. In this case, the factor k has value $k = 1 - \frac{1-0{,}4}{1-0{,}3} = \frac{1}{7}$ and probabilities of death are $p_{1,1} = p_{0,1} + \left(1 - p_{0,1}\right) \times k = \frac{1}{5} + \left(1 - \frac{1}{5}\right) \times \frac{1}{7} = \frac{11}{35}$ and $p_{1,2} = p_{0,2} + \left(1 - p_{0,2}\right) \times k = \frac{2}{5} + \left(1 - \frac{2}{5}\right) \times \frac{1}{7} = \frac{17}{35}$. In this case, if we calibrate the survival probabilities, we want to reduce the expected value by calibrating, so we get that the factor l holds $l = \frac{1-0{,}4}{1-0{,}3} = \frac{6}{7}$. The calibrated survival probabilities are equal to $1 - p_{1,1} = 1 - p_{0,1} \times l = 1 - \frac{1}{5} \times \frac{6}{7} = \frac{31}{35}$ and $1 - p_{1,2} = 1 - p_{0,2} \times l = 1 - \frac{3}{5} \times \frac{6}{7} = \frac{17}{35}$, and probabilities of death are $p_{1,1} = \frac{4}{35}$, $p_{1,2} = \frac{18}{35}$, which, again, are different results pointing out that this version of multiplicative scaling is not symmetric.

After we have devoted a lot of space to the symmetry of the task, let us discuss the other properties of this method.

Firstly, it is worth mentioning the advantage of this method, that it exactly achieves the desired expected value. However, it comes at the cost of disrupting the marginal distributions. Although it does not directly reflect the relationship between the covariate and the dependent variable, the changes in the individual values are not random, but are always directly based on the magnitude of the initial value.

It also respects the zero probabilities. In the case of a simple multiplication of probability by the relevant factor, this is obvious (because if we multiply zero by any constant, we get zero again). The same holds, if we want to reduce the expected value by calibration, as is clear from the formula $p_{1,i} = p_{0,i} + p_{0,i} \times l$. In case we want to increase the expected value, let us look at the breakdown of this transformation in case of zero probabilities $p_{0,i}$. We get $p_{1,i} = p_{0,i} + \left(1 - p_{0,i}\right) \times k = 0 + (1-0) \times k = k = 1 - \frac{1-p_1}{1-p_0}$. Therefore, $p_{1,i} = 0$ is true only if $p_1 = 0$, which, however, will not happen because we want the calibrated value $E[X_1] = N \times p_1$ for fixed N to be higher than $E[X_0] = N \times p_0$, whereas p_0 belongs to interval $[0,1]$.

This method is computationally effective because it is performed in linear time.

For this method, it is interesting to look at the distribution deviation index, DDI. If we multiply all values by the same constant, it results equal to zero because $DDI = \sum_{i=1}^{n} \frac{N_i}{N} (S_i - RO_i)^2 = \sum_{i=1}^{n} \frac{N_i}{N} (S_i - RO_i)^2$. Similarly, if we reduce the expected value by calibration, this index is zero because $S_i = \frac{1}{N_i} \sum_{j=1}^{N_i} p_{1,j} = \frac{1}{N_i} \sum_{j=1}^{N_i} p_{0,j} (1 + \frac{p_1 - p_0}{p_0})$, $O_i = \frac{1}{N_i} \sum_{j=1}^{N_i} p_{0,j}$, $R = \frac{p_1}{p_0}$, i.e., $S_i - RO_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(p_{0,j} + p_{0,j} \frac{p_1}{p_0} - p_{0,i} - p_{0,j} \frac{p_1}{p_0}\right) = 0$. However, this property does not apply to increasing the expected value. In such case, $S_i - RO_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(p_{0,j} + \left(1 - p_{0,j}\right) \frac{p_1 - p_0}{p_0} - p_{0,j} \frac{p_1}{p_0}\right) = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(1 - \frac{p_1}{p_0}\right) (2p_{0,j} - 1)$. Therefore, index DDI approaches zero if the resulting expected value differs from the initial very little, or when the initial values are around one-half.

Summary of Strengths and Weaknesses

The main advantage of this method is that it strictly adheres to the determined expected value. Unfortunately, in a more complicated version, when it changes values proportionally, separately for each individual, it does not preserve the individual distribution. Furthermore, it is computationally efficient, but its asymmetry can be a problem.

4.5.2 Suitability for Application in NEMO

Calibration coefficients are calculated outside Prophet by simply comparing the Prophet results and the calibration targets.

A certain complication occurs the moment we want to implement the calculated calibration coefficients in the model. Since we are unlikely to have monthly calibration targets (but rather annual or five-year targets), we will need to make the appropriate conversion. Again, it will be necessary to take care of the cases already mentioned in the comments on the previous methods; for a detailed explanation we refer again to chapter 4.8.3.

4.5.3 Final Evaluation

Multiplicative scaling is a very simple method, easy to implement and computationally not time-intensive. Its main disadvantage is the difficulty in using probability calibration. However, this is a very good method for calibrating salaries, partly because it is already being used in some form.

4.6 Sidewalk Method

4.6.1 Method Description

This method was introduced by O’Donoghue and Li as Sidewalk Method (O'Donoghue, et al., 2014). Its first version was implemented in CORSIM (a model developed in the US to model a social security system) where it is used to limit statistical errors in the results (Anderson, 2019). Individuals who may have the same event at the same time are examined by the model one by one, adding up their probabilities. Whenever a person's contribution results in an integer threshold being exceeded, the event occurs for that person. This technique ensures that the simulated number of events will differ by a maximum of 1 from the expected value. However, this is not a calibration method as such, since it is not possible to move away even intentionally from the expected value based on microdata.

The DYNACAN model (Canadian social security model, built on the basis of CORSIM) uses a version of this method, which, in addition to limiting the randomness of the model, also acts as a calibration technique (Anderson, 2019). Again, the probabilities of a certain event for individual persons are gradually added up, but this time the events are based on a random simulation. The total number of events occurring is then compared to the expected number. If these two values move away from each other, the algorithm will begin to slightly adjust the probabilities for other persons to increase the probability of the result which approximates them again. Both sums persist from year to year, so if there is a difference in one year, the algorithm will try to equalize it in the following year.

It is suitable to use the logit function for probability adjustment, because it is able to convert any real number to probability. The following function is a simple way to take into account the model's evolution to date:

$$p_1 = \mathrm{logit}^{-1}\left(\mathrm{logit}(p_0) + k\,\frac{C - D}{N}\right),$$

where:

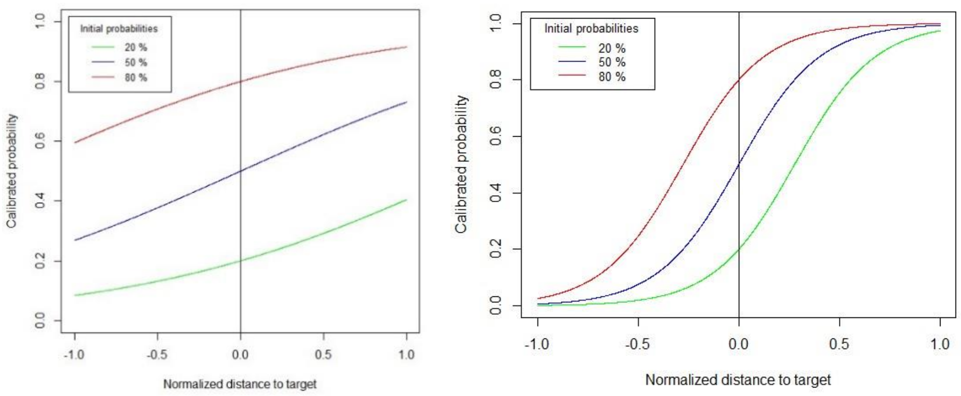
- p_0 is the probability of the event before the calibration;
- p_1 is the probability of the event after the calibration;
- C is the current calibration target, i.e., the number of events that should have occurred according to the calibration target for persons processed by the model;
- D is the current result of the model, i.e., the number of events that actually occurred in the model for persons processed by the model;
- N is the size of the population subject to the calibration;
- k is the coefficient that determines the speed of the calibration.

Please note that the values C and D vary for each person.

Let us now describe in more detail how the probabilities change under this calibration method. If the calibration target and the number of event realizations match perfectly in the model, there is no adjustment to the probability. Similarly, if their difference from population size is negligible, the probability will change only slightly. The more the values vary from each other, the more we adjust the probabilities, and the dependence of the magnitude of the adjustment on the magnitude of the difference is approximately linear at the beginning and then it slows down (because the probability cannot get out of the interval $[0,1]$). The rate at which the probability increases is determined by the coefficient k. A lower k means lower probability sensitivity to model randomness. We recommend selecting this version if the calibration target is not too far from the model result before calibration. On the other hand, a high k will make it possible to achieve also calibration targets more distant from the original model result. The price for this is a more pronounced deviation of probabilities even when there is no good reason for it, because the calibration target is relatively well consistent with the model result so far (and the differences between the calibration target and the actual results are mainly due to the randomness of the model).

These dependencies are illustrated in the following charts. In both cases we capture the dependence of the calibrated probability of the event on the standardized distance of the previous model results from the calibration target (i.e., on the value $\frac{C-D}{N}$). We calibrate three persons: the green curve corresponds to the default probability of 20%, the blue one to probability of 50% and the red one to probability of 80%. The calibrated probability equals the initial probability in all cases when the distance of the calibration target from the previous model results is zero. In the charts, these points are indicated by a vertical black line. We present charts for $k = 1$ and $k = 5$; however, in the event of implementation, it makes sense to consider higher values as well.

The code with which these charts can be generated in R software is presented in B.2.



Sidewalk method with parameter 1 (left) and parameter 5 (right)

The choice of the population size N remains to be commented. Generally, we will use the number of all persons for whom the realization of a random variable is permissible (i.e., for example, excluding those who do not meet the legal requirements for the transition or who passed away in that year). This convention works well if we calibrate each person only once (e.g., calibration of employment in one year or calibration of transition to employment or unemployment after graduation). If we perform calibration repeatedly in each modeling period, it may be appropriate to set N separately for each year (and use the same k for all years). If we used the total number of calibrated cases over all periods, the fraction $\frac{C-D}{N}$ in the first periods would be low and the calibration would only have a real effect in later periods.

The opposite happens if the population size changes significantly during the projection. For example, if a large proportion of the population becomes extinct in one particular period, there may be a difference $C - D$ from past periods which may be high relative to the remaining population. The probabilities will then suddenly be adjusted much more strongly than in the previous period. In such case, it may be preferable to select N as the number of calibration cases for the entire projection or, as the case may be, transfer from the past period only a portion of the difference $C - D$ or not transfer at all.

So far we have assumed that for the choice of N we know the number of persons for whom random selection will take place. This figure is easy to find if all persons are counted in parallel. If not (which is the case, for example, of the NEMO model), it is possible, nevertheless, to estimate the number of cases. An error, if any, in estimating N will have the same effect as if we had chosen exactly the right N and a slightly different k. Therefore, as long as the error remains low (e.g., up to 10%), there is no significant model distortion.

This method can even be used to simultaneously calibrate two interconnected events, such as employment and unemployment. In such case, we will want to keep an overview of the occurrences of both events so far and their consistency with the calibration target separately. At times when only one of the events is relevant (e.g., transition from employment to inactivity), the procedure will remain unchanged. If both events are related to a certain event at once, we simply insert one term for each of them into the calibration equation, with the correct sign. For example, we would use the following formula for the transition from unemployment to employment:

$$p_1 = \mathrm{logit}^{-1}\left(\mathrm{logit}(p_0) + k\frac{C_z - D_z}{N} - k\frac{C_n - D_n}{N}\right) = \mathrm{logit}^{-1}\left(\mathrm{logit}(p_0) + k\frac{C_z - D_z - C_n + D_n}{N}\right).$$

The notation remains the same as above: p_0 is the initial probability of transition, p_1 is probability after calibration, k coefficient expressing the calibration speed, N the number of persons for whom the transition may occur, C_z and C_n calibration targets for the number of employed and unemployed; and D_z and D_n the numbers thereof so far. We therefore increase the probability of transition if we have a lack of employed individuals and a surplus of the unemployed individuals relative to the calibration target, otherwise we reduce it. If both counts are deviated toward the same side, the probability will deflect such as to reduce the larger of the two differences between the results so far and the calibration targets.

The sidewalk method is also available in the LIAM2 and JAS-Mine platforms. In both cases, however, it is rather a complementary option, the documentation of LIAM2 directs the user to the alignment by sorting method (see chapter 4.7), while the developers of JAS-Mine prefer Bi-Proportional scaling (see chapter 4.8). According to our information, the method is implemented in both models only in its simple version which controls the dispersion of results, but does not serve for calibration as such.

4.6.2 General Evaluation

To better talk about the pros and cons of the sidewalk method, we divide it into cases where we are using version number one (a change in probabilities based on exceeding an integer limit) and version number two (with a random change in transition probabilities).

Version 1

The first version of this calibration method offers the advantage of its straightforwardness. This is manifested in the efficiency of the calculation, which only runs one loop across all persons in each period and requires a single, linear, auxiliary calculation.

As mentioned above, this method maintains distribution in terms of the expected value which differs by no more than 1 from the expected value before calibration. If we have data arranged randomly (or, more precisely, in an order which does not depend on the size of the probabilities of individuals), the probability that we indicate the event for that individual is directly proportional to the magnitude of their respective transition probability. Individual partial distributions are therefore also maintained. Similarly, one can also see that this method maintains symmetry.

The first version maintains zero probabilities. Let us note that they must be defined really equal to zero. Once a probability is defined as a very small but nonzero number, it also increases the interim total and may cause it to exceed the integer value.

Version 2

The second version of the sidewalk method is significantly more complex than the first version. It will therefore be somewhat difficult to implement, but it is still not a difficulty that would make it impossible for MoLSA to use the method.

Computation efficiency is reduced again when using the second version, since each transition probability (at all relevant times for all individuals) needs to be recalculated and, in addition, at every step we look for the value of the fraction $\frac{C-D}{N}$. However, we are still in linear complexity.

The second version, however, compensates these disadvantages by responding more flexibly to the distance from the calibration target. The disadvantage is that this version does not preserve the original distribution. When the calibration target differs significantly from the number of simulated events, there are steep (illogical) changes of the transition probability for individual persons. However, since this method converges to the calibration target, it also approximates distributions that are similar to the initial distributions before the calibration. The speed and smoothness of this convergence depends on the choice of parameter k, on knowledge and choice of parameter N, as well as on the variability of the initial distributions and the a priori distance from the calibration target. The exact discussion is presented above in the method description.

Let us now open the discussion on the preservation of the relationship of the covariate and the response variable. Of course, it is a fact that individual probabilities can change step-wise independently of their regressors, but there is another view that will be more interesting in terms of using this method by MoLSA. If we do not look at the probability of occurrence of the event for an individual, but rather the number of events for a larger group (for example, groups of men or women, age groups, etc.), the sidewalk method will respect that classification. The reason is the aforementioned convergence of the method.

Although this method, as described above, adjusts zero probabilities and may increase their value, it is not difficult to introduce a simple condition into the implementation that will prevent the recalculation thereof. A problem could only arise if some probabilities were equal to a small positive number instead of an exact zero. The value of the logit function could then be noticeably increased for them, and maintaining zero would be violated.

If we found a suitable division of the population into groups to calculate the Distribution Deviation Index (DDI), this index can serve us not only for comparison with other calibration methods, but also for checking which choice of parameter k is suitable for our case.

Summary of Strengths and Weaknesses

The main strengths of using the sidewalk method include the fact that the method is comprehensible, easy to implement and has low demands for computation capacity, and always approximates the calibration target for a suitable parameter value. Moreover, the first version maintains symmetry.

As for weaknesses, it is necessary to stress that the method does not guarantee preservation of the original distribution of individuals; for some individuals, it can model sudden transitions between individual statuses that do not have a logical interpretation. However, for larger groups, this distribution is maintained.

4.6.3 Suitability for Application in NEMO

This method can be implemented directly into the NEMO model and performed in a single run. No auxiliary tools created in other programs are needed and no auxiliary runs are needed.

The calibration input will be a table with probabilities of calibrated events occurring and the number of persons involved in the calibration (for example, with the employment rate and size of active population in a given cohort). These values can be determined from external projection. The table may typically depend on the year of birth, sex and calendar year, but it is also possible to imagine a division according to other criteria.

In the NEMO model, we then introduce a variable of the extended formula type, in which we will observe the current number of events (if we want to divide the calibration by certain groups or calendar years, we will keep an array of the appropriate dimensions in the variable). We will upload this value whenever a calibrated event is to be decided and complete it with data from the input table.

- A complication is that events in the NEMO model typically occur in a number of ways. Calibration must therefore be included in all places where decisions on the event are made. In general:
- If the event occurs deterministically, we count it in the number of events that occur, but we will not calibrate the probability in any way. Such behavior is shown, for example, by the recovery of a sick employee – the duration of the sickness is determined at the commencement of the sickness and is not further adjusted (at least as long as the person remains employed).
- If the probability of an event depends on a certain occurrence, we will not change the probability of the occurrence as such, but we will change the probability of the event itself. For example, we will calibrate the probability of leaving employment as a result of retirement, but not the probability of retirement as such.
- If the probability of an event is determined based on a transition matrix that is applied without additional conditions every month, we will adjust it.
- Since we typically want to calibrate the number of people in a particular status, we will need to include also persons who remain in that status. The probability of staying in a certain status does not appear in the model itself, but the model uses probabilities of transition from a calibrated status to another (for example, if we calibrate the number of employed individuals, we also adjust the probability of transition from employment to unemployment). These will be modified again using the principles described in the previous three points (i.e., taking into account the dependence, if any, on an event or fixed duration of the status). Since this time we calibrate the non-occurrence of the event, while we have considered its occurrence so far, the sign in the calibration function changes and the result is that the probability of leaving the status will be calculated as
$$p_1 = \mathrm{logit}^{-1}\left(\mathrm{logit}(p_0) - k\,\frac{C-D}{N}\right).$$
The meaning of all symbols remains the same, C and D therefore still relate to the number of persons who are or should be in the calibrated status at the end of the period (i.e., after application of the transition probability).

Individuals must be calculated in random order to perform the method correctly. (For example, if the model were to run the persons starting out as employed first and then unemployed persons, the employment would be too high after running the first persons and the model would apply unnecessarily strong adjustments.) This cannot be easily accomplished in Prophet. However, random ordering can be easily done in an external tool or incorporated into data preparation procedures in DCS.

The main challenge in implementing the method in Prophet would be to find all the places in the model that need to be adjusted. Individual adjustments in the code would be rather simpler.

4.6.4 Final Evaluation

The sidewalk method can perform calibration within a single model run in Prophet, does not require creation of external tools or too complex input structure. Its main disadvantage is the fact that the way a particular individual is calibrated depends on the order in which the individuals enter the calculation. Nevertheless, the method preserves the probability distribution of the calibrated variable to a large extent. At the same time, it is possible to calibrate several variables at once. Therefore, we consider it to be one of the best options for the calibration of state variables, in particular employment or, as the case may be, unemployment.

4.7 Alignment by Sorting

4.7.1 Method Description

The alignment by sorting method was also described by O'Donoghue and Li (O'Donoghue, et al., 2014) and one of its versions is used in the LIAM2 model. As the only method in this study, it always achieves perfect match with the calibration target; its application should be therefore considered especially for transitions that occur only for a small group of individuals or with a low probability but which are nevertheless essential for the outcome of the projection. The basic principle is very simple: we calculate the probabilities of the event for all individuals and sort them from highest to lowest. If the calibration target indicates that the event should occur in M cases, we select M individuals with the highest probability and assign the event to them.

This basic version has one significant disadvantage: when a group of persons subject to the same calibration target contains persons with different probabilities, persons with a high probability are always selected, while persons with a low probability are never selected. This happens very often because the calibration target is usually in aggregate values. For example, if this method were used to calibrate the total number of newly awarded disability pensions in the population, all new disability pensioners would fall into the oldest years because the probability of disability in models typically depends mainly on age. This is obviously at odds with reality. A numerical example of this characteristic is provided in the evaluation section.

The first way to get rid of this constraint is to include a random component in the sorting, i.e., sort not by probability alone, but by the sum of probability with a random number generated from uniform distribution over the interval $(0,1)$. As a result, a person with a low probability can be chosen for realization of the event if a high random number is generated for his or her.

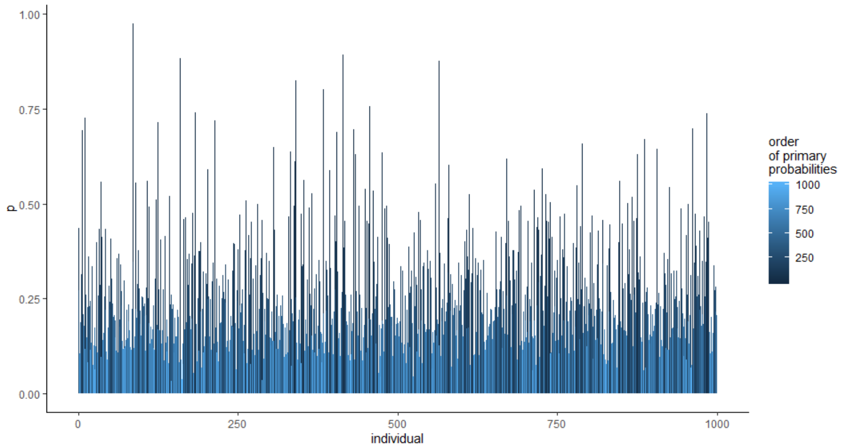
The second option is to insert a random value in the inverse logit function argument and instead of a simple uniform distribution we choose a distribution based on logit transformation; specifically, we sort the persons according to $logit^{-1}(logit(p) + x)$, where p is the initial probability and x is the value generated from uniform distribution. As a result, all resulting values will belong to the interval $(0,1)$.

The alignment by sorting calibration technique is also implemented in the LIAM2 model (de Menten, et al., 2019), an open-source platform for creating and managing microsimulation models which is currently developed by researchers from the Federal Planning Bureau in Belgium, CEPS/INSTEAD and Inspection Générale de la Sécurité Sociale in Luxembourg. The user can specify any function on the basis of which persons will be sorted and selected: the model then enters the input probabilities of the event for individual persons, sorts the persons according to the results, and selects the specified number of the highest results. Random number generation can also be part of the function. In this way, it is possible to replicate any of the three alignment options mentioned above, with the model developers recommending the third, logit option. The user can also specify a group of persons who are always or never selected – for example, when calibrating the number of unemployed individuals, we may want to strictly exclude old-age pensioners. This mechanism is particularly useful if the calibration is to be applied to several non-overlapping groups.

Note that the basic form of this method works best if each person has a slightly different probability. If the probability can only attain a few values in the model (for example, if a modeling method has been chosen where the probability of the birth of a child depends solely on age and the highest education level achieved), there will be a larger group of persons who all have the same probability of an event, but out of this group the event is only supposed to occur for some. It will then be necessary to apply a random selection. Conversely, if the probability of the event is sufficiently refined (for example, if, in the modeling approach, the probability of finding a job decreases steadily with each passing day of unemployment), we will avoid this problem. Similarly, this problem will no longer exist if we enrich the method by generating random numbers.

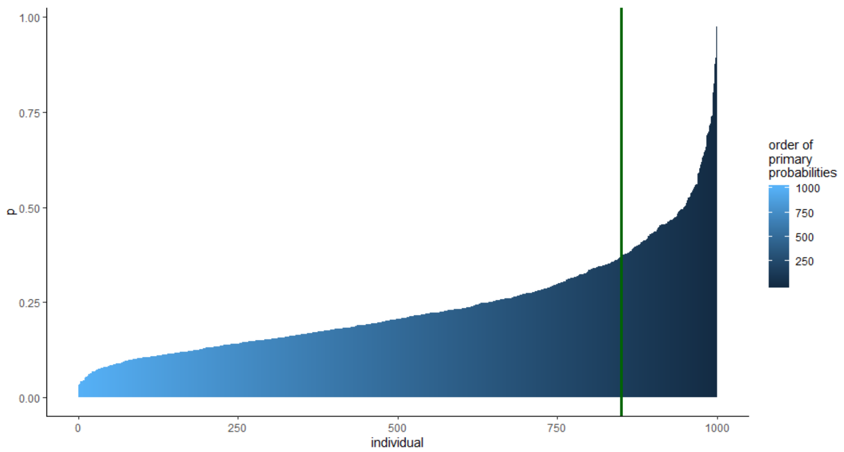
4.7.2 General Evaluation

Let us begin with the example we promised in the section describing this method. The code for the following charts is presented in Appendix B.4. For 1000 individuals, we simulate probabilities from a lognormal distribution with zero expected value and standard deviation equal to one half re-scaled to the interval [0,1]. For each individual, we remember which highest value was assigned to him. This is represented in the charts by the shade of the color. The first chart shows the values of the generated probabilities.



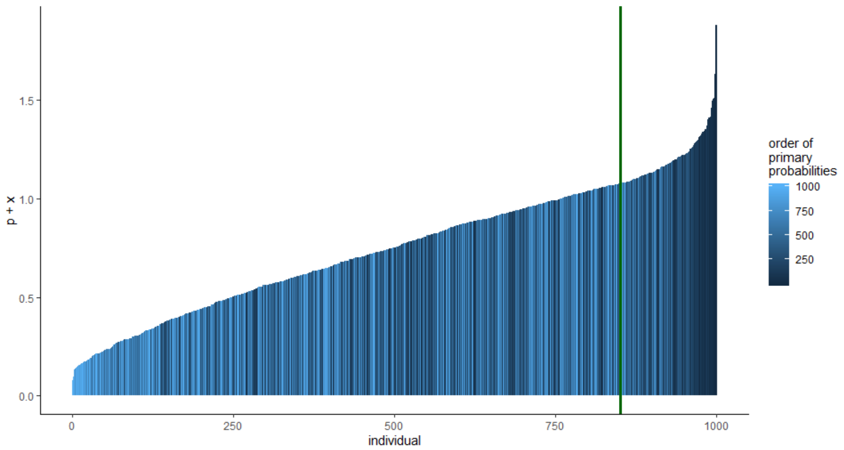
Primary probabilities

The dark green vertical line in the following charts indicates the limit of the number of individuals for whom an event occurs (M); in this example we chose the value 150. Now let us see how the first method handles this data. First, it sorts the individuals in ascending order by probability and then decides that the event will occur for those who are to the right of the green line in the chart, i.e., select the desired quantity of M individuals. We will notice that this result also corresponds to the selection of the “darkest” individuals.



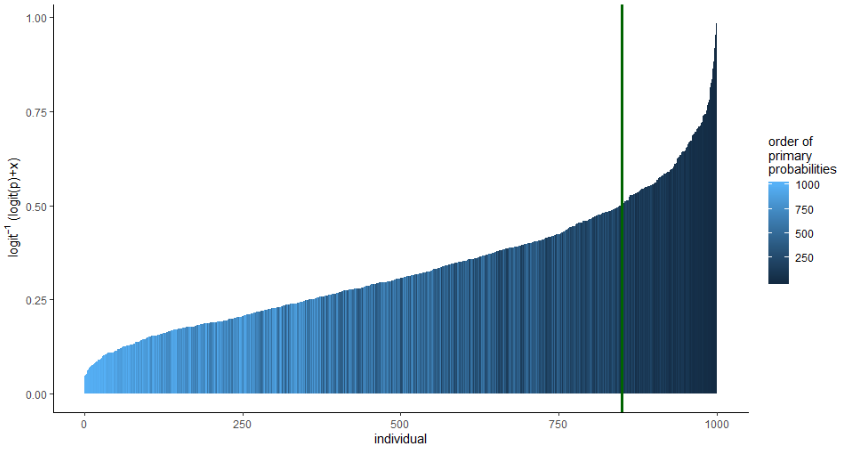
Ordered primary probabilities

In another version, we first add a random number to the probabilities of these individuals from the interval (0,1). The adjusted and sorted probabilities are shown in the third chart. Again, the event occurs for individuals who are to the right of the green curve in the chart. Note that in this case, a relatively small number of individuals were chosen who had a high probability in the original distribution (i.e., they are colored in dark shades).



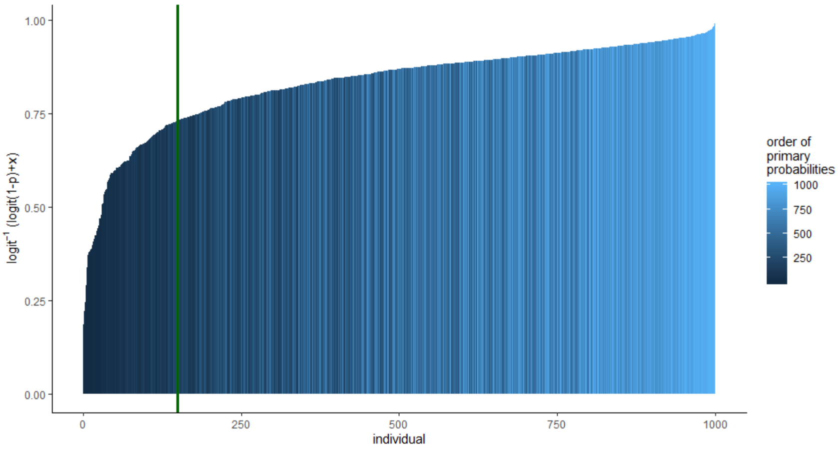
Ordered probabilities after adding x

Finally, the last method first transforms probabilities using the formula $logit^{-1}(logit(p)+x)$ and subsequently sorts and selects individuals to the right of the green line, which is shown by the following chart. The chart also shows that, compared to the previous one, this variant of the method selects a larger number of individuals with a high initial probability. This is the behavior we would expect in the real situation. From this point of view, therefore, this version is more beneficial than simply adding a random variable x.



Ordered probabilities after logit transformation

This method maintains the symmetry of the task. In the simplest version it is obvious, in the more complex ones it deserves a small comment. Since we select random variable x from a uniform distribution, the addition of their realizations can be viewed as an operation that does not change the order of the probabilities. In addition, the function $logit^{-1}(logit(p)+x)$ in the argument p is a strictly increasing function. Therefore, symmetry is maintained in this case, too. The last chart of this section illustrates the symmetric role of the logit version of the example above.



Ordered probabilities after logit transformation in symmetric task

A big disadvantage of the method, on the other hand, is the fact that (except for the first version in which we do not bring any randomness), it does not meet the requirement for maintaining zero probabilities. Even if an individual has a zero probability, the other two versions can (with a low probability) model the situation that the event occurs.

As mentioned in the introduction, the great advantage of this option is the precise adherence to the calibration target. However, it pays for this by violating the distribution of individuals and, moreover, is unable to maintain a relationship with the covariate.

In terms of computing, the alignment by sorting method is not very expensive. Its complexity will correspond to the complexity of the sorting algorithm which is basically $Nlog(N)$ for a population with the size of N.

4.7.3 Suitability for Application in NEMO

The implementation of this method in Prophet encounters major obstacles. The main reason is that Prophet always runs the whole career path of one person before moving on to the next person, while forgetting all data except the so-called reporting variables before moving on to the next person. The variable to be calibrated is therefore not known to all calibrated persons at any time. It is therefore not possible to sort them accordingly and select those with the highest probability.

The theoretical way to overcome this constraint is to record the values of the variable in an external file, identify the individuals to be transitioned using an external tool, write their IDs in a table, and use the table in a new run. However, this approach has two basic pitfalls. The first is the size of the external table to be recorded in – writing values for each person in the population of about ten million (active at a given moment, i.e., without the dead and unborn) puts significant requirements on the computing power.

The second and more serious constraint is the fact that only one period can be successfully calibrated using one pair of runs. As soon as we calibrate the first period, we change the numbers of persons entering the next period as well as the probabilities of their transitions. The values we listed in the first run will no longer be valid, and it will therefore not be possible to make the correct selection on the basis of these values. For each additional period, it would be necessary to run the model again, and this would increase the computing time disproportionately.

We therefore do not consider the method suitable for use in Prophet.

4.7.4 Final Evaluation

The alignment by sorting method significantly undermines the underlying probability distributions and its implementation in the NEMO model would be challenging. Therefore, it cannot be recommended for the MoLSA needs.

4.8 Bi-Proportional Scaling

4.8.1 Method Description

This method is proposed by Stephensen (Stephensen, 2016) and it is implemented as the main calibration method in the JAS-Mine platform. According to its author, it fulfills all eight criteria which he has set for calibration methods in the same article; in particular, it can also be used in a situation where a decision is to be made among several possible states (while other methods usually require binary decisions).

In order to perform this calibration method, we create a large table where the number of rows is equal to the number of persons whose probabilities we are calibrating, and the number of columns is equal to the number of possible output statuses. Each table element determines the probability that a certain person (given by a row number) will make a transition into a certain status (corresponding to a column number). So we want to achieve a situation where the sum of each row is 1 (because it is a sum of probabilities and we include all the statuses that the person can make a transition to) and the sum of each column will correspond to the calibration target for the corresponding status (then the expected value of the transitions given by this table will correspond to the calibration target). At the beginning of the calibration we enter the default transition probabilities in the table – i.e., the sum of each row will be 1, but the column totals will differ from the calibration targets.

The calibration itself is then performed by iterative repetition of two steps. In the first step, we multiply each of the columns by a coefficient such that the totals within the columns correspond to the calibration target (we will have one coefficient for each column in the table). However, this distorts the totals within rows. In the second step, on the other hand, we multiply each row by a coefficient so that the totals within the rows again produce 1 (this time we will have as many coefficients as there are rows in the table). We will repeat these two steps until we reach a state where, even after the second step, the column totals approach the calibration target. Since convergence is very fast in the normal situation (see the method evaluation section), it is possible to set a relatively strict goal, for example: "The sum of deviations of modeled counts from calibration targets across all statuses must not exceed 0.01% of the total population." It is also recommended to limit the number of iterations of the algorithm to ten, for example, and issue a warning if the required accuracy cannot be achieved in this number of iterations.

The algorithm converges for all achievable assignments (Deeparnab, et al., 2018). An analysis of which calibration targets are achievable is presented in the evaluation section.

4.8.2 General Evaluation

We have stated that the method can only give results for some assignments. The first condition is that the total sum of the calibration targets (i.e., sum of the number of persons we want to have in each status after the calibration) must be equal to the size of our population, i.e. the total number of rows of the transition matrix. If it is not the case, the calibration cannot be successful – the method is unable to add or remove persons who were not initially in the system.

Another example of a non-achievable assignment can be the extreme situation where the transition probabilities are expressed by the identity matrix, i.e., we have the same number of persons as the target states, each person only makes a transition to one of them and each person makes a transition to different one. In a population of three people, the transition matrix would look like this:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Imagine that the calibration target is for an average of 1.2 persons to make a transition to the first state, 0.8 persons to the second state, and for the number of transitions to the third state to remain unchanged at 1. In the first step of the algorithm, we would multiply the entire first column by 1.2, the second by 0.8, and the third by 1 to get the following matrix

$$\begin{pmatrix} 1.2 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore, only one person was affected in each column. In the second step of the algorithm, we adjust the sums in each row to 1. So, we divide the first row by 1.2, the second by 0.8 and the third by 1, and we get to the identity matrix we started with. Thus, the algorithm never converges. The reason here is that no adjustment affects more than one person.

From here we see the second necessary condition for the algorithm to be successful: no calibration target shall be greater than the number of persons with a nonzero value in that column in the starting matrix. Zero values will remain equal to zero and nonzero values may increase to a maximum of 1 (because we count with probabilities), therefore, for each status, there is an upper limit for the number of people we can move into it.

Let us now comment on the speed of convergence of this method, i.e., the number of iterations needed to achieve the required accuracy. It is possible to find estimates in the literature (Deeparnab, et al., 2018) of the rate at which the required number of iterations increases as the number of persons calibrated increases. According to these results, it cannot be ruled out that up to several million iterations may be needed for a single calibration, depending on the required accuracy, so caution should be warranted. However, it would be a mistake to reject the method based on this information alone, because in MoLSA conditions, calibration may be easier than in the most difficult scenario (while the study provides an estimate valid in general) and because the estimate for the worst-case scenario in the article may not be optimal.

Therefore, we tested the method using an R script, which we implemented and used for test data. With 10 million people, six possible statuses, and a permitted deviation of no more than 1,000 persons in all states together, the calibration was performed in a maximum of five iterations and lasted less than 10 seconds (a four-core work notebook with a 2 800 GHz processor and 8GB of memory was used). This result alone cannot be considered conclusive because it does not guarantee that there are no initial settings for which significantly more iterations would be required. However, it suggests that the method is indeed quite effective under normal circumstances.

The indisputable advantages of this method include its convergence for almost all cases. Furthermore, if the probability of a particular state for a certain person is zero, the above-described algorithm is not able to re-scale that zero value to a non-zero value – the algorithm only adjusts nonzero probabilities of states.

Talking about a symmetric role may be irrelevant in this case. As mentioned in its implementation, the bi-proportional scaling method has the advantage of being able to calibrate the transition probabilities to a greater number of possible states simultaneously. To talk about symmetry, we would have to find a variable that attains only two different values and, at the same time, is independent of all others. Even for our classic example of the number of survivors and non-survivors, it is necessary to realize that the condition of independence also affects other statuses; for example, only an individual who is also a survivor can be employed.

For each individual during the algorithm we multiply his or her individual transition probabilities by different numbers that are not (at least linearly) dependent on each other; this shows that this method disrupts the form of the primary distributions of individuals.

4.8.3 Suitability for Application in NEMO

It is not possible to implement this method in the current NEMO model, as it calculates individual persons gradually: first the whole life path of the first person, then the whole life path of the second person, and so on. Therefore, Prophet never knows the transition probabilities of all persons at the same time and therefore cannot perform this calibration method. It would therefore be necessary to create an auxiliary tool in some other program (e.g., in R), upload all the necessary transition probabilities into it, perform the calibration there, and transfer the results back to the model in Prophet.

Since the initial number of people in each status can be determined directly from the model point database and transition probabilities from the input tables entering the NEMO model, there is no need to run Prophet at all during the preparation for calibration. All necessary tasks can be performed directly in the external tool.

Since not all statuses are present in external projections, the analyst has some freedom in setting the calibration targets. We recommend choosing a simple method here, for example, to take over all the numbers that are explicitly available in the external projection and set the others in the same ratio as predicted by the NEMO model in a given period.

In its basic form, the method assumes that each line will represent one person. Such a matrix would be very extensive for the NEMO model, and if calibration occurred frequently (e.g., every month) and there were many possible transitions, we might run into computation capacity constraints. We will partially solve this problem by grouping individuals with the same profile in terms of transition probabilities. If two people have the same age, education, sex, current position in the labor market and all other predictors, their transition probabilities will also be the same, and it will therefore be possible to count them all on the same row. We then have to multiply the entire row by the number of people included, so instead of the transition probabilities we will present the expected value of the transitions to the particular status, and the total number of persons as the required row sum. Even if there were several thousands of such groups, the process would be significantly accelerated.

We have already seen that performing the calibration in R is very simple and computationally efficient. Similarly, adjusting the calculation in Prophet using the calibration results would not be difficult. In this method, all columns and rows are gradually multiplied by some coefficients. We can always multiply the coefficients associated with a certain row to obtain the resultant multiplier for that row, and then we do the same for columns. Then we obtain a calibrated value for each transition probability (i.e., every element of the table) by multiplying the original value by the resultant multiplier of the row and the resultant multiplier of the column. We can summarize these multipliers in a table that will enter the NEMO model and then adjust the transition probabilities directly while the model is running. Therefore, once we transfer the appropriate transition probabilities to R, the rest of the assignment is not complicated.

The most difficult part of this assignment is its beginning, namely the extraction of transition probabilities for individual persons. The transition can happen in several different ways. In some cases, the number of months a person remains in a particular status is determined upon transition to that status, and the transition probabilities are then zero until that time expires. In addition, the person always returns to the original status after this period. In this way, for example, the transition to and from a sickness is modeled. In other cases, the transition probabilities strongly depend on the occurrence of a certain event or are triggered automatically by it; for example, a child birth event always triggers a transition to a child care status. Finally, some transitions may occur without an event and their incidence is tested every month (for example, transition from employment to unemployment). We must take all these cases into account in an appropriate way.

Event-independent transitions are simple in this respect. Their probabilities are based solely on the person's data (such as age, education or initial state) and are not further modified in any way. So we can simply include all starting and ending states in the matrix intended for calibration and determine the transition probabilities based on the input tables.

In fact, event-related transitions are a composition of two transitions to target states, a transition given by the event and transition conditioned upon the event. So we can express them by the transition probabilities to two possible target states – for example, an employed person may, on the one hand, make a transition to a non-working old-age pensioner (i.e., the retirement event also caused a transition from employed to unemployed), and, on the other hand, to the state of an employed old-age pensioner (i.e., this time the retirement event did not cause a change in the employed state). The initial probabilities of these transitions are obtained by multiplying the probability of the event itself (for example, retirement) and the probability of a follow-up change (for example, leaving a job upon retirement). Conversely, we can decompose the calibrated probabilities into the probabilities of the individual sub-events.

We use a similar procedure for statuses that have a predetermined duration. Some of the people in this status will make a transition to another status because the duration of the status has expired, others may make a transition based on a different transition probability. When creating a matrix intended for calibration, it is necessary to put these movements together in a common transition matrix, and it is also necessary to divide the number again after the calibration back into two different effects. Unlike the previous case, this distribution may not be unambiguous if both effects cause a person to make a transition to the same combination of states (for example, transition from employed student to inactive student). It is then up to the calibration analyst to decide what part of the change to attribute to which effect. The number of transitions caused by the expiration of a fixed duration can be adjusted by changing the parameters of the probability distribution from which they are generated – for example, the duration of employment of working students is generated from an exponential distribution with a parameter of 12 months. Decreasing this parameter will shorten the generated times and thus increase the frequency of transitions. (It is necessary to bear in mind that such a parameter adjustment affects more than one time point. However, we assume that calibration will only be performed at certain milestones with greater spacing over time, for example every five years. If it is possible to set the probability distribution parameter for each milestone separately, the calibration performed for one milestone will not have a significant effect on the calibration in subsequent periods due to this spacing, and therefore this influence need not be taken into account.)

We can therefore see that we are able to calibrate all the basic transition modes found in the model using the method of bi-proportional scaling . Nevertheless, a few more considerations need to be made. The method assumes that for each person, all the states he or she can enter are listed (the sum in each row is 1 or, more precisely, it is equal to the number of people the row represents). Mortality should therefore be also included. However, we take it directly from external sources and do not want to further modify it. Therefore, deaths will not be included in the list of possible statuses. In reality, we will calibrate the transition probabilities conditioned upon the person's survival in the period. This also corresponds to the design of the NEMO model. We are making a minor error by doing this if mortality varies for different groups within the same cohort - for example, for healthy people and disability pensioners. However, we do not expect this simplification to have a significant impact.

For a similar reason we also exclude disability pensioners from the set of states. Again, this is a simplification, as the event of disability commencement entails a change in some transition probabilities. However, it occurs so rarely that the impact on the overall results of the cohort should remain within acceptable limits.

An important aspect to consider is the choice of the time period. The problem would be simple if we calibrated a new transition matrix for each month. However, this would not be acceptable, on the one hand, for capacity reasons, and on the other hand, we do not have such fineness in external projections. So we choose a longer period so that we know the numbers of persons in individual states at the end of the period (e.g., 5 years) from the external projections. To determine the default transition matrix resulting from the model, we upload the transition matrices valid for our cohort in all periods monitored from the input data and multiply them (transition probabilities may change with age). We will calibrate the transition matrix for this longer period and then we will "take the root" of it again to get a monthly matrix. From a mathematical point of view, however, it is not guaranteed that this "taking the root " can be done. In this respect, it would be necessary to carry out a more in-depth analysis of whether the specific features of the NEMO model allow for this calculation.

Another way to get the monthly transition matrix is to convert external projection data to monthly transitions. If we know that N unemployed people should be added during the year, we can assume a uniform increase and calibrate the transition matrix so that $N/12$ are added in the first month. We will then use this matrix throughout the period until the next milestone when we calibrate the new transition matrix. This procedure is, of course, inaccurate: already in the first month the number of persons in individual states will change slightly; in the second month, the same transition matrix will be applied to a different initial state and the increments in the number of persons in each state will change. The split between the model and the external projection will increase over time, and it would be necessary to examine more closely whether it remains within acceptable limits. Of course, the differences can be mitigated by choosing a more appropriate distribution of the annual change between months, and it makes sense to try to find such a version at least for the important transitions (involving many people). Even then, however, the external projection cannot be expected to be replicated perfectly.

We can see that while on the theoretical level this method is very simple and straightforward, there are significant challenges when connecting it with the NEMO model. These can be dealt with, but the question remains whether the advantages of the method justify the demand of the process.

4.8.4 Final Evaluation

Bi-proportional scaling is a method with a solid theoretical foundation and good mathematical characteristics. However, its implementation for MoLSA would be very complicated due to specific properties of the NEMO model and Prophet in general.

5 Evaluation of the Methods from MoLSA Point of View and Recommendation of the Most Suitable Method

5.1 Selection of the Most Suitable Methods for Individual Parts of the Calibration

In chapter 4 we have analyzed several calibration methods and evaluated each one of them from the theoretical perspective as well as in terms of implementation in the Prophet system. In chapter 3.1 we decided which variables we want to calibrate (both generally and for each of the two important external projections (i.e., the AWG projection and the projection by the Czech Fiscal Council), and we also know that a different method may be appropriate for each variable. Now we will summarize all these inputs and use them as a basis for selecting the methods we recommend for the specific needs of MoLSA.

Overall, the calibration will consist of the following steps:

- Preparatory phase where we take input tables from the external projection (see 4.1);
- Calibration of employment and unemployment using the sidewalk method (see 4.6);
- Income calibration using multiplicative scaling (see chapter 4.5
- Verification of calibration results.

By choosing this sequence, we ensure that later steps only minimally affect the quantities calibrated in the previous steps. In the preparatory phase, we align the initial population, birth rate, mortality and disability rates in the first place. A number of other variables depend on their values, but they themselves are not affected by them in the NEMO model. In the second phase, we are focusing on the calibration of employment and unemployment, which depend on the previous variables, but not on the average income. Lastly, we calibrate the average income, which directly or indirectly depends on all other quantities. By choosing this sequence, we significantly reduce the need to go back to the previous steps and perform calibration iteratively.

The main pitfall in this sequence of adjustments is the probability of retirement. This depends on employment, but it also affects employment itself. Therefore, we carry it out both in the first and second step. We explain this procedure in the following chapters.

5.1.1 Preparatory Phase of the Calibration

During the preparatory phase, using the procedures described in 4.1, we take the following inputs from the external projection:

- Mortality and disability rates;
- Birth rate;
- Initial population;
- Number of old-age pensioners.

In this way, we achieve the reconciliation or approximation of important values, which further appear in the projections as key variables. The process of taking the data requires a relatively simple recalculation or the data can be taken directly.

For the initial population, we assume that calibration will not be necessary because the initial numbers of employed individuals and old-age pensioners and the total initial population are known from the statistics of the Czech Statistical Office, so there is a chance that the external projection will build on the same numbers as the NEMO model. Should MoLSA ever work with a specific projection where this initial consistency is breached, it will be possible to adjust the total number of persons and their composition to match the external projection, using one of the following approaches:

- Model point scaling: in each cohort for which the size of initial population in the external projection is known, the cohort size in that external projection and in the NEMO model is compared. The number of persons in model points is adjusted accordingly. If e.g., the external projection has 20% more persons in a given cohort, in the NEMO model the number of persons in the model point (INIT_MEM_IF) will be set to 1,2 for persons falling in this cohort.

- If the model user needs to maintain the approach that one model point always represents exactly one participant, it is possible to adjust the size of the population by removing or duplicating, as the case may be, some appropriately selected model points.

Even after taking the inputs, the differences due to the randomness of the NEMO model will remain between the projections. At the same time, we will observe a deviation in the number of old-age pensioners exceeding the impact of randomness because this projection uses simplifications to model the number of pensioners which should not be taken to the NEMO model. However, there will be an approximation of the average retirement age, which we will improve even further performing calibration of employment and unemployment

5.1.2 Calibration of Employment and Unemployment

We will calibrate employment and unemployment using the sidewalk method described in chapter 4.6. It has many advantages: it is not too hard to implement and does not require preparatory runs or creation of an external tool. It also exhibits acceptable mathematical properties. The method is illustrated on a model created in MS Excel attached hereto as Appendix B.3.

Another important advantage of the method is its ability to calibrate unemployment together with employment. Therefore, we will calibrate these two variables together whenever they are available in the external projections.

An important decision in terms of the method implementation is the choice of the coefficient k determining the speed of the calibration. Generally, we want to find the lowest k such that the calibration targets can still be achieved (with sufficient accuracy). However, solving this problem analytically is not easy, but instead we recommend trying a few values first in the prototype and then in the live model and finding k empirically. The coefficient value can then remain stable also for further calibrations until the model changes significantly; therefore, it is recommended to repeat its derivation only in connection with major legislative changes, major interventions in the modeling approach (for example, a significant expansion of the set of information known about each individual), or always after a few years pass.

During the selection of the coefficient, it is also necessary to monitor the extent to which the probability for individual persons has changed by the calibration. A useful guide in this respect may be to create a histogram of the probability difference before and after the calibration, expressed in percentage points (i.e., for example, one histogram column may correspond to a probability increase between 10 and 20 percentage points). If there are many significant increases as well as many decreases, it means that the method significantly distorts the life paths of individuals and k is too high. In such a case, it may make sense to lower the coefficient even if it means that the aggregate values will not be calibrated so well. (Nevertheless, we expect that practically the reduction of k in such situation will not substantially impair the accuracy of the calibration.)

None of the other methods for calibration of employment can be recommended. Calibration by bi-proportional scaling (chapter 4.8) and iterative model runs (chapter 4.3) are difficult to implement and compromises might be necessary. The methods of multiplicative scaling (4.5) and refinement of average values (4.4) are not suitable for calibrating the probabilities of interrelated events depending on multiple factors. Calibration method using the residual population (4.2) only works when we assign the same value to all members of a certain group, and finally, the alignment by sorting method (4.7) completely breaches the relationship between the explanatory and the dependent variable.

As the adjustment of the retirement probabilities described in the previous chapter was made on the basis of the model's results before the calibration of employment which itself affects retirements (via insured periods), it is now necessary to set these probabilities again. We will use the same method as in the beginning. Therefore, the average retirement ages will approximate.

5.1.3 Calibration of Incomes

We recommend using the multiplicative scaling described in chapter 4.5 for income calibration. This method is itself easy to implement and MoLSA already uses it to calibrate incomes: the average income in each year is multiplied by the so-called residual wage inflation to match the values from the external statistics. Thus, it suffices to extend the existing functionality so that the income can be calibrated not only for the whole population in a given calendar year, but also by age cohorts. This is not a difficult modification.

Moreover, multiplicative scaling, as the only one of the calibration methods described, perfectly preserves the form of the probability distribution of incomes resulting from the model before calibration. Its only disadvantage is that it weakens the relationship between the inputs entered into the model (specifically wage inflation and development of the person's income during his or her career) and the simulated income. This, however, is common to most methods. To avoid this weakening, it would be necessary to choose a very complicated procedure, for example, to use iterative model runs (chapter 4.3and change the employment rate of persons in each income grade so that the employment calculated for each cohort would still correspond to the values taken from the external projection, but the average income would change. However, there is no guarantee that such an effort would lead to the goal without introducing another type of inconsistency into the calculation. Therefore, we do not recommend this alternative.

Since the income of a person does not affect any other significant variable other than the amount of pension (which we do not want to calibrate, see chapter 3.1) and the average income in the cohort, on the other hand, depends on who is at what point in his or her life path, we will include the calibration of incomes as the last step after the calibration of all other variables, in particular employment.

5.2 Algorithm of the Complete Calibration

Now that we know which calibration method we want to use for what part of the problem, we finally proceed fully with the practical aspects and write down the sequence of steps that need to be taken during the calibration. In the process, we will distinguish the procedure for the two important external projections mentioned at the beginning of this study.

5.2.1 Reconciliation with the AWG Projection

Defining the groups and the outputs of the projection

Because the simulation of the AWG projection is performed by cohorts according to age and sex, we divide the NEMO model into cohorts using an SP code which represents the respective sex ($\mathrm{SEX\_MP}$) and year of birth ($\mathrm{initial\_year}$ – INIT_AGE_MP). The lower limit of the interval for the year of birth is calculated as $\mathrm{initial\_year} - MAX$(INIT_AGE_MP).

To calibrate the incomes, two variables must be included in the output variables – in the first one, the monthly incomes are added up for each currently employed individual from the cohort, in the second variable, the number of months in which the individual was employed is recorded. The values of these variables will then be used to calculate the average monthly income.

As the main output of the projection is the amount of pension expenditures by type of pension, variables in which the addition of the pensions paid for individual types of pensions for the corresponding calendar year occurs will be determined as further output parameters of the model.

Taking the input tables from the preparatory part of the calibration

Tables for mortality and birth rates by age and sex will be taken from the demographic projection of EUROSTAT, which is one of the main inputs of the AWG projection.

The overall population data used by the AWG projection will be compared with the inputs of the NEMO model. If necessary, we will adjust the input population by omitting or duplicating randomly selected individuals with an appropriate profile.

The probabilities of disability, broken down by sex and age, are calculated as the ratio of the number of newly awarded disability pensions to the population reduced by existing disability pensioners. All necessary data represents inputs to the external projection, which can be directly taken from it. The calculated probabilities will be uploaded to the input tables *Morb_females.fac* and *Morb_males.fac* which are broken down by age and calendar year of the projection.

As input, the model also loads the table of cohort probabilities of disability termination, the values of which are calculated from the AWG projection data as follows:

- We add up the number of new pensioners in year $x + 1$ and surviving pensioners from year x;
- We subtract the number of disability pensioners that AWG projected in year $x + 1$; we assume that this number will be lower than the above-mentioned sum;

- The differences obtained in each cohort represent terminated disabilities. By comparing the number of terminated disabilities from a certain year with the number of disability pensioners at the beginning of that year, we get the probabilities of termination of disability.

In table *ret_age.fac,* we set the retirement age such that the average retirement age observed in the NEMO model corresponds to the effective retirement age used in the AWG projection. Furthermore, in table *retirement.fac*, we set the probabilities of early retirement and retirement postponement for each cohort so that, after applying these probabilities to a population of persons who meet the statutory retirement conditions in that age, we get the numbers corresponding to the NEMO model run used for routine reporting (i.e., before any calibration). To calculate the values in both tables, we will use the NEMO model run before calibration. A more detailed description of this procedure is presented in chapter 4.1.

Calibration of employment and unemployment

The calibration target is the number of employed and unemployed individuals in each cohort by age and sex, which can be read directly from the AWG database. We enter these values into the input table, which will be used as a basis for calibration in the NEMO model using the sidewalk method. Thus, in all follow-up outputs from Prophet, the numbers of employed and unemployed individuals will be reconciled with the external projection.

After this step, it is necessary to repeat the calibration of the retirement probabilities in the same way as performed at the beginning of the calibration.

Calibration of the average income

At the input, the table of average income is uploaded into the model from external data, which will be used to calculate the scaling coefficients. The tables of average income broken down by age and sex can be obtained directly from the AWG projection.

After the first model run, we calculate the average monthly income from the output database of results for each cohort and each calendar year of the projection, i.e., we divide the values of the variable representing the sum of average incomes by the corresponding number of months worked.

Then we will compare this output table with the table of external average incomes. This will produce the scaling coefficient values that we will upload into the model as an additional input table. Then we start the second run, in which we multiply the average income of the cohort in each year by the corresponding scaling coefficient. This run already produces the calibrated values.

5.2.2 Reconciliation with the Projection of the Czech Fiscal Council

Defining the groups and the outputs of the projection

Using the SP code, we define cohorts by sex and birth year.

The same variables as in the case of calibration against the AWG projection are determined as model output variables. We will list the variables needed for the calibration of incomes and variables capturing the total volume of benefits paid according to the type of benefit.

Taking the input tables from the preparatory phase of the calibration

Since the projection of the Czech Fiscal Council is based on the demographic Projection of the Population 2018 – 2100, the tables of:

- mortality, disability and birth rates,
- count of the initial population,

broken down by age and sex should correspond to the available data from the databases of the Czech Statistical Office or the Czech Social Security Administration. However, it is necessary to compare them.

In the NEMO projection, the number of disability pensioners is controlled both by the probability of occurrence of disability and by the probability of its termination. Since the projection of the Czech Fiscal Council only provides the total cohort numbers of disability pensioners and the number of new or terminated pensions is unknown, there is generally an infinite number of solutions for these two variables that produce the required total number of pensioners. Therefore, we will leave the probabilities of termination of disability unchanged from the setting used in the basic projections of MoLSA and calculate the probabilities of occurrence thereof as follows:

1. Number of disability pensioners in year $x+1$: $disabled_{x+1}$ will be calculated as:

$$disabled_x * (1 - prob_{termination}) + prob_{occurence} * (population_{x+1} - disabled_{x+1})$$

(The choice of indices, i.e., which values are taken from which year, needs to be carefully considered and specified consistently with other relevant conventions used in the NEMO model.)

2. In this equation we know all values except the probability of occurrence of disability (the numbers of individuals in the population and the numbers of disability pensioners are taken from the external projection), and so it can be easily calculated.

To reconcile the number of old-age pensioners, we will proceed similarly to the AWG projection, the difference being that we set the retirement age value to correspond to a mother with two children for all women. For a more detailed procedure we refer to chapter 4.1.

Calibration of employment

The projection of the Czech Fiscal Council does not directly report the number of employees, but the number can be derived from the total income of the pension system and the average wage. In this way, however, we will only find an aggregate target for men and women; information on employment by cohorts is not available.

Therefore, we propose spliting this aggregate target into cohort targets before the first run. We will rely on the results after the first model run, i.e., after taking the input tables from the previous points, but before the next calibration. The employment rates for the calendar year implied by this calculation will be all multiplied by the same coefficient so that the overall employment rate of the male or female cohorts corresponds to the overall calibration target for men or women, read from the projection of the Czech Fiscal Council. We will do the same for unemployment.

Subsequently, the calibration proceeds similarly as in the case of AWG, including re-calibrating the probabilities of retirement in the same way as used at the beginning of the calibration.

Calibration of the average income

The sequence of steps during the calibration of the average income remains the same in reconciliation with the projection of the Czech Fiscal Council as in the case of reconciliation with the AWG projection, see chapter 5.2.1 section.

6 Estimated Timeframe and Complexity of the Proposed Solution

At the end of the study, we provide our estimate of the time and financial resources required to implement the recommended calibration method in the current NEMO model. All of the estimates below are based on the following assumptions:

- Calibration will be performed to one of the two external projections mentioned in Appendix Appendix A.
- The calibration procedure recommended in the previous chapter will be selected without further extensions. The recommended methods have been chosen with regard to the expected efficiency and difficulty of implementation, therefore, the choice of some other methods can change the difficulty even by orders of magnitude.
- A complete NEMO model and its run including the input preparation tools will be available as a starting point.
- Functionality of the NEMO model will not change significantly from the version valid at the time of this study;
- The calibration must not change ("break") the calculation formulas and functionalities implemented in the model (for more information see chapter 2.3).

Calibration of a model similar to NEMO is a difficult task that represents a great deal of uncertainty for the solving team. Even with all theoretical preparation, it is not possible to fully predict how the individual decisions will affect the various monitored states during calibration. If the implementation objectives were set in advance in the form of the required degree of approximation of the projections, it would present a risk for the solving team that no external contractor might be willing to undertake.

Therefore, we recommend splitting the implementation into three parts:

- In the first part, the inputs to the NEMO model will be adjusted so that they approach to the assumptions of the external projection; at the same time, the NEMO model will be modified to be ready to process these new inputs in adjusted formats;
- In the second part, the selected calibration methods will be technically implemented into the NEMO model and all necessary technical tools will be prepared;
- In the third part, the employment, unemployment and average income will be calibrated using the methods implemented in the previous point. This will therefore include also a search for suitable parameters of the used methods and analysis of the results.

We recommend that a project committee composed of MoLSA experts and the external contractor work together on the project. Instead of setting fixed calibration targets already at the beginning of the project, we recommend first setting indicative targets and then, first at the end of the aforementioned first and second phases and then during the third phase (after the first iteration of parameter setting and calibrations), we recommend that the project committee work together to assess with what expectations to proceed to the next phase, in the light of the results so far, and whether the planned approach or objectives need to be adjusted. The work on the first two phases and a more thorough analysis of the portfolio associated with it may bring an insight which will imply the need to adjust the initial expectations, for example, in terms of achievable approximation to the calibration targets.

At the same time, note that external support is, in our opinion, mainly necessary in the second phase, because it is the only one that directly relates to the development of new tools and implementation of new functionalities in Prophet. The other phases mostly contain the preparation of the assumptions and repeated runs of the model, which are tasks for which MoLSA analysts are fully technically equipped.

Upon these considerations, we present an estimated implementation schedule divided into the above-mentioned three phases. The timing is only approximate, depending mainly on the availability and quality of the input data and on the size of the implementation team.

The range of the estimates in the first and third phases is based mainly on the fact that any calibration requires reruns and subsequent detailed analyses of the results of the entire model, and in the case of the NEMO model, one run of this model alone (without potential input adjustments) lasts approximately one day.

Step	Comment	Duration (calendar weeks)
Phase 1		
Specification of calibration targets	Specification of the calibration approach chosen and its confirmation with the Client. Selection of the external projection which will be the calibration target (e.g., AWG). Mapping of data requirements for calibration of the target values (=what values and in what segmentation are needed for the steps below; specification of the sources thereof; is any source missing? What will we replace it with?). Specification of target values for the individual variables monitored, projection years and segments (=collecting values from the source into tables in a uniform format for further processing). Setting the initial indicative targets for the calibration rate. Approval of the specified values with the Client.	6
Taking the inputs - DCS - size and composition of the initial population, birth rate and immigration	Existing population: target for each segment will be divided into 2 sub-segments: individuals from amended INEP, new individuals. The numbers of persons from INEP will be calibrated if necessary. The numbers of "new individuals" will be calibrated by appropriately adjusting the target NI numbers and their target distribution in age categories. Future births and immigrants: target numbers determined according to target birth rate/immigration. Re-run of all DCS for the preparation of model points, including design of new SPCODEs for tracking results by key segments, incl. random sorting of MP. Checking the results (= checking achievement of targets) and discussion thereof with the Client.	6
Taking the inputs from the external projection	Preparation of tables for Prophet (taking the mortality rates, additional calculation of disability rates and retirement age assumptions). Run with calibrate inputs, checking the results and discussing them with the Client. The project committee will jointly evaluate with what expectations and approach to proceed to the next phase.	4
Phase 2		
Implementation of the sidewalk method for the calibration of employment and unemployment	Specification of model functionality modifications and approval thereof with the Client. Model functionality modification – implementation of the calibration method in the code. Preparation of tables for Prophet (parametrization of the method, targets of employment/unemployment /inactivity). Determining the range of achievable values: run without calibration (after the previous calibration steps), run with maximum calibration ($k = +\infty$).	5

Implementation of the multiplicative scaling method for the calibration of salaries	Specification of model functionality modifications (entering the scaling factors) and approval thereof with the Client. Model functionality modification – implementation of the calibration method in the code. Preparation of tables for Prophet (parametrization of the method, targets).	2
Delivery of the model	Delivery of the modified model to the Client. The project committee will jointly evaluate with what expectations and approach to proceed to the next phase.	1
Phase 3		
Calibration of employment using the sidewalk method	Determination of the target level of calibration (*k*), run with the target calibration Checking the results and discussing them with the Client.	4
Calibration of salaries using the multiplicative scaling method	Run with the target calibration. Checking the results and discussing them with the Client. The project committee will jointly evaluate, based on the initial results of this phase, with what expectations and approach to complete the calibrations.	2
Checking the overall results and completing the calibrations	Completion of the calibrations according to the conclusions of the previous paragraph. Detailed analysis of calibration impacts not only on calibrated but also on non-calibrated variables – in particular retirements, pension amounts, etc.; across the segments as well as in total. Approving the results with the Client.	8
Additional aspects	Based on experience with similar tasks, we expect additional tasks or obstacles to come up during more thorough analyses of data, the resolution of which cannot be planned in advance of the project. We therefore recommend allowing for sufficient resources.	6
Documentation	Documentation of the final approach.	2
Delivery and training	Training of the Client in the technical execution of all steps, demonstration of the use of all tools developed (preparation of tables, etc.) and functionalities in the model.	2
Total		**48**

Table 3: Schedule of implementation of the proposed solution

According to our estimates, an experienced team is able to perform all tasks at the following prices if the above conditions are met:

- The first phase approximately for 3 million CZK not inclusive of VAT;
- The second phase approximately for 2.5 million CZK not inclusive of VAT;
- The third phase approximately for 5 million CZK not inclusive of VAT.

In all cases, this is the price at which the implementation of the relevant phase would be delivered in full by an external supplier.

These price estimates are based on the assumption that the implementation will be carried out within the above time frame by a team of the following composition (unless otherwise stated, we assume for each of the persons full capacity during the relevant delivery phase):

- Phase 1
    - Methodology specialist
        - Task: based on the assumptions of the calibration method and the method of preparation of the assumptions used in NEMO, prepares the specifications of the calibration targets, develops the approach methodology and checks the correctness of its implementation.
        - Appropriate qualification: experience with development or operation of a microsimulation model of pension or social system of similar scope.
    - DCS modeling specialist
        - Task: adjusts the DCS input preparation programs as necessary
        - Appropriate qualification: experience in developing programs in DCS in the range of tens of millions of data records. This person will not need to be involved during the entire delivery of this phase. Therefore, if the Methodology specialist has the necessary qualifications, this activity can also be provided by the Methodology specialist in cooperation with one of the analysts.
    - Two analysts
        - Task: technically create tools for taking over inputs based on the Methodology specialist's specification.
        - Can be occupied by junior staff.
- Phase 2
    - Methodology specialist
        - Task: based on the knowledge of the model and selected calibration methods, leads the development of the technical specification and subsequently checks the correctness of the implementation according to this specification.
        - Appropriate qualification: 5 years in a position working with the Prophet system, experience with the development or operating a microsimulation model of similar scope.
    - Senior Prophet modeling specialist
        - Task: implementation of specified methods into the NEMO model.
        - Appropriate qualification: more experienced staff (3-5 years in a position working with the Prophet system), as it is necessary to correctly identify all places in the current code where changes need to be made, taking into account the structure of interdependencies between variables.
    - Junior Prophet modeling specialist
        - Task: support of the senior specialist in implementation and testing.
        - Appropriate qualification: at least 1 year in a position working with the Prophet system.
- Phase 3
    - Methodology specialist
        - Task: based on the modeling method used in NEMO and the calibration effects, he or she determines target levels of calibration, designs approaches and analyzes to evaluate the success of the calibration, he or she is able to interpret the results and evaluate how to modify the approach.
        - Appropriate qualification: experience with the development or operation a microsimulation model of pension or social system of a similar scope.
    - Two analytics
        - Task: create tools to modify inputs and check results, run model runs, perform analyses, and verify that results meet expectations.

- Appropriate qualification: experience with processing outputs from the Prophet system.

Bibliography

Anderson, J. M. 2019. Models for Retirement Policy Analysis. *Society of Actuaries.* [Online] 12 4, 2019. https://www.soa.org/resources/research-reports/2000-2006/research-models-for-retirement-policy-analysis/.

—. 2019. Models for Retirement Policy Analysis. *Society of Actuaries.* [Online] December 4, 2019. https://www.soa.org/resources/research-reports/2000-2006/research-models-for-retirement-policy-analysis/.

Bonin, H., Reuss, K. and Stichnoth, H. 2015. Life-cycle Incidence of Family Policy Measures in Germany: Evidence from a Dynamic Microsimulation Model. *DIW Berlin.* [Online] 2015. https://www.diw.de/documents/publikationen/73/diw_01.c.510616.de/diw_sp0770.pdf.

Browne, J., et al. 2014. Modelling Work, Health, Care and Income in the Older Population. *Institute for Fiscal Studies.* [Online] 2014. https://www.ifs.org.uk/uploads/publications/wps/wp201412.pdf.

de Menten, G and Bryon, G. et al. 2019. LIAM 2 User Guide. *LIAM 2 Website.* [Online] December 2, 2019. https://liam2.plan.be/documentation/0.12/processes.html.

de Menten, G. and Bryon, G. et al. 2019. LIAM 2 User Guide. *LIAM 2 Microsimulation made EASIER.* [Online] 12 2, 2019. https://liam2.plan.be/documentation/0.12/processes.html.

Deeparnab, C. and Sanjeev, K. 2018. Better and Simpler Error Analysis of the Sinkhorn-Knopp Algorithm for Matrix Scaling. *Cornell University, arXiv.* [Online] 2018. https://arxiv.org/pdf/1801.02790.pdf.

Deeparnab, Chakrabarty and Sanjeev, Khanna. 2018. Better and Simpler Error Analysis of the Sinkhorn-Knopp Algorithm for Matrix Scaling. [Online] 2018. 10.4230/OASIcs.SOSA.2018.4.

Hlaváček, M. and Kubíček, J. 2019. Projekce důchodového systému. *Informační a podkladové studie ÚNRR.* [Online] 6 2019. https://unrr.cz/vydavame/studie/.

Hlaváček, Michal and Kubíček, Jan. 2019. Projekce důchodového systému. *Informační a podkladové studie ÚNRR.* [Online] June 2019. https://unrr.cz/vydavame/studie/.

Ministerstvo financí České republiky. 2018. Katalog datových prvků a dalších položek používaných v Informačních systémech o platu a služebním příjmu (ISPSP). *Ministerstvo financí České republiky.* [Online] 12 2018. https://www.mfcr.cz/cs/o-ministerstvu/informacni-systemy/is-o-platech#12769.

Ministerstvo Financí. 2017. The 2018 Ageing Report: Economic and Budgetary Projections for the EU Member States (2016-2070). *Pension projections of the Czech Republic.* [Online] 10 2017. https://ec.europa.eu/info/publications/economy-finance/2018-ageing-report-economic-and-budgetary-projections-eu-member-states-2016-2070_en.

Ministry of Finance. 2017. The 2018 Ageing Report: Economic and Budgetary Projections for the EU Member States (2016-2070). *Pension projections of the Czech Republic.* [Online] October 2017. https://ec.europa.eu/info/publications/economy-finance/2018-ageing-report-economic-and-budgetary-projections-eu-member-states-2016-2070_en.

O'Donoghue, C. and Li, J. 2014. Evaluating Binary Alignment Methods in Microsimulation Models. [Online] 2014. https://doi.org/10.18564/jasss.2334.

—. 2014. Evaluating Binary Alignment Methods in Microsimulation Models. *Journal of Artificial Societies and Social Simulation.* [Online] 2014. https://doi.org/10.18564/jasss.2334.

Scheil-Adlung, X. and Sandner, L. 2010. The case for paid sick leave. *World Health Organization.* [Online] 2010. https://www.who.int/healthsystems/topics/financing/healthreport/SickleaveNo9FINAL.pdf.

Stephensen, P. 2016. Logit Scaling: A General Method for Alignment in Microsimulation models. *International Microsimulation Association.* [Online] 2016. https://www.microsimulation.org/IJM/V9_3/4_IJM_2015_18_Stephensen_edit_final.pdf.

—. 2016. Logit Scaling: A General Method for Alignment in Microsimulation models. [Online] 2016.

Wittenberg, R. and Hu, B. 2015. Projections of Demand for and Costs of Social Care for Older People and Younger Adults in England, 2015 to 2035. *Personal Social Services Research Unit.* [Online] 2015. https://www.pssru.ac.uk/pub/DP2900.pdf.

Appendix A Important External Projections

A.1. Ageing Working Group (AWG)

Based on the document entitled *The 2018 Ageing Report* published by the European Commission, the projections for pensions will be run by the Member States of the European Union using their own national pension system model(s) and using commonly agreed assumptions of AWG. The report was published in 2018 and its base year for the projections is 2016.

This report is the same for each Member State and includes projections of 145 variables. In order to complete this report, Member States will provide a detailed description of the projections in their reporting sheets.

The reporting sheet of the Czech Republic named *final_country_fiche_cz.pdf* contains inputs, outputs and assumptions, which can be divided into three main areas.

The first one is a demographic projection of EUROSTAT, divided by age and sex. The current projection foresees a decline in the population in the long run.

Projection of the labor market, Cohort Simulation Model (CSM), is used as the second important input. It includes mainly the following variables:

- Projection of the rate of economic activity of women/men/persons aged
- Projection of the rate of employment of women/men/persons aged (by combining information on economic activity and employment, unemployment data can also be obtained)
- Mean age of the labor force
- Projection of the proportion of working individuals divided by age in the labor force
- Projection of the average retirement age of men and women
- Projection of the average pension insurance period (women, men and total)
- Projection of the pension duration
- Projection of the effective retirement age
- At present, the effective retirement age is above the statutory retirement age. However, future developments assume that the effective age will be lower than the statutory threshold, which will add new numbers of inactive individuals (we assume that a person will leave the labor market when he or she has reached the effective age). We also assume that these inactive individuals will become early retirement pensioners, even if they are penalized for that, if they are entitled to (early) old-age pension.
- In addition, we expect that, in line with this trend, more and more people will tend to withdraw their capital savings from the third pillar and therefore opt for a pre-retirement system which will become increasingly popular.
- Projection of the number of employees (women, men and total)
- Projection of the personal assessment base (women, men and total)
- Projection of the share of personal assessment bases and the average salary (women, men and total)
- Projection of the average pension insurance period (women, men and total)
- Projection of the calculation base coefficient (women, men and total)
- Number of contributors to pension insurance
- Projection of the amount collected for the pension insurance from:
    - the employee
    - the employer

The AWG projection is also complemented by several additional assumptions and inputs:

- Number of existing pensions divided by type of pension, age and sex
- Number of new pensions divided by type of pension, age and sex
- Average pension benefit divided by type of pension, age and sex
- Average newly awarded pension benefit divided by type of pension, age and sex
- Number of new pensions (divided by type of pension) for a given combination of personal assessment base and insurance period
- Valorization of pensions and the valorization index. It is done annually in January. It is set in such a manner that the basic pension amount is 9% of the gross average monthly wage (the old percentage before the law modification in 2018).
- Assumption of zero taxation of pensions (the law grants most pensions an exemption from tax)
- Consumer price index (CPI) of pensioners – it is assumed to be the same as the cost of living index of pensioners' households
- Projection of the gross and net pension expenditure as a percentage of GDP and total in EUR divided by type of pension. A distinction is made between the flat-rate component of gross expenditure and the component dependent on income.
- Projection of the disability rate by age groups
- Projection of the benefits ratio (the ratio of the average pension benefit to the average wage)
- Projection of the compensation rate (the ratio of the newly awarded pension benefit to the average gross wage at the time of retirement)
- Projection of the number of pensioners
- Projection of the proportion of pensioners (men and women) and inactive population
- Projection of the gross expenditure on newly recognized old-age pension and early retirement pension benefits (women, men and total)
- Projection of the number of newly awarded old-age pension and early retirement pension benefits (women, men and total)

Total expenditure on a particular type of pension is calculated according to the following formula:

$$pen\_e_t = \sum_{g,s} \left(pen_t^{g,s} - npen_t^{g,s}\right) \cdot pen\_v_{t-1}^{g,s} \cdot (1 + ind_t) + npen_t^{g,s} \cdot npen\_v_t^{g,s},$$

where:

- g is the population generation (divided by calendar year)
- s is the respective sex (male/female)
- (pen) is the number of pensions calculated as

  $$pen_t^{g,s} = pen\_s_t^{g,s} \cdot pop_t^{g,s},$$

  where $pop_t^{g,s}$ is the value of the respective population and $pen\_s_t^{g,s}$ is the proportions of the respective pension by age, which are calculated on the basis of the conditional probability that the individuals defined by the pair (g, s) become recipients of the pension in question
- $(npen)$ is the number of newly awarded pensions, which is defined as

  $$npen_t^{g,s} = pen_t^{g,s} - pen_{t-1}^{g,s} \cdot (1 - \varepsilon_t^{g,s}),$$

  where ε is the specific mortality rate determined by sex and generation
- (pen_v) is the average value of the respective benefit. It is calculated as the weighted average of the average value of the benefit from the previous period and the average value of the newly awarded benefit:

$$pen\_v_t^{g,s} = \frac{pen_t^{g,s} - npen_t^{g,s}}{pen_t^{g,s}} \cdot pen_{v_{t-1}}^{g,s} \cdot (1 + ind_t) + \frac{npen_t^{g,s}}{pen_t^{g,s}} \cdot npen_{v_t}^{g,s}$$

- $(npen\_v)$ is the average value of the newly awarded benefit, defined as follows:

$$npen\_v = frc + erc,$$

  where frc is the assessment base and erc is the percentage. For the calculation of the percentage, we assume that the reduction limit develops simultaneously with salaries.

  The basic input for the calculation of the percentage is a table that divides the population according to the assessment base and the period of participation in pension insurance. For each combination of these two variables we know the number of persons who have achieved them and, at the same time, we can calculate the corresponding percentage. The average percentage is determined by simple weighing of all values by the numbers of people read from the table. The values in the table are a part of the CSM projection on which the AWG projection is based.

- (ind_t) is the statutory valorization factor

A.2. Czech Fiscal Council

In the first step, the projection of the Czech Fiscal Council (Hlaváček and others, 2019) deals with the number of recipients of pension benefits and in the following step it deals with the projection of the amount and volume of pension benefits paid.

Projection of the number of old-age pensions

The projection of the number of old-age pensioners is based on the demographic structure of the population which is determined by a projection of the Czech Statistical Office (Population Projection 2018 – 2100 document, specifically).

The main input for the projection of the number of old-age pensions is the (revised) rate of retirement. It is the ratio of the number of old-age pensioners at the end of the calendar year according to the information from the Czech Social Security Administration (ČSSZ) to the number of people of the given age as of 1 January of the following year according to the information from the Czech Statistical Office, reduced by the number of disability pension recipients of the given age – this is a momentary situation at the end of the year and the movements, if any, are not recorded. The number of disability pension recipients is also projected within the projection of number of disability pensions.

Retirement rates depend on the age of the person and on the age at which the person reaches (or has reached) the statutory retirement age. If there are two different retirement ages in a particular year, we consider that the retirement age for that year is the weighted average of the two retirement ages (where the weights are the number of months during which each retirement age existed). This dependence of the retirement rate on the distance from retirement age will be called the retirement curve. For further calculations, we will use the retirement curve calculated as the average of the retirement curves for the period 2013 – 2017. This period is suitable both in terms of the availability of data and the condition of the economy – it included periods of economic boom and the period of ending recession; the average values should therefore not be significantly burdened by the effects of the economic cycle.

Furthermore, retirement rates are divided by sex. For the calculation of the retirement age of women, the assumption is that each woman has two children.

For the period of increasing retirement age (until 2030), the retirement age curves for the currently valid retirement age in the given year are used for the projection of the number of old-age pensioners. From 2030, pension curves are used that relate to the personally relevant retirement age.

A certain number of pensions cease to exist every year on the basis of mortality assumptions. The number of newly awarded pensions is calculated by additional recalculation from the total number of pensions in a given year, the total number of pensions in the previous year, and the number of pensions that ceased to exist. This division is important for determining the amount of pensions.

The following effects are only considered implicitly in the projection (they are already part of the pension rates) or are not considered at all:

- Early retirement pensions

- Potential effect of pre-retirement from the third pillar (for example, some people may use pre-retirement instead of retiring early)
- Unification of the retirement age for women (cessation of the majority of preferential deals for women according to the number of children raised).

Projection of the number of disability pensions

Disability rates are divided by age cohort and are based on historical disability rates and the estimated postponement of the retirement age. The result is the following values:

- Up to 55 years as the average of disability rates in the period 2015 – 2017 or, as applicable, only for the year 2017 if this value was lower
- For the age of 55 years, we assume disability of 15.5% for women and 15.3% for men
- Two years before reaching the retirement age – maximum of the disability rates. It is determined as the maximum of the disability rates from 2017 (approximately 18% for women and approximately 20% for men).
- A uniform increase is assumed for the age of 55 years up to two years before reaching the retirement age
- For the age of 64 years, the disability rate is at the level of the average for the period 2013 – 2017
- For the age of 65 years and higher, a zero disability rate is assumed because disability pension is converted into old-age pension upon retirement. This limit is shifting as the retirement age increases.

These disability rates will only give the total number of disability pensioners. For the purpose of classification thereof into individual disability degrees, we assume that the share of individual disability degrees in the total number of disability pensioners will remain unchanged throughout the projection period.

Projection of the number of survivors' pensions

For orphan's pensions, it is assumed that 1.75% of people aged 0 to 21 receive it.

Widow's and widower's pensions paid separately are projected based on the assumption that the proportion of these pensions for persons over 21 years of age will be constant. It is not modelled for younger persons. The relevant coefficient is determined as the average over the last three years available, with values of 0.18% for men and 0.8% for women.

In the case of widow's and widower's pensions paid in conjunction with an old-age pension or disability pension, an age-specific rate is applied (separate for men and women) from the sources of MoLSA and the Czech Statistical Office for the year 2017. However, it is adjusted for the postponement of the statutory retirement age (until 2030) and prolongation of life expectancy (which will reduce the number of pensions).

Results of the projection of the number of pensions

The projection of the number of pensions is carried out for the following versions of the demographic projection:

- Medium
- High
- Low
- Medium with zero migration balance
- Medium with the so-called tied retirement age (the same retirement age for men and women is set so that, for those who reach it, the period they spend in retirement is 25% of the total life expectancy).

Amount of the old-age pensions

The amount of newly awarded pensions is determined by reference to the ratio to the average wage. This ratio is part of the model's assumptions. For men, a stable ratio of 46.6% is expected for the period 2017-2050, followed by a decline to 44.8%, which will take place in the period 2050 - 2055. The ratio will then remain at this value until the end of the projection. For women, this ratio is gradually increasing from a baseline of 39.6% to 44% in 2030. It will remain at this level until 2050, then drop to 42.2% between 2050 and 2055 and remain at this value until the end of the projection.

The valorization of pensions already awarded is based on real wage growth, inflation rate, and the cost of living index of senior households. As a result, it is assumed that the pension is annually increased by real wage growth increased by 0.3 percentage point.

We assume that the average amount of lapsed pensions is 95% of the average amount of old-age pensions.

By linking all these assumptions, we will get the ratio of the amount of old-age pension to the average wage. The projection results range between 38% - 40.1% with a significant increase in the period 2030 – 2040. Based on this development and the development of the number of old-age pensioners, we will get the projection of old-age pension expenditure as a percentage of GDP.

Disability pensions

To determine the amount of disability pensions, a constant ratio between the average disability pension of a given degree and the average old-age pension is assumed, with the starting year for determining the ratio being 2018.

Survivors' pensions

The amount of survivors' pensions is determined as a percentage of the average old-age pension over the last three years. The percentages used are as follows:

- 51.2% for orphan pensions
- 57.2% for widower's pensions paid separately
- 64.8% for widow's pensions paid separately
- 16.2% for widower's pensions paid concurrently
- 21.3% for widow's pensions paid concurrently

Projection of pension system income

The income of the pension system is directly based on the development of wages and salaries, which is taken from the Long-Term Macroeconomic Projection of the Czech Republic produced also by the Office of the Czech Fiscal Council. The share of wages and salaries in GDP is expected to increase gradually during the projection, from the current 8.7% to approximately 9.5%.

Projection of the labor market

The labor market is modeled only indirectly in the projection of the Czech Fiscal Council, through the income of the pension system and the retirement rate. However, there is an assumption available in the model of the average wage and the total income of the pension system. From there, the number of employed individuals can be determined very easily. It would be more difficult to derive the number of unemployed individuals from the projection.

Appendix B R code

B.1. Verification of the refinement of average values

The refinement of average values method is described in Chapter 4.3. This chapter introduces a process where we start with a uniform distribution U on the interval $[0,1]$, add a certain random variable X and require the sum to be uniformly distributed again. Functionality of this method was verified by a simulation in R, where we choose normal distribution and beta distribution for X. Here is the code that can fully replicate this verification.

The code requires the *tmvtnorm* and *truncnorm* packages, which allow to simulate from a constrained normal distribution (i. e. simulate from a normal distribution, and whenever the simulated value does not lie within the desired interval, in our case $[-1,1]$, the value is not used and a new one is simulated) and bimodal distribution. If these packages are available, the code can be executed simply by copying it to the R console.

```
# A script to verify the uniform distribution of the results for the re-
finement of average values method

set.seed(1234)

# We use package tmvtnorm to simulate the constrained normal distribution

require("tmvtnorm")


# Case 1 - normal distribution

n <- 1000000 # number of simulations

U <- runif(n,0,1) # uniformly distributed random variable

X <- rtmvnorm(n,mean=0,sigma=1,lower=-0.5,upper=0.5) # normally distributed
random variable


# Illustration of properties of random variable X

# To visualize variable X, we firstly need to order its values by size.
That is the reason to depict X[order(X)] in the graph

plot(qnorm, type="l", main="Percentile function of the normal distribu-
tion",xlab="Percentile", ylab="Value of the percentile",xlim=c(0,1))

hist(X[order(X)], breaks=seq(from=-1, to=1, by = 0.02),freq=F, ma-in="His-
togram of normal distribution", xlab="Value of the distribution",
ylab="Probability density")


# Calculation of the resulting distribution

V <- U + X # the sum before modification - its values are from -0.5 to 1.5
and it is not uniformly distributed

# Modification of values that are not within the interval [0,1]

for(i in 1:n){

 c=U[i]+X[i]

 V[i]<-if(0<=c& c<=1) c else if(c<0) -c else 2-c}


# Depiction of QQ graph - compare with uniform distribution of U
```

```
qqplot(U,V, type="l", main="QQ-plot of distribution created from a normal
", xlab="Quantiles of uniform distribution", ylab="Quantiles of the result-
ing distribution")


# If V is uniformly distributed, it should be true that the q-th quantile
of V is equal to q. Therefore, we calculate this difference for several
different values of q and determine the maximum absolute difference.
s <- seq(from=0.01,to=0.99,by=0.01)  # calculated quantiles
max(abs(quantile(V,s)-s))       # maximal difference


#Testing by K-S test
ks.test(V,"punif",0,1)


## Testing other distributions of X
require("truncnorm")
X2 <- c(rtruncnorm(n/2, a=-0.5, b=.5, mean=-0.1, sd=.5),
    rtruncnorm(n/2, a=-0.5, b=.5, mean=0.1, sd=.5))
for(i in 1:n){
 c=U[i]+X2[i]
 V[i]<-if(0<=c& c<=1) c else if(c<0) -c else 2-c}
ks.test(V,"punif",0,1)


X3 <- c(rtruncnorm(n/2, a=-0.5, b=.5, mean=-0.4, sd=.5),
    rtruncnorm(n/2, a=-0.5, b=.5, mean=0.4, sd=.5))
for(i in 1:n){
 c=U[i]+X3[i]
 V[i]<-if(0<=c& c<=1) c else if(c<0) -c else 2-c}
ks.test(V,"punif",0,1)
```

B.2. Creating graphs illustrating the Sidewalk Method

In Chapter 4.5, we describe a calibration method in which the model goes through the population one person at a time and always adjusts the probabilities of transition to a particular state according to how many people have already entered that state during the projection. The formula contains a coefficient k that determines the sensitivity of the calibrated probability to the magnitude of the difference between the calibration target and the number of transitions to date – a higher k means a higher sensitivity. To illustrate the impact of choice of k, two graphs are provided. These can be generated in the R software using the code below. One can run the code by simply uploading it to the R console.

```
sidewalk_coef <- function(cf=1, prob=c(0.2,0.5,0.8)) {


 s <- seq(-1,1,0.001)
 logit <- function(x) {return(log(x/(1-x)))}
 logitinv <- function(x) {return(exp(x)/(exp(x)+1))}
```

```
 plot(s, logitinv(logit(prob[1])+cf*s),type="l", ylim=c(0,1), col="green",
xlab=paste0("Normalized distance to target"), ylab="Calibrated probabil-
ity", main=paste0("Sidewalk Method using coefficient ", cf))
 points(s, logitinv(logit(prob[2])+cf*s),type="l", col="blue")
 points(s, logitinv(logit(prob[3])+cf*s),type="l", col="red")
 abline(v=0)}

sidewalk_coef(cf=1)
sidewalk_coef(cf=5)
```

B.3. Illustration of the sidewalk method in MS Excel

The sidewalk method described in Chapter 4.5 is further illustrated on a prototype in MS Excel, in which we calibrate transition probabilities according to the current state.

Excel demonstrates the sidewalk method at 1000 random model points with three possible positions in the labor market: employment, unemployment and inactivity. At the same time, we are also introducing the "in household" status that makes it impossible to switch to unemployment (and thus only allows transitions between employment and inactivity according to adjusted probabilities). The methodology of these two functionalities is the same as that of most transitions that occur in the current Prophet microsimulation model, so the prototype provides a good idea of how the calibration method will be translate into the NEMO model. Initial states were generated randomly according to the expected distribution. Initial transition probabilities were taken from the NEMO model.

The projection is started on the "Results" sheet using the "Run simulation" button. Each model point is projected for the next 30 years. The transition probabilities are adjusted according to the current state and the parameter determining the calibration speed.

The overall results can be seen in the "Result" sheet. Excel allows to test the calibration speed beyond the projections on the "Cal_speed" sheet. For a given probability and calibration speed, one can see a change in the probability of transition depending on the distance of the results from the calibration target.

MPSV_sidewalk_method_illustration.xlsm

B.4. Illustration of calibration by alignment by sorting

In Chapter 4.7.2, we gave an example of the alignment by sorting calibration method, which selects the individuals for whom it simulates a given event by sorting all persons from the population with respect to probabilities of the event. Specifically, we present three variants, the first one works with probabilities as such, the second one adds to each of them a random realization of the uniform distribution, and the last one is transformed using the logit and inverse logit functions. The following code is written for the R software, requires the *ggplot2* and *tidyr* libraries, and draws a graphical representation of the usage of each of the variants in the above mentioned example. Moreover, it illustrates that this method retains symmetry when using the logit transformation.

```
N=1000 #number of persons in modeled population
M=150 #the number of individuals for which an event occurs
people_prob=rlnorm(N, meanlog = 0, sdlog = .5)
people_prob=people_prob/max(people_prob)-min(people_prob/max(peo-
ple_prob))/2

# sorting individuals according to their initial probabilities
```

```
col_by_ord=c(1:N)
for(i in 1:N) {col_by_ord[i]<-match(i, order(people_prob, decreasing =
TRUE))}

x<-runif(N, min = 0, max = 1)

logit <- function(y) {return(log(y/(1-y)))}
logitinv <- function(y) {return(exp(y)/(exp(y)+1))}

library(ggplot2)
library(tidyr)

# An areay of primary probabilities and "their order"
df_1 <- data.frame(
 prob=people_prob,
 ord=col_by_ord)

#generated probabilities
data_graph<-ggplot(data=df_1, aes(x=c(1:N), y=people_prob, fill=ord)) +
 geom_bar(stat="identity")+
 theme_classic()+
 labs(title= "Primary probabilities",y= "p", x = "individual",fill =
"order\nof primary\nprobabilities")
data_graph

#ordered primary probabilities
df_2 <- df_1[order(df_1$prob, decreasing = FALSE), ]

data_graph_sorted<-ggplot(data=df_2, aes(x=c(1:N), y=df_2$prob,
fill=df_2$ord)) +
 geom_bar(stat="identity")+
 theme_classic()+
 labs(title= "Ordered primary probabilities ",y= "p", x = "individual",fill
= "order of\nprimary\nprobabilities")+
 geom_vline(xintercept = N-M,
       color = "Darkgreen", size=1.3)
data_graph_sorted

#adding x
df_3 <- df_2
df_3$prob<-df_3$prob+x
```

```
df_3<-df_3[order(df_3$prob, decreasing = FALSE), ]

data_graph_sorted_x<-ggplot(data=df_3, aes(x=c(1:N), y=df_3$prob,
fill=df_3$ord)) +
 geom_bar(stat="identity")+
 theme_classic()+
 labs(title= "Ordered probabilities after adding x",y= "p + x", x =
"individual",fill = "order of\nprimary\nprobabilities")+
 geom_vline(xintercept = N-M,
       color = "Darkgreen", size=1.3)
data_graph_sorted_x

#logit tranformation
df_4 <- df_2
df_4$prob<-logitinv(logit(df_2$prob)+x)
df_4<-df_4[order(df_4$prob, decreasing = FALSE), ]

data_graph_sorted_x_logit<-ggplot(data=df_4, aes(x=c(1:N), y=df_4$prob,
fill=df_4$ord)) +
 geom_bar(stat="identity")+
 theme_classic()+
 labs(title= "Ordered probabilities after logit transformation",y=
bquote('logit'^-1~'(logit(p)+x)'), x = " individual ",fill = "order
of\nprimary\nprobabilities")+
 geom_vline(xintercept = N-M,
       color = "Darkgreen", size=1.3)
data_graph_sorted_x_logit

#symmetric task for logit transformation
df_4_sym <- df_2
df_4_sym$prob<-logitinv(logit(1-df_2$prob)+x)
df_4_sym<-df_4_sym[order(df_4_sym$prob, decreasing = FALSE), ]

data_graph_sorted_x_logit_sym<-ggplot(data=df_4_sym, aes(x=c(1:N),
y=df_4_sym$prob, fill=df_4_sym$ord)) +
 geom_bar(stat="identity")+
 theme_classic()+
 labs(title= " Ordered probabilities after logit transformation in symmet-
ric task",y= bquote('logit'^-1~'(logit(1-p)+x)'), x = "individual",fill =
"order of\nprimary\nprobabilities")+
 geom_vline(xintercept = M,
       color = "Darkgreen", size=1.3)
data_graph_sorted_x_logit_sym
```

Deloitte.