Applications of microsimulation modelling

TÁRSADALOMBIZTOSÍTÁSI KÖNYVTÁR

Applications of microsimulation modelling

A selection of papers presented during the 2016 European meeting of the International Microsimulation Association in Budapest

Edited by
Gijs Dekkers and József Mészáros

GIJS DEKKERS • Federal Planning Bureau of Belgium; CESO University of Leuven; and LISER Luxembourg. President of the International Microsimulation Association.

JÓZSEF MÉSZÁROS • Central Administration of National Pension Insurance of Hungary, Director General

ISBN 978 963 693 766 9

TABLE OF CONTENTS

GIJS DEKKERS
Introduction 7

CONFERENCE PAPERS

GIJS DEKKERS – KAREL VAN DEN BOSCH
Prospective microsimulation of pensions in European Member States 13

ZUZANA SIEBERTOVA – NORBERT SVARDA – JANA VALACHYOVA
Improving the validity of microsimulation results: lessons from Slovakia 34

MIROSLAV ŠTEFÁNIK
VZAM_microsim_1 – the EU-LFS based microsimulation model to estimate the Slovak labour market skills needs 65

PETRA STEIN – DAWID BEKALARCZYK
On the prognosis of third generation migrants’ occupational status in Germany – a dynamic microsimulation 95

JUSTIN VAN DE VEN
The simulator of individual dynamic decisions, sidd 128

KRISZTIÁN TÓTH
Simulating the long-term labour market effects of economic transition in Hungary: the MIDAS_HU microsimulation model 153

Introduction

GIJS DEKKERS

In between the years of the World Conferences, the International Microsimulation Association IMA aims to organise regional conferences. In 2016, the Brookings Institution organised a American conference in Washington DC, and a two-day European conference was organized by the Central Administration of National Pension Insurance (CANPI), in September 22th and 23rd, Budapest, Hungary. First of all, the editors would like to thank the many people at both institutions who made these conferences possible. Next, we would like to thank all those that submitted and presented their work during this conference.

This book contains a selection of papers presented during the 2016 European meeting of the International Microsimulation Association IMA in Budapest, Hungary. Its intended audience is not so much those that are professionally involved in microsimulation modelling and model development. Rather, we aim to show to those involved in policy making the importance of (dynamic) microsimulation, actually but also –by presenting some work on new models– potentially. Hence, besides quality, these papers are being selected on the basis of whether they are readable to a non-technical audience.

Obviously, a book on microsimulation models and modelling cannot start without having spent a word on what microsimulation is. Where macro- or meso-economic models simulate relations between aggregates, such as national account entries or aggregates on the sectoral level, microsimulation models start modelling on the level of the individual agent, which in social sciences is often an individual in his or her household. As a result, microsimulation models focus on the relations and the data beyond the aggregates and can therefore be used to model the entire distribution of a variable, often described by indicators of (re) distribution, inequality and poverty. Of course, also aggregates can be derived from the simulation results. These distributional effects can have profound impacts on the averages. For example, in a previous study (Dekkers et al., 2015) found that a policy measure that restricts access to early retirement in Belgium had no impact on the retirement age of women, and this in contrast to men. The reason turned out to be that the distribution of the careers of women is highly bimodal, with one group of women having such short careers that they could not go into early retirement before nor after the reform. The other group had such long careers that they were eligible to early retirement, both before and after the reform. So, albeit for opposite reasons, neither groups were affected by the reform. Obviously this result would not have been possible to simulate should the career length of women be based on some average figure, as would have been the case in a macro- or meso-level model. Atkinson (2009, 42) makes the case for microsimulation in policy-oriented research by discussing the difference between household income and national income, and the importance of considering the median besides just the mean in simulations.

In recent years, especially static microsimulation with EUROMOD has gained the most attention, and not only in applied research in the EU member states. And rightfully so – the contribution of EUROMOD to the development and visibility of the field of microsimulation cannot be overestimated. See Sutherland (2014); Lelkes and Sutherland (2009) for an overview. However, maybe somewhat obscured, quite important investments have been done in dynamic microsimulation models and modelling. This book hopes to provide a brief testimony to those developments.

Before turning to a brief discussion of the papers in this book, allow us to dwindle an instant longer on what is meant by 'dynamic' in dynamic microsimulation. The reason is that the word has various meanings in recent literature. The standard interpretation is that "dynamic microsimulation" is used to describe models where the characteristics or attributes of individuals are changed (Harding, 1996, 4). These changes can be in comparative-static perspective, where the endogenous characteristics change between scenarios simulated by the model. For example, many structural behavioural models simulate the impact of changes to the fiscal system or social security benefits to labour supply, comparing the results to the original dataset i.e. in comparative statics.

Models usually are referred to as dynamic when the individual characteristics change over time. Models of this type are also known as "dynamic ageing models". So, individuals age, their labour market state changes, they build up entitlements to a social security system, et cetera. These models therefore directly model time via the risks that endogenous variables change. Another strand of models use static ageing techniques, whereby the weight of the individuals in the dataset is changed to meet joint or marginal distributions of auxiliary variables. These can reflect alternative scenarios, the actual characteristics of small geographical areas for which a full sample unavailable, but they can also be projections produced by (semi-)aggregate demographic or economic models. These latter static ageing models therefore mirror a notion of time, however without modelling it directly. In most recent overview papers (See Li, et al., 2014; Harding, 1996; Williamson et al., 2009), these static ageing models therefore are not considered dynamic, because the individual characteristics of individuals do not change over time. Static ageing models nevertheless do include some loose notion of time, since the auxiliary data is of a forward-looking nature, i.e. are projections made by other models or approaches. This way, these models are classified as static models, but "used for short-term projections" (Li et al., 2014, 63). However, Dekkers and Belloni (2009) differentiated between simulation properties and the technical characteristics of a model. The simulation properties "have consequences for the […] research problems that can be covered by a model, as well as the […] assumptions that a model makes when handling a specific research problem" (op. cit., 6). They describe the difference with technical characteristics of a model, such as the modelling structure or the programming language. Dekkers (2015) then argues that, under certain well-known conditions, the approaches of static and dynamic ageing become equivalent and the difference between the two therefore becomes more a technical characteristic than a simulation property. Hence, without wanting to challenge the above traditional classification, we believe it might be relevant only for the modellers themselves and not necessarily for those that use the simulation results of these models in a context of policy assessment. Put differently, if the way time is modelled is a technical characteristic rather than a simulation property, then the question how time is modelled might be primarily relevant for modellers, whereas the question if time is included at all might be relevant to the users of the output of these models. Hence, Dekkers and Belloni (2009) group static-ageing models that model time indirectly under dynamic models. In a recent working paper, Ugo Colombino (2015) appears to make the same case.

This book illustrates this point of view by completing chapters on 'full' dynamic models with a chapter that 'dynamising' an otherwise static model to make short-term projections. Furthermore, in the first chapter, Gijs Dekkers and Karel Van den Bosch describe dynamic pension microsimulation models in (semi-) public sector organisations of EU member states. They consider models with dynamic as well as static ageing.

Zuzana Siebertova, Norbert Svarda and Jana Valachyova describe the lessons learned in the process of building a static-ageing microsimulation tool for Slovakia. This chapter, previously published in the IMA's house journal, the International Journal of Microsimulation, has nevertheless been included in this volume, because it shows how simulation results can be substantially improved by using calibrated weights. They also calibrate weights to match population totals in two projection years – thereby making the model dynamic in the loose sense of the word. This way, their model becomes an efficient and valuable tool for policy assessment in Slovakia.

Miroslav Štefánik simulates prospective skills needs of the Slovak labour market, and explicitly models and matches labour supply and demand in reduced form. As such, the model is an innovation that can be used to project skills mismatches on the labour market. He sketches a situation in which the skills mismatch in the Slovak labour market would increase dramatically, especially in manufacturing. Resolving these mismatches might require investments in specific vocational training and education programs.

Petra Stein and Dawid Bekalarczyk use dynamic microsimulation techniques to reveal the longitudinal occupational status of third generation immigrants in Germany over a period of about 40 years. By blocking the development of the age of the immigrants, changes of 'performance levels' are argued to be the main determinant of the increase of the employment proportions among third-generation immigrants over time.

Most dynamic-ageing microsimulation models use reduced-form behavioural models, for example in the form of logistic regressions. Although conveniently simple, at least in principle, they are by definition a reflection of the past, which might not be appropriate to describe a future situation, especially in the case of policy scenarios. Justin Vandeven presents an open-source, structural dynamic microsimulation model that can be used for policy assessment in a dynamic context. This model is based on a theoretical framework for labour/leisure and consumption/savings decisions that lies below the policy context of the model; which means that this framework can be used for policy assessment.

Finally, Krisztián Tóth uses so-called labour market profiles, essentially groupings of entitlement acquisition data or career characteristics over more recent decades, to increase the validity of the labour market modules in the dynamic Hungarian model MIDAS_HU. He shows that this approach improves the fit of the labour market block of the model.

References

Atkinson, T., 2009, An Enlarged role for Tax-benefit models. In: Lelkes, Orsolya, and Holly Sutherland, 2009, Tax and Benefit Policies in the Enlarged Europe: Assessing the Impact with Microsimulation Model. Public Policy and Social Welfare, European Centre Vienna; Ashgate: Farnham, UK, chapter 2, pp. 33-46.

Colombino, Ugo, 2015, The ISTAT microsimulation models. In: Rivista di Statistica Ufficiale, 2015 No 2. 5-15. http://www.istat.it/it/files/2015/10/1-The-Istat-Microsimulation-Models.pdf [05/03/2017]

Dekkers, Gijs, Raphaël Desmet, Nicole Fasquelle, Saskia Weemaes, 2015, The social and budgetary impacts of recent social security reform in Belgium. In Ioana Salagean, Catalina Lomos & Anne Hartung, "The young and the elderly at risk: Individual outcomes and contemporary policy challenges in European societies", Intersentia. ISBN 978-1-78068-343-0. Chapter 6, pp. 129-158.

Gijs Dekkers, 2015, The simulation properties of microsimulation models with static and dynamic ageing – a guide into choosing one type of model over the other. International Journal of Microsimulation, 8(1), pp 97-109.

Harding, Ann, 1996, Introduction. In: Harding, Ann (Ed.) Microsimulation and Public Policy. Contributions to Economic Analysis No 232. Bingley, UK: Emerald, Chapter 1, pp. 1-20.

Dekkers, Gijs, and Michele Belloni, 2009, A classification and overview of micro simulation models, and the choices made in MIDAS. In: Gijs Dekkers, Hermann Buslei, Maria Cozzolino, Raphael Desmet, Johannes Geyer, Dirk Hofmann, Michele Raitano, Viktor Steiner, Paola Tanda, Simone Tedeschi, Frédéric Verschueren, 2009, What are the consequences of the European AWG-projections on the adequacy of pensions? An application of the dynamic micro simulation model MIDAS for Belgium, Germany and Italy, ENEPRI Research Report No 65, AIM WP4. Brussels: Centre for European Policy Studies, pages 5-24.

Lelkes, Orsolya, and Holly Sutherland, 2009, Tax and Benefit Policies in the Enlarged Europe: Assessing the Impact with Microsimulation Model. Public Policy and Social Welfare, European Centre Vienna; Ashgate: Farnham, UK.

Sutherland, Holly, Multi-Country Microsimulation. In: Cathal O'Donoghue (Ed). Handbook of Microsimulation. Emerald Insight. Contributions in Economic Analysis series. Chapter 4, pp. 77-106.

Williamson, Paul, Asghar Zaidi and Ann Harding, 2009, New Frontiers in Microsimulation Modelling: Introduction. In: Asghar Zaidi, Ann Harding and Paul Williamson, New Frontiers in Microsimulation Modelling. Public Policy and Social Welfare, European Centre Vienna; Ashgate: Farnham, UK, chapter 1, pp. 31-50.

Conference Papers

Prospective microsimulation of pensions in European Member States

GIJS DEKKERS[1] – KAREL VAN DEN BOSCH[2]

The authors wish to thank Raphaël Desmet, Olle Sundberg, Anna Kwiecińska, Jan Škorpík, Ádám Rézmovits, Boris Majcen, Nataša Kump, Heikki Tikanmäki, Howard Redway, Marco Cacciotti, Dennis Fredriksen, Nils Stølen, Sorin Varna, Concepció Patxot Cardoner, Henri Bogaert, Didier Blanchet and Vidija Paštukienė for providing necessary information on the various models.

1. Introduction

Many public or semi-public research agencies and ministries in EU member states use microsimulation models (MSMs) to make projections of future pensions. As will become clear below, for good reasons, most of these models employ dynamic ageing techniques, though some use a static ageing method. The aim of this chapter is to discuss these models, many of which were not mentioned in the overview by Li and O'Donoghue (2013). Many models used outside of academia in EU member states are less well known, because they are less often documented in English, and their results are not often published in international journals and reports. Moreover, we include models employing dynamic ageing techniques, as well as those using a static ageing method, as both serve similar purposes. This enables us to bring out the advantages of dynamic models. As a secondary goal, this chapter aims to demonstrate the benefits of dynamic microsimulation for policy assessment by presenting some projections of the future adequacy of pensions for three countries. These are based on the hypotheses

---

[1] Federal Planning Bureau; CESO KU Leuven, Belgium, and LISER, Luxembourg.

[2] Federal Planning Bureau, and CSB,Antwerp, Belgium.

of the Working Group on Ageing Populations and Sustainability (AWG) of the European Council's Economic Policy Committee (EPC).[3] .

In the next section of this paper, we first explain why microsimulation of pensions can be useful. An overview of MSMs developed by public or semi-public agencies is presented in section three. A number of characteristics of these models, which are related to their use, are discussed in section four. In section five we show some results of such models for three EU countries, Belgium, Hungary and Sweden. Section six concludes.

2. Why microsimulation of pension systems?

The point of departure of all MSMs is a survey or an administrative dataset representing actual individuals at a certain point in time (Dekkers *et al.* 2010). The use of empirical micro data is what distinguishes microsimulation from agent-based modelling (Mannion *et al.,* 2012, 4.1.). Microsimulation models also differ from (semi-)aggregate budgetary models in that they simulate the impact of policy measures or events for every individual in the sample, and therefore go "beyond the averages", i.e. they are able to produce indicators that depend on the distribution of the variable of interest, e.g. pensions.

Obviously, if a MSM is to produce projections, it needs to include some notion of time. This is typically achieved by using dynamic ageing techniques which means that the model simulates the behaviour of micro-units over time (Li and O'Donoghue, 2013). Contrary to static microsimulation models, of which EUROMOD is the most well-known, or static-ageing models, a dynamic microsimulation model (DMM) builds up complete synthetic life histories for each individual in the dataset, including data on mortality, labour market status, retirement age, savings and so on. This is important in the case of Bismarckian public pension systems, or any pension system (e.g. Notional Defined Contribution schemes) in which pension benefits at retirement are determined by the labour or contributions history. For such pension systems, DMMs are about the only models that can be used to assess the future adequacy as well as financial sustainability of a system.

By contrast, static ageing techniques (like, for instance, used in Brewer *et al.*, 2011) involve updating the starting dataset, not through changing the individuals' characteristics, but rather by changing the weights through which sample results represent population parameters. This reweighting is done in such a way

---

[3] See "http://europa.eu/epc/working-group-ageing-populations-and-sustainability_en" for information about the AWG and the SPC.

that the weighted distribution matches exogenous aggregates, e.g. population projections by age and gender. So an ageing population is represented by increasing the weight of older individuals in the starting dataset, while younger individuals get a lower weight. In some static models (and also in some dynamic models), earnings and benefits are uprated using projected productivity growth rates and indexation hypotheses, respectively. However, static ageing techniques have important limitations. Future pensioners will differ in many ways from the current pensioners, having had different careers in a different economy. For instance, the next generation of retirees will have experienced the effects of the recent economic crisis during their working lives. It is not at all clear how this could be simulated by changing weights. Li and O'Donoghue (2013) state that with static microsimulation models "it is not possible to conduct analyses that require life event histories such as the simulation of pensions." Dekkers (2015) compares the simulation properties of microsimulation models with static and dynamic ageing, and concludes that the former are close to the latter only under rather special circumstances. Otherwise, valid indicators of the future adequacy and distribution of pensions can be only be produced by dynamic models.

It is for this reason that dynamic microsimulation models have traditionally been used for the analysis of pensions and pension reform in a context of demographic ageing. The original dynamic microsimulation model, DYNASIM, developed between 1969 and 1976 by, among others, Orcutt, Caldwell and Wertheimer (1976), was updated to DYNASIM2 in 1983 "to improve its ability to forecast retirement income and the effects of government policy on retirement income into the future" (Ross, 1991, 121). Another early model in the US, PRISM, was developed in 1980 specifically to assess the impact of alternative retirement income policies for the President's Commission on Pension Policy (Ross, 1991, 122; see also Webb *et al.*, 1990). Also today, the evaluation of pensions and pension reform is the most popular use for dynamic microsimulation models. Li *et al.* (2014, table 10.1, p. 307) assess 86 dynamic models worldwide, of which 40% are used for this purpose. The second most popular use is the study of inequality and redistribution; 15% of all models are used for this purpose. In practice, however, most pension models also assess inequality and redistribution, so there is considerable overlap.

3. An overview of semi-public models in EU member states

The goal of this chapter is to discuss models used for policy assessment in public or semi-public research agencies or ministries in EU member states. This is not to downgrade the contributions by models developed in academic settings, but the latter models are generally better known through international publications.

Although some models such as SESIM, and PENSIM II are an exception, models that are developed outside academics are usually less extensively documented in English and less often described in international scientific publications. Hence, information on these models rarely spreads further than the circles of developers. This paper provides information on dynamic microsimulation models that EU member states use for pension policy assessment outside pure academics and tries to identify some broad general characteristics. One must keep in mind that the boundary between models used by public research agencies (including ministries) and research institutions is not always clear.[4] Finally, the definition of what constitutes a microsimulation model is sometimes blurred.[5]

This paper uses the more general overviews of DMMs presented by Li and O'Donoghue (2013) and Li *et al.* (2014). But most information came from the exchange of emails between one of the authors on the one hand and developers or team leaders involved in the country models on the other. For this, the authors expresses their gratitude to the latter.

Table 1 presents a list of microsimulation models in public or semi-public research institutions and ministries in EU member states.

---

[4] The Luxembourg model MIDAS_LU and the Portugese model DYNAPOR, both developed in LIAM2 are not included, for example. This is not so much the result of their setting (the former is being developed in an academic setting, while the latter is the result of a collaboration between the University of Lisbon and the Ministry of Social Affairs), but mainly because it is work in progress (*cf. infra*). The model PENSIPP (Bozio, 2016), on the contrary has been used extensively in the evaluation in pension reform, but is not included either because it is developed purely in an academic setting. Not only is this also the case for the cohort-type model Gameo (Courtioux and Lignon, 2015), but it also does not model pensions (Blanchet *et al.*, 2015, 7) . So, this model is not included in the above list either. In Slovenia, the IER together with the Faculty of Economics, University of Ljubljana is currently also working on a ModGen-based model for Slovenia, DYPENSI. This model is not included in the list, mainly because it is work in progress to validate the results of the IER model.

[5] For example, the Lithuanian Ministry of Social Security and Labour uses a model called LICM (Lithuanian Individualized Cohort Model). Although this model, developed in Visual Basic for Applications includes some microsimulation-elements, one user described it as an "individual cohort model", based on Markov chains projection, a simplified model, and not a microsimulation model like the Polish, Czech or Romanian models (Paštukienė, 2016). As another example, Blanchet et al. (2011, 102) name a number of models used in by semi-public bodies in France (PRISME, Ariane, Vénus, Promess), but of that list, only the models PRISME and DESTINIE are seen as microsimulation models "*stricto sensu*", so they are the ones included in that list, together with French models that appeared afterwards.

Table 1 : A non-exhaustive overview of microsimulation models used for projections of pensions in EU member states

Country	Model name	Kind of model (see text)	Rev'd in Li et al. (2013)?	Software	Institution	Base dataset	Observations
AT	Unknown	Cross-sect.	N		Sozialversicherung des Bundesministerium für soziale Sicherheit		
BE	MIDAS-BE	Cross-sect.	Y	LIAM2	Belgian Federal Planning Bureau (FPB)	sample from administrative data	304,000 individuals
CZ	NEMO	Longitudinal	Y	Prophet	Ministry of Labour and Social Affairs	Sample from administrative data (2009), complemented by survey data	10% of total population
ES	DyPeS	Longitudinal	N	ModGen	Instituto de Estudios Fiscales IFS	„Muestra Continua de Vidas Laborales"Longitudinal administrative data	1.2M individuals
FI	ELSI	Cross-sect.	N	APL	Finnish Centre for Pensions	Sample of dministrative data, 2008	250000
FR	TRAJECTOIRE	Cross-sect.	N	SAS	DREES Direction de la recherche, des études, de l'évaluation et des statistiques	Administrative data, 2009	2.7% sample of individuals born between 1942 and 2009
FR	DESTINIE	Cross-sect.	Y	Perl/R & C++	INSEE	Financial assets survey	37,000 individuals
FR	PRISME	Cross-sect.	N	SAS	CNAV	5% sample from administrative data of contributors and pensioners in the „regime general"	about 200,000 individuals

Country	Model name	Kind of model (see text)	Rev'd in Li et al. (2013)?	Software	Institution	Base dataset	Observations
FR	Aphrodite	Cross-sect.	N	R	French Ministry of Finance		
HU	MIDAS-HU	Cross-sect.	N	LIAM2	Central Administration of the National Pension Insurance (CANPI)	Administrative data, 2012	2M individuals (20% of population)
IT	T-DYMM	Cross-sect.	N	LIAM/LIAM2	Economic and Financial Analysis and Planning, Office II	Adminstrative data, linked to EU-SILC survey data	55,708 individuals
IT	FaMiMod	Static weighting	N	Stata	National Institute of Statistics ISTAT	Adminstrative data, linked to EU-SILC survey data	
PL	ZUS-model	Longitudinal	N	Prophet	Polish Social Insurance Institution ZUS		
RO	SimPro-Vision 3	Longitudinal	N	Prophet	National House of Public Pensions (CNPP)	Administrative data on entire population	22.5M individuals
SE	SESIM3	Cross-sect.	Y	Visual Basic	Ministry of Health and Social Affairs	Longitudinal Individual Data Base (LINDA), 1999	100,000 individuals (Li) or 308,000 individuals
SI	IER model	Static weighting	N	Stata	IER	Administrative data, combined with survey data	115,000 individuals (5.8% of the population)
UK	PENSIM II	Cross-sect.	Y	GENESIS(SAS)	Head of Model Development - Unit Department for Work and Pensions	Household Panel Survey and Lifelong Labour Market Database	400,000 individuals

It is notable that nearly half of the MSMs in Table 1 were not mentioned in Li and Donoghue (2013). Almost all are dynamic models (either of the longitudinal or of the cross-sectional kind, see below), where the characteristics of individuals in the dataset are being changed over time. The IER model in Slovenia and the Italian FaMiMod are the only models in the above list that uses static ageing techniques, in which the ageing of the population is simulated by increasing the weight of older individuals in the starting dataset, while younger individuals get a lower weight. In the Slovenian model, this reweighting reflects changes in the age distribution and also in the educational profile of the population, both by gender (Majcen et al., 2012, section 3.4., page 13 and further, and page 44). Furthermore, earnings and benefits are uprated using productivity growth rates and indexation hypotheses, respectively. In the recently developed Italian model FaMiMod, the weights reflect changes in the demographic structure and the employment status of the resident population. Furthermore, uprating is based on "National Accounts average growth rates of personal incomes, broken down whenever possible by source, geographical area and economic sector of activity" (Cozzolino *et al.*, 2015, 27).

The models shown and discussed in this paper are all used for the analysis of pensions and pension reform. Given that pensions depend on earnings, this implies that earnings of the working population are also simulated. However, there is a difference in the scope of the models apart from pensions and earnings. Most models, such as PRISM, TRAJECTOIRE and the models that use longitudinal ageing (DyPeS, NEMO, SimProVision and ZUS) simulate only earnings and pensions. Other models include also other sources of income such as unemployment benefits, means-tested minimum benefits, child benefits and other sources of income. This is, for example, the case with MIDAS-BE, MIDAS-HU, PENSIM II, DESTINIE, and SESIM. This information makes it possible simulate household-income for all households, which in turn enables these models to make projections of poverty risks and inequality for the whole population.

There are a number of collaborative ties between the various models, especially the more recent ones, and also between those and the models that currently are in development. Often these ties are the result of the use of a common software. The Polish model ZUS, the Czech model NEMO and the Romanian model SimProVision were all developed by Deloitte. Although used and maintained independently, they share a common basis and structure. The first generation of the DMM MIDAS for Belgium, Germany and Italy were developed in LIAM (Dekkers et al., 2010b). The Italian model was later updated and expanded with a structural labour market module to become the current version of T-DYMM. Meanwhile, the open-source development tool LIAM2 was developed by the Belgian Federal Planning Bureau in collaboration with the Luxembourg Institute of Social and Economic Research (LISER) (De Menten *et al.*, 2014). This tool was used to create the current version of MIDAS-BE and the Hungarian model MIDAS-HU, and is also applied to the Italian model T-DYMM. LIAM2 is also used to develop MIDAS-LU in Luxembourg and a Portugese model called DYNAPOR. Finally, the Slovenian work-in-progress model DYPENSI is being based on the Spanish DyPeS framework.

4. Characteristics of DMMs

4.1 Data used

A key element in any microsimulation model is the data which it uses. According Li and O'Donoghue (2010, table 2, page 12; see also Li *et al.*, 2014), because of "legal and privacy reasons", most dynamic models are based on survey or census data, or a combination of the two. As Table 1 shows, this is less the case for the models reviewed in this chapter, as they are have been developed or used mainly by ministries or (semi-)public organisations. Indeed most of them use a dataset consisting of administrative records, often complemented by (imputed) information from survey or census data. Survey data are used to impute household structure and/or educational attainment level and to estimate behavioural equations. Almost all models use large samples, of up to 4 million individuals, as in the case of the French model PRISME, though in proportional terms the samples amount to less than 10% of the population. The Romanian SimProVision model uses the fact that longitudinal ageing does not require that all individuals are in memory at the same time to use an exceptionally large dataset that includes data for the entire population alive at September 2014, or about 22,5 million individuals.

The Polish model ZUS is based on a compound administrative dataset. Though the model is capable of simulating the entire Polish population, for most applications a 10% sample is used. Survey data are used not in the starting dataset, but to estimate transition rates and their trends over time. The Spanish model DyPeS has as its starting dataset a sample of about 1.2 million individuals in 2007, drawn from the "Muestra Continua de Vidas Laborales" (MCVL; Fernández-Díaz *et al.*, 2013, 3) longitudinal administrative registers. This is about 4% of the Spanish population of workers and pensioners. The Swedish model SESIM is based on the LINDA database of 1999. This is a 308 thousand individuals' sample of administrative records, or about 3.5% of the Swedish population (Flood *et al.*, 2012, chapter 3). This dataset is completed by imputations based on survey data. For example, LINDA being administrative, it only contains data on fiscal households. Also, LINDA only contains records of individuals living on Swedish territory. Economic households and information on pension rights of those living abroad are being imputed on the basis of the Household Income Survey. The French model TRAJECTOIRE is based on a 2009 administrative dataset covering all generations born between 1942 and 2009. For each generation, a representative sample of about 2.7% of each cohorts' population is drawn (Duc *et al.*, 2013, 10). The other French model in our list, PRISME, describes the participants and pensioners of the "regime general" (general scheme; Albert (2009, page 2)) and uses a sample of 5% of administrative data, which in its version of 2009 contains about 4 million individuals. The sample on the active population is renewed every 2 years, but data on the retired is updated every 6 months (Albert (2009., page 4). The UK model Pensim2 is based upon a combination of administrative data (the Lifelong Labour Market Database (LLMDB) which contains a sample of 400,000 individuals or 1% of the population in 2001) and survey data (Family Resources Survey and British Household Panel Survey (Emmerson et al, 2004, 14; O'Donoghue *et al.,* 2010, 67). The Belgian model MIDAS_BE uses an all-administrative dataset of 304,000 individuals in 2001, which equals about 3% of the Belgian population (Dekkers *et al.*, 2015). This dataset is currently being replaced by an administrative dataset of almost 602,000 individuals in 2011. The individual records in this dataset are linked with Census data in order to assess, among other things, the educational attainment level. The Czech model NEMO is based on an administrative database for December 2009 from the Czech Social Security Administration (CSSA). This dataset is completed by survey-datasets from the Ministry of Labour and Social Affairs (MLSA) and the Czech Statistical Office (CSO) (Ministry of Labour and Social Affairs, 2012, appendix 3, 150). The latter two are used for imputing household structures and educational attainment levels. The combined starting dataset contains the full Czech population, but for efficiency reasons a 10% sample is used. The Slovenian model IER is also based on a combination of administrative and survey data for a sample of 115,000 individuals (Majcen *et al.*, 2012, 64), which is a bit less than 6% of the population. Survey data is used to describe household structure and educational attainment level. As in Belgium, these characteristics are not imputed but individual records are linked using a unique individual registration number. The Finnish model ELSI is based on an administrative dataset of approximately 250,000 individuals in 2008 (Tikanmäki, *et al.*, 2014, 14), or about 5% of the Finnish population. The data for the Hungarian MIDAS_HU model also consists of administrative records for almost 2 million individuals or 20% of the Hungarian population in 2012 (Dekkers *et al.,* 2015). Retrospective information prior to 1997 has been imputed using a different and not-linked administrative dataset. The Italian model T-DYMM use an administrative dataset, called "Administrative SILC" and which consists of administrative data from the INPS (National Institute of Social Security) with EU-SILC survey data collected by ISTAT (National Institute of Statistics). The information from the latter is mainly used for educational attainment level, the structure of economic households and marital status. All in all, it consists of 55,708 individuals or about 0.09% of the Italian population. However, a new version of T-DYMM is being constructed in LIAM2, which will be based on a dataset of nearly 150,000 observations. The Italian model FaMiMod is based on the EU-SILC dataset only.

The general conclusion is therefore that, in contrast to the conclusion drawn by Li *et al.* (2014) on a broader sample of models, nearly all European models discussed in this paper are completely or mainly based on administrative records. If survey data is used at all, it typically is used to impute household structure and/or educational attainment level and to estimate (reduced-form) behavioural equations.

4.2 Longitudinal vs. cross-sectional models

As explained above, in DMMs the characteristics of the individuals in the dataset are being changed over time. So individuals go through the various stages of life between (and including) birth and death, while earning an income when working and receiving a pension benefit at retirement. Any DMM simulates *n* individuals from period 0 to *T*. An important distinction between DMMs pertains to the order in which this is done. In so-called longitudinal models, the lifetime of one individual *i* is simulated for all the years t=0..*T* from start to finish before the simulation of the next individual *i*+1 is started. In contrast, models that use what is called "cross-sectional ageing" first simulate all individuals 1..n for a certain period and then move on to the next period. Sometimes the former type of models are also referred to as 'case-based' models, since one case is simulated after another, in contrast to 'time-based' models where one period is simulated after another (Fernández-Díaz *et al.*, 2013, 10).

The advantage of longitudinal ageing is that one needs only to have one individual (and all previous years) in the memory of the computer at any time. Hence the simulation of interactions between individuals is difficult. This is problematic for two reasons. The first is that one wants to be able to simulate not just the transition of individuals between various states, but also household formation and dissolution through such events as marriage, widowhood, divorce, leaving the parental home etc. Information on the household situation is needed because many of the EU indicators of (pension) adequacy are based on equivalent household income. This includes the At-Risk-of-Poverty rate, inequality indicators such as the Gini and the Income Quintile Share Ratio (European Union, 2012, annex 1, page 141). The second reason why the lack of interaction can be problematic is that probabilities of an event happening (say moving from one state to another) at any moment *t* might be dependent on an aggregate, for example the proportional number of people in either state at *t-1*. This is obviously very difficult in longitudinal-ageing models, whereas it comes naturally in cross-sectional ageing models.

Four models in the above list that use longitudinal ageing are the Spanish DyPeS model (Fernández-Díaz *et al.*, 2013), the Czech model NEMO (Fialka, et. al., 2011), the Polish ZUS-model and the Romanian model SimProVision (CNPP, 2015). It is interesting to note that the last three models (all developed by Deloitte) simulate households indirectly through so-called "model points". A model point consists of one individual, the "main character", but also contains information on "auxiliary individuals" such as family members (Ministry of Labour and Social Affairs, Actuarial Report on Pension Insurance, 2012, page 59 and page 142). The full simulation of the model is only done for the main individuals, whereas the information simulated for auxiliary individuals is limited for the purpose of simulating characteristics that have an effect on the pension of the main individual, and in order to simulate a proxy of household income as the basis of equivalent income[6]. The problem of 'aggregates' is resolved in the Polish ZUS-model and the Romanian SimProVision through pre-runs. The Czech model, in contrast, uses background software to perform what they call "dynamic simulation". As far as we know, the technical details of both approaches remain undocumented.

To our knowledge, all the other models in the above Table 1 use cross-sectional ageing techniques. This allows to have all individuals being simulated in any future period at the same time. Thus interactions and the simulation of household income is – at least conceptually – straightforward. The drawback is that these models are very demanding in terms of memory and CPU. So the numbers of individuals that are simulated is often lower in these kinds of models.

4.3 Time and order of events

Another typical characteristic of the above models is that they make prospective simulations in discrete time steps, a month in the case of the Polish, Romanian and Czech models; 3 months in the case of the French model PRISME, but mostly a year. The exception is the Spanish DyPeS model, which simulates in continuous time, although "some of the events happen only once a year"

---

[6] Note that, in the starting dataset, any auxiliary individual in any model point is a main character in another. But during simulation, this link is broken, and the 'two' individuals are simulated independently.

(Fernández-Díaz *et al.*, 2013, 10)[7]. Discrete simulation involves updating the characteristics of each individual in each time interval. In continuous-time microsimulation models, an individual runs parallel risks on 'competing events'. An individual has at the same time a 'risk' of falling ill and of losing her job, for example. The order of simulation of these events is therefore irrelevant. In discrete time models, the modeller needs to choose a specific ordering of the events. This simulation of one event at a time in a fixed order with conditional transitional probabilities, is also known as recursive simulation.

The particular order imposed may have an impact on the aggregate number of events simulated in a period. First of all, some events may make other events impossible. Second, the probability of an event taking place may be affected by the other event. Take, for example, the event of dying and of giving birth. If birth is simulated prior to death a woman can give birth and die in the same period. If, on the other hand, death is simulated prior to giving birth, that woman will die and not give birth. In the first scenario, children will be born (and will be orphans), while they will not be born at all in the second scenario. This difference will be reinforced if giving birth increases the immediate probability of death of the mother, because a woman will be more likely to die in the first scenario than in the second one. This example shows that the order of the processes can over time result in quite different simulation results. This problem could be partially circumvented by making the probability of an event in one period dependent on events in the previous period. In our particular example, this could mean simulating mortality prior to fertility, but at the same time giving women that have given birth *in the previous year* a higher mortality risk in the current year. But even in that case, the order is to a great extent arbitrary and therefore debatable, while it can change the simulation results a model produces.

In most models, demographic processes are simulated before labour market processes, which in turn precede the simulation of the various earnings and benefits, and the (pension) adequacy indicators. Hence the implicit assumption is that demographic processes affect labour market processes in the same period, while earnings, benefits and poverty risks are the result of the labour market state (or inactive state) the individual occupies at any moment in time, but that these earnings and benefits do not affect the labour market state in the same period (but possibly in later periods). Of course, this does not preclude that the potential wage or benefit, given a certain labour market transition, are calculated

---

[7] The Slovenian work-in-progress model DYPENSI, also developed in ModGen, shares this feature with the Spanish DyPeS. Basically, it is a continuous-time model, but labour market transitions are simulated monthly while household formation and dissolution and net immigration are simulated annually. Contrary to DyPeS, however, DYPENSI is a cross-sectional ageing model.

for an individual, and that these potential incomes influence the transition probabilities. This is in fact important to model the retirement decision. In many cases, a potential retirement benefit is being calculated, based on the previous career of the individual or the points gathered or contributions made, depending on the system in effect. This potential benefit becomes actual when the individual retires. So for a working individual, information on the simulated earnings and potential pension benefits is available, and the confrontation between the two can easily be used as a proxy for the implicit tax on staying in the labour market (OECD, 2003, Gruber and Wise, 2007), and hence be included in the simulated retirement decision of the individual.

5. An application

In this section, we present a recent application of dynamic microsimulation on the EU level. The impact of demographic ageing, in conjunction with the pension policies of Member States, on the future financial sustainability of pension systems is a focal point of attention on the European level. But a sensible assessment of financial sustainability cannot do without taking into account the social impact of ageing, and that of current and future pension policies. The Working Group on Ageing Populations and Sustainability (AWG) of the European Council's Economic Policy Committee (EPC) uses two indicators of the future adequacy of pensions: the benefit ratio (the ratio between the average pension benefit and the economy-wide average wage) and the replacement rate at retirement (the average first pension as a share of the economy wide average wage at retirement) (European Commission, 2015a, pp. 91-94). These indicators are based on aggregate figures, and, as noted in the EPC's report, have important limitations. In the context of the European Council's Social Protection Committee (SPC), the SPC-AGE working group has developed another indicator of the adequacy of pensions, viz. theoretical replacement rates (TRR) (European Commission, 2015b, p. 112). These are calculations on an individual basis, for an assumed hypothetical worker, of the level of pension income in the first year after retirement, as a percentage of individual earnings at the moment of retirement. Since the characteristics (level of earnings, full or incomplete career, etc.) of the hypothetical worker can be changed, the TRRs also give an indication about variations in adequacy across retirees.[8] However, prospective values of

---

[8] An important limitation of the TRRs, as given in the latest report of the SCP (European Commission, 2015b), is that they are given only for the year 2053, i.e. for a person who started work in 2013 at the age of 25, and retired at age 65. Of course, TRRs could be computed for other years as well.

key indicators, such as the at-risk-of-poverty rate among the 65, or the income inequality within the same group, are not available.

As a demonstration of how microsimulation models could fill an information gap about future pension adequacy, teams of Sweden (SESIM), Belgium (MIDAS-BE), and Hungary (MIDAS-HU) agreed to use their models to simulate possible developments in the at-risk-of-poverty rate (AROP) and income inequality (Gini coefficient and the Income quintile share ratio S80/S20) among pensioners, the elderly (aged 65 and more), the working and, finally, the population as a whole. As they took as much as possible the projections and hypotheses of the AWG into account, the results show the implications of the latter on prospective poverty risks and income inequality. The full report (Dekkers *et al.*, 2015) also contains a more detailed description of the simulation results for a number of scenarios. In this chapter we will limit ourselves to a brief discussion of the AROP results for pensioners.

The following figures present the trends in the AROP rate, using 60% of median equivalent household income as the poverty threshold and applying the modified OECD equivalence scale (1, .5, .3), for Belgium, Sweden and Hungary. These results were simulated using the abovementioned DMMs. The Benefit ratio and Gross replacement rate are the aggregate results reported by the AWG.

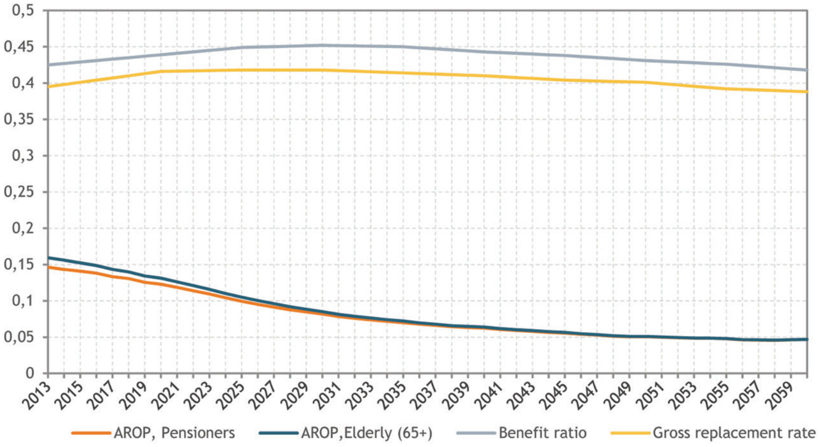


Figure 1: AROP, benefit ratio and gross replacement rate for Belgium

Note: Benefit ratio and gross replacement rate are only given for years 2013 and 2020-2060 at five-year intervals. Values for other years are interpolated by authors.

Sources: AROP: MIDAS – base scenario. Benefit ratio and gross replacement rate: European Commission (2015a)

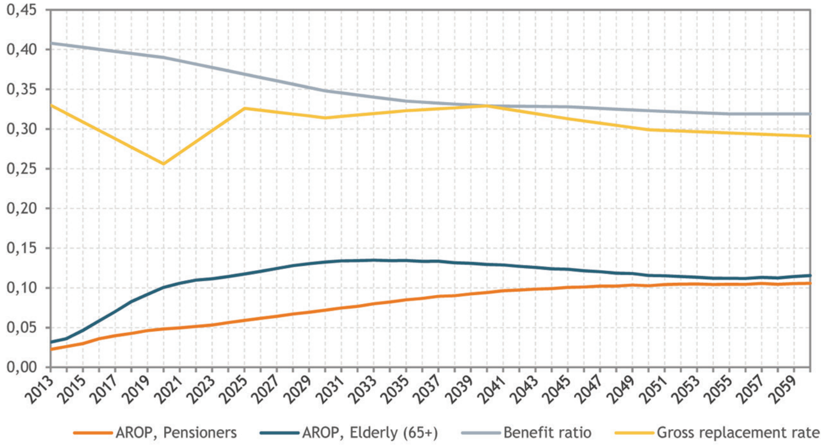


Figure 2: AROP, benefit ratio and gross replacement rate for Hungary

Note: Benefit ratio and gross replacement rate are only given for years 2013 and 2020-2060 at five-year intervals. Values for other years are interpolated by authors.

Sources: AROP: MIDAS_HU – base scenario. Benefit ratio and gross replacement rate: European Commission (2015a)

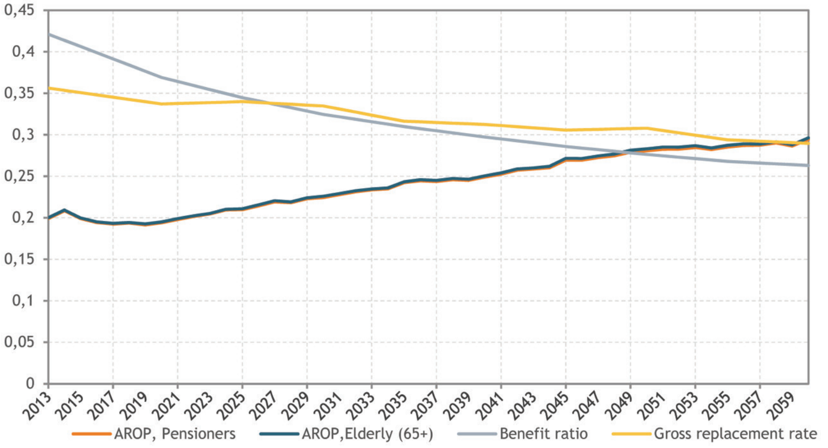


Figure 3: AROP, benefit ratio and gross replacement rate for Sweden

Note: Benefit ratio and gross replacement rate are only given for years 2013 and 2020-2060 at five-year intervals. Values for other years are interpolated by authors.

Sources: AROP: SESIM – base scenario. Benefit ratio and gross replacement rate: European Commission (2015a)

As the simulation results are consistent with the budgetary projections in the base scenario of the AWG (or at least as much as possible), it is possible to interpret the simulation results on poverty in Figures 1 to 3 while simultaneously considering the AWG projections on the budgetary consequences of ageing, as well as the benefit ratio and replacement rate in these three countries. In Belgium, the increasing dependency ratio is expected to cause gross public pension expenditures to increase by 3.5 %-points of GDP between 2013 and 2060 (see Federal Planning Bureau (2014, Table 10, page 19)). The increasing number of pensioners is not mitigated by reductions of the relative benefit levels, as indicated by the rather stable evolution of the benefit ratio and the gross replacement rate. In fact, as Figure 1 shows, the poverty risk among pensioners and the elderly would strongly decrease, mainly because of increases in the minimum pensions, the increasing employment rate of women, and a lower productivity growth rate in the first half of the simulation period. In Sweden, on the other hand, gross public pension spending would decrease by 1,5 percentage points over the same period (Ministry of Finance Sweden, 2015). The fact that the AWG assumes a fixed retirement age, whereas the Swedish NDC system decreases the individual's pension account value as life expectancy increases, causes the benefit ratio and replacement ratio to decrease and the risk of poverty among the pensioners to increase. Finally, pension expenditures in Hungary are expected to decline by 0.1%-point over the period 2013-2060 (Hungarian Ministry for National Economy, 2015, table 8a, p. 16). Also, the benefit ratio decreases over the whole simulation period (while the replacement rate has a less clear-cut trend), mainly due to the shorter service years of future cohorts entering retirement. This in turn results in an increasing poverty risk among pensioners. However, for the population aged 65 or more, the AROP is projected to rise until the 2030s, and then starts to decline. The group of pensioners does not coincide with that of persons aged 65+: many elderly people do not draw a pension, while an important minority of pensioners are below 65. The extent of overlap varies considerably between countries, and is also not constant across time.

Note that the above results are based on the base-variant of the AWG. The AWG also developed a number of sensitivity tests "in order to quantify the responsiveness of projection results to changes in key drivers," (European Commission, 2015, p. 48). The impact of these tests have also been simulated for Belgium and Sweden. It would go too far to discuss all variants and their simulation results; suffice to say that these results for Belgium show that increasing the employment of older workers not only results in an important reduction of pension expenditure relative to the base scenario, but also considerably reduces the poverty risk among the elderly.

6. Conclusion

This chapter discussed microsimulation models used for policy assessment in public or semi-public research agencies or ministries in EU member states. Many of these do not result in academic publications or reports that are easily accessible, and therefore were not listed in the overview by Li and O'Donoghue (2013). Almost all of these are dynamic simulation models, in the sense that these build up synthetic life histories of individuals, although static microsimulation models, where the ageing process is simulated by changing weights, are also used. Though different in many ways, they share common characteristics, of which their use of administrative data as the primary database is the most remarkable one. Survey data play a supplementary role, to impute missing data, or to estimate behavioural equations. In this they differ from the more academically oriented dynamic microsimulation models surveyed in Li and O'Donoghue, which rely to a greater extent on survey and census data. Furthermore, there are extensive ad-hoc collaborative ties between various teams, partly driven by various EU-funded research projects. As a result, even though the ambition of a network for dynamic microsimulation has not been realized yet (Dekkers and Zaidi, 2011), collaboration between various teams is now more intense than ever.

Projection of key indicators of the social adequacy of future pensions, such as the at-risk-of-poverty rate, and the Gini and S80/S20 indicators of income inequality among pensioners, is only possible using a microsimulation model. Thus, these models can provide information on projected pension adequacy. This is illustrated by applying dynamic microsimulation models for Sweden, Hungary and Belgium, taking the budgetary projections of the Working Group on Ageing Populations and Sustainability (AWG) into account. Results show that the information on pension adequacy provided by microsimulation models have a clear added-value with respect to currently used indicators of the future adequacy of pensions, i.e. the benefit ratio and the gross replacement rate. The latter do not reflect some developments with important effects for the future incomes of the retired, e.g. the increased labour participation rate of women.

References

Brewer, Mike, James Browne and Robert Joyce (2011) *Child and Working-Age Poverty from 2010 to 2020*. The Institute of Fiscal Studies, October 2011. IFS Commentary C121. http://www.ifs.org.uk/comms/comm121.pdf [10/07/2015]

Courtioux, Pierre, and Vincent Lignon, 2015, Homogamie éducative et inégalités de revenu salarial : une perspective de cycle de vie. *Economie et Statistique* n° 481-482, pp. 149-184. http://www.insee.fr/fr/publications-et-services/sommaire.asp?reg_ id=0&ref_id=ECO481 [12/8/2016]Cozzolino, Maria, Carlo Declich, Marco Di Marco, Gaetano Proto, Fabrizio Solari, Paola Tanda (2015) Modelling Social Security, Direct Taxes and Cash Benefits. *Rivista di Statistica Ufficiale*, 2015 No 2. 27-66. http://www.istat.it/it/archivio/171133 [08/02/2017]

Cuvilliez, Julia, and Thomas Laurent, 2016, Le modèle de microsimulation dynamique des retraites Aphrodite. Documents de Travail de la DG Trésor. Document de Travail Numéro 2016/04 (July 2016), http://www.tresor.economie.gouv.fr/File/426608

Dekkers, Gijs, Hermann Buslei, Maria Cozzolino, Raphael Desmet, Johannes Geyer, Dirk Hofmann, Michele Raitano, Viktor Steiner, Paola Tanda, Simone Tedeschi, Frédéric Verschueren, 2010b, The flip side of the coin: the consequences of the European budgetary projections on the adequacy of social security pensions. *European Journal of Social Security*, 12(2), June 2010, pp. 94-120.

Dekkers, Gijs, Raphaël Desmet, Greet De Vil, 2010, *The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model*, Working Paper 10-10, May, 3rd 2010, Brussels: Federaal Planbureau. http://www.plan.be/publications/publication_det.php?lang=nl&TM=30&IS=63&KeyPub=946 [09/04/2015]

Dekkers, Gijs, Raphaël Desmet, Nicole Fasquelle, Saskia Weemaes, 2015, The social and budgetary impacts of recent social security reform in Belgium. In Ioana Salagean, Catalina Lomos & Anne Hartung, *The young and the elderly at risk: Individual outcomes and contemporary policy challenges in European societies*, Intersentia. ISBN 978-1-78068-343-0.

Dekkers, Gijs, Raphaël Desmet, Ádám Rézmovits, Olle Sundberg, and Krisztián Tóth, 2015, *On using dynamic microsimulation models to assess the consequences of the AWG projections and hypotheses on pension adequacy: Simulation results for Belgium, Sweden and Hungary*. Paper presented to the SPC Working Group on Ageing Issues (SPC-WG-AGE), Brussels, March 27th, 2015. http://www.plan.be/publications/publication_det.php?lang=nl&KeyPub=1466 [10/07/2015]

Dekkers, Gijs, Asghar Zaidi, 2011, The European Network for Dynamic Microsimulation (EURODYM) – a vision and the state of affairs. International Journal of Microsimulation, Vol. 4(1), Spring 2011, pp. 100-105. http://microsimulation.org/IJM/V4_1/Volume%204%20Issue%201/6_NOTE_Dekkers_Zaidi_IMA_2011_EURODYN_revised_last_time_FINAL.pdf [13/05/2015]

Dekkers, Gijs, 2015, The simulation properties of microsimulation models with static and dynamic ageing – a guide into choosing one type of model over the other. *International Journal of Microsimulation*, 8(1), pp 97-109.

de Menten, Gaëtan, Gijs Dekkers, Geert Bryon, Philippe Liégeois, Cathal O'Donoghue, 2014, LIAM2: a new open source development tool for discrete-time dynamic microsimulation models, *Journal of Artificial Societies and Simulation*, 17(3)9. http://jasss.soc.surrey.ac.uk/17/3/9.html [07/05/2015]

Duc, Cindy, Laurent Lequien, Félix Housset and Corentin Plouhinec, 2013, Le modèle de microsimulation TRAJECTOiRE – TRAJECoire De Carrières TOus RÉgimes. Direction de la recherche, des études, de l'évaluation et des statistiques DREES, document de travail Série Sources Et Méthodes n° 40, May 2013. http://fulltext.bdsp.ehesp.fr/Ministere/Drees/SerieSourcesMethodes/2013/40/seriesource_method40.pdf [10/04/2015]

European Union, 2012, *Pension Adequacy in the European Union 2010-2050*. Report prepared jointly by the Directorate-General for Employment, Social Affairs and Inclusion of the European Commission and the Social Protection Committee. http://ec.europa.eu/social/BlobServlet?docId=10421&langId=en

European Commission, 2015a, The 2015 Ageing Report – Economic and budgetary projections for the 28 EU Member States (2013-2060)., *European Economy* 3/2015. http://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf

European Commission, 2015b, The 2015 Pension Adequacy Report: current and future income adequacy in old age in the EU, Volume 1. Luxembourg: Publications Office of the European Union. http://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=7828&visible=0.

Federal Planning Bureau, 2015, *Economic Policy Committee's Ageing Working Group. Belgium: Country Fiche 2014*, Federal Planning Bureau, REP_COUNTRYFICH2014_10912, November 27th, 2014.

Fernández-Díaz, F., C. Patxot, G. Souto, 2013, DyPeS : A Microsimulation model for the Spanish Retirement Pension System. Fedea, Documento de Trabajo 2013-6. http://documentos.fedea.net/pubs/dt/2013/dt-2013-06.pdf [07/05/2015]

Fialka, Jiři, Aleš Krjedl, Petr Bednařik, 2011, *Summary Based On The Final Project Report Of The Dynamic Micro-Simulation Model Of The Czech Republic*. Deloitte, Actuarial & Insurance Solutions. http://www.deloitte.com/ASSETS/Dcom-CzechRepublic/Local%20Assets/Documents/Deloitte_-_Summary_Report_%20Dynamic_Microsimulation_Model_of_Czech_Republic.pdf [21/04/2015]

Flood, Lennart, Fredrik Jansson, Thomas Pettersson, Tomas Pettersson, Olle Sundberg and Anna Westerberg, 2012, SESIM III – A Swedish Dynamic Microsimulation model. http://www.sesim.org/ [06/05/2015]

Gruber, Jonathan, and David Wise, 2007, Introduction. In Gruber and Wise (Eds), *Social Security Programs and Retirement around the World*, National Bureau of Economic Research Conference Report. Chicago: the University of Chicago Press, pp. 1-42.

Holzmann, R. and R. Hinz, 2005, "Old-Age Income Support in the Twenty-first Century: An Inter-national Perspective on Pension Systems and Reform", Web Version of 18 February 2005, World Bank.

Hungarian Ministry for National Economy, 2015, Country Fiche On Pension, Hungary, 22nd January, 2015.

Li, Jinjing, Cathal O'Donoghue, 2012, Simulating Histories within Dynamic Microsimulation Models. *International Journal of Microsimulation*, 5(1), 52-76.

Jinjing Li, Cathal O'Donoghue, and Gijs Dekkers, 2014. Dynamic Microsimulation. In: Cathal O'Donoghue (Ed). *Handbook of Microsimulation*. Emerald Insight. Contributions in Economic Analysis series.

Boris Majcen, Jože Sambt, Mitja Cok, Tomaž Turk, Gijs Dekkers, Vladimir Lavrac, Nataša Kump, 2012, *Development Of Micro-Simulation Pension Model: Linking The Modules Within Graphic Interface*. IER Working Paper No. 69, 2012. Institute for Economic Research, Ljubljana, Slovenia.

Bozio, Antoine, 2016, PENSIPP – dynamic micro-simulation model of the French pension system. https://github.com/abozio/PENSIPP [04/08/2016]

Mannion, Oliver, Roy Lay-Yee, Wendy Wrapson, Peter Davis and Janet Pearson, 2012, JAMSIM: a Microsimulation Modelling Policy Tool. *Journal of Artificial Societies and Social Simulation* 15(1), 8. http://jasss.soc.surrey.ac.uk/15/1/8.html [13/05/2015]

Ministry of Finance Sweden/Regeringskansliet, 2015, *The pension system and pension projections until 2060*, Swedish pension fiche prepared for the AWG projections for age-related public expenditure 2015.

Ministry of Labour and Social Affairs, 2012, Actuarial Report on Pension Insurance, http://www.mpsv.cz/files/clanky/14299/PMZ_2012_en_final.pdf [22/ 04/2015]

National House of Public Pensions (CNPP), 2015, Final report on micro-simulation model of the evolution of public pension system of Romania. Development of a national pension model for policy-making purposes. https://www.cnpp.ro/documents/10180/567439/Raport%20final_model_EN.pdf [06/04/2016]

OECD, 2003, OECD (2003) *Labour Force Participation of Groups at the Margin of the Labour Market: past and future trends and policy challenges*, Paris: Organisation for Economic Co-operation and Development OECD, Economics Department for Working Party No 1 on Macroeconomic and Structural Policy Analysis, ECO/CPE/ WP1(2003)8, September 2003.

Orcutt, Guy, Steven Caldwell, Richard Wertheimer III, 1976, *Policy Exploration Through Microanalytic Simulation*. Washington D.C.: The Urban Institute.

O'Donoghue, Cathal, Howard Redway, John Lennon, 2010, Simulating Migration in the Pensim 2 dynamic microsimulation model. *International Journal of Microsimulation*, 3(2), 65-79. http://microsimulation.org/IJM/V3_2/Volume%203%20Issue%20 2/5_IJM_51%20Proof.pdf [06/05/2015]

Ross, Christine, 1991, DYNASIM2 and PRISM: Examples of Dynamic Modeling. In Constance Citro and Eric Hanushek (Eds.) *Improving for Social Policy Decisions – the uses of Microsimulation Modeling*, Vol. II – technical papers. Washington D.C.: National Academy Press.

Statistics Canada (2015) ModGen (Model Generator) http://www.statcan.gc.ca/eng/microsimulation/modgen/modgen [11/05/2015]

Tikanmäki, Heikki, Hannu Sihvonen and Janne Salonen, 2014, *Microsimulating Finnish earnings-related pensions*. Finnish Centre for Pensions Working papers, 02/2014.

http://www.etk.fi/en/gateway/PTARGS_0_2712_459_770_3439_43/http%3B/content.etk.fi%3B7087/publishedcontent/publish/etkfi/en/julkaisut/research_publications/working_papers/microsimulating_finnish_earnings_related_pensions_7.pdf [23/04/2015]

Webb, Randall, Richard Michel and Anne Bergsman, 1990, The Historical Development of the Transfer Income Model (TRIM2), in: Gordon Lewis and Richard Michel (Eds.), *Microsimulation Techniques for Tax and Transfer Analysis*, Washington D.C.: The Urban Institute Press, Chapter 2, 33-75.

Albert, Christophe, Caroline Berteau-Rapin and Alessandra Di Porto, 2009, PRISME - A dynamic microsimulation model for the French pension scheme CNAV. Working Paper by the Statistiques, recherches et prospective de la Caisse Nationale d'Assurance Vieillesse. http://www.statistiques-recherches.cnav.fr/images/prospective/Prisme_english_version2009.pdf [19/5/2016]

Paštukienė, Vidija, 2016 (personal communication, 11 April 2016 10:27).

Blanchet, Didier, Sophie Buffeteau, Emmanuelle Crenner, and Sylvie Le Minez, 2011, Le modèle de microsimulation Destinie 2 : principales caractéristiques et premiers résultats. *Économie Et Statistique* No 441–442, 2011, pp. 101-121. http://www.insee.fr/fr/ffc/docs_ffc/ES441F.pdf [05/08/2016]

Didier Blanchet, Cyrille Hagneré, François Legendre et Florence Thibault, 2015, Introduction - Microsimulations statique et dynamique appliquées aux politiques fiscales et sociales : modèles et méthodes. *Economie et Statistique* n° 481-482, pp. 5-30. http://www.insee.fr/fr/publications-et-services/sommaire.asp?reg_id=0&ref_id=ECO481 [12/8/2016]

Sungard (2016), Prophet Procuct Sheet. http://www.prophet-web.com/assets/Datasheets/New-Brand-Datasheets/12_2015/0511INSProphet-Professional-Product-sheetWEB.pdf. [10/08/2016]

Improving the validity of microsimulation results: lessons from Slovakia*

ZUZANA SIEBERTOVA – NORBERT SVARDA –
JANA VALACHYOVA

ZUZANA SIEBERTOVA
Council for Budget Responsibility
I. Karvasa 1, 813 25 Bratislava, Slovakia
e-mail: siebertova@rrz.sk

NORBERT SVARDA
Council for Budget Responsibility
I. Karvasa 1, 813 25 Bratislava, Slovakia;
Faculty of Mathematics, Physics and Informatics
Comenius University, Bratislava, Slovakia
e-mail: svarda@rrz.sk

JANA VALACHYOVA
Council for Budget Responsibility
I. Karvasa 1, 813 25 Bratislava, Slovakia
e-mail: valachyova@rrz.sk

ABSTRACT: This paper summarizes the lessons learned in the process of building a microsimulation tool tailored to country-specific conditions and involving a maximum degree of user control. The objective to construct a model useful in the process of budgeting and fiscal forecasting has been achieved by paying attention to policy simulation details as well as to the representativeness of the underlying micro-dataset. The validity of simulated results improved significantly after the input database sample has been reweighted in such a way that the new weights replicate, among other factors, the earned income distribution and selected age cohorts directly. Innovative approaches in bringing the model closer to legislation as well as data highlight the benefits of having more user control compared with standardized microsimulation tools.

KEYWORDS: microsimulation, calibration, EUROMOD, tax and transfer policy, Slovakia
JEL classification: C81, C83, C88, I38, H24.

---

*Published in the International Journal of Microsimulation in 2016

1. Introduction

Microsimulation modelling techniques are increasingly used to study the effects of reform policies contributing thus both to policy debates and academic literature (Figari *et al*., 2015). This is largely possible due to availability of highly useful and user-friendly standardized tools. EUROMOD – a fine example of such a tool – has become a benchmark for conducting microsimulation-based policy analyses for European countries.

If the appeal of the microsimulation models is to go beyond providing guidance on the design of policy reforms and be used in the budgeting process as a tool to assess an actual or proposed policy reform, policy makers must have confidence in the simulated results. The fundamental prerequisites of such a credible microsimulation model are high level of precision in the simulated policies together with high degree of representativeness of underlying micro-data. To achieve that, the interaction of the user with the model might have to go beyond the functionalities offered by standardized tools.

This paper summarizes the lessons learned in the process of building a microsimulation tool tailored to country-specific conditions and involving a maximum degree of user control. The tool is characterized by fine attention to detail and increased accuracy in important categories. The latter has been achieved by application of a recently developed approach to sample weights calibration on the underlying dataset. It is shown that the new approach applied in general improves the fit between simulated output, underlying data and official statistics. Improved fit has been documented convincingly for the simulations of payroll taxes and majority of family related benefits. The SIMTASK (**SI**mulation **M**odel of **TA**xes and transfers in **S**lova**k**ia) model itself is used to evaluate the impact of legislative changes in the areas of taxes and benefits in Slovakia but the exercise contains valuable lessons for users of standardized tools such as EUROMOD wishing to obtain an enhanced degree of user control.

Whilst the model alone is designed to assess static effects of policy changes, the benefits of maximum user control through modelling in Stata become more apparent when incorporating the model into other labour supply (Siebertova *et al*., 2015) and general equilibrium models (Horvath *et al*., 2015). In those cases, it can be readily used for the evaluation of the long-run consequences of tax and benefit reform strategies too.

In SIMTASK, the emphasis has been put on details, namely on precise adaptation of valid legislation in order to achieve the highest precision in policy simulations. The modelling of benefits whose amount and duration is conditional on unobserved factors - such as the material needs, unemployment and maternity leave benefit is a particular strength and contribution of the work summarized in this paper.

For a microsimulation model to provide trustworthy assessment of budgetary and distributional effects of tax and transfer system reforms, it is crucial that the underlying data are representative with respect to the income distribution. As survey data rarely comply with this requirement, some input data corrections are shown to be beneficial. In addition, for the precise simulation of family related benefits, the correct representation of children in the corresponding age cohorts is essential. Therefore, we choose the approach of recalibrating the sample weights.

The importance of retaining information from the original weights is often stressed in the literature, there are, however, situations where re-weighting is required because the original weights supplied with the data do not adequately represent key analytical groups required for the analysis (O'Donoghue and Loughrey, 2014). This, we believe, is very relevant in our case. Moreover, there is no guarantee that sample weights calibrated to match demographic population totals produce appropriate revenue, expenditure and income distribution results (Creedy, 2004).

In our re-calibration, compared to original sample weights, the new weights allow for more detailed control of small children age categories and the earned income distribution is taken into account as a calibration factor directly. We show that this approach improves the match between simulated output and official statistics.

The paper is structured as follows. Section 2 describes the micro-level data used and the re-weighting method applied to sample weights of the underlying input dataset. Section 3 summarizes the tax and benefit system in Slovakia. Section 4 gives an overview of SIMTASK's development process and describes major differences compared to the existing Slovak EUROMOD modules. Section 5 presents validation and provides a discussion of the simulation results. We also report the implications of these methodological improvements for income distribution and inequality indicators. Section 6 concludes.

2. Data

A necessary precondition for the development of a microsimulation model is the existence of a suitable micro-dataset containing information preferably both on individuals and households. Usually, household survey data are used for these types of analyses; use of the administrative (or census) data is rather scarce.

The national version of the EU-SILC survey, abbreviated as SK-SILC, was selected as a base dataset for the tax-benefit microsimulations. Currently, it does best at meeting the data requirements for a microsimulation model of tax and transfer system when compared to other datasets that are available. In contrast to the EU-SILC, the SK-SILC dataset includes more variables that are country specific. The EU-SILC is an annual survey that has been conducted in Slovakia since 2004, it is collected by the Statistical Office of the Slovak Republic on behalf of EUROSTAT. Survey questions are focused on the income and living conditions of different types of households, as well as on the individual demographic characteristics, education, health status, employment, housing conditions and deprivation measures.

2.1. Re-weighting of the input SK-SILC dataset

Microsimulation tax-benefit system models are frequently used for the assessment of the effects of actual system reform policies as well as ex-ante simulations of reforms. Therefore, the input dataset should reflect the actual economic and social conditions as closely as possible. As survey data are available with a time lag (usually 2-3 years), the reference period of the input dataset and the baseline tax-benefit system might not refer to the same period. The approach frequently applied in the literature is to uprate the original market income by appropriate growth factors and to re-weight the sample to account for selected demographic and labour market changes. For a survey on current methodologies, see O'Donoghue and Loughrey (2014) and Figari *et al.* (2015).

The SK-SILC dataset is calibrated (such that sample weights are adjusted to match the known population totals in selected categories) and integrated weights (such that cross-sectional household weights and personal weights equal) are provided by the Statistical Office of the Slovak Republic. The calibration is an optimization procedure undertaken at two levels (household and individual) by using CALMAR2, a SAS macro developed by INSEE (LeGuennec and Sautory, 2002). By employing this macro, calibration uses 21 different categories in one strata and is performed on a number of household members (5 categories), 6 age categories divided by gender (together 12 categories) and 4 variables describing labour market status of a person (employees, unemployed, self-employed and pensioners). Stratification is based on NUTS3 level (8 regions). Condition of having integrated weights implies that simultaneous calibration should be applied, but there is no guarantee that the calibration process converges and the result is a kind of an approximate solution (as argued by Glaser-Opitzova *et al.*, 2014 an exact simultaneous solution has never been found).

As it is frequent in most survey data, SK-SILC dataset does not correctly represent the income distribution (both labour and non-labour) when compared to the official statistics. The most important component of individual labour income are gross earnings from employment (it constitutes more than 85% of aggregate income share in SK-SILC dataset and around 90% in the official data).

When the distribution of gross earnings from employment of individuals implied by SK-SILC (using original weights) is compared to the official statistics that can be retrieved from the official administrative data (in Slovakia it is a Social Security Agency (SSA) database), low-income groups and high-income groups are under-represented (the latter in fact missing) and incomes around the average wage substantially over-represented. Gross income from employment is the most important element that enters to the computation of the income tax base and its correct representation is highly important for the validity of any tax and transfer microsimulation model. Given this argument, it is beneficial to use the gross earned income as an additional calibration factor. Income as a calibration factor has been used in Slovenian EU-SILC, where calibration included also employee cash or near cash income (Inglic *et al.*, 2013). A similar idea has been described also by Creedy and Tuckwell (2003) who calibrated the weights of the New Zealand Household Economic Survey to take into account the number of recipients of several social transfers directly.

An improvement along these lines is provided by the calibration software "Calif" that has been recently developed by the Slovak Statistical Office. This tool allows to consider more categories in the calibration procedure due to the fact that it is able to find an approximate solution of the optimization problem, it works with several optimization methods, takes into account stratification and computes integrated weights. For the detailed description, mode of use and the documentation on Calif, see Vlacuha and Frankovic (2015). As an input to calibration procedure when using Calif, the number of individuals in the defined categories in the underlying SILC dataset and corresponding population totals obtained from administrative sources are needed. External administrative statistics on demographic categories, labour market status and household composition comes from the Statistical Office of the Slovak Republic, while income distribution is matched to the individual dataset of the SSA. Details are documented in Table 1.

Table 1 Sources of external controls for calibration

Calibration category	Source of the official statistics
Age cohorts (0, 1-3, 4-16, 17-25, 26-45, 46-retirement age, over retirement age). Categories over 16 years considered separately for gender (male, female)	Statistical Office of the Slovak Republic (Demographic Balances of Population)
Labour market status (employees, unemployed, self-employed)	Statistical Office of the Slovak Republic (Labour Force Survey)
Household size (members 1, 2, 3, 4, 5+)	Statistical Office of the Slovak Republic (Population and Housing Census 2011)
Gross income from employment distribution (deciles 1-2, 3-5, 6-8 and 9-10)	Database of individual records from Social Security Agency (SSA)

Majority of calibration categories that are used when computing original sample weights were also considered when re-calibration using Calif is performed. The difference between the two is in the definition of age cohorts and that the earned income has been directly taken into account as an additional calibration factor in the latter approach. When controlling for the labour market status, we use 3 categories (employees, unemployed and self-employed). We left out the category of pensioners (used in the original weighting scheme), since this is highly correlated with the corresponding old-age cohort, see also Kump and Navicke (2014).

In the calibration procedure, 7 age categories (0, 1-3, 4-16, 17-25, 26-45, 46-retirement age, over retirement age) were considered. Age cohorts over 16 years are matched to population totals such that also gender has been considered, i.e. separately for males and females. The idea of using extended age categories was to correctly represent newborn (age 0), small children (1-3) and youth (up to 25) population sub-groups that are essential controls for the simulation of family related benefits (child's birth grant, maternity benefit, child benefit and parental allowance).

Based on the earned income distribution of individuals identified in SILC dataset and income distribution of individuals that can be retrieved from administrative source SSA, extra categories for calibration were formed. To make the corresponding income categories comparable, in both underlying datasets the gross income from employment has been considered. For the calibration procedure two pieces of information were needed. First, decile points of income distribution given by data in SSA were computed. These decile points were used as threshold values also in SK-SILC dataset. In the second step, in every income decile (defined by decile points computed in SSA dataset) number of individuals were counted both in SSA and SK-SILC datasets and 4 additional calibration categories (in 8 strata) were constructed by grouping several deciles together (deciles 1-2, 3-5, 6-8 and 9-10).

The weights were re-calibrated separately for every strata, i.e. independently for 8 regions by using a linear bounded optimization method. When using the linear bounded method, the upper and lower bounds for the exit rates should be set. We decided to start initially with wider bounds (0.1 for the lower bound and 10 for the upper) and to gradually reduce them (similarly like Creedy, 2004 or Kump and Navicke, 2014). At the same time, we checked whether calibrated weights match the population totals and control for the standard deviation of weights (that should be low). Based on these three criteria, we discussed the estimated results and chose the set of calibrated weights. Consequently, newly estimated calibration weights correct the earned income distribution in a way that it sufficiently matches the official statistics (see Figure A1 in the Appendix).

The SK-SILC dataset corresponding to income reference period 2011 reports 15,440 individuals living in 5,291 households and SK-SILC referring to 2012 contains 15,426 individuals in 5,402 households. Table 2 presents descriptive statistics of the grossing-up weights and population estimates of the samples weighted by original weights and using weights computed with a new calibration tool. In addition, in Table A1 in the Appendix we present the descriptive statistics of main demographic and income related variables.

Table 2 Descriptive statistics of grossing-up weights in SK-SILC samples

Policy year	**2011**	**2011**	**2012**	**2012**	**2013**	**2014**
Underlying SILC dataset	2011	2011	2012	2012	2012	2012
Grossing-up sample weight	Original	Calibrated	Original	Calibrated	Original	Calibrated
Mean	349.5	350.0	350.4	350.0	351.2	351.8
Std. Dev.	126.0	468.7	132.0	482.5	521.3	581.1
Minimum	108.7	10.9	119.9	12.0	12.0	12.0
Maximum	1,226.1	5,505.1	1,083.9	4,481.0	5,974.1	8,961.1
Dataset characteristics						
Individuals	15,440	15,440	15,426	15,426	15,426	15,426
Households	5,291	5,291	5,402	5,402	5,402	5,402
Projected population	5,395,519	5,396,355	5,404,664	5,398,917	5,417,340	5,427,220
Projected households	1,911,664	1,909,248	1,852,059	1,852,027	1,852,059	1,853,409

Source: Author’s calculations using SK-SILC.

In order to test the predictive accuracy of the SIMTASK when the income reference period of the underlying input dataset and the simulated tax and transfers system refer to different time periods, we applied a two-step nowcasting method. As an underlying dataset we used the latest SILC survey available to us at the time of writing with the income reference period 2012. In the first step, we uprated income variables (including all labour and non-labour income variables listed in Table A1 in the Appendix) in the dataset by the corresponding growth factors. In the next step, we applied a variant of static ageing technique and re-weighted the input dataset to account for changed population structure (both demographic and labour-market status). Calibration of weights has been performed by comparing the data from the uprated dataset to the external statistics from the target policy year. In the last two columns of Table 2 we present the new calibration weights that were estimated and that are later used in simulating taxes and transfers to test for the accuracy of SIMTASK in policy years 2013 and 2014.

2.2. SK-SILC versus official statistics

The SK-SILC dataset is largely representative of the country population. However, as it is frequent in survey data, SK-SILC might also over-represent or under-represent certain population groups. Particular limitations are inspected in details below, in such a way that SK-SILC data are compared to the appropriate official statistics using both the original and calibrated weights. Tables displayed below suggest that in most aspects the newly calibrated weights helped to improve the fit. These comparisons are also highly instructive in later assessment of simulations.

Table 3 presents the ratios of the number of individuals in the selected age cohort in the input SK-SILC database to external benchmark. While 2011 and 2012 SK-SILC datasets weighted with original weights underestimate the number of new-born (age 0) and small children (under 3 years), using the calibrated weights where we directly control for the number of children in certain age groups leads to almost perfect fit. For the prime age and retirement age cohorts, datasets using calibrated weights match demographic statistics closely both in 2011 and 2012.

Table 3 Validation of weighting schemes: representation of gender and age cohorts

	Original weights		**Calibrated weights**	
	2011	**2012**	**2011**	**2012**
Female	1.00	1.00	1.00	1.00
Age cohort				
0	0.55	0.78	1.00	1.04
0-3	0.63	0.84	1.00	1.01
0-16	0.82	0.96	1.00	1.00
0-26	0.98	1.04	1.00	1.00
Prime age	1.03	1.01	1.00	1.00
Retirement age	1.08	1.04	1.02	1.01

Source: Authors' calculations using SK-SILC and Slovak Statistical Office

Note: Ratios display number of individuals in SK-SILC dataset (weighted) to population by gender and in the respective age cohort.

Prime age: 15-64 years. Retirement age: males 62+, females 58+ in 2011 and 59+ in 2012.

Data on representation of the economic activity of Slovak population is shown in Table 4. The reported ratios document that based on these criteria, SK-SILC dataset reflects the official statistics very well, the only exception being the group of employees. Comparing the two weighting schemes, the number of employees is originally significantly oversampled, but when calibrated weights are used the number of employed gets well closer to the official statistics (in both years).

Table 4 Validation of weighting schemes: representation of economic activity

	Original weights		**Calibrated weights**	
	2011	**2012**	**2011**	**2012**
Employed	1.01	0.97	0.96	0.96
Employee	1.20	1.15	1.08	1.08
Self-employed	1.01	0.96	1.02	1.01
Unemployed	0.97	0.94	0.99	0.99
Economic active pop.	1.01	0.97	0.97	0.97
Economic inactive pop.	0.99	1.03	1.04	1.04
Population total	1.00	1.00	1.00	1.00

Source: Authors' calculations using SK-SILC and LFS.

Note: Ratios display number of individuals in SK-SILC dataset (weighted) to LFS in the respective category.

In Table 5, the different sources of income reported in SK-SILC are related to the official statistics given by SSA. A comparison is provided with respect to the ratios of reported aggregate amounts of income as well as in terms of the ratios of the number of individuals receiving certain type of income.

Table 5 Validation of weighting schemes: representation of income

	Amounts (ratios)				**Individuals (ratios)**			
Weights	**Original**		**Calibrated**		**Original**		**Calibrated**	
Income from	2011	2012	2011	2012	2011	2012	2011	2012
Employment	1.16	1.02	0.97	0.95	1.08	1.05	0.93	0.95
Agreements	0.43	0.34	0.43	0.34	0.36	0.38	0.33	0.36
Self-employment *	2.84	2.73	2.98	2.78	1.34	1.25	1.36	1.31
Employment and agreements	1.13	0.99	0.94	0.91	0.97	0.94	0.97	0.94

Source: Authors' calculations using SK-SILC and SSA.

* Validation of income for self-employed is only indicative. SK-SILC reports for self-employed the value of profit/loss in the current year, while the SSA database reports the assessment base which is based on the value of return in the year t-2 (inconsistency both in variable and time).

Note: Amounts (ratios) display aggregate amount of income of individuals in SK-SILC dataset (weighted) to aggregate income computed by using records from SSA. Individuals (ratios) display aggregate number of individuals in SK-SILC dataset (weighted) to aggregate computed by using records from SSA.

The overall picture does not differ in 2011 and 2012; the number of people reporting an income from employment is only slightly undersampled and matches relatively well with the administrative data from SSA. Those declaring an income from agreements (temporary employment contracts) are significantly under-represented and this applies to both original and calibrated weights. On the other hand, the number of self-employed individuals compared to SSA statistics is substantially oversampled. It should be noted that comparing the number of self-employed to the statistics of SSA is not completely correct. SSA database is primarily a dataset of paid social insurance contributions providing information on gross income. In the case of self-employed persons, SSA dataset captures only those individuals who pay SIC (social insurance contributions) which is a subset of the total number of registered self-employees. However, it is instructive to show these ratios, since our simulations of taxes and social security contributions are validated against the statistics provided by SSA.

Aggregate income from employment approximately matches the aggregate amount documented by SSA, while the income from agreements is substantially underreported. Since the volume of agreements makes approximately only 5%

when compared to the income from employment, the total effect of employment and agreements matches SSA dataset well.

Note that aggregate income from self-employment should be validated with caution and the results proposing substantial over-reporting in the input data are only indicative. The reason is that SK-SILC reports for the self-employed the value of profit/loss in the income reference period, while the SSA database reports the legislatively correct assessment base which is based on the value of declared return in the year t-2 (i.e. there is an inconsistency both in variables that are equated and time aspect). However, relative weight of self-employed in the labour market is rather low, as they constitute only 7% of the total population.

The main non-simulated benefits and pensions, which serve as an input to later simulations, are inspected in Table 6. Maternity benefit recipients are substantially undersampled when the original weighting scheme has been applied. Using the calibrated weights makes the number of recipients to match well in 2011, but overestimate in 2012. Since the eligibility for the maternity benefit is up to approximately 7 months after the child's birth, the reported ratios match with the good fit of the youngest age cohort of new-born children in SK-SILC in 2011 and small oversampling in 2012 as it is documented in Table 3.

Table 6 Non-simulated benefits and pensions

	Amounts (ratios)				**Recipients (ratios)**			
Weights	**Original**		**Calibrated**		**Original**		**Calibrated**	
Benefits	**2011**	**2012**	**2011**	**2012**	**2011**	**2012**	**2011**	**2012**
Maternity	0.63	0.60	1.22	1.27	0.57	0.61	1.01	1.28
Sickness	0.31	0.34	0.50	0.59	n.a.	n.a.	n.a.	n.a.
Pensions								
Old age	1.16	1.09	1.09	1.08	1.10	1.04	1.04	1.03
Disability	0.87	0.88	1.03	1.10	0.84	0.86	1.03	1.10
Widow/er	0.98	1.05	0.97	1.06	1.06	0.99	1.01	0.97
orphans	0.83	0.82	0.93	0.57	0.69	0.68	0.79	0.52

Source: Authors' calculations using SK-SILC and SSA.

On the other hand, the demographic group of elderly is represented well in both input samples. This subsequently mirrors in the share of old-age pension beneficiaries close to one. Orphans are undersampled in the input data when both weighting schemes are used, while disability pensioners are slightly underestimated/overestimated when original/calibrated weights are applied. Widows and widowers well approximate the figure addressed by SSA.

Table 6 summarizes also the information on the aggregate amounts of paid benefits and pensions: data in input datasets are compared to the external statistics recorded by SSA. Not surprisingly, old-age pension payments are slightly overestimated, but match relatively well. Other non-simulated benefit and pension payments are in general underestimated when original weighting has been used. The gap between official records and input data is extreme in the case of sickness benefits, where aggregate payments reported in SK-SILC reached around 30% of the official statistics using the original weighting scheme. The gap has been slightly reduced with the calibrated weights to around 50%. Maternity benefit payments represent around 60% of the official SSA records with the original weights, while using the calibrated weights leads to overestimation. Both these ratios are in line with the number of recipients reported above.

3. The tax and benefit system in slovakia

3.1. Taxes and social insurance contributions

The Slovak tax system is largely unified; all important components are set at the state level. Taxation of income is conducted at an individual level and it is levied on gross income including wages, income from business activities, fringe benefits, capital incomes (dividends excluded), interest and rental income. Joint taxation of married couples is not possible. Social and health insurance contributions and social benefits are exempt from the tax base, i.e. the tax base is given as gross earnings net of employee social and health insurance contributions.

All relevant parameters needed to compute personal income tax (PIT) are available in the SK-SILC data - both those which are related to individual and household level. During the years 2009 to 2012 PIT amounts to a 19% flat tax rate with a non-taxable allowance. From 2013, two tax brackets were introduced and incomes exceeding the threshold are taxed by 25% rate.

Tax expenditures that are deducted from the tax liability in the PIT and that are incorporated in SIMTASK include:

a. Basic tax allowance: tax allowance each individual can apply, the amount of the allowance is based on the legally defined minimum subsistence level. A progressive reduction in basic tax allowance is applied when annual gross earnings exceed about 18,000 euros (approximately twice the Slovak average yearly gross wage) and it influences around top 10% of tax payers.
b. Spouse tax allowance: an individual may be entitled to a spouse tax allowance if the income of spouse satisfies certain conditions (earnings under a certain level).

c. Employee tax credit (ETC): the amount depends on employee's income and on the period he has been working (at least 6 months). It is targeted at low-income groups who have to pay health and social insurance contributions.
d. Child tax credit: one spouse may claim an allowance for each child in the household if the child satisfies certain conditions (e.g., aged under 18 or aged under 26 and in full time education or aged under 26 when physically or mentally disabled and not receiving disability pension). This tax credit can be received, if the parent annually earns at least 6 times the minimum wage. If the credit exceeds the tax liability, the excess is paid to the taxpayer.

The Slovak social insurance system is made up of two components; namely social insurance contributions and health insurance contributions. The assessment base for contributions is narrower compared to the PIT base since capital income is not considered. Up to 2012 maximum assessment base differed based on the type of insurance and employment contract. Effective from 2013, assessment bases for social and health insurance contributions of employees were unified. For the self-employed, the computation of the assessment base was redefined.

a. Social insurance contributions (SIC)
   Both employers and employees pay an unemployment, sickness, disability, and an old age insurance, but different percentages from the social insurance assessment base. In addition, employers also pay contributions to a reserve solidarity fund, accident insurance and guarantee insurance. The self-employed are treated differently; they pay sickness, disability and old age insurance and contributions to the reserve solidarity fund.
b. Health insurance contributions (HIC)
   These contributions are paid by employers, employees and self-employed. The percentage to be paid is different for the three categories of payers.

3.2. The social system

The Slovak benefit system consists of three components, termed as contributory, social assistance and poverty, and state social support.

a. Contributory benefits include old-age pension, early old-age pension, disability pension, widow's and widower's pension, orphan's pension, sickness cash benefit, benefit for nursing a sick relative, equalization allowance, maternity benefit, and unemployment insurance benefit.
b. Social assistance program covers material need benefit.

c. State social support includes several programs, namely child birth grant, additional birth grant, multiple birth benefit, child benefit, additional child benefit, parental allowance, funeral benefit, scholarships for pupils in elementary school, scholarships for students in secondary school, and social scholarships for university students.

4. Tax benefit simulations

4.1. Existing models

EUROMOD has been the only model available for the Slovak tax-benefit system microsimulations, which could be used equally by government agencies and the academic community. It is an EU-wide tax-benefit microsimulation model that can simulate individual and household tax liabilities and benefit entitlements according to policy rules valid in the respective EU states. EUROMOD is developed and maintained by the Institute for Social and Economic Research at the University of Essex, in collaboration with national teams. For its current state and details of the project, see Sutherland and Figari (2013). EUROMOD for Slovakia is well documented in the EUROMOD Country Report, for a detailed overview of application rules and payable eligibility, see Porubsky *et al.* (2013) or Strizencova and Hagara (2014). In this analysis, EUROMOD version G2.0+ is used.

The Slovak EUROMOD runs on SK-SILC data and the simulated policies currently include personal income tax, all health and social insurance contributions paid by employers, employees and self-employed. Benefits that are fully simulated include family related programs, namely child birth grant, child benefit including additional child benefit and parental allowance. Means-tested material needs benefit and contributory unemployment insurance benefit are simulated partially under simplifying assumptions. Simulations of other benefits, which may impact both individual and household incomes, are not included due to the lack of information on previous employment and contribution history. In particular, these include sickness benefits and disability pensions. Old-age pensions are not simulated since there is no information on contribution period.

4.2. SIMTASK

A new microsimulation model SIMTASK has been developed such that the set-up of the EUROMOD model has been taken as a template and an independent program that runs in software Stata has been created. It is important to stress that a primary intention has not been to replace the existing Slovak EUROMOD, which is a simple and transparent static tax-benefit calculator with the advantage of cross country comparability and that can be also linked to other models. Rather, the objective has been to expand its use and to tailor it directly to the principal demand of having a simulation tool that can be easily incorporated into other models to provide an accurate-enough evaluation of measures for the process of budgeting and fiscal forecasting. Besides the considerations about the type of microsimulation model that was needed in terms of its capability to include behavioural responses, the mode of operation, i.e. how easy it is to incorporate and handle with it in such a model setup, where the convergence could be achieved only after numerous iterations, has been an issue too.

All tax and benefit instruments in the SIMTASK model are simulated in the same order as in Slovak component of EUROMOD (further referred as a baseline model). In addition, SIMTASK includes the simulation of the length of the eligibility period to a maternity benefit and a substantial extension of simulation of material needs benefit.

In the baseline model setup all benefit instruments are simulated on a yearly basis. Based on predefined eligibility requirements, it is tested if an individual is entitled to receive a certain benefit. An assignment is provided if the predefined conditions are met and subsequently the corresponding amount is simulated. For example, conditional eligibility to an unemployment benefit is checked (among other conditions, an individual should not receive parental allowance) and parental allowance is simulated prior to unemployment benefit. In other words, subsequent entitlement to certain transfers is ruled by the order of simulation policies. However, this procedure does not take into account possible variability that can occur during the whole period of one year – such that an individual might be eligible for several transfers that are available to him/her subsequently, if these transfers are paid for shorter period than one year.

In order to allow an individual to receive different benefits during the annual period in SIMTASK, eligibility to selected transfers is simulated on a monthly basis depending on the predefined requirements. This approach could be applied thanks to the fact that information on month of birth of an individual is recorded in SK-SILC dataset. The simulations of benefits for shorter periods than one year are already available for some countries within EUROMOD for example in Estonia (Vork and Paulus, 2014) and it would be possible to implement this approach also in Slovak EUROMOD. Consequently, knowing the month of the year when a child was born, it is possible to accurately allocate family related benefits. This applies particularly to family related and unemployment benefits, which are simulated in the following order:

- maternity benefit: the length of the eligibility period is simulated, which is 8 months (or 10 months in case of multiple births, or 9 months for lonely parent). The amount of benefit is presently not simulated because of lack of information on contribution history to health insurance.
- parental allowance: the length of the eligibility period is simulated, entitlement ends when the child reaches 3 years of age. Entitlement is possible up to 6 years in case of child's unfavourable health condition, but there is a lack of information to simulate this. The amount needs not to be simulated - it is a fix payment.
- unemployment benefit: the length of the eligibility period is simulated, maximum is 6 months.

Minor modifications of tax-benefit system simulations used in SIMTASK (as compared to Slovak EUROMOD) are detailed in Siebertova *et al*. (2014). Two major modifications were implemented and these apply to the simulation of material needs benefit and unemployment benefit.

The material needs benefit (MNB) is a means tested transfer that is intended for families with income below the minimum subsistence level. The actual benefit amount is calculated as a difference between the eligible maximum of MNB - composed of social benefit, health care allowance, housing allowance, activation and protection allowance - and the income of individuals living in a household. In our simulation, we include a more precise specification of the assessed income computation (compared to baseline). Social benefit and health care allowance are set as fixed amounts, these are not simulated. Furthermore, we include a different computation of the protection allowance: in our implementation, it is based on the set of predefined eligibility conditions. The essential is the change in the definition of an individual allocation to the activation allowance. In the baseline model, activation allowance is assigned to all those, who are not eligible to receive protection allowance. However, this approach is not based on a valid legislation and as a result, it largely overestimates the assignment of the activation allowance. On the contrary, in our approach we define a set of eligibility conditions that an individual needs to fulfil in order to be entitled to draw this allowance. This gives us a set of people who potentially might take part in activation works. In the next step, we randomly draw from this predefined group a subset of individuals (who will be finally assigned to activation works participation), such that the ratio of those who participate in activation works to the total number of those who receive MNB equals, when compared to the official statistics. In 2014 this "random draw" procedure is applied also to the basic allowance due to legislation changes.

The unemployment insurance benefit is a contributory transfer aimed to compensate temporarily for the income loss due to unemployment. In our detailed adaptation (as compared to the current version of Slovak EUROMOD) we provide a more precise simulation of eligibility period on a monthly basis, this is possible also thanks to the more precise simulation of the length of the maternity benefit.

5. Model and validation of simulation results

Validation of model outputs, i.e. comparison of computed results with reality, is a useful approach to test the overall relevance and weak points of the microsimulation model. There are several possible approaches how to validate results produced by a microsimulation model. We adopt an approach frequently used in the academic literature, where baseline systems are validated and tested at aggregate macro level such that simulated outputs are compared to the external official statistics. In this section we also show the validation of the predictive performance of SIMTASK. Finally, we provide an overview how the whole income distribution, inequality and poverty indices are affected by the new weights and refined simulation.

5.1. Aggregate validation

In this section we demonstrate that refinement of simulations as well as reweighting of the input dataset leads to improved validity of aggregate results. We show that different approach to the simulation of the material needs benefit (compared to EUROMOD) significantly improves accuracy of the results. Increased precision of the simulation of parental allowance, which is the most important transfer paid to families (in the sense of total volume of payments) is gained by using calibrated weighting scheme and SIMTASK that takes into account duration of benefit take-up less than one year. The same argument applies to the simulation of personal income and payroll taxes where the improved accuracy has been achieved mainly due to application of the calibrated weighting scheme.

Total expenditures and the number of beneficiaries of those transfers that are not simulated, but act as inputs to SIMTASK model, are compared to the official statistics in section 2.2 above. In the next step we look in detail at transfers that are simulated both by EUROMOD and SIMTASK and compare the simulation results to the official statistics in 2011 and 2012. To make simulation results comparable, we use the same underlying SK-SILC datasets when running EUROMOD and SIMTASK.

When validating results with respect to the total number of people, a concept of "unique occurrence" has been used. This applies to the aggregate number of benefit recipients, tax payers, unemployed, employed, self-employed or persons with agreement contracts. By construction, the SK-SILC dataset should include every person receiving a given benefit, paying taxes or having an employment contract during the income reference period. Therefore, the statistics on "unique occurrence" should better correspond to the reality that is reflected in SK-SILC than the average monthly number, which is the statistics usually reported by administrative sources.

The choice of an appropriate external statistics has been considered also regarding the aggregate validation of estimates of tax and different contributions revenues. The official statistics on PIT, SIC and HIC revenues published by the Ministry of Finance mirrors the payments received during the income reference period, which might be distorted by the sum of unpaid contributions. Therefore, PIT, SIC and HIC revenues are calculated directly using the administrative Social Security Agency database that contains individual records of payments on monthly basis. Note that this corresponds better to simulated aggregates by SIMTASK that represent liabilities that should be paid, rather than actually received payments.

Finally, we provide a simulation exercise where the predictive ability of SIMTASK is tested. Based on 2012 input data we simulate tax and transfer systems valid in 2013 and 2014 and verify simulation results against the official statistics. A summary on the aggregate validation of the main simulated benefits from EUROMOD and SIMTASK with original and calibrated weighting schemes against the external official statistics is depicted in Table 7. Comparing the results in columns "original" and "calibrated" shows the disparities that arise due to different weighting schemes used. On the other hand, comparing "EUROMOD" and "SIMTASK" (with the same weighting) document distinctions that appear due to refinements in simulations.

Table 7 Simulated benefits: Ratios of aggregate amounts

Model	**EUROMOD**		**SIMTASK**	
Weights	**Original**	**Calibrated**	**Original**	**Calibrated**
	2011			
Unemp.benefit	0.54	0.46	0.55	0.47
Parental allowance	0.78	1.18	0.64	0.99
Child benefit	0.87	1.00	1.08	1.16
Child birth grant	0.61	1.10	0.61	1.10
Material needs benefit	1.24	1.82	0.83	1.19
	2012			
Unemp.benefit	0.38	0.37	0.40	0.38
Parental allowance	1.09	1.07	0.87	0.96
Child benefit	0.97	0.98	1.17	1.14
Child birth grant	0.80	1.08	0.81	1.08
Material needs benefit	1.46	1.92	1.00	1.27

Source: Authors' calculations using SIMTASK and EUROMOD, official statistics SSA (unemployment benefit), other benefits COLSAF (Central Office of Labour, Social Affairs and Family).

Note: Numbers display ratios of aggregate number of payments to individuals computed by EUROMOD and SIMTASK to aggregate amounts referred by official statistics.

Table 7 shows that simulation results are substantially improved by using the calibrated weights. Aggregate validation of total amounts of family related benefits, namely parental allowance and child birth grant, shows that simulations using original weighting schemes are underestimated in EUROMOD and SIMTASK model when compared with the official statistics in 2011. Using the calibrated weights made the corresponding ratios get closer to one. The reported underestimation of these transfers directly mirrors undersampling of new-born and small children in SK-SILC with original weights. Since calibrated weights directly control for the correct number of children, this has led to the improved validation results.

The aggregate amount of child benefit payments is overestimated in SIMTASK. Using original weights payments are overestimated by 8% in 2011 and 17% in 2012, when using calibrated scheme overestimation is 16% in 2011 and 14% in 2012. This imprecision arises due to the broader definition of eligibility condition that is applied. According to valid legislation, also parents of university students (up to 26 years of age) studying in an internal form are eligible to receive child benefit. It is not possible to distinguish between internal and external form of university study in the input dataset. When we adjusted the simulated output and took into account that internal form of study applies to around 70% of university students, the resulting numbers approached the official statistics closely. In the EUROMOD simulation of child benefit, all university students irrespective of form of study are excluded from the eligibility condition.

Validation results for the material needs benefit differ substantially based on the weighting scheme that has been used. When using the original weighting, the income distribution has not been taken into account and low-income earners in input datasets were under-sampled. This translated into undersampling of the number of recipients of MNB, leading to ratio 73% and 81% of the official statistics in 2011 and 2012, respectively (see Table A2 in the Appendix). Overall, this resulted into underweighting of the aggregate amount of this benefit in 2011 (83% of the official statistics). Using the calibrated weights, more weight has been placed on low-income earners who are also the most likely material needs benefit recipients, and finally this led to overestimation of this transfer in terms of the amount of benefits received (19% and 27% in 2011 and 2012, respectively). These results are in line with the evidence documented in the empirical literature suggesting considerable non-take-up of means tested benefits (Wiemers, 2015 or Matsaganis *et al*., 2008). In Slovak EUROMOD, MNB transfer is simulated differently compared to SIMTASK (see section 4.2.1). Overestimation of several components of MNB leads to even more pronounced overestimation of both the total payments and the number of recipients.

In the simulation of the unemployment benefit, only the length of the eligibility period is simulated but not its allocation to recipients. However, the number of recipients declared in the input dataset is only around 45% of the official statistics (both in 2011 and 2012). Therefore, the total number of recipients as well as aggregate amount of payments of unemployment benefit is substantially underestimated when both weighting schemes are applied.

Aggregate validation of total number of recipients of main simulated benefits leads to comparable results as those presented in previous paragraphs. Detailed results can be found in Table A2 in the Appendix.

The aggregate sum of payroll taxes compared to the official statistics is more precise when using calibrated weights and SIMTASK. Detailed output related to personal income tax and social (SIC) and health (HIC) insurance instruments is depicted in Table 8.

The aggregate sum of tax liabilities (including tax credits and tax allowances) shows almost perfect fit to the official statistics (1.03 and 0.95 in 2011 and 2012, respectively) when using SIMTASK with calibrated weights. A difference in the validation of simulations of SIC for employees and employers and HIC can be observed when the two weighting schemes are compared - using the calibrated weights leads again to the almost perfect fit. SIC paid by self-employed should be interpreted differently and results documented here are only indicative. The reason is an inconsistency in variables that are equated; profit/loss of self-employed reported in SK-SILC versus the assessment base for SIC in the official SSA database that is based on the performance two years prior to the income reference period.

Table 8 Personal income tax and social insurance contributions: Aggregate amounts

Model	**EUROMOD**		**SIMTASK**	
Weights	**Original**	**Calibrated**	**Original**	**Calibrated**
	2011			
Personal income tax	1.17	1.05	1.14	1.03
Social insurance contrib. (SIC)				
SIC: Employer	1.24	1.03	1.23	1.02
SIC: Employee	1.25	1.04	1.25	1.04
SIC: Self-employed	1.72	1.77	1.44	1.50
Health Insurance contrib. (HIC)				
HIC: economic active pop	1.19	1.01	1.18	1.00
HIC: economic inactive pop.	1.02	1.09	0.94	1.02
	2012			
Personal income tax	0.94	0.97	0.91	0.95
Social insurance contrib. (SIC)				
SIC: Employer	1.09	1.01	1.09	1.01
SIC: Employee	1.11	1.02	1.11	1.02
SIC: Self-employed	1.63	1.67	1.38	1.41
Health Insurance contrib. (HIC)				
HIC: economic active pop.	1.07	1.01	1.06	0.99
HIC: economic inactive pop.	1.03	1.06	0.96	0.99

Source: Authors' calculations using SIMTASK and EUROMOD, official statistics Ministry of Finance (PIT and HIC), SSA (SIC).

Note: Numbers display ratios of aggregate number of recipients computed by EUROMOD and SIMTASK to aggregate number of recipients referred by official statistics.

5.2. Validation of the predictive accuracy

SIMTASK is designed so that it can be used also for ex-ante evaluation of the proposed legislative reforms of Slovak tax and social system. In order to test for the predictive accuracy of SIMTASK we have performed a simulation exercise. We show that simulation results match the official statistics adequately and the observed discrepancies are qualitatively not different from those reported in our previous ex-post simulations.

As it has been already outlined above, we proceed in two steps. First, selected income variables in the input SK-SILC dataset (income reference year 2012) were uprated with the corresponding growth factors to refer to 2013 and 2014, respectively. In the next step, new weights in the uprated datasets were calibrated to match the population totals in 2013 and 2014 using the selected socio-demographic groups, groups defined based on economic activity and labour income distribution defined in terms of calibration factors.

Table 9 Simulated benefits (calibrated weights)

	Aggregate amounts (I)				
	2013	**2014**		**2013**	**2014**
Unemp. benefit	0.41	0.48	Child birth grant	1.03	1.21
Parental allowance	0.94	0.92	Material needs benefit	1.32	1.15
Child benefit	1.15	1.15			
	Recipients (II)				
	2013	**2014**		**2013**	**2014**
Unemp. benefit	0.48	0.54	Material needs benefit	1.07	1.05
Parental allowance	0.97	0.97	Housing allowance	1.10	1.03
Child benefit	1.08	1.09	Activation allowance	1.59	1.31
Child birth grant	0.98	1.21	Protection allowance	1.07	0.86

Source: Authors' calculations using SIMTASK, official stat. SSA (unemployment benefit), COLSAF (other benefits).

Note: Numbers display ratios of aggregate amount of payments to individuals (I) and number of recipients (II) computed by SIMTASK to official statistics.

Aggregate validations of simulation of transfers, tax and social security contributions are summarized in Table 9 and Table 10. Overall picture is comparable to validation statistics of simulations for 2012 when calibrated weights have been used. This is not a surprise since the same underlying input dataset has been used, although weights were calibrated differently using the updated external statistics. To sum up, observed departures from the official statistics (either under- or over-sampling) are similar both in direction and magnitude to those reported for 2012.

Table 10 Personal income tax, social and health insurance contributions (calibrated weights)

	2013	**2014**		**2013**	**2014**
Personal income tax	1.00	0.99	SIC employer	0.99	1.00
HIC: economic active pop.	1.04	1.06	SIC employee	1.00	1.00
HIC: economic inactive pop.	0.96	1.02	SIC self-employed	1.42	1.71

Source: Authors' calculations using SIMTASK, official stat. Ministry of Finance (PIT an HIC), SSA (SIC).

Note: Numbers display ratios of aggregate amount of payments computed by SIMTASK to aggregate amounts computed by using official statistics. SIC stands for social insurance contribution and HIC for health insurance contributions.

5.3. Impact on income distribution, inequality and poverty measures

In the following part we present a comparison of indicators of income distribution, inequality and poverty reported by Eurostat and estimated by EUROMOD and SIMTASK. Results published by Eurostat are reported for reference, as they are not directly comparable to the estimates by EUROMOD and SIMTASK. Several reasons may explain differences between computed results. In particular, although Eurostat results are also based on SILC data and use the original weighting scheme, equalised households' disposable income definition includes different components compared to definitions used by EUROMOD and SIMTASK (that are comparable). On top of that, some income sources that enter to the computation of disposable income may have different values due to the fact that they are simulated in EUROMOD and SIMTASK.

We can, however, provide a meaningful comparison of inequality measures calculated on data generated by different simulation tools. This should give the reader a flavour of the impact of the weights calibration and of the closer match of the model with legislation on simulated inequality measures. Indicators in EUROMOD and SIMTASK were computed using the same methodology like Eurostat. In particular, results were calculated on the basis of the total equalised disposable income attributed equally to each member of the household. Disposable income is defined as a sum of all monetary income sources of all household members net of paid payroll taxes. Household members are equalised by weighting using the modified OECD equivalence scale that assigns a value 1 to the household head, 0.5 to each additional adult member and 0.3 to each child under 14.

The distribution of equalised disposable income by deciles is reported in Table 11. Results show that shares of disposable income are very similar when estimated by EUROMOD or by SIMTASK and when using the same weighting scheme. Differences can be noticed when outcomes using original and calibrated weights are contrasted. To sum up, differences between results from Eurostat, EUROMOD and SIMTASK are small, whereas estimates based on calibrated weighting scheme appear to be closer to Eurostat results.

Table 11 Shares of equalised disposable income by deciles

	EUROSTAT	**EUROMOD**		**SIMTASK**	
		Original	**Calibrated**	**Original**	**Calibrated**
	2011				
Decile 1	3.5	3.9	3.7	3.8	3.5
Decile 2	5.7	6.1	5.7	6.0	5.6
Decile 3	6.9	7.5	7.1	7.1	6.8
Decile 4	7.9	8.2	8.1	8.0	7.7
Decile 5	8.6	8.9	8.8	8.9	8.6
Decile 6	9.5	9.8	9.4	9.8	9.5
Decile 7	10.7	10.6	10.5	10.9	10.7
Decile 8	12.1	11.9	11.8	12.1	12.1
Decile 9	14.1	14.0	14.4	14.0	14.6
Decile 10	21.0	19.1	20.5	19.3	20.8
	2012				
Decile 1	3.6	3.9	3.5	3.9	4.0
Decile 2	5.7	6.2	5.9	6.1	5.4
Decile 3	6.8	7.4	7.1	7.3	7.0
Decile 4	7.7	8.2	8.0	8.2	7.9
Decile 5	8.7	9.0	8.7	9.0	8.7
Decile 6	9.7	9.8	9.6	9.9	9.7
Decile 7	10.9	10.8	10.7	10.9	10.8
Decile 8	12.3	11.9	12.0	12.1	12.3
Decile 9	14.3	13.9	14.3	14.0	14.4
Decile 10	20.3	18.8	20.1	18.7	19.9

Source: EUROSTAT and authors' calculations using SIMTASK and EUROMOD.

Note: The ratio of disposable income in the corresponding decile to the population. Computed for individuals based on household disposable income and equalised by the modified OECD equivalence scale.

In Table 12, some income inequality and poverty measures are presented. Differences in estimated indices by EUROMOD and SIMTASK that can be attributed to refined simulations are small (when EUROMOD and SIMTASK with the same weights are compared). Disparities are larger when the results based on two weighting schemes are compared. In accordance with disposable income distribution presented in Table 11, when calibrated weights are used, results are closer to official figures reported by Eurostat.

Table 12 Income inequality and poverty rates

	EUROSTAT	**EUROMOD**		**SIMTASK**	
Weights		**Original**	**Cali-brated**	**Original**	**Cali-brated**
	2011				
GINI	25.7	23.3	25.6	23.5	26.0
S80S20 ratio *	3.8	3.3	3.7	3.4	3.9
At risk of poverty rate **					
Total popul.	13.0	12.1	13.4	12.1	13.5
Females	12.8	12.0	13.5	12.0	13.6
Males	13.1	12.3	13.3	12.2	13.3
	2012				
GINI	25.3	22.4	24.4	22.6	24.7
S80S20 ratio *	3.7	3.2	3.6	3.3	3.6
At risk of poverty rate **					
Total popul.	13.2	11.5	12.9	11.7	13.1
Females	13.2	11.3	12.4	11.5	12.8
Males	13.3	11.6	13.3	11.9	13.4

Source: EUROSTAT and authors' calculations using SIMTASK and EUROMOD.

* The ratio of total income received by the top 20 % of population to that received by the bottom 20%.
** Percentage of population below 60% of median equalised income.

Measures are computed for individuals based on disposable household income and equalised by the modified OECD equivalence scale.

6. Conclusion

This paper provides a summary on the construction of the Slovak tax and transfers microsimulation model SIMTASK. This model has been built up due to the Slovak Council for Budget Responsibility's (CBR) need to have a model being able to assess the static effects of policy changes as well as the long-run consequences of tax and benefit reform strategies. Therefore, a microsimulation model, which works in the common environment and thus can be easily incorporated as a part of more complex models used within CBR was developed.

A number of challenges were addressed during the process of development. First, we considered issues related to the simulation of social structures themselves, i.e. we identified possible improvements (compared to Slovak component of EUROMOD) such that the national tax and benefit system can be replicated as closely as possible. At this point, a major task was to precisely replicate the valid legislation in the corresponding years. At the same time, we inspected the used micro dataset in great detail and we compared it with appropriate administrative statistics. We re-weighted the input data sample such that the new calibrated weights replicate, among other factors, also the earned income distribution and selected age cohorts directly. Hence, the validity of simulated output was interpreted further in light of differences between simulations using original and new weighted survey data on one side and the official statistics on the other side.

We conclude that weight calibration considerably improves the fit of the model with respect to important income tax and social security contributions categories. However, some distortions when using calibrated weights result too. These involve mainly non-simulated transfers with a lower number of recipients (and in lower total volumes) such as benefits for orphans, disabled or maternity benefits. Weights calibration helps SIMTASK to become a more convincing tool to simulate and evaluate ex-post and ex-ante the impact of selected tax and transfer system policies. However, we showed that re-weighting is not a panacea and the focus of further analysis that user is interested in is of relevance.

Acknowledgement

We would like to thank the editor of this journal and two anonymous referees for their helpful comments. We are also grateful to Michal Horvath from the University of York and to our colleague Matus Senaj for their help and support.

References

Callan T, Keane C, Walsh J R and Lane, M (2010) ‘From Data to Policy Analysis: Tax-Benefit Modelling using SILC 2008’, Papers WP359, Economic and Social Research Institute (ESRI).

Creedy J and Tuckwell, I (2003) ‘Reweighting the New Zealand Household Economic Survey for Tax Microsimuilation Modelling’, Treasury Working Paper Series 03/33, New Zealand Treasury.

Creedy J (2004) ‘Survey Reweighting for Tax Microsimulation Modelling’. In *Research in Economic Inequality*, Volume 12 (ed. by Y. Amiel and J. Bishop), pp. 229-249. New York: AI Press.

Figari F, Paulus A and Sutherland H (2015) ‘Micro-simulation and Policy Analysis’, in A.B. Atkinson and F.Bourguignon (Eds.) *Handbook of Income Distribution*, Vol. 2B Amsterdam: Elsevier, pp. 2141–2221.

Frankovic B (2013) ‘Calibration of weights of statistical surveys in R language’, *Forum Statisticum Slovacum*, 19-37.

Glaser-Opitzova H, Ivancikova L and Frankovic B (2015) ‘The Calibration of Weights Using Calmar2 and Calif in the Practice of the Statistical Office of the Slovak Republic’, European conference on quality in official statistics Q2014, Vienna.

Horvath M, Senaj M, Siebertova Z and Svarda N (2015) ‘The End of the Flat Tax Experiment in Slovakia’, CBR Working paper 4/2015, Council for Budget Responsibility.

Immervoll H, Lindström K, Mustonen E, Riihelä M and Viitamäki H (2005) ‘Static data ageing techniques. Accounting for population changes in tax-benefit microsimulation’, EUROMOD Working paper, EM7/05, Institute for Social & Economic Research, University of Essex.

Inglic R, Seljak R, Intihar S (2013) ‘Final Quality Report, EU-SILC-2010, Slovenia’, Statistical Office of the Republic of Slovenia.

Kump N and Navicke J (2014) ‘Re-weighting EUROMOD for demographic change: an application on Slovenian and Lithuanian data’, EUROMOD Working Paper, EM13/14, Institute for Social & Economic Research, University of Essex.

LeGuennec J and Sautory O (2002) ‘CALMAR2: une nouvelle version de la macro CALMAR de redressement d’échantillon par calage’, *Actes des Journées de Méthodologies*, INSEE, Paris.

Matsaganis M, Paulus A and Sutherland H (2008) ‘The take up of social benefits’, Research Note 5/2008, Social Situation Observatory, European Commission.

Porubsky M, Machlica G and Strizencova K (2013) ‘EUROMOD Country Report, Slovak Republic (2009-2012)’, Institute for Social & Economic Research, University of Essex.

O’Donoghue C and Loughrey J (2014) ‘Nowcasting in Microsimulation Models: A Methodological Survey’, *Journal of Artificial Societies and Social Simulation*, vol. 17(4), p. 1-12.

Siebertova Z, Svarda N and Valachyova J (2014) ‘A Microsimulation model of the Slovak Tax-Benefit System’, CBR Discussion paper 4/2014, Council for Budget Responsibility.

Siebertova Z, Senaj M, Svarda N and Valachyova J (2015) ‘To Work or Not to Work? Updated Estimates of Labour Supply Elasticities’, CBR Working paper 3/2015, Council for Budget Responsibility.

Statistical Office of the Slovak Republic (2014), EU SILC 2012 SR UDB version 20/01/2014.

Statistical Office of the Slovak Republic (2014), EU SILC 2013 SR UDB version 23/07/2014.

Strizencova K and Hagara E (2014) ‘EUROMOD Country Report, Slovak Republic (2009-2013)’, Institute for Social & Economic Research, University of Essex.

Sutherland H and Figari F (2013) ‘EUROMOD: the European Union tax-benefit microsimulation model’, *International Journal of Microsimulation* 6(1), 4-26.

Vlacuha R and Frankovic B (2015) ‘The Calibration of Weights by Calif Tool in the Practice of the Statistical Office of the Slovak Republic’, *Romanian Statistical Review 2*.

Vork A and Paulus A (2014), ‘EUROMOD Country Report – Estonia’, Institute for Social & Economic Research, University of Essex.

Wiemers J (2015) ‘Endogenizing take-up of social assistance in a microsimulation model A case study for Germany’, *International Journal of Microsimulation* 8(2), 4-27.

APPENDIX

Table A1 Summary statistics SK-SILC 2013 (Income reference period 2012)

	Mean (Std.Dev)				**Number of individuals**		
	SK SILC		**Admin. Data**		**SK SILC**		**Admin. Data**
Weights	**Original**	**Calibrated**			**Original**	**Calibrated**	
Labour income (monthly)							
Gross wage employment	692.97 (369.69)	722.59 (530.82)	805.00 775.60	* **	1,983,176	1,788,927	1,881,598
Income from self-employment	722.89 (558.06)	709.50 (540.66)	382.43	**	347,397	361,935	277,125
Other payments made by employers	17.24 (20.29)	20.79 (29.89)			315,322	278,861	
Income from agreements	70.16 (154.41)	68.96 (93.76)	143.94	**	364,458	341,729	955,330
Non-labour income (monthly)							
Unemployment benefit	271.20 (164.21)	259.96 (148.52)	312.00		58,064	60,399	143,896
Maternity benefit	431.87 (179.48)	451.57 (168.34)	443.00		35,203	71,317	24,221
Child birth grant	60.46 (22.24)	62.95 (23.17)	153.67		38,020	70,298	56,994
Child benefit (incl. additional child benefit)	23.49 (12.27)	23.34 (11.19)	37.78		779,739	738,748	688,344
Parental allowance	200.71 (78.99)	200.43 (71.97)	195.87		140,111	164,278	142,274
Material needs benefit	132.63 (120.49)	122.99 (103.3)	132.62		102,402	118,566	183,091
Nursing allowance	150.05 (100.47)	140.28 (89.04)	142.22		32,411	41,248	58,700
Sickness and nursing benefits	78.84 (81.78)	101.03 (96.78)	209.70		107,431	139,511	119,092
Disability pension	272.53 (112.53)	261.98 (98.64)	260.90		201,729	257,721	227,801
Old-age pension	375.44 (112.07)	374.58 (108.71)	367.00		1,072,056	1,062,519	980,863
Early retirement pension	348.33 (107.57)	325.15 (104.03)	374.00		16,765	16,862	24,404

Non-labour income (monthly)							
Widow's pension	153.16 (97.51)	156.99 (100.6)	237.00		339,647	346,049	336,877
Orphan's pension	159.09 (65.77)	146.41 (55.18)	126.00		19,479	16,972	26,923
Sample size	15,426	15,426					

Source: Author's calculations using SK SILC, official statistics Statistical Office, SSA and COLSAF.

Note:
*Mean monthly wage reported by the Statistical Office.
**Means computed by using data from SSA.

Table A2 Simulated benefits: Ratios of aggregate number of recipients

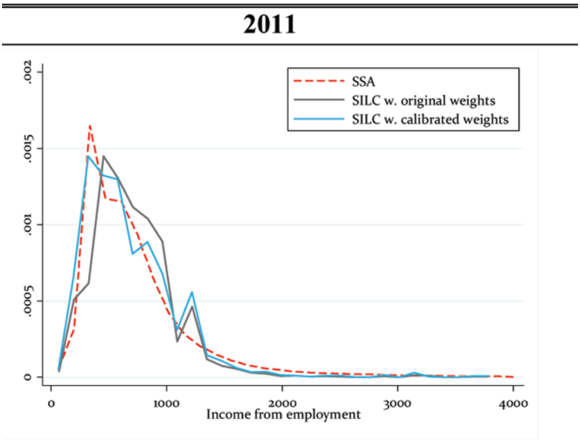
Model	EUROMOD		SIMTASK	
Weights	**Original**	**Calibrated**	**Original**	**Calibrated**
	2011			
Unemp. benefit	0.50	0.44	0.51	0.45
Parental allowance	0.68	1.03	0.67	1.03
Child benefit *	0.88	1.00	1.08	1.15
Child birth grant	0.59	1.07	0.59	1.07
Material needs benefit	1.00	1.36	0.73	1.01
Housing allowance	1.92	2.62	0.79	1.15
Activation allowance	4.68	6.47	1.00	1.50
Protection allowance	1.41	1.87	0.58	0.68
	2012			
Unemp. benefit	0.38	0.41	0.40	0.43
Parental allowance	0.88	0.87	0.87	0.99
Child benefit *	0.96	0.94	1.15	1.08
Child birth grant	0.75	1.02	0.75	1.02
Material needs benefit	1.07	1.43	0.81	1.04
Housing allowance	2.12	2.78	0.82	1.10
Activation allowance	5.25	6.83	1.25	1.56
Protection allowance	2.06	2.80	0.73	0.92

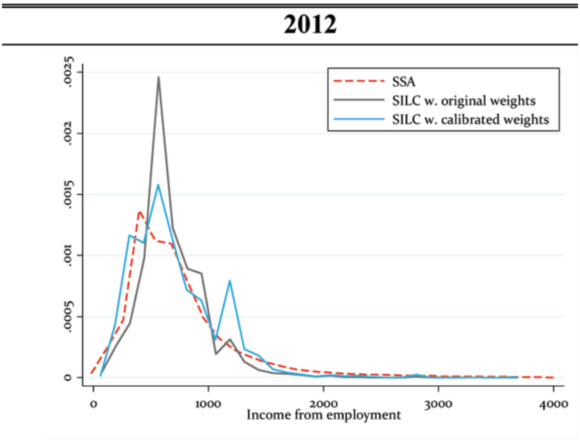
Source: Authors' calculations using SIMTASK and EUROMOD, official statistics SSA (unemployment benefit), COLSAF (other benefits).

*Official statistics on child benefit recipients is taken as the average of monthly data over the year. Official statistics on other benefits is the total number of individual recipients (incidence).

Note: Numbers display ratios of aggregate number of recipients computed by EUROMOD and SIMTASK to aggregate number of recipients referred by official statistics.

Figure A1 Income distribution SK-SILC and Social Security Agency (SSA)





Source: Authors' calculations using SK-SILC and SSA.

VZAM_microsim_1 – the EU-LFS based microsimulation model to estimate the Slovak labour market skills needs[1]

MIROSLAV ŠTEFÁNIK[2]
Institute of Economic Research, Slovak Academy of Sciences[3],
Šancova 56, 811 05, Bratislava, Slovakia, email: miroslav.stefanik@savba.sk

ABSTRACT: This paper describes a possible way of re-designing a skills needs forecasting model into a microsimulation model simulating a chain of discrete events in time. VZAM_microsim_1 is an elementary microsimulation model employing Labour Force Survey microdata to simulate future skills needs of the Slovak labour market. VZAM_microsim_1 simulates demographic processes, educational attainment, economic activity, exits to retirement and matching labour supply and demand. The demand side distinguishes 18 economic sectors. The supply side recognises educational level, and field of individuals, complemented with their background individual characteristics, such as gender, age, or employment history. Several advantages of the microsimulation approach, in comparison to traditional manpower requirement modelling approaches, can be identified in predicting the replacement demand. The mechanism used for matching labour supply and demand in terms of skills relies dominantly on previous employment history. Two scenarios are suggested based on the strictness of employers´ decisions in the hiring process. Comparing the figures from the strict employment scenario to the complete employment scenario provides an indication of expected future skills mismatch in economic sectors. Slovak manufacturing appears to be the most problematic regarding skills mismatch between the demanded and supplied labour.

KEYWORDS: skills mismatch, replacement demand, manpower requirement modelling, labour supply predictions

JEL classification: J21, J24, J11, J26

---

[1] Author would like to acknowledge the great support he received from Philippe Liegeois, who was guiding him into the methodology of microsimulation in LIAM 2. Marek Radvanský provided valuable comments to this draft text of the paper.

[2] The research leading to these results has received support under the European Commission's 7th Framework Programme (FP7/2013-2017) under grant agreement n°312691, InGRID – Inclusive Growth Research Infrastructure Diffusion.

[3] This work was supported by the Slovak Research and Development Agency under the contract No. APVV-14-0324.

1 Introduction

In this paper, we introduce the first version of VZAM_microsim, a model developed to forecast labour supply and analyse skills mismatch at the Slovak labour market. The model employs the European Union Labour Force Survey (EU-LFS) microdata published by EUROSTAT. VZAM_microsim_1 was redesigned from a previous version of VZAM (Workie et.al., 2012, which was based on semi-aggregated EU-LFS data, following the concept of manpower requirement model, adopted in the methodology of CEDEFOP´s forecasts of Skills Supply and Demand in Europe (CEDEFOP, 2009).

In the following text, first, the concept of the manpower requirement model and related literature is presented. The second subsection of the introductory chapter briefly deals with selected microsimulation models which inspired the redesigning of the original VZAM model. The second chapter describes the new modular structure of redesigned VZAM_microsim_1 by modules and particular processes. Results are presented in the third chapter, together with the results of the now-casting variant of the model. The fourth chapter highlights some of the advantages of the microsimulation approach in skills needs forecasting.

1.1 Concept of the manpower requirement model

Under the Europe 2020 Strategy, the European Commission (EC) launched the initiative of New Skills for New Jobs (COMMISSION OF THE EUROPEAN COMMUNITIES, 2008). The urge for matching educational systems according to the labour market needs became even more pressing during the post-crisis period[4].

The European Centre for the Development of Vocational Training (CEDEFOP) has produced several rounds of EU-wide skills forecasts (Skills Supply and Demand in Europe, 2010 and 2012). These present a continuation of the tradition of manpower forecasting pioneered by Parnes (1962) and further elaborated at the Maastricht Research Centre for Education and the Labour Market (ROA)[5] or the Warwick Institute for Employment Research[6] (WIER).

Typically, skills needs forecasting is done by combining outputs from a macroeconomic model, with (what we call here) semi-aggregated data from a La-

---

[4] The positive aspects of information brought by existing labour market forecasts are highlighted for example in: (Sexton, 2002) or (Barnow, 2002) in (Neugat, M., Schömann, K. (Eds), 2002)

[5] See for example: (Willems E., de Grip A., 1993) or (ROA, 1995)

[6] See for example: (Wilson, R.A. et al, 2007)

bour Force Survey. The macroeconomic model is usually based on a general equilibrium methodology[7]. Such models are supposed to provide information about the level of employment (demand) in economic sectors, under some assumptions, for example, if the economy would perform at the level of its potential output.

The supply side of existing manpower requirement models usually focuses on the disaggregation of the predictions of employment in sectors. For this purpose, mostly EU-LFS based semi-aggregated data is used to disaggregate total employment into narrower subgroups (based on occupational, educational or age group). Semi-aggregate data is employed either in a cohort component model, in line with Willems and de Grip (1993), or in a timeline structure to extrapolate trends in that particular segment (CEDEFOP, 2009, p. 70).

The manpower requirement model became widespread in two waves[8]; first in the Seventies, after being inspired by Parnes (1962), with a consequent renaissance in the Nineties[9]. In the last years, the stream of research on manpower requirement modelling transformed into skill needs forecasting, covering under the umbrella of upskilling and skills mismatch research.

One of the strongest heritages of the Nineties´ discourse is the distinction between expansion and replacement demand for labour. Expansion demand for labour is a result of expansion or restriction of employment in that particular segment of the labour market. It can be positive (new jobs are appearing) or negative (existing jobs are diminishing) and is mostly driven by macroeconomic and structural factors. It is therefore usually modelled in a macroeconomic model.

Replacement demand appears when an existing job needs to be filled-in because of a worker exiting the job. Several types of exits are accounted for, to retirement, other forms of economic inactivity, or mobility between labour market segments. Replacement demand is mostly driven by individual factors, as it is strongly determined by age-specific life situations. Willems and de Grip (1993) therefore introduced a cohort component method to count replacement demand from the differences in the age-cohort between two consequent time periods. Quantifying replacement demand in this way grasps all types of exits; those related to retirement, inactivity, inter-occupational mobility, as well as mortality.

---

[7] Input-Output model in (Parnes, 1962), or more sophisticated CGE models elaborated later.

[8] A nice historical overview of the manpower requirement modelling discourse in Austria can be found in Lassnigg (2002).

[9] For an overview of the second wave initiatives see for example: Hughes (1993), Heijke (1994), OECD (1994), Heijke et al. (1998) and Neugat, M., Schömann, K. (Eds), (2002).

This methodology is strongly limited by the structure of the data employed. For anonymization related reasons, EU-LFS microdata is usually provided coded in five-years age groups. The age cohort is therefore limited to five-year age groups and the time period, used to count the difference, has to be five years[10]. In this paper, an alternative solution is proposed to quantify replacement demand, based on simulating discrete time events on the level of individuals.

1.2 Microsimulation based labour market modelling

Recent increase in office computers´ computing performance, together with an increase in the quality of data provided for research purposes, opened new opportunities for microsimulation modelling. It has resulted in the development of more sophisticated microsimulation models. Some of them are running on EU-wide datasets, such as EUROMOD (Sutherland, 2007), (Sutherland and Figari, 2013). EUROMOD is a static EU-wide tax-benefit microsimulation model employing microdata from the European Union Survey on Income and Living Conditions (EU-SILC). Since EUROMOD, several country-specific EU-SILC data based microsimulation models emerged, mostly designed to simulate the impact of ageing on the system of social benefits and public finance. Another example is the MIDAS model designed to test the Belgian public pension scheme (Dekkers et.al, 2010), with a related initiative of developing a universal software tool for microsimulation modelling – LIAM 2 (Bryon et.al, 2015). LIAM 2, thanks to its universal applicability, was also used in redesigning VZAM into VZAM_microsim_1.

Norwegian labour market model MODAG is combining a macroeconomic model with administrative data on the supply side (Bjørnstad et.al, 2010). Being designed for a similar purpose, to explore possible skills imbalances in the Norwegian labour market, it takes advantage of the high-quality administrative data on Norwegian population and educational attainment. Initiatives of the Austrian Public Employment Office in designing a microsimulation model to assess Austrian labour skills needs are mentioned in (Lassnigg, 2002, p. 4). Nevertheless, none of the microsimulation models known to the authors to this date was employing EU-LFS data for the purpose of skills needs forecasting.

In the Slovak context, one microsimulation model was developed employing the EU-SILC data by the Slovak Council for Budgetary Responsibility (Siebertova et.al, 2014). Its purpose is, alike the most of the EU-SILC based models

---

[10] Such methodology was, on Slovak data, applied for example in Mikločovič, et.al. (2015) or Radvanský and Miklošovič (2016)

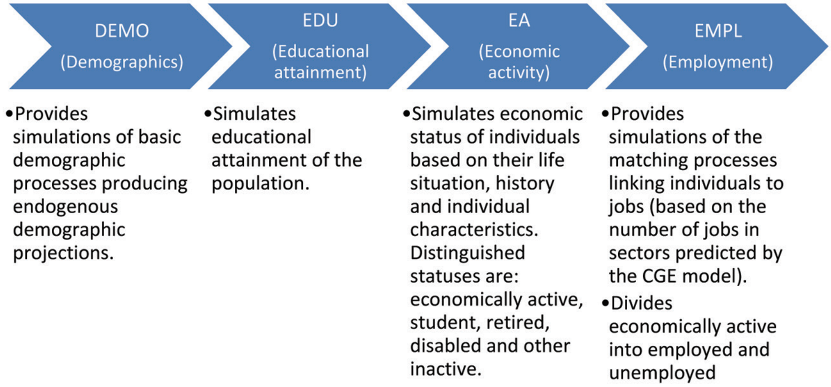
available, to provide simulations of the Slovak tax-benefit system. It thus joined the dominant wave of microsimulation modelling in Europe, responding to the EUROMOD inspired initiatives.

2 Description of the supply side of the VZAM_microsim, version 1

The VZAM model was originally developed to forecast future structure of employment on the Slovak labour market, in terms of economic sectors, occupations and educational groups. In its previous versions, it followed the methodological approach of traditional manpower forecasting models (Workie, et.al, 2012). On the demand side, an equilibrium based macroeconomic model (CGE) was used to produce projections of employment in economic sectors – expansion demand (Miklošovič, et.al., 2015). The supply side adopted the concept of a stock-flow model to anticipate future developments in the educational structure of the population. Replacement demand was modelled in a separate module employing semi-aggregated LFS data. It was originally designed based on the architecture of the EU-wide model developed under the cooperation of CEDEFOP (CEDEFOP, 2012). Here we are describing an example of how could the supply side in this approach be translated into a microsimulation model.

We adopt the original purpose of the model but try to introduce a methodological improvement in switching from semi-aggregate EU-LFS data to microdata. Our microsimulation model relies on a picture about the structure of labour supply drawn from the EU-LFS in 2014. The original dataset is inflated, using population weights, to produce an individual level file of the size of the Slovak population, representing its structure in terms of gender, age, region and educational structure.

Individuals in this file are subject to processes defined in the model. Processes are defined in four, relatively autonomous, consequently executable modules. The scheme below displays the modules of the model in their sequential order.



- Provides simulations of basic demographic processes producing endogenous demographic projections.

- Simulates educational attainment of the population.

- Simulates economic status of individuals based on their life situation, history and individual characteristics. Distinguished statuses are: economically active, student, retired, disabled and other inactive.

- Provides simulations of the matching processes linking individuals to jobs (based on the number of jobs in sectors predicted by the CGE model).
- Divides economically active into employed and unemployed

Scheme 1: Modules of VZAM-microsim 1

2.1 Module DEMO

The demographic module DEMO shelters demographic processes related to the dynamics of Slovak population. Because of anonymization purposes, the original LFS microdata was provided in five-year age groups. For this reason is the first process related to the disaggregation of these five-year age groups into one year age groups. The disaggregation is done assuming random choice assignment from a five-year age group to a one-year age group using proportions from the official population statistics[11] for 2014 (the year of the survey). To provide a clearer idea about how age is assigned based on the information about the five-year age group, an example of the script transforming the first age group (0-4 years old) to one year age groups follows:

```
age: if(ageg==2,
        choice([0, 1, 2, 3, 4],
        [0.1932, 0.1941, 0.1969, 0.2136, 0.2022]),
```

This process is special because it only runs in one period for 2014 and thus complements the attribute age into the original LFS 2014 file (inflated).

---

[11] Data based on Census 2011 updated by reporting on births and deaths. Published by the Statistical Office of the Slovak Republic: http://bit.ly/2pVZeWY

2.1.1 Ageing

After age is disaggregated into one-year age groups in the LFS 2014 file, processes running annually during the whole forecasted period (2015-2025) can be simulated. First, it is ageing. This is a deterministic process described by the function:

age: age + 1

2.1.2 Birth

Birth is simulated using age-specific fertility tables[12]. Here we rely on a national source of data using the predictions produced by Slovak Demographic Research Centre, despite that Eurostat also provides the age-specific fertility rates under the EUROPOP 2013 projections. Female between 15 and 50 years of age give birth in the proportions provided in the age-specific fertility tables. Within a one-year age group giving birth is distributed randomly, no other individual characteristics are taken into account. For example education level of the mother could play a role here, but is not accounted for, despite that the architecture of the model and the software tool used would technically allow such distinction.

2.1.3 Death

Analogously death is distributed randomly within a one-year age group based on age-specific mortality tables[13]. Age-specific mortality tables align the proportions of dying within the age group for males and females. Here also gender is taken into account as age-specific mortality tables are produced separately for men and females. This is a standard in demographic predictions. Also here the architecture of the model, and the software tool used, allow for further precision distinguishing other relevant individual characteristics, such as region or education level which might influence mortality.

---

[12] Two alternative sources of age-specific fertility and mortality tables can be used for Slovakia:

EUROPOP 2013 published by EUROSTAT: http://ec.europa.eu/eurostat/data/database

Slovak National Demographic Projections published by the Demographic Research Centre: http://www.infostat.sk/vdc/en/index.php?option=com_wrapper&view=wrapper&Itemid=43

[13] Also here, two alternative sources of age-specific mortality tables can be used for Slovakia:

EUROPOP 2013 published by EUROSTAT: http://ec.europa.eu/eurostat/data/database

Slovak National Demographic Projections published by the Demographic Research Centre: http://www.infostat.sk/vdc/en/index.php?option=com_wrapper&view=wrapper&Itemid=43

Migration related functions are missing in the demographic module of the current version of the model. There are several reasons for this decision. First, current immigration into Slovakia is marginal and one of the lowest within EU, so the model may produce relatively realistic figures for the mid-term period even without taking into account for migration. Because of the expected demographic development, migration should start to play a role. One of the ambitions of the model is to show, and eventually quantify, the need for labour to be "imported" into the country. This presents another reason for abstracting from migration.

2.2 Module EDU

VZAM_microsim_1 works with a strongly simplified module on educational attainment of the population. International applicability of the model was a priority here. For this reason, we decided to rely on LFS data and avoid employing, more detailed but internationally incomparable, national specific statistics on numbers of students by type level and field of study. The current setting of the EDU module is also partly limited by the rationale of the scenario. Performed simulations would like to show where would skill mismatches become the bulgiest if the structure of education provided remains unchanged.

With respect to international applicability and preserving the educational structure, we have adopted the following simplifying assumptions. All individuals, regardless of their level of education achieved, remain students until the age of 20. When reaching the age of 20, individuals´ education level is assigned randomly, regardless of other individual characteristics. This assignment keeps the overall proportions of particular levels of education the same as was the educational assignment of the age group of 30-34 years old. The age group of 30-34 years old is the youngest age group, whose educational attainment seems to be finalised, based on the EU-LFS data[14]. Similar procedure is followed for fields of education within each of the educational levels.

Module EDU, thus, has two consequent processes. First the process of educational level assignment and second the process of assigning field of education within each of the educational levels. Analogously as in simulating educational levels, assignment to a field of education is also random keeping the proportions of fields within each of the educational levels based on the age group 30-34.

---

[14] The main model (based on EU-LFS 2014 microdata) is using the proportions from 2014. For the now-casting variant of the model (based on EU-LFS 2011 microdata) we have pooled the observations for this age group from EU-LFS rounds 2011 to 2013. A newer version of the ISCED classification of education was applied in the 2014 EU-LFS microdata.

2.3 Module EA

“EA” is an abbreviation for economic activity. In fact, this module distinguishes several types of economic statuses:

- Economically active
- Students
- Retired
- Disabled
- Other inactivity

At the beginning of each period, economic status is set to unknown. In cases when favouable, economic status from the previous period is used to model economic status in the current period. The order of the statuses assigned also plays a role. Age-related statuses are modelled first, starting with students and retired.

2.3.1 Students

As already mentioned, each individual before reaching 21 years of age is a student, and therefore economically inactive.

2.3.2 Retired

Retired are those who reached the retirement age and were not assigned to a special group of working retired. Retirement age in Slovakia is, based on the valid legislation, supposed to grow together with the mean life expectancy[15]. Retirement age is gender specific. Besides reaching retirement age, individuals´ may leave for retirement even earlier, two years before reaching the legal retirement age. Existing Slovak legislation enables such option, and a separate process also covers it in the model. Early retirement is assigned based on a modelled probability of early retirement estimated on the EU-LFS 2011-2014 microdata using the economic sector, status in the previous period and other individual characteristics as explanatory variables. The complete results of the probit model used to count the early retirement score can be found in the online annexe[16]. The overall proportion of early retired is aligned based on age relatively to retirement age (retirement age -1, and -2). Proportions are based on Census 2011 data.

---

[15] It was implemented into the model as an exogenous variable based on a study of the Slovak Council for Budget responsibility: http://www.rozpoctovarada.sk/svk/rozpocet/300/vypocet-dochodkoveho-veku

[16] Annex 1: http://ekonom.sav.sk/uploads/work/economic_statuses_score_estimations.txt

2.3.3 Disabled

Disability is defined based on the probability of being disabled which is a product of a probit estimation on EU-LFS 2011-2014 microdata[17]. Furthermore, the proportions of disabled are aligned based on its real occurrence in 2011-2014 EU-LFS distinguishing five-year age groups.

2.3.4 Other inactivity

This category is supposed to cover all other forms of economic inactivity. Explicitly it works with maternity leaves using a deterministic function of a maternity leave. Each woman takes a maternity leave for the period of three following years after giving birth. This is the maximum maternity leave supported by the Slovak legislation and remains the most widespread pattern of behaviour.

Maternity is complemented with other forms of economic inactivity which are modelled, analogously to disability and working retired, based on a probability model and aligned to fit the overall proportion of this group in the reference population. The results of the estimation of the probability of being inactive can be found in the online annexe[18].

2.3.5 Economic activity

Individuals are, in each period, considered to be economically active, if they do not fulfil the conditions necessary to be assigned to one of the above-listed forms of economic inactivity. If an individual is neither student, neither retired, neither disabled or in other form of inactivity; he is assumed to be economically active.

Moreover out of retired, the model identifies cases when working after reaching retirement age appears. This is based on a separate function constructed like in the case of early retirement. By counting the working retirement score based on a probit model estimated on EU-LFS 2011-2014 microdata, when taking into account sector of previous employment and other individual characteristics (complete results of the estimation can be found in the online annexe[19]). In the second step, the proportions of those working after reaching retirement age are aligned relatively to (individuals´ age minus the retirement age) based on early retirement occurrence from Census 2011.

---

[17] Annex 2: http://ekonom.sav.sk/uploads/work/economic_statuses_score_estimations.txt

[18] Annex 3: http://ekonom.sav.sk/uploads/work/economic_statuses_score_estimations.txt

[19] Annex 4: http://ekonom.sav.sk/uploads/work/economic_statuses_score_estimations.txt

Economic status is reconstructed anew in each period. For its reconstruction information about the previous status is taken into account, where feasible. In simulating the economic status, we are trying to copy the reality as much as possible to provide realistic information about the expected future development of the supply of labour in Slovakia. In the next step economically active are divided between employed and unemployed based on a matching function, within a consequent EMPL module. In reconstructing the matching processes, the ambition is not as much to stick to the reality but to provide information which is a result of theoretically defined scenarios relying on reasonable assumptions.

2.4 Module EMPL

The EMPL module simulates employment of those who were assigned as economically active. This group is divided between employed an unemployed based on the result of two matching processes. Each of the matching processes relies on different assumptions. Their outcome presents two polarised, theoretical scenarios. In the first "strict employment" scenario, matching process is done keeping the educational proportions of employed within each of the economic sectors from the period 2011-2014. In the second, "complete employment" scenario, demand for labour is satisfied in the full extent filling in all of the jobs generated based on the macroeconomic model. Figures provided by the two of the scenarios thus border the interval in which possible employment should be developing.

In the case of both scenarios, information about the employment in a particular economic sector is inherited from previous periods. This assures a high share of reproduction, keeps the model close to the reality (because that is the "modus operandi" in most employment contracts), but restricts voluntary inter-sectoral migration.

Within the model, thus, individuals´ only leave employment in their economic sector in cases of:

- replacement due to retirement, disability or other inactivity (maternity leave, etc.);
- if the overall employment in their economic sector is shrinking

(In this case individuals are, within the period, first assigned as unemployed and eventually reassigned to a different sector based on one of the result of two alternative matching procedures.)

2.4.1 Strict employment scenario

In the case of economic sectors with a positive increase in labour demand[20], workers are being complemented out of the pool of unemployed and graduates. In the case of the strict employment scenario individuals suitable for the economic sector are selected in two steps. First, a pool of possibly suitable candidates is selected aligning the structure of this pool based on the educational structure[21] of employed in the sector during the reference period[22]. Consequently, in the second step, individuals are picked for employment out of the pre-selected pool based on the probit score counted based on the estimated probability of being employed in that particular economic sector[23].

If a graduate or an unemployed individual is possibly employable in more than one of the economic sectors he chooses the one with a higher average wage. It is assured by ordering the sectors in the matching process based on average hourly wage.

Table 1: Economic sectors in the order as prioritised by the model

Order in matching	Code in the model	NACE code	Label
1	10	J	Informatics
2	11	K	Financial services
3	4	D	Electricity
4	13	M	Professional activities
5	2	B	Mining
6	6	F	Construction
7	8	H	Transportation
8	3	C	Manufacturing
9	12	L	Real estate activities

---

[20] Expansion demand + Replacement demand > 0

[21] Distinguishing a combination of 5 educational levels and ISCED 97 3-digit classification of fields of education in the now-casting model and ISCED 11 3-digit classification of fields of education in the main model. This is because of the data availability, as the new ISCED 11 classification was adopted in EU-LFS in 2014.

[22] In case of the now-casting model the reference period is 2011-2013. In case of the main model the reference period is 2014.

[23] In estimating the probability of being employed in a particular sector we use individual characteristics, such as gender, age-group, education, region, as well as past employment experience of individuals. Complete results of probit estimates for all 18 economic sectors can be found in the online annex at: http://ekonom.sav.sk/uploads/work/economic_sector_employment_score_estimations.txt

Order in matching	Code in the model	NACE code	Label
10	16	P	Education
11	18	R/U	Other service activities
12	5	E	Water & Waste
13	17	Q	Health
14	15	O	Public administration
15	1	A	Agriculture
16	7	G	Wholesale & Retail
17	9	I	Accommodation
18	14	N	Administrative and support service activities

2.4.2 Complete employment scenario

In the complete employment, scenario matching is done applying only the second step of the above-described, matching mechanism. In the case of an economic sector with positive labour demand, additional workers are complemented out of the pool of unemployed and graduates without any strict restrictions regarding their education. Individuals are prioritised based on the same probit score, estimating the probability of being employed in the particular sector.

In this scenario, either employees are perfectly flexible in adjusting their formal education by lifelong learning, or employers are perfectly flexible regarding their requirements towards newly hired employees. In other words, there is no skills mismatch between the demand and supply of labour.

3 Overview of the results

In this section, we present the first results of VZAM_microsim_1 model. The main objective here is to show the structure of information which can be reported based on the results of the simulations. Moreover, we are trying to provide information about the reliability of presented results by adding information from a now-casting variation of VZAM_microsim_1. In this variation, the same model is applied on EU-LFS microdata for Slovakia from 2011. The figures of indicators published can be thus confronted with the real EU-LFS figures for 2013[24]. Such confrontation provides us with some idea about the reliability of the results of the main model. Results of the main model are published for the period up to 2025.

[24] Here we use the LFS 2013 figures because of the change in the ISCED classification of education in 2014 from ISCED 97 to ISCED 11.

3.1 Economic status

In the case of economic inactivity statuses of individuals, the correspondence to the reality is of high importance because the logic of our approach requires a realistic picture about the supply of labour. In the case of the total numbers of students, retired, disabled and inactive, the correspondence of model predictions to the reality is satisfactory. The highest error of the predictions can be observed in the case of other forms of inactivity. VZAM expects the number of individuals in this group to be 232 876 in 2013, while the actual EU-LFS figures for 2013 show 238 667 individuals in this subgroup. It presents a 2.43% error of the predictions. When assessing the errors of the prediction, one needs to take into account that EU-LFS is a survey which is also related to some amount of sampling error.

Table 2: Results of the VZAM_microsim_1 predictions for 2013 and EU-LFS data for Slovakia in 2013

	EU-LFS 2013	**VZAM_ S1**	**Difference S1**	**% difference S1**	**VZAM_ S2**	**Difference S2**	**% difference S2**
Employed	2 329 249	2 378 030	-48 781	-2.09%	2 603 698	-274 449	-11.78%
Unemployed	386 142	367 198	18 944	4.91%	141 530	244 612	63.35%
Student	1 309 129	1 286 519	22 610	1.73%	1 286 519	22 610	1.73%
Retired	980 222	965 093	15 129	1.54%	965 093	15 129	1.54%
Disabled	167 431	166 695	736	0.44%	166 695	736	0.44%
Inactive	238 667	232 876	5 791	2.43%	232 876	5 791	2.43%
Total	5 410 840	5 396 411	14 429	0.27%	5 396 411	14 429	0.27%

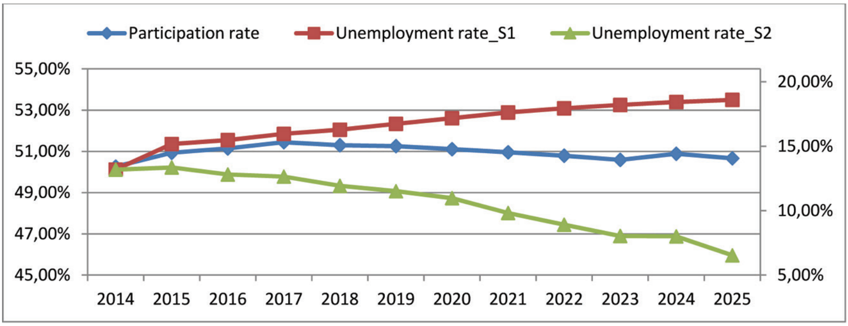
Note: VZAM_S1 stands for the strict employment scenario, VZAM_S2 stands for the full employment scenario

Source: EU-LFS 2013, VZAM_microsim_1 (now-casting variant)

In the case of economic inactivity statuses, the figures are the same for both scenarios (strict and full employment) because the scenarios are applied in the matching process and thus influence only the process of simulating employment. Here the correspondence to the reality is of secondary importance because the simulated figures provide information about hypothetical development under adopted assumptions. Nevertheless, in the case of now-casting, the strict employment scenario seems to be providing figures closer to the reality. A possible interpretation for this would be that the additional assumption of perfect flexibility in the hiring process behind the full employment scenario is not realistic. Employers are rather strict in the requirements they are applying towards newly hired employees. Alternatively, from the other side, employees are not perfectly flexible in responding to employers´ requirements in terms of their qualification.

In the case of the restricted employment scenario, unemployment is expected to grow up to 499 974 in 2025. In the case of the full employment, total unemployment would drop to 175 722 in 2025. In the mid-term, up to 2025, the restricted employment scenario would probably lose its correspondence to the reality, as the adjustment of the market to supply-demand imbalances can be expected to strengthen. Future development should, thus, remain in the borders of these two scenarios.

Graph 1: Expected development of the main labour market indicators based on VZAM_ microsim_1



Note: Participation rate on the left axis, Unemployment rate on the right axis

Source: EU-LFS 2014, VZAM_microsim_1 (main model)

Moreover, VZAM_microsim_1 results allow following more detailed results within economic sectors, as described in more detail in the consequent sections on replacement demand and skills mismatches.

As regarding other economic statuses, the number of students is expected to decline in the period up to 2025, which is in line with other studies (Herich, 2012, 2013a, 2013b) (Lichner and Štefánik, 2015). The number of retired is, in contrast, expected to grow. These changes are driven mainly by the demographic change.

Table 3: Number of persons based on their economic status in 2014 and 2025

	EU-LFS 2014	**VZAM_S1 2025**	**VZAM_S2 2025**
Employed	2 363 001	2 188 592	2 512 844
Unemployed	358 909	499 974	175 722
Student	1 289 767	1 156 778	1 156 778
Retired	996 081	1 040 971	1 040 971
Disabled	164 312	193 258	193 258
Inactive	243 417	226 873	226 873
Total	5 415 487	5 306 446	5 306 446

Source: EU-LFS 2014, VZAM_microsim_1 (main model)

A slight growth is also expected in the case of disabled persons and persons in other forms of economic inactivity. Also here it is results of the relative increase of older age-groups in the population and also is in line with existing studies (Radvanský and Lichner, 2014).

3.2 Replacement demand

Another of the indicators possibly constructed from the modelled results is related to replacement demand in particular sectors. Replacement demand can be seen *as a job opening arising because of people leaving the workforce* (CEDEFOP, 2012, p. 65). In a microsimulation model, individuals can be observed with all their predefined attributes and also attributes´ related history. Thanks to this, the numbers of individuals employed in the economic sector during the previous period and retired in the current period can be extracted. When looking only at exits to retirement, the average retirement exit rate is a share of those leaving for retirement to the total employment in the sector during the previous period. Values of the retirement rate vary from 2.45% in Agriculture to 0.33% in Mining. Based on the EU-LFS, the average retirement rate was 1.05% in 2013.

Table 4: Replacement demand due to exits to retirement in 2013, by sector

Economic sector	EU-LFS: total employment in t-1	EU-LFS: retirement	EU-LFS: retirement rate	VZAM: total employment in t-1	VZAM: retirement	VZAM: retirement rate	Retirement rate diff in p.p. of total employment	Retire-ment rate diff in %
Agriculture	75 381	1 850	2.45%	74 970	1 481	1.98%	-0.44%	-18.52%
Mining	12 649	42	0.33%	12 575	101	0.80%	0.40%	108.68%
Manufacturing	570 320	4 913	0.86%	568 140	6 288	1.11%	0.19%	20.28%
Electricity	24 334	571	2.35%	24 203	256	1.06%	-1.39%	-57.64%
Water & Waste	26 108	473	1.81%	25 938	409	1.58%	-0.46%	-23.39%
Construction	240 714	3 142	1.31%	239 448	2 639	1.10%	-0.26%	-19.30%
Wholesale & Retail	289 858	1 893	0.65%	289 039	1 476	0.51%	-0.16%	-24.47%
Transportation	156 975	2 162	1.38%	156 252	1 083	0.69%	-0.79%	-53.16%
Accommodation	97 163	762	0.78%	96 970	832	0.86%	0.06%	7.60%
Informatics	61 063	572	0.94%	60 901	293	0.48%	-0.69%	-57.47%
Financial services	51 879	520	1.00%	51 761	213	0.41%	-0.63%	-60.95%
Real estate activities	16 030	106	0.66%	15 959	133	0.83%	0.23%	37.59%
Professional activities	72 045	481	0.67%	71 770	593	0.83%	0.06%	7.87%
Administrative and support service activities	61 472	707	1.15%	61 168	606	0.99%	-0.87%	-52.48%
Public administration	184 790	1 785	0.97%	184 098	1 455	0.79%	-0.27%	-26.43%
Education	157 399	1 878	1.19%	156 796	2 495	1.59%	0.34%	28.23%
Health	161 284	2 235	1.39%	160 730	1 055	0.66%	-0.77%	-54.07%
Other service activities	69 496	317	0.46%	69 278	621	0.90%	0.43%	90.93%
Total	2 328 961	24 408	1.05%	2 319 996	22 029	0.95%	0.10%	10.37%

Source: EU-LFS 2013, VZAM_microsim_1 (now-casting variant, complete employment scenario)

At the first glance, the predictions of replacement demand produced by VZAM_microsim_1 scenarios, differ substantially. The reasons behind are various, but first one should be aware that these differences are counted out of small shares, which means that even in a case of a substantial divergence of the modelled retirement rate from the reality (108,58%), the final bias in terms of total employment is only marginal. The highest relative difference from real figures can be observed in the case of the Mining sector. In absolute terms, the bias caused here presents only 0,4% of the total employment. Moreover, Mining is a small sector, with the numbers of employees exiting to retirement being a marginal slice of the employment in the sector; a sample-based survey may run into troubles in covering such a small subgroup. If we consider that the numbers reported from EU-LFS as well as those employed in VZAM_microsim_1 are based on sampling weights. In 2013 there was probably only one individual leaving the Mining industry for retirement.

Sampling related problems thus may substantially influence the correspondence of the simulation results to the reality, as both, simulation source data, as well as the observed reality, are subject to sampling error. A total error of 10.37% representing a difference between 1.05% and 0.95% p.p. of total employment could, in this context, be considered as acceptable.

Exit to retirement is not the only type of exit from employment individuals may happen to meet. Together with the exit to disability and to other forms of inactivity, exit to retirement represents the total replacement, which results to the quantification of the replacement demand in VZAM_microsim_1.

Exit from employment to study is not possible within the assumptions of the model. A substantial share of exits to retirement is, within the model, done via exits to disability, because these occur more often in higher (pre-retirement) age. Individuals are thus exiting from disability to retirement after reaching the retirement age. These cases are not visible when displaying only the employment – retirement transitions.

Table 5: Replacement demand due to all types of exits (retirement + disability + inactivity) in 2013, by sector

Economic sector	EU-LFS: total employment in t-1	EU-LFS: retirement	EU-LFS: retirement rate	VZAM: total employment in t-1	VZAM: retirement	VZAM: retirement rate	Retirement rate diff in p.p. of total employment	Retirement rate diff in %
Agriculture	75 381	2 678	3.55%	74 970	4 544	6.06%	2.51%	70.60%
Mining	12 649	42	0.33%	12 575	204	1.62%	1.29%	390.33%
Manufacturing	570 320	15 726	2.76%	568 140	15 434	2.72%	-0.04%	-1.48%
Electricity	24 334	898	3.69%	24 203	501	2.07%	-1.62%	-43.92%
Water & Waste	26 108	613	2.35%	25 938	636	2.45%	0.10%	4.45%
Construction	240 714	4 921	2.04%	239 448	4 665	1.95%	-0.10%	-4.70%
Wholesale & Retail	289 858	7 981	2.75%	289 039	7 699	2.66%	-0.09%	-3.26%
Transportation	156 975	3 198	2.04%	156 252	3 079	1.97%	-0.07%	-3.27%
Accommodation	97 163	2 787	2.87%	96 970	5 284	5.45%	2.58%	89.95%
Informatics	61 063	1 044	1.71%	60 901	1 537	2.52%	0.81%	47.55%
Financial services	51 879	1 370	2.64%	51 761	1 598	3.09%	0.45%	16.92%
Real estate activities	16 030	179	1.12%	15 959	663	4.15%	3.04%	272.44%
Professional activities	72 045	2 484	3.45%	71 770	2 963	4.13%	0.68%	19.75%
Administrative and support service activities	61 472	707	1.15%	61 168	1 895	3.10%	1.95%	169.33%
Public administration	184 790	4 225	2.29%	184 098	4 696	2.55%	0.26%	11.57%
Education	157 399	4 270	2.71%	156 796	5 979	3.81%	1.10%	40.58%
Health	161 284	3 943	2.44%	160 730	4 702	2.93%	0.48%	19.66%
Other service activities	69 496	889	1.28%	69 278	2 767	3.99%	2.71%	212.23%
Total	2 328 961	57 955	2.49%	2 319 996	68 846	2.97%	0.48%	19.25%

Source: EU-LFS 2013, VZAM_microsim_1 (now-casting variant, complete employment scenario)

When looking at replacement rate, accounting for exits to retirement as well as to disability and other inactivity, the total bias in comparison to reality almost doubles (from 10.37% to 19.25%). Nevertheless, in absolute terms, the divergence is between 2.49% measured in EU-LFS in 2013 versus 2.97% simulated by the model. These figures are produced for the complete employment scenario; in the strict employment scenario, the comparison to the reality is more favourable.

Table 6: Replacement demand due to all types of exits (retirement + disability + inactivity) in 2014 and 2025, by sector

Economic sector	**EU-LFS: total employment in 2013**	**EU-LFS: retirement 2014**	**EU-LFS: retirement rate**	**VZAM: total employment in 2024**	**VZAM: retirement in 2025**	**VZAM: retirement rate**
Agriculture	75 381	3 065	4.07%	75 240	4 077	5.42%
Mining	12 649	319	2.52%	8 821	278	3.15%
Manufacturing	570 320	13 450	2.36%	445 652	13 699	3.07%
Electricity	24 334	1 179	4.84%	27 622	1 157	4.19%
Water & Waste	26 108	796	3.05%	30 710	974	3.17%
Construction	240 714	4 420	1.84%	225 199	5 396	2.40%
Wholesale & Retail	289 858	8 158	2.81%	245 742	7 888	3.21%
Transportation	156 975	2 473	1.58%	162 209	5 210	3.21%
Accommodation	97 163	2 932	3.02%	118 208	6 060	5.13%
Informatics	61 063	300	0.49%	64 840	1 107	1.71%
Financial services	51 879	1 180	2.27%	63 169	1 686	2.67%
Real estate activities	16 030	207	1.29%	16 448	743	4.52%
Professional activities	72 045	2 275	3.16%	82 713	4 057	4.90%
Administrative and support service activities	61 472	985	1.60%	67 723	2 137	3.16%
Public administration	184 790	4 123	2.23%	202 748	6 817	3.36%

Economic sector	EU-LFS: total employment in 2013	EU-LFS: retirement 2014	EU-LFS: retirement rate	VZAM: total employment in 2024	VZAM: retirement in 2025	VZAM: retirement rate
Education	157 399	6 342	4.03%	130 015	6 522	5.02%
Health	161 284	3 697	2.29%	158 347	6 320	3.99%
Other service activities	69 496	2 903	4.18%	74 489	2 987	4.01%
Total	2 328 961	58 804	2.52%	2 199 895	77 115	3.51%

Source: EU-LFS, VZAM_microsim_1 (main model, strict employment scenario)

During the period of the forecast, up to 2025, simulated replacement ratio is expected to grow from 2.52% in 2014 up to 3.51% in 2025. The absolute number of persons to be replaced annually on the labour market will grow from 58 804 to 77 115.

3.3 Skills mismatch

Simulated results offer several options for constructing indicators of skills mismatch. They can be based either on unemployment rate based on the educational group (level or field), or the difference between employment in the strict and complete employment scenario. Here we focus on the latter option.

The difference between the two scenarios presents two opposite extremes, based on two polarised, hypothetical situations. In the first scenario (strict employment) employers are keeping the same educational structure of employed in the economic sector in hiring. In the second scenario (complete employment scenario), employers in sectors are hiring to the full extent of the demand for labour estimated by a macroeconomic equilibrium model, regardless of the education of the newly hired persons. The difference between these two scenarios indicates the skills mismatch with respect to employment in economic sectors.

Table 7: Difference in the number of employed between strict and complete employment scenarios, by sector

Economic sector	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**
Agriculture	4 951	3 128	1 398	776	0	0	77	0	0	0	0
Mining	0	0	0	0	0	0	0	0	0	0	0
Manufacturing	20 443	36 972	50 567	65 547	78 813	92 081	108 873	122 755	134 120	136 119	145 939
Electricity	0	0	0	0	0	0	0	0	0	0	0
Water & Waste	0	0	0	0	0	0	0	0	0	0	0
Construction	8 074	9 421	10 862	12 228	13 097	13 832	14 693	15 432	15 571	13 010	12 629
Wholesale & Retail	5 387	10 963	14 995	20 508	25 119	29 361	35 870	40 381	43 826	44 013	46 157
Transportation	897	0	0	0	0	175	509	1 186	1 504	0	19
Accommodation	5 941	7 560	9 247	12 553	15 720	18 957	22 666	26 085	28 575	29 842	31 800
Informatics	0	0	0	0	0	0	0	0	0	0	0
Financial services	0	0	0	0	0	0	0	0	0	0	0
Real estate activities	0	0	0	0	0	0	0	0	0	0	0
Professional activities	657	786	622	1 838	3 294	4 980	4 896	6 300	7 480	7 742	9 088
Administrative and support service activities	0	0	0	0	0	0	0	0	0	0	0
Public administration	3 784	5 098	4 652	5 689	5 311	4 255	11 488	10 106	7 917	2 673	0
Education	0	0	0	0	0	1 183	4 164	8 329	14 318	21 918	38 372
Health	0	0	0	1 427	3 124	5 878	10 396	15 369	21 464	27 315	39 623
Other service activities	0	0	0	0	0	0	0	0	0	0	625
Total	50 134	73 928	92 343	120 566	144 478	170 702	213 632	245 943	274 775	282 632	324 252

Source: VZAM_microsim_1 (main model)

Based on the difference between the two scenarios, Slovak labour market was missing over 50 thousand workers with specific qualifications in 2015. The unsatisfied demand for labour could eventually climb to 324 252 skilled workers in 2025. Such situation would occur, if employers would not adjust their requirements based on the available supply of labour and unemployed and graduates would not be willing to change their qualifications after attaining their formal education in initial (pre-career) education and training. Such situation is, of course, extreme and in reality, the demand for skilled workers would be more flexible. Nevertheless, our goal here is to show areas where skills mismatch will become the most urging in the future. The scenarios were designed for this goal.

The highest skills mismatch can be expected in Manufacturing. This is true in absolute terms (145 939 in 2025) as well as in relative terms (25.01% of the total employment in 2025).

Table 8: Relative difference in the number of employed between strict and complete employment scenarios, by sector

Economic sector	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Agriculture	5.72%	3.66%	1.65%	0.93%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%
Mining	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Manufacturing	3.69%	6.60%	8.93%	11.50%	13.75%	16.00%	18.74%	21.02%	22.92%	23.30%	25.01%
Electricity	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Water & Waste	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Construction	3.53%	4.05%	4.62%	5.15%	5.48%	5.77%	6.11%	6.40%	6.47%	5.43%	5.30%
Wholesale & Retail	1.88%	3.79%	5.15%	7.02%	8.60%	10.06%	12.22%	13.74%	14.97%	15.13%	15.97%
Transportation	0.58%	0.00%	0.00%	0.00%	0.00%	0.11%	0.31%	0.73%	0.92%	0.00%	0.01%
Accommodation	4.82%	5.95%	7.08%	9.39%	11.52%	13.64%	15.97%	18.04%	19.48%	20.12%	21.19%
Informatics	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Financial services	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Real estate activities	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Professional activities	0.84%	0.98%	0.76%	2.20%	3.87%	5.75%	5.65%	7.15%	8.36%	8.53%	9.84%
Administrative and support service activities	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Public administration	1.75%	2.31%	2.09%	2.56%	2.42%	1.97%	5.20%	4.66%	3.74%	1.29%	0.00%
Education	0.00%	0.00%	0.00%	0.00%	0.00%	0.84%	2.95%	5.85%	9.84%	14.36%	23.05%
Health	0.00%	0.00%	0.00%	0.83%	1.81%	3.41%	5.96%	8.69%	11.88%	14.64%	20.17%
Other service activities	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.81%
Total	0.93%	1.37%	1.71%	2.23%	2.68%	3.17%	3.97%	4.59%	5.14%	5.31%	6.11%

Source: VZAM_microsim_1 (main model)

In the case of some sectors, the additional demand for workers is more strongly driven by replacement demand. This is the case of Agriculture (in the earlier stages of the prediction period) or Education (in the latter stages of the prediction period). These are sectors with rather declining total employment, but relatively older workforce at the beginning of the prediction period.

Additional demand for labour in Health sector is driven by both expansion demand (the result of the ageing of the population), as well as replacement demand (the result of the relatively old workforce).

Slovak economy heavily relies on export-oriented industries, such as motor vehicles or electrical equipment manufacturing. Our simulations show that these would be the economic sectors most harmed by the skill mismatch to be expected. The potential harm to the whole Slovak economy could, therefore, be substantial.

4 Discussion on the potential of a microsimulation approach in skills needs forecasting and possible next steps

VZAM_microsim version 1 presents the first version of a microsimulation model employing EU-LFS microdata to simulate skills mismatches on the labour market. There are several alleys where the first version can be improved. In the first version, the priority was more on international comparability, than on complexity of the model. In this version, VZAM_microsim can be quite easily adjusted to any of the countries covered by the EU-LFS survey. A significant increase in the precision of the results could be gained from switching from survey data to administrative data, or eventually CENSUS 2011 data. Complementing the simulations using additional attributes of individuals, such as occupations (eventually also the region and other) would be probably the very first next step. This could be done in a similar matter as it was done with economic sectors. Such step should not be technically too demanding if we adopt the assumption of fixed occupational structure within each of the sectors. Other main challenges for the future remain in:

- ➢ refining the model in various areas where inspiration can be found in EU-SILC based models, for example in:
    - implementing households as an entity,
    - sophisticating the simulation of pre-career educational attainment using administrative data
    - simulating economic activity based on the change of potential income out of work and social benefits
    - implementing migration

- making the matching process dynamic by allowing adjustments of employers´ preferences in hiring new employees based on the available supply
- incorporating lifelong learning, work experience and re-skilling
- allowing iterations between the microsimulation model and macroeconomic equilibrium model

Despite these remaining challenges, several advantages of applying a microsimulation approach to the manpower requirement model can be stressed, based on the experience in re-designing the labour supply side of VZAM.

Linking the forecast of the inflow to the LM (graduates) with the demographic part of the prediction

The sole fact, that demography is endogenous in this version of VZAM enables designing scenarios based on different demographic assumptions. On top of that, a direct link to the educational process can be designed. Demographic predictions are of high accuracy in comparison to predicting other social processes. Educational system predictions can be built on this accuracy, employing administrative data on the numbers of students. It would enable educational projections distinguishing narrower categories in terms of level and field of education. Moreover designing variant scenarios in applying alternative educational policies becomes possible. The model would show variant implications for all endogenous processes defined in the model, as in the case of VZAM would be skills mismatch on the labour market.

Economic (labour market) status can be modelled in a chain of consequent decisions

The decision whether to be economically active is usually a result of several consequent sub-decisions. These can be modelled on an individual level, taking into account individual characteristics as well as supra-individual variables referring to the situation in the economy and the labour market. Individual´s history presents a strong explanatory variable as well.

The concept of a structural probit model[25] can be employed, taking into account the tax-benefit system of the country, its possible changes (variant scenarios) with related labour market consequences.

---

[25] As applied for example in: Breunig, R. and J. Mercante (2010)

Replacement demand can be based on cross section or panel data with annual change

Willems and de Grip (1993) introduced a cohort component method based approach in modelling replacement demand, which was since adopted in several applications (including CEDEFOP, 2012). Switching from cohort-component approach to a microsimulation approach allows for possible several methodological improvements. Exits to retirement, disability or inactivity can be distinguished and modelled based on more transparent assumptions. Mortality and migration could become endogenous in the model.

Replacement demand figures can be produced annually and verified in a nowcasting scheme as it is done in this paper. Starting from Willems and de Grip (1993), most of the studies dealing with replacement demand pay attention to the occupational structure. The first version of VZAM_microsim instead differentiates economic sectors, because we have assessed the sectoral structure to be the most relevant in linking the demand and supply side of the model. Adding the attribute of occupation, with related processes, to the simulations should technically not be a problem. Especially if we consider, that the demand side of the model produces figures of labour demand within sectors for qualification levels.

Matching process can be grasped in a more intuitive and more precise way

In constructing skills mismatch indicators, existing manpower requirement models either rely on a simple labour supply-demand matching mechanism, or avoid matching at all producing time-change type of indicators. Also here, switching to a microsimulation approach enables new options. If the educational categories distinguished within the model are detailed enough, basically if they include some information about the field of education, matching based on a detailed conversion table (education to the sector, or education to sector X occupation) can be done. Such conversion tables can be constructed either based on an objective definition for each of the categories or, as in the case of VZAM_microsim_1, based on the empirical incidence of educational subgroups within economic sectors or occupations.

On top of that, variant scenarios of the matching process can be designed, for example, based on contra-extreme assumptions, as it was done in the case of VZAM_microsim_1.

5 References

BARNOW, B.S. (2002): ‘Occupations and skills in the United States: Projection methods and results through 2008’, in M. Neugart and K. Schömann, (eds.), Forecasting Labour Markets in OECD Countries, Measuring and Tackling Mismatches, Cheltenham, UK: Edward Elgar, 2002

BJORNSTAD, R., GJELVIK, M. L., GODOY, A., HOLM, I., STOLEN, N. M. (2010): Demand and supply of labor by education towards 2030 - Linking demographic and macroeconomic models for Norway, Reports 39/2010, Statistics Norway. Oslo. 2010, Online: http://www2.warwick.ac.uk/fac/soc/ier/research/forecast/forecastingsymposium/countryforecasting/norway/rapporter_2010-39.pdf

BREUNIG, R. and MERCANTE, J. (2010): “The Accuracy of Predicted Wages of the Non Employed and Implications for Policy Simulations from Structural Labour Supply Models”, The Economic Record, The Economic Society of Australia, vol. 86(272), pages 49-70, 2010

BRYON, G., DEKKERS, G., DE MENTEN, G. (2015): LIAM2 User Guide - Release 0.11.0, Federal Planning Bureau, November 24, 2015, Online: http://liam2.plan.be/download/LIAM2UserGuide-0.11.0.pdf

CEDEFOP (2009): Future Skill Needs in Europe, Medium-Term Forecast, Background Technical Report. Luxembourg: Publications Office of the European Union, 2009. ISBN 978-92-896-0506-9.

CEDEFOP (2010): Skills supply and demand in Europe, Medium-term forecast up to 2020. Thessaloniki: European Centre for the Development of Vocational Training, CEDEFOP, 2010,

CEDEFOP (2012): Skills supply and demand in Europe (Methodological framework). Research paper No. 25. Luxembourg: Publication Office of the European Union, 2012. ISBN 978-92-896-1112-1.

COMMISSION OF THE EUROPEAN COMMUNITIES (2008): New Skills for New Jobs, Anticipating and matching labour market and skills needs, Brussels, COM(2008) 868/3, 2008

CORVERS, F. and HENSEN, M. (2005): Forecasting regional labour-market developments by occupation and education, In: van Oort F.; van Wissen L. and Thissen M. (Eds). A survey of spatial-economic planning models in the Netherlands. Theory, application and evaluation, Netherlands Institute for Spatial Research, Rotterdam, 2005

DEKKERS, G. DESMET, R., DE VIL, G. (2010): The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model, Federal Planning Bureau. WORKING PAPER 10-10.

EUROSTAT (2015a): EU Labour Force Survey Database, 2015 release, 2015.

EUROSTAT (2015b): EU Labour Force Survey Database User Guide, EUROSTAT, November 2015.

Heijke, H. (ed.) (1994): Forecasting the labour market by occupation and education. Dordrecht: Kluwer Academic Publishers, 1994.

HEIJKE, H.; BORGHANS, L. (eds) (1998): Towards a transparent labour market for educational decisions. Aldershot: Ashgate, 1998.

HERICH, J. (2012): Prognóza vývoja ukazovateľov materských a základných škôl do roku 2025. Bratislava: Ústav informácií a prognóz školstva.

HERICH, J. (2013a): Prognóza vývoja ukazovateľov gymnázií do roku 2025. Bratislava: Ústav informácií a prognóz školstva.

Herich, J. (2013b): Prognóza vývoja ukazovateľov stredných odborných škôl do roku 2025. Bratislava: Ústav informácií a prognóz školstva.

HUGHES, G. (1993): Projecting the occupational structure of employment in OECD countries. Paris: OECD, 1993 (OECD Labour Market and Social Policy Occasional Papers No 10).

MIKLOŠOVIČ, T., RADVANSKÝ, M., HVOZDÍKOVÁ, V. (2015): Strednodobá prognóza sektorového dopytu po práci so zameraním sa na vzdelanostnú úroveň, In: Trh práce na Slovensku 2016+. Bratislava. 2015. ISBN:978-80-7144-1, online: http://ekonom.sav.sk/en/publications/-p327

LASSNIGG, L. (2002): Projections of qualifications and occupations in Austria – short-termism, macro perspective and emphasis on the supply side. In: Neugart, M.; Schömann, K. (eds). Forecasting labour markets in OECD countries; Measuring and tackling mismatches. Cheltenham: Edward Elgar, 2002.

LICHNER, I., ŠTEFÁNIK, M. (2015): Prítok absolventov stredných a vysokých škôl na trh práce, In: Trh práce na Slovensku 2016+. Bratislava. 2015. ISBN:978-80-7144-1, online: http://ekonom.sav.sk/en/publications/-p327

NEUGART, M.; SCHOMANN, K. (eds) (2002): Forecasting labour markets in OECD countries; Measuring and tackling mismatches. Cheltenham: Edward Elgar, 2002.

OECD (1994): Employment outlook. Paris: OECD, 1994.

PARNES, H.S. (1962): Forecasting educational needs for economic and social development. Paris: OECD, 1962.

RADVANSKÝ, M., LICHNER, I. (2014): Impact of ageing on long term care demand and supply in Slovakia : NEUJOBS country report. Radvanský M., Lichner I. In NEUJOBS working paper, 26.02.2014, no. D 12.2, p. 1- 26. Online: http://www.neujobs.eu/search/node/Radvansky

RADVANSKÝ, M., MIKLOŠOVIČ, T. (2016): Estimating long-term structural and educational changes of labour demand using CGE model - case of Slovakia, EuroMod 2016, conference proceedings, April 2016, Online: http://ecomod.net/conferences/ecomod2016?tab=downloads

ROA. (1995): The labour market by education and occupation to 2000. Maastricht: Research Centre for Education and the Labour Market, 1995.

SEXTON, J.J. (2002): A review of occupational employment forecasting for Ireland, In: M. Neugart and K. Schömann, (eds.), Forecasting Labour Markets in OECD Countries, Measuring and Tackling Mismatches, Cheltenham, UK: Edward Elgar, 2002

SIEBERTOVÁ, Z., ŠVARDA, N., VALACHYOVÁ, J. (2014): A Microsimulation model of the Slovak Tax-Benefit System, Council for Budget Responsibility, Discussion Paper No. 4/2014

SUTHERLAND, H. (2007): ‘EUROMOD: the tax-benefit microsimulation model for the European

GUPTA, A., and A. HARDING (Eds.) (2007): Modelling Our Future: population ageing, health and aged care, International Symposia in Economic Theory and Econometrics, Vol 16. Elsevier.

SUTHERLAND, H. FIGARI, F. (2013): EUROMOD: the European Union tax-benefit microsimulation model, INTERNATIONAL JOURNAL OF MICROSIMULATION (2013) 6(1) 4-26

WILLEMS E., DE GRIP A. (1993): Forecasting replacement demand by occupation and education, International Journal of Forecasting, Volume 9, Issue 2, 1993

WILSON, R.A. et al. (2007): Skills in England 2007: key messages. Coventry: Learning and Skills Council. 2007

WORKIE TIRUNEH, Menbere - BUJŇÁKOVÁ, Tatiana - HORVÁT, Peter - LICHNER, Ivan - ŠTEFÁNIK, Miroslav. (2012): Predvídanie potrieb trhu práce v SR : teoretické východiská a empirické výsledky = Forecasting labour market needs in the Slovak republic : theoretical foundations and empirical results. Bratislava : Ekonomický ústav SAV, 2012. 181 s. ISBN 978-80-7144-204-2

On the prognosis of third generation migrants' occupational status in Germany – a dynamic microsimulation

PETRA STEIN[1] – DAWID BEKALARCZYK[1]

Abstract: This article introduces the modelling of the future development of third generation immigrants' occupational status in Germany for a period of 30-40 years. It focuses on the connection of this development with the change of the ethnical and socio-structural population composition, the latter being formed by demographical change and, especially, by migration history. These changes of the population composition are assumed to influence the state of integration of the third generation (composition effects) in addition to causal mechanisms. The method being used is the dynamic microsimulation, which – by its modularized structure and stochastic extrapolation technique – allows to model multidimensional processes that cannot be simply formulized by mathematical functions. First findings suggest that using this approach the demographic development can be decently reproduced. Furthermore, a weak composition effect was identified.

Key Words: Microsimulation, future projection, panel analysis, migration research, demography

1. Introduction

Through mass immigration, Germany has evolved to a country of immigration since the 1960s. Currently, most immigrant groups in Germany are significantly disadvantaged in many areas of life. The extent of these disadvantages varies depending on the observed group of migrants or the considered attribute. There are important disadvantages at almost all stages of the educational career (choice of the kindergarten: B. Becker 2010, B. Becker & Biedinger 2016; school enrolment: Tuppat & B. Becker 2014; transition to secondary school: Kristen & Dollmann 2010, Dollmann 2016; graduation: Gresch & Kristen 2011). Furthermore, in numerous studies strong disadvantages can be found for various indicators

[1] University of Duisburg-Essen, Institute of Sociology, Lotharstr. 65, 47057 Duisburg, Germany

of occupational status that are in large part due to time preceding educational disadvantages (Diehl et al. 2009, Imdorf 2011, Damalang & Haas 2006, Brück-Klingberg et al. 2011, Seibert & Solga 2005, Sürig & Wilmes 2011, Kalter et al. 2011, Basilio & Bauer 2010, Sürig & Wilmes 2011, Erlinghagen & Scheller 2011, Hunkler 2016).

Ethnic disadvantages do not occur only temporarily immediately after immigrating. They are often reproduced in the second generation (Kalter et al. 2011: 257, Kalter 2008a, Diehl et al. 2009: 49, Treichler 2014: 208, Olczyk et al. 2016). However, these studies show that, compared to the first generation, in the second generation most groups were able to reduce their drawbacks significantly. Whether such findings indicate a transgenerational establishment of inequality between migrant groups and the societal mainstream or a process, in which ethnic inequality *does* reduce, but rather slowly, is discussed controversially (Esser 2008). This question remains a fundamental issue in the field of migration research (Alba 2008, Portes & Rumbaut 2006, Zhou 1999).

The project presented here shall contribute to the answer of this question. First results are shown in the course of this paper. The project is called "Longitudinal modelling of the future development of occupational status in the third generation of migrants using the dynamic microsimulation", which is funded by the German Research Foundation. Main goal is to model the middle-term (30-40 years) future integration development of third generation migrants living in Germany. The specific feature of this modelling approach is that the development of integration is connected to the demographically induced change in the ethnic and socio-structural population composition. This allows studying the effect of the interaction of causal mechanisms with composition effects on the development of (labor) integration (in the third generation of immigrants). The focus lies on the third generation, since it is the first generation whose parents were born in Germany. Thus, the integration development in this generation is particularly ground-breaking for the long-term integration development.

The research question can be examined only with a future projection, since the majority of members of the third generation of immigrants is just growing up or is not even born yet. A broad empirical database is available to model the development of relevant mechanisms for integration. Likewise, natural changes of the population can be traced well with official data. Furthermore, since demographic processes (if one excludes migration) are relatively sluggish, they can be reliably projected into the future. The future projection is realized using the dynamic microsimulation.

Firstly, the state of migration research is outlined in "Theoretical Background", considering in particular the integration of immigrants living in Germany. This section reveals immediately that a future projection of the integration development in the third generation of migrants is necessary. In "Methodical

Considerations", the potential and the limitations of a simulative prediction method in the context of the present research question are discussed and the chosen method (dynamic microsimulation) is introduced. In the following sections, the simulated model is specified in the light of previous theoretical considerations and empirical analyses. A presentation of the results for a first simulated scenario follows. The paper closes with conclusions and an outlook regarding further work in this research project.

2. Theoretical Background

How ethnic inequality respectively integration/assimilation[2] of migrants in Germany will develop in the future is controversially discussed in the field of migration research. Subject of the basic research in this area is the question whether almost universal patterns of incorporation exist for migrants in a host country, which can be applied to diverse migration situations and thus allow a reliable forecast of the future development of integration. However, until now, no clear forecast can be derived from basic studies:

In traditional integration approaches (in the context of immigration to the United States: Park 1928, Eisenstadt 1953, Taft 1953, Gordon 1964) assimilation was – despite some restrictions – the only expected outcome of the incorporation process of migrants (Kalter 2008b). However, this automatism is doubted since at least the empirical finding of "anomalies" in the United States

---

[2] Ethnic inequality exists when the distribution of a "vertical" attribute significantly differs between migrant groups or between particular groups of migrants (also: combined to a large group of "people with a migration background") and the autochthone population (examples of the countless studies about ethnic inequality: Alba et al. 1994, B. Becker & Biedinger 2006, Kalter 2008a, Siegert & Roth, 2013). The term "ethnic" is misleading, as he does not necessarily refer to "ethnic groups" (for a critical inspection see Wimmer 2008) in the German migration research, but to analytically disjunct migrant groups. Here such a view is followed, in which it is not assumed that these groups have a sense of community and a separate identity. The terms assimilation and integration are inconsistently used and rarely precisely delineated. Hartmut Esser is one of the few who strictly systematizes these terms (1980, 1999, 2001, 2009). In his logic, assimilation can be regarded as a special case of integration, where immigrants give up caring and expanding their origin-specific capital (such as the native language and contacts with people of the same origin). Since it is not necessary in this project to distinguish between assimilation and other forms of integration, both terms are used interchangeably. Further, since migrant groups are focused and no single individuals, ethnic inequality can be understood on such a group level as the absence of integration/assimilation.

in the 1960s (Zhou 1999: 197). In this context, competing approaches evolved (e.g. Alba 2008, Portes & Rumbaut 2006, Zhou 1999; for an overview see Esser 2008, Ballarino & Panichella 2013), from which different predictions can be derived.

All these concepts are characterized by a relative openness in respect of the future development of the integration of migrants. Therefore, a "successful" integration, in the sense of the absence of ethnic inequality, is from this perspective only *one* possible outcome of a contingent process, which is subject to certain conditions and mechanisms. These mechanisms can in turn be influenced by different, sometimes hardly predictable events and developments (e.g. political interventions). Projections of the future development of integration should therefore always be based on *scenarios*. A precondition to create well-founded scenarios is a previous empirical research of the mechanisms that favor integration or vice versa solidify ethnic inequality.

The most important attribute from the perspective of the research on social inequality (Kalter & Granato 2002, Lindenberg 1986) is the occupational status (Esser 1999, 2001). For this reason, this attribute serves as the main dependent variable when integration mechanisms are analyzed. The occupational status itself mainly depends on the formal education acquired in the educational system, whereas the acquisition of such certificates in turn heavily depends on ethnic and social origin (Kalter et al. 2011).

However, it is not enough to reduce the pool of relevant receiving-country-specific resources to education certificates and parental resources. In addition, the receiving-country-specific *cultural* and *social* capital is also able to explain a relevant part of ethnic inequality referred to occupational status (social capital: Granato 2009, Hunkler 2010, Schacht et al. 2014, Kalter 2006; language as a cultural capital: Esser 2006, Gresch & Kristen 2011, Hunkler 2010, Dustman & Soest 2002, Kempert et al. 2016). Moreover, to get a nearly complete picture, the "institutional side" (school, employers) has to be considered, because institutions can contribute to ethnic inequality – either through support (reduction of inequality) or through direct or indirect forms of discrimination (increase of inequality: Radtke 2004, Gomolla & Radtke 2009, Fereidooni 2011, Imdorf 2011, Sprietsma 2009, Fibbi et al. 2006, Bertrand & Mullainathan 2004, Diehl & Fick 2016).

To sum up, the future integration on the labor market results from the interaction of different mechanisms. This interaction has to be modelled as precisely as possible when creating a prediction model. Only when the constructed scenarios are based on empirically well-funded knowledge the scenario-conditioned forecast may turn out to be reliable.

However, the future development of the occupational status of migrants does not only depend on mechanisms that influence the professional status at the individual level causally. Demographic processes[3] can also have an impact on this development by causing composition effects (Kalter & Granato 2002). When considering particular generations of migrants, it is often not considered that their age structure and their composition by countries of origin can change over time. This is especially true for Germany, since larger groups of migrants have immigrated in different time slots (Gresch & Kristen, 2011: 213). In addition, these gradually immigrated groups differ significantly in terms of education. Following this and some plausible assumptions[4], the composition of a succeeding generation (like the second or third one) with different "successful" migrant groups logically changes over time. Therefore, even if fertility rates, life expectancy and the distribution of educational and professional success would remain perfectly constant within migrant groups, the average performance of this generation would still change with time. This is meant by composition effects here. Thus, when modelling the future integration development, composition effects triggered by demographic processes have to be considered as well. Only then, it is possible to obtain a forecast result which is generated by the interplay of causal mechanisms on individual level with composition effects.

The focus regarding the future integration development lies on the *third* generation of migrants. For this generation, as opposed to the second generation, the effect of *social origin*, which is central for education and occupational status, is no longer affected by the necessity to transfer capital acquired in the country of origin by the parents. This is because parents of members of the third generation are already born in Germany. Due to the significant distance of this generation to the act of migration, the integration in the third generation is an important stage in the intergenerational integration development.

Empirical analyses with respect to the performance of the third generation in the education system can be found, inter alia, in Kristen & Dollmann (2010), B. Becker (2011) and Hans (2015). However, to our knowledge no study concerning *the occupational status* of this generation exists. This is most likely because the majority of the third-generation members is still growing up or is not yet born. Nevertheless, there is already an enormous amount of empirical information (social origin, intermediate stages in the education system), which makes it

---

[3] About the demographic transition in Germany see e.g. Kluge et al. 2014, Birg 2003, Sommer 2007, Fuchs 2009, Statistisches Bundesamt 2009.

[4] A strong effect of social origin, which is exhaustively documented in Germany, needs to be assumed also for migrants (e.g. Kalter et al. 2011), along with some demographic regularities (e.g., that in all groups the mean age when becoming the first and following children is similar and that there is no selective and significant childhood mortality – which is not the case in Germany).

possible to estimate the distribution of professional status (and its development) in the third generation reliably by a future projection.

Finally, it should be noted that there are three main motives to conduct a future projection based on current integration trends here: Firstly, we are, as described above, interested in the *future* development of integration of persons living in Germany with a migrant background for theoretical reasons. Secondly, the focus lies on migrants from the *third* generation. Since their members are still too young on average, to analyze their achievements on the labor market with existing data, knowledge about the integration development in this generation can only be obtained via a forecasting method, especially because demographically caused composition effects also influence this development.[5] Thirdly, our intention is to show that a future projection, in particular a microsimulation, is an appropriate method for studying relevant sociological questions.

3. Methodological Considerations

3.1 Why Projection Instead of Analyzing Existing Data?

The third generation of immigrants, who came in the course of the labor recruitment in the 1950s and 1960s or during the wave of resettling from Eastern Europe to Germany, has a low average age. Most of the third-generation members are children, adolescents, or not born yet. In addition to that it is difficult to identify members of the third generation in popular datasets in Germany.[6] Only a few data sources allow such an identification requiring high efforts to connect the cases among each other. Therefore, an analysis of the development of occupational status of the third generation with empirical data is actually not possible. With a future projection, however, there is already a way to make statements about this development – even now. This option is rarely used or even considered in the social sciences because of existing reservations (Gilbert & Troitzsch 2005). Particularly, the attempt to make statements about processes and states in the future, is viewed skeptically (ibid.). This is caused, inter alia, by the poor fit of statistical models compared with empirical data – especially in the

---

[5] In fact, the National Education Panel Survey (NEPS; Blossfeld et. al 2011) constitutes a new, very detailed data source for investigating education. Due to its used instruments, it also allows to identify migrants from the third generation. However, the panel is quite new and needs some more years (waves) to achieve a level of accumulated information that makes it valuable for conducting analysis with it.

[6] It is not sufficient to ask, whether the parents are born in Germany (as it is the case for the second generation). Information about the grandparents are needed.

field of sociology. Often, a significant amount of unexplained variance remains, which prevents a prediction with an acceptable level of safety. In addition, many developments that are of particular interest for sociologists can be significantly affected by unforeseen events and policy interventions.

This skepticism is caused by the incorrect assumption that with prediction methods a real prediction of the future is sought. However, prediction models can also be understood as quasi-experimental designs, which measure the extent of the response to a stimulus (ibid.: 26). This can help to understand how a development could look like when the conditioning factors of these development have such a complex interplay that a reliable prediction of this development is not trivially achieved by simple logical reasoning or calculating.

This paper focuses on the interplay of demographic processes that can trigger composition effects in conjunction with developments on the individual and group level. With the here presented approach it is possible to examine the overall performance in the third generation of immigrants for the case that the effects of causal mechanisms (e.g., mediated by political interventions) interact with the demographically induced change in the composition of this generation. Since the processes, particularly the demographic transition (Birg 2003), are assumed to be dynamic and non-linear, the result of such an interaction cannot be calculated with a simple model of multivariate statistics, which requires committing oneself to known mathematical functions and probability distributions.

In addition, this approach allows for the first time to systematically test various controversially discussed hypotheses about the future integration development of migrants living in Germany, accounting for demographic dynamics.

3.2 Methodological Requirements

The main goal is to model the future development of occupational status of third generation migrants, which results from the interaction of causal mechanisms at the individual level with the demographically induced change in the composition of this generation. Since occupational status and demographic change are subject to processes at the individual level (language acquisition, education, labor market performance, bearing children, dying, etc.), the method must allow for modelling causal mechanisms at the individual level. In addition, although developing independently to a certain degree, these processes have to be modelled jointly. Therefore, a modularization of processes should be possible.

Additionally, a *development* should be projected and not just a *state* at a point in the future. Therefore, the extrapolating method must allow for recursive updating over multiple time points. Since the relevant causal mechanisms are formulated at the individual level, the projection should result in a longitudinal data set at the individual level, which can be aggregated for particular time points within the update period. Only then, causal effects can be separated from compositional effects.

All these criteria are fulfilled by the *period-oriented dynamic*[7] *microsimulation* (Sauerbier 2002, Gilbert & Troitzsch 2005, Leim 2008, Hannappel & Troitzsch 2015). Alternative equation-based projection methods, such as conventional statistical methods (e.g., structural equation modelling) or macro simulations are not suitable, because they require to specify the relationship between the professional placement and its mechanisms by a connected set of mathematical functions in advance. Furthermore, a macro simulation reaches modelling limits fast when partitioning an aggregate into subgroups (Gilbert & Troitzsch 2005). This is incompatible with a big set of connected/nested hypotheses at the individual level.

The more popular method, the agent-based simulation, appears to be inappropriate in the present case (for a review see for example Flache & Mäs 2015, Gilbert & Troitzsch 2005: 172-198, Bornmann 2010). Although from a broad point of view an agent-based simulation can be also seen as a microsimulation (Spielauer 2011: 13), because both have a common methodological ground (Li & O’Donoghue 2013: 25), some crucial differences can be derived from “typical applications” of both (some attempts exist to connect both, e.g. Klabunde et. al 2015). These differences arise from the historical fact that both approaches developed in almost total ignorance of each other (Spielauer 2011: 13). Typically, in agent-based simulations the entities to be simulated are agents who can choose an action from a set of competing actions depending on their observations, who can react on characteristics of other agents and who can interact with other agents. Often, the starting dataset is a network which consists of connected agents (e.g. Flache & Mäs 2015: 499) or it represents individuals who live in an (artificial) area (e.g. in segregation models, Bornmann 2010). Mostly, those starting datasets (individuals, networks, areas) are synthetic and the number of cases at micro level is, compared to typical applications of microsimulations, considerably smaller. In sociology, agent-based modeling is used to test middle-range theories, in which certain social mechanisms on individual level are assumed to generate social phenomena at macro level (Bornmann 2010). These models allow to examine, if a result that would follow from the explanation to be tested, really arises. In contrast, in microsimulations potentially unexpected outcomes of scenarios are predicted to describe a possible development that cannot be attained by simple calculations.

---

[7] In paragraph 3.3 it is explained what we mean with “period-oriented”.

Our approach clearly fits into the typical use of microsimulations. It focuses on developments at group level (groups of migrants by country of origin and generation), which are subject to results of acting individuals, rather than to interactions between individuals. Certainly, these results are highly dependent on interactions and negotiations between individuals. However, the quantitative empirical analysis of the nature of interactions (starting, maintaining, finishing interactions) is not well advanced in the social sciences. Therefore, it is not possible to determine detailed quantitative parameters for a simulation without a large amount of arbitrary decisions. It is furthermore important for the present research question to extrapolate a dataset which is large and representative for the population living in Germany, containing up to hundred attributes per person. This is not a main feature of agent-based simulation of networking actors. The potentially possible implementation of micro interactions into large datasets would greatly increase computational costs and complexity (Li & O'Donoghue 2013: 25). Additionally, big empirical datasets normally contain unobserved heterogeneity (Stein 1997) to a certain amount (for example the existence of unknown migrant subgroups that are much more successful on the labor market than the rest of the corresponding "supra group"). Although unobserved, in our case where composition effects matter, it is better to maintain this heterogeneity during the simulation than to exclude it from the very beginning, what is typically the case in a small artificial dataset for agent based modeling.

3.3 The Period-Oriented Dynamic Microsimulation

A big advantage of a microsimulation in general is the possibility to implement very detailed empirical information which allows to mimic the "reality" (especially non-linear dynamics) accurately. A large dataset at the individual level (e.g., n = 200,000; several hundred variables) can serve as starting data, such as a microcensus from official statistics. Given that the extrapolation is organized by separate modules, a high number of extrapolation parameters can be specified, which can also be estimated empirically.

In the microsimulation, primarily attributes of *individuals* from starting data are extrapolated, but it is also possible to model processes at meso- and macro-level as well as cross-level-feedback-processes. The extrapolation works stochastically/non-deterministically, which also makes the modelling more realistic. Therefore, with this stochastic approach it is for example possible to implement the error variance from a regression model into the simulation as a random distribution, which affects the simulated values (with good reason).

The extrapolation of an attribute *y* is organized in a so-called *module*. Core of a module is an updating algorithm (Leim 2008: 38 f.). It guarantees, inter alia, that only persons pass through the module of *y*, whose values of *y* may change. This can prevent that, for example, women between 80 and 90 years go through the birth module – or men. If relevant for *y*, the algorithm also checks to which subpopulation a person belongs. This is the case, when assumptions exist, on which attributes *y* depends. The whole population is then divided into subpopulations according to the combinations of the values of these attributes. The parameters for extrapolating are then estimated for each subpopulation separately (e.g., probability to die for men aged 75 or for women aged 80).

The stochastic updating technique bases on cumulative relative frequencies for the values of *y*, estimated from empirical data. These frequencies are interpreted in the simulation as cumulative probabilities (Hannappel & Troitzsch 2015: 461 f.). If assumptions exist, on which attributes *y* depends, then for each subpopulation (each combination of the values of the independent variables) a separate frequency distribution of *y* is estimated – either directly or with the help of regression models.[8]

In a simulation run, values of *y* are assigned to each individual. If these values can be regarded as events, then it is simulated for each individual whether one of the events has happened or not. In order to realize this, usually a number from a uniform distribution in the interval [0,1] is drawn randomly for each individual. In the next step it is checked, in which co-domain of the cumulative probability distribution the random number falls. The corresponding value of *y* is then assigned to the individual. It is the new simulated *y*-value for the time point at which the simulation actually proceeds.

Finally, the update algorithm reorganizes the dataset subject to the changes during the simulation (for example, when two persons move in together as a simulated event, they have to get a new household id and to be deassigned from the old ones). With this step, the extrapolation of one-time interval (usually one year) is completed. The dataset, which results from this extrapolation, now acts as starting dataset for the next run. This process can be repeated until the desired number of simulated time points is reached (Hannappel & Troitzsch 2015: 466).

If equally long time intervals are updated successively, as just described, this is called *period-oriented*, whereas in the *event-driven* approach, which is not used here, the time until the occurrence of an event is simulated. From this follows that "time" is treated as discrete in the case of a period-oriented microsimulation (whereas it is treated as continuous in event-oriented models). The result is an artificial panel dataset on individual level, which can be analyzed with common panel analytical techniques. The term *period-oriented* and

---

[8] In the case of continuous variables, with regression models the value of the variable itself can be estimated instead of the *probability* for a certain value.

its counterpart *event-driven* originate in the German scientific community (e.g. Hannappel & Troitzsch 2015, Leim 2008) and are therefore used in this paper. In the international discussion, a period-orientied microsimulation would be conceptually categorized as a microsimulation with a discrete modelling of time and a cross-sectional ageing process (Dekkers & Belloni 2009: 8-10).

With *dynamic* models in the context of microsimulations it is meant that the updated dataset can change in its size and composition (Leim 2008: 30 f.). Further, it is possible to model feedbacks from different levels, which can change the behavior of individuals (e.g., policy interventions). In contrast, in a static microsimulation only the characteristics at the individual level and the weights of the cases are manipulated, which leads to new total values, as for example tax receipts (Flory & Stöwhase 2012). Concatenations of processes and demographically caused changes of the size of a dataset cannot be modelled with this method.

Historically, microsimulations have been applied to answer a broad range of questions concerning economic and social subjects. As the method of choice, it has been frequently used to project results of governmental policy changes in the context of the development of pensions, health, demographics, and poverty issues (for an overview see Li & O'Donoghue 2013). Examples for applications of a dynamic microsimulation in Germany are MIDAS Germany (Dekkers & Belloni 2009), AVID (Schatz 2010), ZEWDMM (Bonin 2013), the extensive work of SfB 3 (Sonderforschungsbereich 3, Galler & Wagner 1996) and the dissertations of Leim (2008) and Hannappel (2015).

3.4 Software to Conduct a Period-Oriented Dynamic Microsimulation

Currently, only a few software packages exist for the implementation of a dynamic microsimulation. First of all, there are some tools and applications, which are thematically highly specific thus not suitable for the purposes of this project (e.g. the microsimulation tool MIST[9] or the "Community Health Advisor"[10], which can be rather seen as an access to an already existing model allowing for modeling some preselected interventions; both tools deal with aspects relevant for the health sector). In case of a complex model with a particular set of requirements, these tools struggle to deliver the necessary freedom for adjustments (Li & O'Donoghue 2013: 32). Here, an own "free" programming is especially

---

[9] https://github.com/scipy-conference/scipy2013_talks/tree/master/talks/jacob_barhak (27.10.2016)

[10] http://www.communityhealthadvisor.org/cha3 (27.10.2016)

needed to expand the extrapolation technique to linear and particularly to so-called linear *mixed* models (Rabe-Hesketh & Skrondal 2012).

Fortunately, there are some programming tools that allow for such a distinct implementation of a self-developed model to a varying degree, like UMDBS[11], Modgen, GENESIS, JAMSIM, LIAM, and LIAM2 (Li & O'Donoghue 2013: 31). Especially LIAM2 (Bryon et al. 2011) delivers an open framework for implementing large-scaled period-oriented dynamic microsimulations, without requiring deep knowledge of the underlying object-oriented program language Python (de Menten et al. 2014: 3). Since the conceptual logic of a microsimulation requires object-oriented programming (Leim 2008: 39), the simulation can be also programmed in the object-oriented language of the statistics package R. Moreover, this yields the advantage that no additional tool for analyses before and after the simulation is needed.

For the present paper, the model was implemented using Stata (for pragmatic reasons). In the further course of the project, we want to switch to R. The programming of a dynamic microsimulation in R is more "from scratch" (existing contributed packages about microsimulation are too specific) compared to LIAM2 for example. On the other hand, we see the potential to contribute with our work to further development of R-packages for microsimulation. Since R is widely used among statisticians in different disciplines, this development could help making microsimulation more popular.

4. Implementation of the Simulation and First Results

4.1 Theory-Based Modelling

First of all, the specification of a microsimulatic prediction model requires the formulation of a base model. This base model determines which observations with which characteristics are included in the simulation and how the characteristics are influencing each other. In a second step, rules need to be implemented that determine how these characteristics will change over time based on the settings specified in the base model (projection into a fictitious future). Theory-based (longitudinal-)analysis with empirical data will be conducted here for both steps to build the prediction model on the basis of valid knowledge from the present and early past. However, some of the extrapolating rules are based on unexaminable bridge-hypotheses. This is due to two reasons: Firstly, it is

---

[11] UMDBS (Sauerbier 2002) suffers from inherent restrictions (the software is rather out-of-date) which limit the scope of application (e.g. it can only process 32,000 observations).

impossible to generate empirical results in some cases – either because there is no existing empirical data on some phenomena in the present or because the data cannot exist yet as the phenomenon itself is going to happen in the future for the first time (like the development of influencing factors on the occupational status of the *third* generation of migrants). Secondly, it is a simulation's main goal to model assumed *to-be-changes* (stimuli) to proof if and how these changes affect the dependent variable(s). As an example, it could be assumed that a political intervention in the future will increase the probability of acquiring higher education for Turkish youths of the second generation stronger than expected on the basis of results from the early past. With the help of the simulation it can then be investigated how this assumed increase would affect this group's labor market performance in contrast to a scenario without this fictitious political intervention.

The base model, which is used here, is illustrated in figure 1. As the German Microcensus works as the starting dataset, the starting population in the simulation matches its population-definitions.[12],[13] The reason to include the whole population and not only the third generation of migrants in the simulation and analysis is, that demographical changes can only be predicted correctly when information on Germany's whole population is given. Furthermore, the inclusion of the second generation is mandatory if effects of social origin on the third generation's performance are to be simulated.

Another reason for simulating the whole population is the "comparative perspective" in the analysis of integration processes. As footnote 1 suggests, advances in integration can only be determined by comparing the (change of the) distribution of an integration indicator for a migrant group or generation with the (change of the) same distribution in other groups or generations and in the autochthone population as reference. This follows from the sociological-oriented integration theory, where "ethnic inequality" is a central point. Thus, we need to compare the simulation results of the third generation to the "Germans" as the reference as well as to the second generation.

Figure 1 shows the characteristics which are included in the simulation and the corresponding effect structures. The occupational status is the main dependent variable. Based on the latest state of research a resource-orientated theo-

---

[12] Total population in private households and institutions with their main and secondary residence in Germany:

Website GESIS: http://missyms.gesis.org/studie/erhebung/studienbeschreibung/ (checked at last 09/08/2016).

[13] Indeed the majority of the simulation is conducted by using the German Socio-Economic Panel (GSOEP). However, the GSOEP's population-definition is almost identical to the microcensus definition (see Spieß 2008 for details).

retical perspective is used to explain occupational status (see e.g. Kalter 2006, Esser 2009).[14] The base model was created on this theoretical perspective and was slightly adjusted and optimized after some empirical analysis. A closer look on the illustrated mechanisms reveals educational and occupational degrees as the main resources (Kalter et al. 2011, R. Becker 2011, Seibert & Solga 2005). Besides these, "softer" receiving-country-specific resources like cultural and social capital (see e.g. Kalter 2006) need to be considered. Above all, institutions (school, employers) are implicitly considered as sources that can contribute to inequality. This is guaranteed by empirical determining the "net" effect (under control of all other influencing factors) of belonging to an immigration group, which is explicitly not interpreted causally (see footnote 9).

The sub-model for "demography" ensures that natural population movements are modelled separately for each ethnical group. Only then, compositional effects can be generated correctly during the process of simulation. Besides the well documented effect of ethnical group membership on childbearing (e.g. Milewski 2010: 97, Birg 2003), the effects specified here strongly follow a sociologically based microsimulative modelling of the demographical change as seen in Leim (2008) and Wolters (2010).

After the base model's specification, the rules for the prospectively simulated developments are determined in the next step. These rules are then organized in modules. Each endogenous variable defines an own module.

---

[14] The effect of the "ethnic group" is not interpreted as causal. It rather represents a catch-all-category which also covers differences due to discrimination against particular ethnic minorities. In the analysis for estimating extrapolation parameters these effects (differences between ethnic groups) are reduced as much as possible in order to explain them with causal mechanisms. The unexplained rest (ethnic residuals, ethnic penalties – Kalter 2006) is, although not causal, included in the simulation to make the model as complete as possible. The implementation of these ethnic-group-effects is also necessary due to technical reasons: In the upcoming scenarios in this project the developments within particular migrant groups will be manipulated. Therefore, for every concerned group an influence parameter is needed.

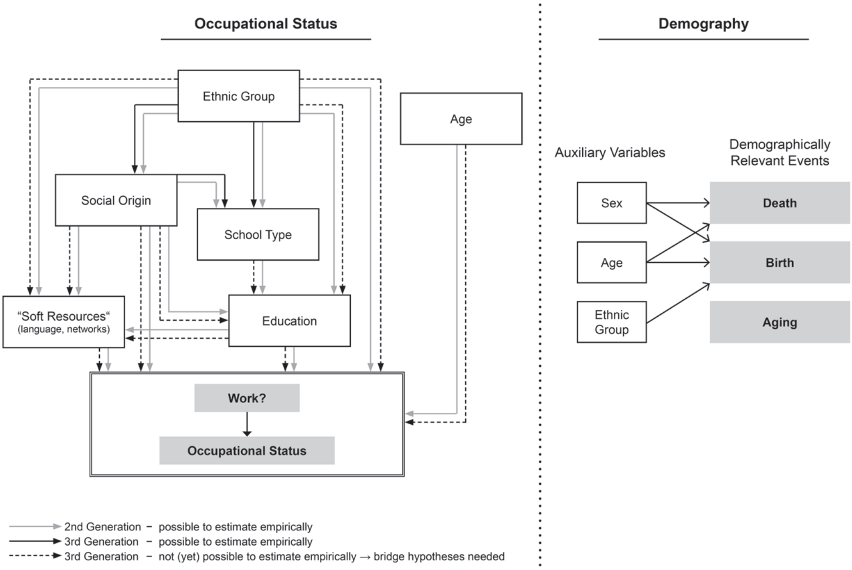


Figure 1: Base model for the microsimulation

For a first scenario, a simple approach is chosen here: Empirical findings regarding the developments in the early past are adopted from longitudinal analysis and serve unmodified as module parameters for groups of observations with the same combinations of attributes (so-called analytical groups) during the extrapolating process in the simulation. Influences, which can be transferred directly from analysis to simulation, are illustrated with solid arrows.[15] However, dashed arrows represent influences which cannot be estimated reliably for the analytical groups. In this context, these are influences in the third generation of migrants which refer to characteristics during the advanced adolescent- and early-adult-ages. Because of this, the unexaminable assumption is formulated, that these influences are identical to those from the second generation. Such a first scenario builds the basis for investigating whether the assumed compositional effects can be reproduced in the simulation or not. It will serve as a *base* scenario to which new scenarios that will represent concrete hypotheses about future integration developments can be compared.

In the next step of this project, which is not yet done, theory-driven scenarios are going to be defined, which can model the developments *within* different groups of migrants and which illustrate processes of mobility between genera-

---

[15] Furthermore, it is possible to estimate all depicted influences for natives as well as for the first generation of migrants. This is not explicitly shown here.

tions of migrants. For example, such developments could arise due to consequences of interventions in education-policy (better support for migrants) or changes of the social climate (like a decrease/increase of discrimination against migrants). With the help of statistical methods, it will then be possible to distinguish in the simulation results between effects of exactly these processes and compositional effects.

For theoretical reasons, immigration processes themselves are excluded in the simulation model (for now). We are focusing on the third generation of migrant groups that already live mainly in the second generation in Germany and immigrated in the 1960's-1990's. However, the offspring of new immigrants who would immigrate during the simulation period in the future would not reach the third generation in a high number before the simulation period ends. Of course, indirect effects are possible, because new immigration, like for example the flow of refugees who came in 2015/2016 to Germany, can influence the chances of the "old immigrants", especially on the labor market. This can be caused by social mechanisms (for example if the acceptance in some segments of the core society towards migrants in general changes) or by political decisions (if for example new laws are passed that concern "old migrants" too). We will account for those processes in the development of our scenarios.

Furthermore, it is also considered to model "emigration" in our future work, if further research analyses lead to the suggestion that emigration is selective concerning the success on the labor market of migrants living in Germany. Emigration can mean that "native Germans" leave Germany or that former migrants will go back to their country of origin (remigration) or to another country. Follow the comparative perspective in the analysis of integration processes, emigration of both groups is – if selective – important for the development of integration of migrants who stay in Germany.

4.2 Empirical Analysis (Extract) and their Implementation into the Simulation

Due to lack of space only one example for results from longitudinal analyses is reported. From those results, extrapolating parameters for the simulation are derived. Table 1 shows regression models for the crucial dependent variable "occupational status".[16] The population under analysis consists of all identified persons with second generation migration background as well as natives as the

---

[16] "Occupational status" has been operationalized with help of the vertical magnitude-prestige-scale (Wegener 1988). The scale was logarithmized due to its skewness.

reference group from the German Socio-Economic Panel (GSOEP), if valid values on all variables exist for these observations. Self-programmed cross-references (connecting children with their parents) were needed to ensure reliable assignments of social origin (education and occupational status). These cross-references use the full potential of the GSOEP-data.

Since occupational placement is, especially due to its crucial age-effect, a time varying continuous variable, random-effects-models were used for analyzing the panel-data (Rabe-Hesketh & Skrondal 2012, Hsiao 2014). Through a decomposition of the residual into a person-specific time-constant part and a time-varying leftover it is considered that a repeated survey of a person cannot be seen as a stochastically independent process referring to a person's first survey. In accordance with integration research's common practice hierarchical modelling techniques are used (see e.g. Kalter 2006): Firstly, the influence of the migrant group (under control of age and sex) is introduced to investigate in the next steps whether this influence can be reduced by adding independent variables derived from resource-theories.[17] Indeed, the consideration of one's own level of education and social origin leads to a distinct reduction of the non-causal effect of belonging to a migrant group (= reduction of group differences). It further confirms the common finding that influences of these resources are highly significant. This is why model 3 (grey) is adopted for simulation purposes. Analogous model estimates exist for all other endogenous variables, whereas the specific method is subject to the depended variable's level of measurement and the determination whether the considered variable is seen as time-constant or time-varying.

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + u_i + \varepsilon_{it} \quad (1)$$

In the meantime, implementing results from such regression models into microsimulation is widespread (see Milne et. al. 2015; McLay et al. 2015 for a general overview; Richiardi & Poggi 2014 for panel models with random effects). This can be seen as an extension of the primary microsimulation approach, which derives extrapolation parameters from cross-tables of empirically investigated frequencies (see e.g. Leim 2008, for details see Hannapel & Troitzsch 2015: 462). Indeed, dependent variables' probabilities can be distinguished (e.g., likelihood of dying depending on one's age) with this approach. However, if too

---

[17] Until now any consideration of "soft resources" is missing. Sophisticated multiple imputation techniques are needed to cope with a severe design-induced item nonresponse to avoid a drastic loss of sample size because different questionnaires with different filter management were used in the GSOEP over the years. This development is not yet completed.

many characteristics are controlled simultaneously, the number of cases in each cell is likely to decrease substantially.

Therefore, the estimated regression equations derived from the empirical panel analysis are implemented into the microsimulation to predict simulated values and to overcome this problem. The stochastic updating-element in the case of a random-effects-model (see equation 1) is defined by randomly drawing both residual components (highlighted in grey) for each individual. The person-specific error u_i is drawn once in the simulation, as long as it is unknown from empirical analyses. This is for example true for persons, who are generated in the simulation first. And the residual ε_{it} is calculated randomly in every simulation wave.[18] Adding both residual values to the fixed part of the equation results in the simulated y-value for each individual.

In contrast to this "linear" method to extrapolate *values* stochastically in the simulation, in the case of categorical dependent module variables, the *probability* for a category or event (for example, the probability to die in the case of the death module) is determined by the estimated equation in a [random-effects] logit/probit model, given meaningful independent variables [and the randomly drawn person-specific intercept (see Richiardi 2014 for the case of binary responses)].

4.3 Modules and Life-Course-Perspective

Figure 2 shows which individuals go through which module at which age in the simulation. Module variables that are unlikely to change once a "crucial age" is reached are kept constant such that the models do not grow unnecessarily complex (this applies mainly for social origin and educational background). Two transitions are modelled referring to schooling: First the transition from elementary to secondary school at the age of eleven. Secondly, individuals who do not already attend higher education schools (Gymnasium) after this transition, are given the chance to change to these schools at the age of 16 in the simulation (this reflects the real situation in Germany).

---

[18] For both error components i.i.d. is assumed and both are normally distributed: $u_i \sim N\,(0;\ \mathrm{s})$ and $\varepsilon_{it} \sim N\,(0;\ \mathrm{t})$. They are independent from each other and independent of the independent variables. s and t are estimated by variances which are calculated from the empirically determined residuals' parts for the analyzed sample.

Table 1: Longitudinal regression of occupational status

Occupational Status		**Random effects models (three variants)**					
		▨ **Chosen for simulation**					
Country of origin	**Germany**	*Reference*		*Reference*		*Reference*	
	Turkey	-0.314	***	-0.063	***	0.019	
	"Foreign Worker"-Countries	-0.275	***	-0.024	*	0.041	***
	Eastern Europe	-0.151	***	-0.059	***	-0.017	
	Other	-0.027	**	-0.013		0.023	
Sex	Male	*Reference*		*Reference*		*Reference*	
	female	-0.054	***	-0.063	***	-0.081	***
Age	Age	0.013	***	0.016	***	0.013	***
	Age^2	0.000	***	0.000	***	0.000	***
Social origin	Education parents			0.026	***	0.008	***
	Occupational status parents			0.003	***	0.002	***
Resources	Education					0.140	***
Dependent variable: magnitude prestige scale logarithmized; own computations; database: The German Socio-Economic Panel (1984-2012); *p < 0.1 **p < 0.05 ***p < 0.01.	Intercept	3.811	***	3.767	***	3.318	***
	sd(Intercept)	0.353		0.327		0.274	
	sd(Residual)	0.155		0.161		0.161	
	ρ	0.839		0.805		0.742	
	Chi2	6,044.670	***	6,134.330	***	15,905.420	***
	r2 within	0.012		0.019		0.020	
	r2 between	0.076		0.157		0.391	
	n (individuals)	39.271		22.151		21.083	
	N (person-years)	263.873		127.755		124.540	

Occupational status is modelled in two steps. First, it is ascertained whether a person is employed or not. Only if this is the case, the occupational placement is simulated. An indirect life-course-perspective follows from the fact that occupational placement is modelled in strong dependence of the individual's age.[19]

---

[19] Another example for considering the life-course indirectly, is educational attainment that is modelled in two steps. In Germany, the school type of the attended school is crucial for the result at the end (Ditton 2010: 250). The parents have to come to a decision very

However, no autoregressive processes are generated by making the actual occupational placement depended on its predicted value, because predicting coherent life-courses is not the aim of this model. It is more important to get reliable results on group level in order to investigate compositional effects further. Modelling autoregressive processes is not mandatory needed for this task. However, adding autoregressive elements to the regression models from table 1 would be possible without much effort.

For details on the organization of the demographic modules "birth" and "death" it is needed to refer to Leim (2008) and Wolters (2010) due to lack of space.

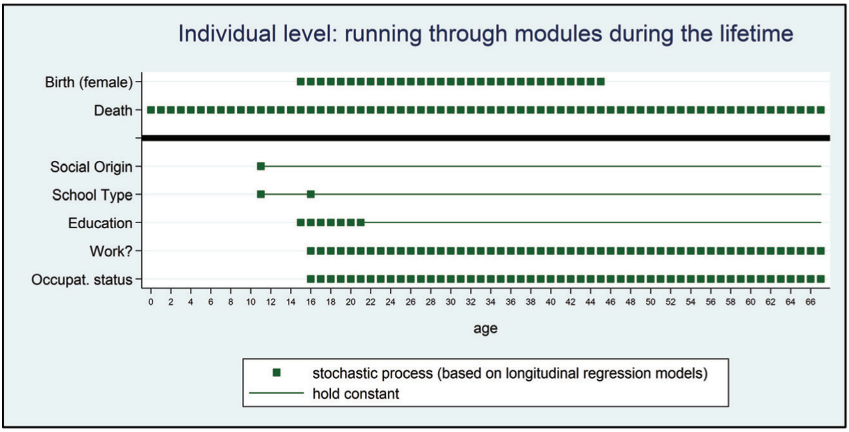


Figure 2: Modules in the microsimulation model

4.4 Starting Dataset for the Simulation

The scientific-use-file of the German Microcensus 2009 serves as the starting dataset. This dataset portrays Germany's demographic structure, especially combinations of country of origin and generation, at a sufficient level. Although there are more recent microcensus-datasets available, this rather old dataset is used, because questions regarding the origin of the participants' *parents* are asked in addition to the standard questions (participants' country of birth, chronology of citizenship) every four years since 2005 (MZG 2005). Thus, the

---

soon, at the children's' age nine or ten (Crul & Vermeulen 2003: 979). But if the highest type of school, the German Gymnasium is not visited by a child, he gets another chance to change to a Gymnasium at the age of 16. This mimics the real situation in Germany.

identification of secondary generation members becomes possible even if their parents do not live in the same household. Pre-analysis showed, that the addition of household-extern information to household-intern information leads to a significant improvement in identifying persons with migration background, who did not migrate themselves. It further protects from rough miss evaluation of intra-generational age-structures. According to the four-year-rhythm, the microcensus 2013 would also allow such analysis, but has not been available in time.

Table 2 shows the marginal distribution of core variables of the dataset and the mean age in each group. Due to capacity reasons, a ten-percent sample is drawn from the "Germans", who only serve as the reference category. It becomes apparent that "Germans" are older on average than individuals with migration background. Differences in age are further noticeable within the group of persons with migration background as well. Unsurprisingly, the average age is lower in numerical higher generations of migrants. Nevertheless, there are also differences within generations. The lower average age of migrants from East-European countries in contrast to people from the former recruitment countries like Turkey provides the basis for the development of compositional effects, because due to this age-differences the offspring (2nd generation) from these groups reaches the working age later. Overall the number of cases shows that the second generation, which is the most important generation for the simulation (because of the crucial effect of social origin for the third generation), is represented in a sufficiently high amount. This even accounts for the particular groups of migrants within this generation.

Table 2: Frequency distribution and mean age by migrant-group and -generation (upper value: mean age; lower value: absolute frequency)

	German	**First generation**	**Second generation**	**Third generation**	**Total**
German	45.34	(-)	(-)	(-)	45.34
	41,875	0	0	0	41,875
Turkish	(-)	43.52	15.36	6.36	29.45
	0	7,753	6,423	806	14,982
Eastern Europe	(-)	42.92	9.67	7.18	35.32
	0	11,222	3,284	39	1,4545
Former „foreign-worker"-countries	(-)	48.53	18.73	7.53	35.47
	0	3,563	2,139	383	6,085
Other	(-)	43.54	13.17	6.85	35.22
	0	22,709	8,017	428	31,154
Total	45.34	43.77	13.90	6.77	38.35
	41,875	45,247	19,863	1,656	10,8641

4.5 Simulation Results

A simulation was conducted with a forecast horizon of 50 years (2010-2060). A time interval compromised one year. Therefore, a longitudinal data set was produced with 50 time points. For space reasons, the general demographic trends are not shown here graphically, but the main points shall be mentioned:

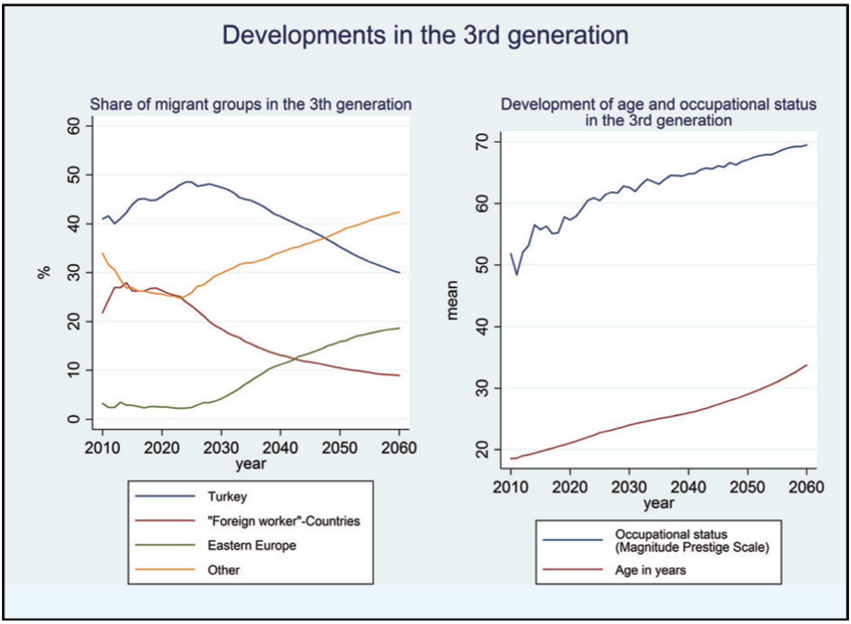
As expected, the population is declining over time. Since the focus is on the third generation of migrants, no new immigration is modelled here. The shrinking of the population (2060 the size of the population decreased to 56.7% of the population in the year 2010) due to low fertility rates is therefore plausible. Concomitantly, the average age is rising in the autochthone population and in all regarded groups of migrants.

The change in the composition of the group with a migrant background *by generations* is also as expected: While at the beginning of the simulation, the first (65%) and the second generation (35%) dominate, the third generation overtakes the first generation in 2050 and almost measures up to the second generation, which is the dominating generation since 2042.

In order to identify composition effects in the third generation, the development of the target variable in this generation must be traced back to the development of the composition of this generation by migrant groups, which differ in their success on the labor market and in the educational system (see figure 3). It becomes apparent that the proportion of Turkish origin migrants and those from the former recruitment countries decreases over time, while the proportion of migrants from Eastern European countries is increasing. The latter are, as pre-simulation longitudinal analyses show, better educated and have a better labor market performance on average, as the former groups. These differences are projected in the simulation into the future. Additionally, this scenario is constructed in a way that no changes of education and labor market performance *within* the migrant groups can appear. This allows to isolate composition effects easily.

But the fact that the average occupational status develops positively (figure 3 on the right) has another reason: the strong (quadratic) effect of age on the occupational status (see table 1). Since the average age of the third generation (also figure 3 on the right) naturally grows during the simulation, the projected occupational status necessarily also increases in time. The ethnic composition effect, which is, as expected, less strong than the age effect, cannot surface in that way.

Therefore, this effect of age was controlled in a next step (this was done in the analysis of already simulated data): The distribution of age was kept constant for each simulation step and the analysis was reduced to a specific age group. The young age group of 18-30 was chosen, since, as already stated, members of the third generation are very young on average, when the simulation starts. In order to keep the age distribution constant within this group, in every simulated wave a subsample of the third generation was randomly drawn so that the age was uniformly distributed per wave. This method is subject to a loss of information. This loss is, however, acceptable in order to achieve a perfect age-control (we work on methods to reduce this loss). Nevertheless, a database that allows acceptable illustrations separated by migrant groups is not reached until 2020, so that the starting point for the presented results is 2020 (figure 4).


Developments in the 3rd generation

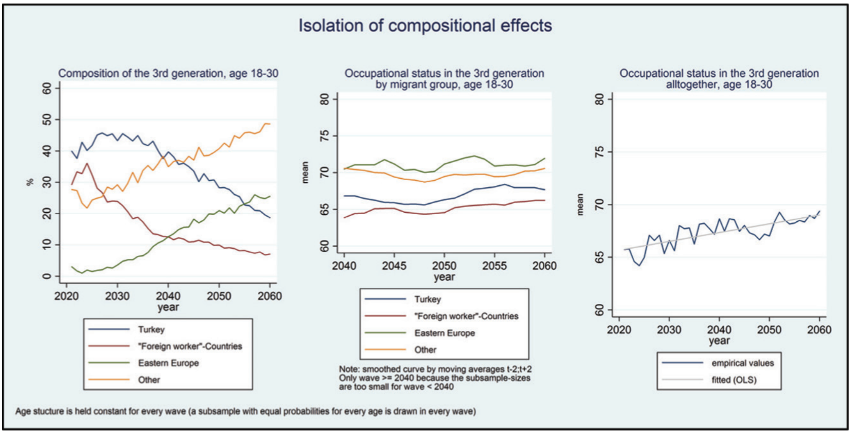
Share of migrant groups in the 3th generation

Development of age and occupational status in the 3rd generation

Figure 3: “Absolute” developments in the simulation

Figure 4 shows the results under control of the age effect. The left graph shows that the development of the shares of the particular migrant groups under age control is similar to the same development without such control (figure 3). The development of occupational status *within* the migrant groups can be seen in the middle graph in figure 4 and appears to be relatively constant, in spite of little fluctuations. This follows directly from the constructed simulation scenario here, in which the transition probabilities for occupational success for a particular migrant group are constant during the whole simulation time. Finally, the right graph shows the migration-group-related composition effect. The positive development of the occupational status over time, which is visible in spite of fluctuations, induced by the use of random experiments in groups with small case numbers, *can solely occur due to the changing composition of the third* *generation* (s. left graph). The OLS regression line, which lies behind the empirical curve, emphasizes the positive effect of time. It is impossible that a trend in *individual* changes caused this result. It solely results from the fact that migrant groups with different educational levels and consequently different labor market performances "enter" the third generation at different time points. In particular, migrants from Eastern Europe, who have immigrated later on average than migrants from the typical recruitment countries, and who therefore reach a significant proportion of the third generation lately in the simulated future, are responsible for the positive slope of the regression line, because they have a relatively good labor market performance (mediated through educational achievements: Segeritz et al. 2010).

This composition effect appears to be not very strong. The regression coefficient b is 0.084 (highly significant at the 0.1% level – despite the fact that, while testing H0: b=0, the aggregated data was treated as individual data, which is dramatically decreasing the sample size). However, the presented simulation here is based on a yet to be optimized database. It is expected that this effect with optimized data (particularly more precise construction of migrant groups) will emerge more clearly.



Note: smoothed curve by moving averages t-2;t+2
Only wave >= 2040 because the subsample-sizes are too small for wave < 2040

Age stucture is held constant for every wave (a subsample with equal probabilities for every age is drawn in every wave)

Figure 4: Composition effects in the simulation

This simulated scenario pictures a very rigid development, in which the affiliation to a migrant group determines the labor market success - mediated by other variables such as education, which also depend on the migrant group membership. This first model presented here should especially provide the basis to easily check whether the assumed composition effects can be (re)produced in the simulation. In the next step, we want to define scenarios that also model developments *within* migrant groups, based on relevant approaches in migration research. Such developments might result for example from education policies (e.g., better support for immigrants) or socio-climatic developments (e.g., reduction of discrimination). Using statistical methods, it will then be possible to distinguish between the effects of these processes within the migrant groups and composition effects. Following this path, the interplay of causal and composition mechanisms can be understood deeply.

5. Conclusions

First results of an ongoing research project about the future development of third generation immigrants' occupational status were presented, using the approach of dynamic microsimulation. In a first scenario, it was shown that the expected demographic developments can be reproduced and, especially, composition effects regarding the development of third generation immigrants can be identified. The composition effect was extracted in two steps: firstly, the development of occupational status *within* migrant groups was kept constant by the scenario definition. Secondly, the longitudinal age distribution was kept constant by the analysis of the simulation results. By doing so, only the change in the composition of third generation immigrants (by specific migrant groups with varying performance levels) can be made responsible for the observed increase of occupational status of the third generation over time. Thereby, it becomes obvious that the sequence of immigration waves in Germany since the 1950s up to the 1990s possesses the potential to produce composition effects, when a generation of migrants is examined as a whole longitudinally.

As the research projects proceeds, some optimizations are planned. Resulting from theoretical considerations, scenarios will be created, which allow for the inclusion of differing assumptions about the future integration development. In particular, theoretically driven hypotheses will be formulated about future developments of integration factors that can affect the development of the occupational status in the third generation. These developments could result from social mechanisms or policy interventions, like for example new regulations in the education policy that improve the performances in the education system within migrant groups – on individual level. The actual scenario, in which developments within migrant groups are kept constant, will serve as base scenario to which these new scenarios will be compared.

As addressed, it is also considered to model "emigration", in case we find more hints in our future research analyses for emigration to be selective concerning the success on the labor market of migrants living in Germany. In addition, the quality of data (for both the starting data and the empirical analyses) will be improved.

Soon the scientific-use-file of the German Microcensus 2013 will be released, which is the most recent dataset of its kind containing additional questions about the parental migration history. Moreover, there are negotiations with the German Federal Office of Statistics about the usage of raw data of microcensus surveys. This is because for the construction of the immigration background and the assignment to a group of migrants as well as to a generation, different information is needed (country of birth, history of citizenship; regarding both the respondent and his parents). However, the values (countries) of the corresponding variables are aggregated differently in the scientific-use-files of the German Microcensus (sometimes, nearly all Western Europe countries are merged to one category, in other cases, the particular countries stand on their own). Common ground in the microcensus is the distinction between four migrant groups as presented in this paper. However, from an integration theoretical point of view, a more sophisticated and precise differentiation of these groups would be essential. If the sample size allows for it, the definition of migrant generations will also become more specific, since there is evidence for considerable differences in the level of integration when a more precise distinction is used (Segeritz et al. 2010).

The regression models resulting from empirical analyses will also be extended. Particularly, "soft resources" such as language skills and the inclusion in inter-ethnical networks will be integrated in the empirical models using multiple imputation methods. For the determination of the probability to give birth, the *education* of prolific women will be prospectively incorporated. Likewise, multiple births will be modelled. Occasionally, the models will be elaborated by the inclusion of additional interaction effects. Furthermore, the models based on the GSOEP have to be replicated employing the microcensus data – where it is possible – to test the robustness of the empirical findings.

Two important steps to validate the quality of the simulation results will follow up these optimizations: With sensitivity analyses, the robustness of the results will be validated as a first step by examining the effects of small changes of the extrapolation parameters on the "overall result". In an unsystematic way this has already been conducted in the presented model and the manipulations only yielded minor effects on the course of labor market integration of third generation immigrants. As a second step an empirical test of the simulation results – a so-called "validation" (Hannapel & Troitzsch 2015: 482 f.) – will be conducted, in which the simulation results for the first years will be compared to available empirical findings. For this task the microcensus 2013 will also be a suitable dataset. Additionally, a reduced form of validation is already possible by adapting the starting dataset based on the microcensus 2009 to the restricted data (no parental migration history for children available, who do not live together with their parents anymore) of the microcensus 2010, and repeating the simulation in consideration of these restricted definitions.

Finally, the simulation will be enhanced to a “real” Monte-Carlo-simulation by repeating the whole simulation several times. Thus, the variance resulting from the random components of the stochastic extrapolation method can be represented and it will therefore be possible to calculate confidence intervals. In the next step, this simulation-inherent variance will be combined with the variance resulting from the fact that the starting data is just a random sample of the interesting population. Actually, we tend to realize this by means of resampling techniques like bootstrapping (Lappo 2016).

References

Alba, R. (2008): Why We Still Need a Theory of Mainstream Assimilation. In: Kalter, F. (Ed.): Migration und Integration. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 48, pp. 37-56.

Alba, R. / Handl, J. / Müller, W. (1994): Ethnische Ungleichheit im deutschen Bildungssystem, Kölner Zeitschrift für Soziologie und Sozialpsychologie 46 (2), pp. 209-237.

Ballarino, G. / Panichella, N. (2013): The Occupational Integration of Male Migrants in Western European Countries. Assimilation or Persistent Disadvantage? International Migration 53 (2), pp. 338-352.

Basilio, L. / Bauer, T. (2010): Transferability of Human Capital and Immigrant Assimilation. An Analysis for Germany. Discussion Paper (4716), IZA Bonn.

Becker, B. (2010): Ethnische Unterschiede bei der Kindergartenselektion. Die Wahl von unterschiedlich stark segregierten Kindergärten in deutschen und türkischen Familien. In: Becker, B. / Reimer, D. (Ed.): Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 17-47.

Becker, B. (2011): Cognitive and Language Skills of Turkish Children in Germany. A Comparison of the Second and Third Generation and Mixed Generational Groups. In: International Migration Review 45 (2), pp. 426-459.

Becker, B. / Biedinger, N. (2006): Ethnische Bildungsungleichheit zu Schulbeginn. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 58 (4), pp. 660-684.

Becker, B. / Biedinger, N. (2016): Ethnische Ungleichheit in der Grundschule. In: Diehl, C. / Hunkler, C. Kristen, C. (Ed.): Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 433-474.

Becker, R. (2011): Integration von Migranten durch Bildung und Ausbildung. Theoretische Erklärungen und empirische Befunde. In: Becker, R. (Ed.): Integration durch Bildung. Bildungserwerb von jungen Migranten in Deutschland. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 11-36.

Bertrand, M. / Mullainathan, S. (2004): Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. In: The American Economic Review 94 (4), pp. 991-1013.

Birg, H. (2003): Dynamik der demographischen Alterung, Bevölkerungsschrumpfung und Zuwanderung in Deutschland. Prognosen und Auswirkungen. In: Aus Politik und Zeitgeschichte. Beilage zur Wochenzeitung ‚Das Parlament' B 20, pp. 6-17.

Blossfeld, H.-P. / von Maurice, J. / Schneider, T. (2011): The National Educational Panel Study. Need, main features, and research potential. In: Blossfeld, H.-P., Roßbach, H.-G., von Maurice, J. (Ed.): Education as a Lifelong Process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft. Sonderheft 14, pp. 5-17.

Bonin, H. (2013): Mikrosimulation ausgewählter ehe- und familienbezogener Leistungen im Lebenszyklus. Gutachten für die Prognos AG. Forschungsbericht. Mannheim: Zentrum für Europäische Wirtschaftsforschung GmbH.

Bornmann (2010): Die analytische Soziologie: Soziale Mechanismen. DBO-Theorie und Agentenbasierte Modelle. In: Österreichische Zeitschrift für Soziologie 35 (4), pp. 25-44.

Brück-Klingberg, A. / Burkert, C. / Garloff, A. / Seibert, H. / Wapler, R. (2011): Does higher education help immigrants find a job? A Survival analysis. Discussion Paper (6), IAB Nürnberg.

Bryon, G. / Dekkers, G. / de Menten, G. (2011): LIAM 2 User Guide documentation MiDaL Project Document.

Crul, M. / Vermeulen, H. (2003): The second generation in Europe. In: International migration review 37 (4), pp. 965-986.

Damalang, A. / Haas, A. (2006): Arbeitsmarkteinstieg nach dualer Berufsausbildung. Migranten und Deutsche im Vergleich. Forschungsbericht (17), IAB Nürnberg.

Dekkers, G. / Belloni, M. (2009): A classification and overview of micro simulation models, and the choices made in MIDAS. In: Dekkers, G. / Buslei, H. / Cozzolino, M. / Desmet, R. / Geyer, J. / Hofmann, D. / Raitano, M. / Steiner, V. / Tanda, P. / Tedeschi, S. / Verschueren, F.: What are the consequences of the European AWG-projections on the adequacy of pensions? An application of the dynamic micro simulation model MIDAS for Belgium, Germany and Italy, ENEPRI Research Report (65), AIM WP4. Brussels: Centre for European Policy Studies, pp. 5-24.

Diehl, C. / Friedrich, M. / Hall, A. (2009): Jugendliche ausländischer Herkunft beim Übergang in die Berufsausbildung. Vom Wollen, Können und Dürfen. In: Zeitschrift für Soziologie 38 (1), pp. 48-67.

Diehl, C. / Fick, P. (2016): Ethnische Diskriminierung im deutschen Bildungssystem. In: Diehl, C. / Hunkler, C. Kristen, C. (Ed.): Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 243-286.

Ditton, H. (2010). Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit. In: Bildung als Privileg. VS Verlag für Sozialwissenschaften, pp. 247-275.

Dollmann, J. (2016): Der Übergang von der Primar- in die Sekundarstufe. In: Diehl, C. / Hunkler, C. Kristen, C. (Ed.): Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 517-543.

Dustman, C. / van Soest, A. (2002): Language and the Earnings of Immigrants. In: Industrial and Labor Relations Review 55 (3), pp. 473-492.

Eisenstadt, S. N. (1953): Analysis of patterns of immigration and absorption of immigrants. In: Population Studies. A Journal of Demography 7 (2), pp. 167-180.

Erlinghagen, M. / Scheller, F. (2011): Migrationshintergrund und Beteiligung an beruflicher Weiterbildung. In: Berger, P. A. / Hank, H. / Tölke, A. (Ed.): Reproduktion von Ungleichheit durch Arbeit und Familie. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 301-325.

Esser, H. (1980): Aspekte der Wanderungssoziologie. Assimilation und Integration von Wanderern, ethnischen Gruppen und Minderheiten. Eine handlungstheoretische Analyse. Darmstadt u.a.: Luchterhand.

Esser, H. (1999): Inklusion, Integration und ethnische Schichtung. In:

Journal für Konflikt- und Gewaltforschung 1 (1), pp. 5-34.

Esser, H. (2001): Integration und ethnische Schichtung. Arbeitspapier (40), Mannheimer Zentrum für europäische Sozialforschung.

Esser, H. (2006): Sprache und Integration. Frankfurt am Main u.a.: Campus-Verlag.

Esser, H. (2008): Assimilation, ethnische Schichtung oder selektive Akkulturation? Neuere Theorien der Eingliederung von Migranten und das Modell der intergenerationalen Integration. In: Kalter, F. (Ed.): Migration und Integration. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 81-107.

Esser, H. (2009): Pluralisierung oder Assimilation? Effekte der multiplen Inklusion auf die Integration von Migranten. In: Zeitschrift für Soziologie 38 (5), pp. 358-378.

Fereidooni, K. (2011): Schule – Migration – Diskriminierung. Ursachen der Benachteiligung von Kindern mit Migrationshintergrund im deutschen Schulwesen. Wiesbaden: VS Verlag für Sozialwissenschaften.

Fibbi, R. / Lerch, M. / Wanner, P. (2006): Unemployment and Discrimination against Youth of Immigrant Origin in Switzerland. When the Name Makes a Difference. In: Journal of International Migration and Integration / Revue de l'integration et de la migration international 7 (3), pp. 351-366.

Flache, A. / Mäs, Michael (2015): Multi-Agenten-Modelle. In: Braun, N. / Saam, N.-J. (Ed.): Handbuch Modellbildung und Simulation in den Sozialwissenschaften. Wiesbaden: Springer VS, pp. 491-514.

Flory, J. / Stöwhase, S. (2012): MIKMOD-ESt. A Static Microsimulation Model of Personal Income Taxation in Germany. In: International Journal of Microsimulation 5 (2), pp. 66-73.

Fuchs, J. (2009): Demografische Effekte auf das künftige Arbeitsangebot in Deutschland. Eine Dekompositionsanalyse. In: Schmollers Jahrbuch (129), pp. 571-595.

Galler, H.P. /Wagner, G.G. / (1986): The Microsimulation Model of the Sfb 3 for the Analysis of Economic and Social Policy, in: Orcutt G. / Merz, J. / Quinke, H. (Eds.): Microanalytic Simulation Models to Support Social and Financial Policy Amsterdam. pp. 227-247.

Gilbert, N. / Troitzsch, K.-G. (2005): Simulation for the Social Scientist. 2nd edition. Open University Press u.a.: Maidenhead, Berkshire u.a.

Gomolla, M. / Radtke, F.-O. (2009): Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gordon, M. M. (1964): Assimilation in American Life. The Role of Race, Religion and National Origin. New York: Oxford University Press.

Granato, N. (2009): Effekte der Gruppengröße auf die Arbeitsmarktintegration von Migranten. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 61 (3), pp. 387-409.

Gresch, C. / Kristen, C. (2011): Staatsbürgerschaft oder Migrationshintergrund? Ein Vergleich unterschiedlicher Operationalisierungsweisen am Beispiel der Bildungsbeteiligung. In: Zeitschrift für Soziologie 40 (3), pp. 208-227.

Hannappel, M. (2015): (K)ein Ende der Bildungsexpansion in Sicht?! Ein Mikrosimulationsmodell zur Analyse von Wechselwirkungen zwischen demographischen Entwicklungen und Bildungsbeteiligung. Dissertation. Social Science Simulations 13. Marburg: Metropolis-Verlag.

Hannappel, M. / Troitzsch, K.-G. (2015): Mikrosimulationsmodelle. In: Braun, N. / Saam, N.-J. (Ed.): Handbuch Modellbildung und Simulation in den Sozialwissenschaften. Wiesbaden: Springer VS, pp. 455-490.

Hsiao, C. (2014): Analysis of Panel Data. Econometric Society Monographs. Cambridge: Cambridge University Press.

Hans, S. (2015): Fack ju Mozart, fack ju Göhte? Hochkulturelle Lebensstiele bei Jugendlichen und die Bedeutung des Migrationshintergrundes. In: Rössel, J / Roose, J. (Ed.): Empirische Kultursoziologie. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 129-161.

Hunkler, C. (2010): Ethnische Unterschiede beim Zugang zu Ausbildung und Erwerb von Ausbildungsabschlüssen. In: Becker, B. / Reimer, D. (Ed.): Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 213-250.

Hunkler. C. (2016): Ethnische Unterschiede beim Zugang zu beruflicher Ausbildung. In: Diehl, C. / Hunkler, C. Kristen, C. (Ed.): Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 597-632.

Imdorf, C. (2011): Wie Ausbildungsbetriebe soziale Ungleichheit reproduzieren. Der Ausschluss von Migrantenjugendlichen bei der Lehrlingsselektion. In: Krüger, H.-H. / Rabe-Kleberg, U. / Kramer, R.-T. / Budde, J. (Ed.): Bildungsungleichheit revisited. Bildung und soziale Ungleichheit vom Kindergarten bis zur Hochschule. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 261-276.

Kalter, F. (2006): Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft. Zugleich ein Replik auf den Beitrag von Holger Seibert und Heike Solga: „Gleiche Chancen dank einer abgeschlossenen Ausbildung? (ZfS 5/2005). In: Zeitschrift für Soziologie 35 (2), pp. 144-160.

Kalter, F. (2008a): Ethnische Ungleichheit auf dem Arbeitsmarkt. In: Abraham, M. / Hinz, T. (Ed.): Arbeitsmarktsoziologie. Probleme, Theorien, empirische Befunde. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 303-332.

Kalter, F. (2008b): Stand, Herausforderungen und Perspektiven der empirischen Migrationsforschung. In: Kalter, F. (Ed.): Migration und Integration. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 11-36.

Kalter, F. / Granato, N. (2002): Demographic Change, Educational Expansion, and Structural Assimilation of Immigrants. The Case of Germany. In: European Sociological Review 18 (2), pp. 199-216.

Kalter, F. / Granato, N. / Kristen, C. (2011): Die strukturelle Assimilation der zweiten Migrantengeneration in Deutschland. Eine Zerlegung gegenwärtiger Trends. In: Bekker, R. (Ed.): Integration durch Bildung. Bildungserwerb von jungen Migranten in Deutschland. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 257-288.

Kempert, S. / Edele, A. / Rauch, D. / Wolf, K. / Paetsch, J, / Darsow, A. / Maluch, J. / Stanat, P. (2016): Die Rolle der Sprache für zuwanderungsbezogene Ungleichheiten im Bildungserfolg. In: Diehl, C. / Hunkler, C. Kristen, C. (Ed.): Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 157-242.

Klabunde, A. / Zinn, S. / Leuchter, M. / Willekens, F. (2015): An agent-based decision model of migration, embedded in the life course. Model description in ODD+D format. MPIDR Working Paper 2015-002, Max Planck Institute for Demographic Research.

Kluge, F. A. / Zagheni, E. / Loichinger, E. / Vogt, T. C. (2014): The Advantages of Demographic Change after the Wave. Fewer and Older, but Healthier, Greener, and More Productive? MPIDR Working Paper 2014-003, Max Planck Institute for Demographic Research.

Kristen, C. / Dollmann, J. (2010): Sekundäre Effekte der ethnischen Herkunft. Kinder aus türkischen Familien am ersten Bildungsübergang. In: Becker, B. / Reimer, D. (Ed.): Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 117-144.

Lappo, S. (2016): Variance estimation of inequality and poverty indicators. Assessing sampling variability in microsimulation. Presentation slides on European Meeting of the International Microsimulation Association, 22-23 September 2016, Budapest, Hungary, https://conference.onyf.hu/wp-content/uploads/2016/09/LAPPO_SAMPO.pdf (30.11.2016).

Leim, I. (2008): Die Modellierung der Fertilitätsentwicklung als Folge komplexer individueller Entscheidungsprozesse mit Hilfe der Mikrosimulation. Marburg: Metropolis-Verlag.

Li, J. / O'Donoghue, C. (2013): A survey of dynamic microsimulation models. Uses, model structure and methodology. In: International Journal of Microsimulation 6 (2), pp. 3-55.

Lindenberg, S. (1986): Individual economic ignorance versus social production functions and precarious enlightenment. Comments on Tullock's View of Rent Seeking in Dictatorships. In: Journal of Institutional and Theoretical Economics 142 (1), pp. 20-26.

de Menten, G / Dekkers, G / Bryon, G. / Liégeois. P. / O'Donoghue, C. (2014): LIAM2. A new open source development tool for discrete-time dynamic microsimulation models. In: Journal of Artificial Societies and Simulation, 17(3)9

McLay, J.M. / Lay-Yee, R. / Milne, B.J. / Davis, P. (2015): Regression-Style Models for Parameter Estimation in Dynamic Microsimulation. An Empirical Performance Assessment. In: International Journal of Microsimulation, 8 (2), pp. 83-127.

Milewski, N. (2010): Fertility of immigrants. A Two-Generational Approach in Germany. Berlin / Heidelberg: Springer-Verlag.

Milne, B.J. / Lay-Yee, R. / McLay, J.M. / Pearson, J. / von Randow, M. / Davis, P. (2015): Modelling the Early life-course (MELC). A Microsimulation Model of Child Development in New Zealand. In: International Journal of Microsimulation, 8 (2), pp. 28-60.

MZG (2005): Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt sowie die Wohnsituation der Haushalte. Mikrozensusgesetz 2005. MZG 2005. Federal Ministry for Justice and Consumerrights, Berlin.

Olczyk, M. / Seuring, J. / Will, G. / Zinn, S. (2016): Migranten und ihre Nachkommen im deutschen Bildungssystem. Ein aktueller Überblick. In: Diehl, C. / Hunkler, C. Kristen, C. (Ed.): Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 33-72.

Park, R. E. (1928): Human Migration and the Marginal Man. In: American Journal of Sociology 33 (6), pp. 881-893.

Portes, A. / Rumbaut, R. G. (2006): Immigrant America. A portrait. Berkeley u.a.: University of California Press.

Rabe-Hesketh, S. / Skrondal, A. (2012): Multilevel and Longitudinal Modeling Using Stata. Continuous Responses. Texas: Stata Press.

Radtke, F.-O. (2004): Schule und Ethnizität. In: Helsper, W. / Böhme, J. (Ed.): Handbuch der Schulforschung. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 651-672.

Richiardi, M. / Poggi, A. (2014): Imputing Individual Effects in Dynamic Microsimulation Models. An application to household formation and labour market participation in Italy. In: International Journal of Microsimulation, 7 (2), pp. 3-39.

Richiardi, M. (2014): Forecasting with Unobserved Heterogeneity. The LABORatorio R. Revelli, http://www.laboratoriorevelli.it/_pdf/wp123.pdf (08.11.2016).

Sauerbier, T. (2002): UMDBS. A New Tool for Dynamic Microsimulation. In: Journal of Artificial Societies and Social Simulation 5 (2).

Schacht, D. / Kristen, C. / Tucci, I. (2014): Interethnische Freundschaften in Deutschland. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 66 (3), pp. 445-458.

Schatz, C. (2010): Altersvorsorge in Deutschland 2005. Methodenbericht Teil II: Fortschreibung und Anwartschaftsberechnung. München: TNS Infratest Sozialforschung.

Segeritz, M. / Walter, O. / Stanat, P. (2010): Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland. Evidenz für segmentierte Assimilation? In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 62 (1), pp.113-138.

Seibert, H. / Solga, H. (2005): Gleiche Chancen dank einer abgeschlossenen Ausbildung? Zum Signalwert von Ausbildungsabschlüssen bei ausländischen und deutschen jungen Erwachsenen. In: Zeitschrift für Soziologie 34 (5), pp. 364-382.

Siegert, M. / Roth, T. (2013): Söhne bevorzugt? Geschlechtsspezifische Unterschiede beim Gymnasialbesuch türkischstämmiger Schülerinnen und Schülern. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 65 (1), pp. 49-72.

Sommer, B. (2007): Der Demographische Wandel in den nächsten Jahrzehnten. In: Höhn, C. / Dorbritz, J. (Ed.): Demographischer Wandel. Wandel der Demographie. Festschrift für Prof. Dr. Karl Schwarz. Schriftenreihe Band 37, pp. 275-300.

Spielauer, M. (2011). What is social science microsimulation? Social Science Computer Review, 29 (1), pp. 9-20.

Spieß, M. (2008): Gewichtung und Hochrechnung mit dem SOEP. German Insitute for Economic Research.

Sprietsma, M. (2009): Discrimination in Grading? Experimental Evidence from Primary School. Discussion Paper (09-074), Zentrum für Europäische Wirtschaftsforschung Mannheim.

Statistisches Bundesamt (2009): Bevölkerung Deutschlands bis 2060. 12. koordinierte Bevölkerungsvorausberechnung. Wiesbaden: Statistisches Bundesamt.

Stein, Petra (1997): Konstruktion und sozialwissenschaftliche Anwendung finiter Mischungen von Kovarianzstrukturmodellen. Köln: Josef Eul Verlag.

Sürig, I. / Wilmes, M. (2011): Die Integration der zweiten Generation in Deutschland. Ergebnisse der TIES-Studie zur türkischen und jugoslawischen Einwanderung. IMIS-Beiträge, Vol. 39. Osnabrück.

Taft, R. (1953): The shared frame of reference concept applied to the assimilation of immigrants. In: Human Relations 6, pp. 45-55.

Treichler, A. (2014): Soziale Ungleichheit auf dem Arbeitsmarkt. Ethnische Diskriminierung oder unzureichendes Humankapital? In: Migration und Soziale Arbeit 14 (3), pp. 205-212.

Tuppat, J. / Becker, B. (2014): Sind türkischstämmige Kinder beim Schulstart im Nachteil? Die Bedeutung genereller und aufnahmelandspezifischer Kompetenzen für die Wahrscheinlichkeit einer Rückstellung. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 66 (2), pp. 219-241.

Wegener, B. (1988): Kritik des Prestiges. Opladen: Westdeutscher Verlag.

Wimmer, A. (2008): Ethnische Grenzziehungen in der Immigrationsgesellschaft. Jenseits des Herder'schen Commonsense. In: Kalter, F. (Ed.): Migration und Integration. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 57-80.

Wolters, U. (2010): Die Prognose der sozio-demographischen Entwicklung mit Hilfe der dynamischen Mikrosimulation, untersucht am Beispiel der Prognose der sozio-demographischen Entwicklung der Bundesrepublik Deutschland bis in das Jahr 2050. Unpublished diploma-thesis, University of Duisburg-Essen.

Zhou, M. (1999): Segmented Assimilation. Issues, Controversies and Recent Research on the New Second Generation. In: Hirschman, C. / Kasnitz, P. / DeWind, J. (Ed.): The Handbook of International Migration. The American Experience. New York: Russell Sage Foundation Publications, pp. 196-211.

The simulator of individual dynamic decisions, sidd

A framework for structural dynamic microsimulation modelling

JUSTIN VAN DE VEN[1]

Abstract: This paper describes a structural dynamic microsimulation model that is designed to analyse the distributional implications of policy alternatives over appreciable periods of time. The model generates panel data describing a wide range of characteristics for each adult in an evolving population cross-section. Savings and labour supply decisions are endogenous in the model, making it useful for exploring the implications of changing incentives implied by policy counterfactuals. The model has been developed to reduce the hurdles associated with use of dynamic programming methods in microsimulation: the model – including associated source code – can be freely download from the internet at www.simdynamics.org.

Key words: structural, dynamic programming, microsimulation, population, cross-section

JEL classifications: C51, C61, C63, H31

1. Introduction

The complexity of the alternative considerations influencing policy design generate demand for analytical tools that identify the likely implications of policy alternatives. Current limitations of computing power and our theoretical understanding of the policy environment mean that there does not exist a single tool that is best for all analytical contexts, resulting in the development of a wide range of alternative methodologies. The first problem that must be solved by anyone seeking to develop analytical capacity in this field is consequently to choose an analytical approach that is fit for purpose. Choose a tool that is too simplistic, and subsequent analysis risks being criticised for lacking realism. Choose a tool that is too complex, and subsequent analysis risks being made impractical. This paper provides a non-technical description of one

[1] Jvandeven@niesr.ac.uk; National Institute of Economic and Social Research, UK & Melbourne Institute of Applied Economic and Social Research, Australia.

such tool; the Simulator of Individual Dynamic Decisions, SIDD. The paper discusses the strengths and weaknesses of SIDD, and best-practices methods of the model's use.

SIDD is an example of a structural dynamic microsimulation model. Working backwards through this description, it is a "model" in the sense that SIDD is designed to generate the logical implications of a set of stylised assumptions about the real-world policy environment. It is a "microsimulation" in the sense that it is designed to reflect differences between individual micro-units that are of interest from a policy making perspective; in the case of SIDD, individual adults. SIDD is "dynamic" because it projects individual adults through time. And it is "structural" because it is designed to project variation of adult decision making based on a theoretical description of how decisions are taken. These terms are explained at greater length in Section 2 of the paper.

Dynamic microsimulation models, like SIDD, are designed for exploring the distributional implications of policy alternatives through time. It is important to explain two terms in the preceding sentence. The term "policy" is meant to extend beyond explicit government directives (like transfer payments), to encompass any change to the modelled environment. This might include, for example, alternative assumptions concerning marital rates, fertility rates, likelihoods of migration, and so on. Furthermore, the phrase "distributional implications" refers to effects that vary by any modelled characteristic that can differ between simulated individuals. Key characteristics over which policy effects are often evaluated include income, wealth, age, birth year, relationship status, and the existence of dependent children. Whereas static microsimulation models are capable of exploring the distributional implications of policy alternatives at a point in time, the distinguishing feature of dynamic microsimulation models is that they can do the same *through time*.

Dynamic microsimulation models suitable for analysing the distributional implications of public policy have been growing in number and sophistication since the ground breaking-work of Orcutt (1957). In a recent review of the associated literature, Li and O'Donoghue (2013) refer to 66 such models, constructed for 19 countries. Recent model development has benefitted from increased computing capacity, greater availability of high quality micro data, development of increasingly sophisticated analytical tools, and the advent of generic software packages (e.g. GENESIS, Edwards, 2010; LIAM, O'Donoghue *et al.* 2009). Nevertheless, developing a dynamic microsimulation model remains challenging, and all models currently in use involve trade-offs of one form or another.

Dynamic microsimulation models are often used to make projections into the very long-run, extending beyond a 20 year time horizon. Prevailing computational, theoretical, and empirical limitations call into question any attempt to generate a forecast of how the world will look so far into the future. Given this limitation, dynamic microsimulation models are primarily useful for conducting controlled experiments that are designed to reveal the influence of a single policy feature of interest. This is commonly achieved by comparing projections from two simulated scenarios, where the only differences between the simulations concern the policy counterfactual of interest. In this context, alternative approaches to dynamic microsimulation adapt the respective models to consider specific subjects. In the case of SIDD, the model's framework is specifically adapted to explore the long-term implications for the population cross-section of a change in the incentives embodied by a policy counterfactual.

How would the distribution of income be affected in the long-term if taxes on labour income of people with low wages were reduced, and off-set by a coincident reduction in the value of contributory state pensions? What are the long-term budgetary implications of alternative assumptions concerning future trends in fertility, wage rates, and longevity? How might savings and employment be affected by the introduction of a subsidised savings scheme targeted toward people at the bottom of the wealth distribution? SIDD is an ideal tool for exploring policy questions like these, within context of a controlled experimental environment.

The remainder of the paper is comprised of four sections. Section 2 extends upon the above discussion to place the SIDD microsimulation model in context of alternative methodological approaches that are currently employed. Section 3 describes specifics of the SIDD framework. Directions for further research are set out in a concluding section.

2. Modelling Methodology

As noted in the introduction, at present there is no single "best" microsimulation approach for exploring the effects of alternative government policies. A range of current best-practice methods currently exist, each with their own relative advantages that adapt them to selected question(s) of policy interest. There are two key dimensions over which such studies vary: the dimensionality of the considered population; and the description assumed for individual behaviour.[2] This

---

[2] A third dimension that is now emerging in the microsimulation literature concerns the approach taken to simulate prices, including the returns to investment and employment. At present, the vast majority of microsimulation models generate prices, without taking account of macro-economic feed-back effects associated with balancing supply and demand. A growing strand of literature seeks to address this limitation; see, for example, Richiardi and Richardson (2016) for a concerted effort to facilitate convergence of (agent-

section provides a high-level overview of the available alternatives concerning these two dimensions, and concludes with discussion of the relative merits of the methodology adopted for the SIDD microsimulation framework.

2.1 The simulated population

The literature can broadly be classified into three classes concerning the in-scope population: selected individual cases; a population group (usually cross-section) at a given point in time; and a population group through time.

Individual case studies

One way to simplify analysis is to focus on a selected set of individual "cases". In regards to tax and transfer policy, the case is usually defined as a person with a selected set of characteristics, and analysis focuses upon the financial implications of the policy of interest either by ex-post observation or ex-ante imputation. As such, the case study approach – sometimes referred to as a "standard simulation" – is capable of considering both existing policies and the implications of counterfactual policy reforms, indicating the ways in which individual circumstances interact with the policies of interest. Many microsimulation models consequently include features that permit isolated case-study analysis: see, for example, the training data feature of EUROMOD.

Steventon *et al*. (2007) use case studies to explore the implications of increasing the trivial commutation for incentives to participate in a state sponsored private pension introduced in the UK in 2008. Steventon *et al*. (2007) focus their analysis on five stereotypical types of individual, and provides a nice example of the essential mechanics that underlie the case study approach. The terms of their analysis are highly stylised to facilitate transparency. The particular cases that are considered for analysis are tailored to the policy reform of interest, focussing upon individuals with broken work histories and modest wages.

By focussing upon a single "type" of individual, case studies can help to clarify the practical implications of alternative reform scenarios. This same feature of the approach, however, makes it ill-suited to exploring population wide effects of policy reform options, as such effects require consideration of a very wide range of alternative case-specific contexts. This has led to the development of microsimulation models for analysing the effects of policy on broad segments of the population.

---

based) macro- and micro-simulation modelling, and Dekkers *et al*. (2015) for an example of a recent effort to integrate macro-side responses in a microsimulation context.

Static microsimulation

Microsimulation models were first applied in an economic context by Guy Orcutt in 1957, and are now commonly used for analysing government policy (Orcutt, 1957). In these models, each micro-unit (also referred to as agent) from a population is individually represented. This facilitates analysis of heterogeneity and diversity within the simulated population. As such, microsimulation models are particularly useful for policy analyses where the effects depend upon individual specific circumstances, or where distributional issues are a focus of interest.

A static microsimulation model is designed to generate output relevant for a single point in time. Variants of this approach that have been adapted to consider the effects of tax and transfer policy commonly start with a reference dataset that describes relevant circumstances for each family unit, with the sample designed to reflect a population cross-section. The information defined by the reference dataset is combined with assumptions concerning agent behaviour and the rules and regulations of an assumed transfer system to generate various measures of interest, including income gross and net of transfer payments, net tax take, employment rates, and so on.

Alternative behavioural assumptions are discussed at length in Section 2.2. The simplest modelling approach is to assume that behaviour remains invariant to changes to the policy environment. In the UK, this approach is adopted for the static microsimulation model TAXBEN developed at the Institute for Fiscal Studies (IFS; see Giles and McCrae, 1995). A similar approach is adopted for the countries of the European Union by the static microsimulation model EUROMOD maintained at the University of Essex (see Sutherland and Figari, 2013). Such models can, however, be fairly easily adapted to permit labour supply responses to be inferred using a theoretical framework, an adaptation that now has a long history. Early examples of this literature include Blundell *et al.* (1984), Duncan (2001), and Creedy *et al*. (2002); for more recent literature see Liegeois and Islam (2013), Capéau and Decoster (2016), and Capéau *et al.* (2016).

Static microsimulation models consequently provide a potentially powerful tool for exploring how policy change is likely to affect a population at a given point in time. By treating each micro-unit of a reference dataset as a separate case study, microsimulation can be used to build up a macro-level picture of the effects of a change to transfer policy. The micro-detail permit a very wide set of analyses to be conducted, from the consideration of population averages and budgetary impact of policy alternatives, through to distributional effects. Static microsimulation models do not, however, allow the analyst to consider the effects of policy through time – for that we need a dynamic microsimulation framework.

Dynamic microsimulation

Dynamic microsimulation models are designed to project a reference population through time. Starting from a static microsimulation model, the simplest approach for projecting the reference population through time is to re-weight each micro-unit in the reference dataset to reflect anticipated temporal variation. This approach is commonly referred to as "static aging", and is usually considered appropriate for projecting population circumstances over limited time-horizons (e.g. up to 5 years).

The problem with using static aging to project a population beyond a short-term horizon is that it becomes increasingly difficult to infer how a change in policy will influence the relative weighting of alternative micro-units reported in the reference dataset. Suppose, for example, that a change in tax and benefits policy had the effect of increasing labour supply. Increased labour supply over a short time horizon can be expected to have a limited impact on the underlying circumstances of individuals, such as their wages and savings. Over longer time horizons, however, both wages and wealth are likely to be influenced by a sustained increase in labour supply, complicating attempts to specify the weighting adjustments required for static aging.

Furthermore, static aging does not permit analysis of the effects of a policy reform on individuals *through* time. A model based on static aging therefore cannot provide information about how a policy is likely to influence alternative individuals during the course of their respective lives.

Dynamic population ageing is designed to address the two key problems identified above for static ageing. Microsimulation models based on dynamic ageing track each individual described by the reference database through time, thereby building up a panel dataset. For example, most dynamic microsimulation models based on dynamic population ageing that are designed to consider the effects of transfer policies start with a reference database that describes marital status, parenthood, employment status, income, and previously accrued measures of wealth. The population is then aged one year with reference to a series of transition equations that describe the dynamic evolution of individual-specific circumstances. Where these equations describe chance events, random draws are taken from the underlying statistical distributions in a process that is commonly referred to as Monte Carlo simulation. The income of each individual in each time period, for example, is usually simulated based on characteristics such as the individual's past income, their accumulated experience, demographic characteristics, and upon a random draw from a log-normal distribution that accounts for unexplained variation. By repeating this procedure over successive years, the life history of each individual in the reference database can be generated. Most dynamic microsimulation models are consequently designed to consider the intertemporal and long term effects of counterfactual conditions, rather than the short-run effects with which static models are typically concerned.

Dynamic population aging introduces two important complications, relative to either static microsimulation models or statically aged models. The first is the need to describe how individual specific circumstances evolve through time. The second complication, discussed in detail in Section 2.2, is that projections of behaviour based on formal economic theory are generally complicated in dynamic contexts where the future is uncertain. The practical limitations implied by this last observation are emphasised by a series of papers, which show that even apparently minor forms of uncertainty can have a very substantial impact on behavioural projections – see especially Kimball (1990), Deaton (1991), and Carrol (1992); Browning and Lusardi (1996) provide a simple worked example. Hence, projecting behaviour on the basis of formal economic theory in dynamic contexts usually requires the use of computationally demanding numerical solution methods. The trade-off when adopting such methods is that, in context of contemporary computing technology, they are unable to account for the diversity of individual specific circumstances that is present in most population cross-sections.

The complications associated with dynamic aging have limited its use in the contemporary literature. Most existing models of this type have mitigated the difficulties involved by omitting economic theory when modelling behaviour, or by adopting stylised theoretical assumptions that help to simplify the associated analysis; see Li and O'Donoghue (2013) for a review. Nevertheless, advances in computing power are now starting to permit development of policy relevant structural dynamic microsimulation models that treat uncertainty as a centrally important feature underlying simulated behaviour. The SIDD model represents an example of the current state-of-the art in this regard, and forms the basis of our subsequent discussion.

2.2 Behavioural assumptions

The public policy literature spans a diverse spectrum of alternative assumptions concerning individual decision making. At one extreme of this spectrum are studies that make exogenous "rule-of-thumb" assumptions about the specific decisions that individuals make. At the other extreme are studies based upon formal models of decision making.

Imposed "rule-of-thumb" behaviour

Studies that exogenously assume behaviour typically make broad generalisations that are easy to describe. Examples of this type of assumption include the proposition that behaviour is policy invariant. This approach is often adopted for evaluations of the impact effects of government policies, especially tax and benefit policy, on the distribution of income. It has the advantage of being transparent, and 'neutral' in the sense that it neither favours nor opposes policy change from a behavioural perspective. These advantages have motivated omission of explicit behavioural equations from many microsimulation models in current use: for example, Li and O'Donoghue (2013) note that fewer than one third of the dynamic microsimulation models they surveyed use behavioural equations to project decisions through time. Nevertheless, omission of behavioural responses leaves a microsimulation analysis open to the criticism that people can and do respond to policy change, which is particularly telling when one objective of policy reform is to influence behaviour in some way.

One way to address concerns regarding the extent to which projected effects of policy depend upon exogenous behavioural assumptions is to conduct associated sensitivity analysis. For example, it might be assumed that a given policy change - say the introduction of an in-work welfare benefit - would increase employment rates for certain population subgroups by a given number of percentage points. This analytical approach has the advantage that it is highly transparent and readily understood. Such an approach, however, remains open to criticism due to the arbitrary nature of the behavioural assumptions that are made. This is especially true when a policy is introduced to motivate a positive behavioural response, in which case there is a temptation in the associated analysis to assume that the positive response is realised. Sensitivity analysis of this sort is also complicated wherever behaviour might vary over multiple dimensions, such as employment and savings, as this increases the range of plausible alternatives open for consideration. The fundamental problem here is that a model based on exogenous behavioural assumptions precludes any predictive information concerning how behaviour will respond to policy alternatives.

Dissatisfaction with the limitations of exogenous behavioural assumptions has led researchers to explore alternative approaches for drawing on observed historical data to inform projections for behaviour in the future.

Reduced-form regressions

The simplest approach for allowing past behaviour to inform views concerning the responsiveness to policy alternatives is commonly referred to as reduced form regression analysis. Reduced form regressions are designed to summarise in an accessible form correlations between observable characteristics that are described by survey data. This methodology includes a range of alternative approaches, from simple linear equations that describe a single decision variable as a function of a set of hypothesised exogenous characteristics, through to a system of equations. The common feature that distinguishes this methodological approach is that the parameters of the equations that are considered for analysis may themselves depend upon factors that are determined by policy makers.

Consider, for example, a simple reduced form analysis that describes labour supply decisions as a function of each individual's age and hourly wage rate. It is reasonable to suppose that any estimated coefficient on the hourly wage rate will be affected by the prevailing tax and transfer policies. This is particularly likely if gross-wages, rather than post-tax and benefit wages, are considered for analysis, because high marginal tax rates drive a wedge between the impact that an hour of work has on gross incomes and take-home pay.

Reduced form methods can be especially powerful tools for exploring behavioural responses to past policy reforms. Blundell *et al*. (2005), for example, use the difference in difference econometric technique to identify the labour market impact of the Working Families' Tax Credit on lone parents in the UK. This estimation method compares the change in behaviour observed between two periods in time of a "treated" population subgroup that was affected by the policy reform of interest, against the change in behaviour observed over the same time period for a "control" population who were not affected by the policy change. The fundamental idea is that by taking the "difference" between the behavioural "differences" observed for the treated and control populations, it is possible to isolate behavioural variation to the policy reform from behavioural variation in context of changes to the decision environment more generally. Using this method, Blundell *et al*. (2005) find that the WFTC and associated reforms increased lone parents' employment by around 3.6 percentage points.

Reduced form methods also provide the researcher with a great deal of flexibility when selecting the analytical specification considered for analysis. The approach allows for the incorporation of a large set of variables that might be difficult to distinguish within a theoretical model of agent behaviour, and can allow the analyst to tailor their study to the data that are available. These advantages have made reduced form regressions a popular first-step in the inclusion of behavioural responses in microsimulation models; for example, all three of the dynamic microsimulation models for the UK cited by Li and O'Donoghue (2013) that include behavioural responses - PenSim2, SAGE, and a model produced at the IFS described in Brewer *et al*. (2007) – are based on reduced form behavioural descriptions.

Nevertheless, the limitations of reduced form analyses are widely understood by the economics profession, especially since Lucas published his critique of the approach in 1976 (Lucas, 1976). The fundamental problem is the unquantified sensitivity of reduced form model parameters to the policy environment, which makes such models inappropriate for considering likely behavioural responses to policy counterfactuals. This has led to the development of so-called structural models, which are explicitly designed to explore behavioural responses to alternative possibilities for policy reform.

Structural models

A structural analysis is founded on a theoretical framework that is assumed to be "structurally stable", in the sense that it remains invariant to policy change. This characteristic of the model makes it suitable for considering behavioural responses to policy reform. Commonly adopted theoretical frameworks depend upon a set of assumptions concerning how decisions are made, and are used to generate functional forms that relate observable characteristics and decisions. Importantly, the parameters of these functional forms are assumed to be fixed through time, and are commonly referred to as "deep parameters". The term "deep" here indicates that the respective parameter represents a fundamental building block upon which the analysis depends.

The most common theoretical framework used to describe behaviour in economics assumes that people make decisions as if to maximise expected (lifetime) utility, given their preferences, constraints, and expectations about the future. Importantly, it is common to assume that expectations are rational, in the sense that they are entirely consistent with the decision making context.

The key advantage of a structural model is that it provides a coherent basis for exploring behavioural responses to policy counterfactuals. The central drawback of the approach is that the theoretical framework upon which the analysis depends introduces a layer of complexity to the analytical problem. The scale of complexity associated with a theoretical framework varies substantively between modelling alternatives. In context of the utility framework, for example, some sets of assumptions permit closed form solutions to be derived for endogenous decisions. In this case a structural model can be parameterised and employed using similar analytical techniques to those of reduced form models discussed above; see, Pylkkänen (2002) for a dynamic microsimulation model based on such an approach.

Unfortunately, very strong assumptions need to be made to derive analytically convenient closed-form solutions for behaviour, particularly where uncertainty concerning future circumstances is likely to be important (e.g. Browning and Lusardi, 1996). Most academic attention has consequently focussed on behavioural models that have no closed-form solution, and which require numerical Dynamic Programming (DP) methods to solve. DP methods are, however, computationally demanding (e.g. Rust, 2008), which has seen this academic literature omit much of the diversity of individual specific characteristics that is a conspicuous feature of the microsimulation literature more generally.

2.3 Advantages of the SIDD model, and best-practice methods of use

SIDD is a structural dynamic microsimulation framework. It is *structural* in the sense that it projects labour/leisure and consumption/savings decisions based upon a theoretical framework of behaviour. It is *dynamic* in the sense that it projects each individual described by a reference dataset forward (and backward) through time, so that it is capable of building up a complete life-history for each. And it is a *microsimulation* model in the sense that it is designed to project the circumstances of an evolving population cross-section. As discussed in sections 2.1 and 2.2, these features of the model endow it with a set of relative advantages, which adapt it for specific subjects of interest.

The fact that SIDD uses dynamic aging to project the evolving population cross-section through time makes the model a powerful tool for exploring the implications of policy counterfactuals for the future, and for interpreting the effects of policy in terms of life-time incomes of individuals present at a given point in time. Furthermore, the fact that the model uses a structural framework to simulate behavioural makes it an appropriate tool for exploring behavioural responses to policy alternatives. The structural framework upon which the model is based also makes it useful for identifying what effects policy alternatives would have on underlying incentives. Such considerations become increasingly important as the projected time-horizon is lengthened, due to feedback effects of behaviour on individual circumstances, and are therefore particularly relevant for dynamic microsimulation models that project circumstances well beyond a short (five year) time horizon.

However, delivering the functionality referred to in the preceding paragraph comes at the cost of a substantial computational burden. This burden has meant that SIDD omits a great deal of the detail that is included in static microsimulation models, or dynamic models based on more stylised behavioural assumptions (either exogenous, reduced form, or closed form decision making). It also tends to take longer to generate simulated output, and require more processing power than other approaches to microsimulation. These considerations imply that, if an analyst is primarily interested in evaluating the effects that a policy reform is likely to have over a short time horizon (up to five years), then a static microsimulation, or a model based on static aging is likely to be more useful than SIDD. Alternatively, if interest is in the medium term implications of policy where behavioural responses are considered to be unimportant, then a dynamic microsimulation with either exogenous behaviour, or behaviour based on reduced form statistical relationships is likely to be most useful.

SIDD is consequently designed to evaluate the impact of a changing economic environment on household circumstances of the evolving population cross-section, measured over at least a five year time horizon. The "economic environment" is defined broadly here so that it includes (but is not limited to) tax and benefits policy, childcare arrangements, housing costs, labour market regulation, the pensions framework, and returns to education. The model is designed so that behaviour responds endogenously to policy change, in a way that takes into account the sorts of uncertainties that people actually face. This is important because endogeneity of this type can be exceedingly difficult to second-guess, and can have a pronounced impact on individual circumstances through time.

The fact that SIDD generates projections for the evolving population cross-section departs from most of the existing literature that explores behaviour in the context of uncertainty. This feature of the model is important, however, because it permits the simulated effects of policy to be mapped onto contemporary population subgroups of interest. It also allows the model to align more closely with other models in common use, including static microsimulation models and dynamic microsimulation models based on alternative behavioural assumptions.

SIDD has been designed in a way that provides substantial flexibility for modelling alternative policy contexts. This approach is designed to avoid complex re-programming when reflecting new country specific contexts, so that the analyst can focus upon the important job of identifying appropriate model parameters. Each alternative adaptation of SIDD is usually given its own name, to distinguish it from others that have been produced. Some examples to date include NIBAX (UK, van de Ven and Weale, 2009), LINDA (UK, van de Ven, 2016a), PENMOD (IRE, Callan *et al.*, 2012), and ITALISSIMO (ITA, 2014; European Commission Grant Number VS/2013/0208).

Nevertheless, SIDD has required the development of bespoke programming code that tends to complicate its validation and use by third-parties. We have sought to mitigate this issue in three ways. First, all inputs to the model are delivered through a Microsoft Excel spread sheet. For the UK variant, LINDA, a series of 'front-end' dialog boxes (user forms) walk an analyst through a selected set of parameters that they are most likely to want to alter, each of which is assigned a plain-English description. Similarly, the model produces a standard set of output for each simulation to a series of Excel spread sheets, designed to facilitate preliminary analyses.

Secondly, the micro-data that the model generates for each simulated population are stored in a standard format that can be read by most statistical packages (comma separated variable, csv, format). These data are crucial, both to facilitate an on-going process of model validation, and because it is impossible, as a modeller, to anticipate all of the statistics that will be useful for any potential analysis.

Finally, the tax and benefit routines that are a central focus of policy interest in the model are provided in a self-contained module (a dynamic link library) that users can alter in whatever way they like, without the need for external consultation. Like the main program, the tax and benefit routines are programmed in Fortran, which has the added advantage of being an accessible language for non-specialists.

i. Best-practice methods of use

Like all dynamic microsimulation models currently in use, SIDD is not adapted to generate forecasts. This is because no attempt has been made to measure the full set of uncertainties associated with projections into the future.

It is useful to distinguish between three types of uncertainty associated with any forecast. There are the uncertainties that are explicitly represented within a model's structure; in SIDD these include the evolution of labour opportunities, investment returns, relationship status, and death. There are uncertainties associated with model parameters; in SIDD there are thousands of parameters governing everything from the likelihood of marriage and divorce to the influence of age on housing costs. And there are 'other' uncertainties, which include anything not explicitly represented in a model; in SIDD, these range from shifts in policy concerning same-sex marriages, through to the outbreak of war with a previously close trading partner. The influence of the first of these types of uncertainty can usually be reflected by a model, the second type is often more difficult as it can require a great deal of data to evaluate in any rigorous way, and the third type of uncertainty – often the most important – is beyond the modelling scope.

Given the above, it is highly advisable to use SIDD to focus on the effects of policy counterfactuals. This is done by evaluating differences between pairs of simulated projections, where the only differences between each simulation in a pair concerns the policy reform of interest. This approach provides a measure of how the policy reform would affect the population in a controlled context where all other features of the economy remained unaltered. From a policy maker's perspective, it focusses upon the influence of those features of the world that the policy maker can affect, ignoring features that are beyond their control. It is simple to use SIDD to generate a measure of the variability of simulated effects to uncertainties that are explicitly represented in the model (the first set described in the preceding paragraph). The second and third sets of uncertainty are generally beyond the scope of analysis.

3. SIDD - the Existing Model Framework

3.1 Overview

SIDD is a dynamic programming model of household sector savings and labour supply decisions that has been developed to reduce the set-up costs associated with using current best practice economic methods of analysis of public policy alternatives. The decision unit of the model is the (nuclear) family, defined as a single adult or partner couple and their dependent children. SIDD considers the evolving circumstances of individual adults and their family units, organised into annual snap-shots during the life-course. Allocations within families are ignored. Decisions regarding consumption, labour supply, and pension scheme participation are endogenous, and are assumed to be made to maximise expected lifetime utility, given a family's prevailing circumstances, its preference relation, and beliefs regarding the future. Preferences are described by a nested Constant Elasticity of Substitution utility function. Expectations are 'substantively-rational' in the sense that they are either perfectly consistent with, or specified to approximate, the intertemporal processes that govern individual characteristics.

The model assumes a small open economy, for which rates of return to labour and capital are exogenously given. Heterogeneous circumstances of adults are limited to the following seventeen characteristics:

year of birth^{d,e}
dependent children^{e}
health status^{e}
potential full-time labour income
private pension wealth
wealth not otherwise defined
age^{d,e}
student status^{e}
carer status^{e}
savings in ISAs
timing of pension access^{d}
survival of reference adult^{e}
relationship status^{e}
education status^{e}
migration status
private pension eligibility
contributory state pensions^{d}

Of the 17 characteristics listed here, four are assumed to evolve deterministically (indicated by a superscript 'd'), and all others may evolve with some uncertainty. Furthermore, nine of the characteristics are simulated exogenously from the decision model (indicated by a superscript 'e'), and all others are endogenous to the decision problem.

Including year of birth in the list of adult characteristics introduces the overlapping generations framework that is necessary to reflect the circumstances of a population cross-section. Age, wage potential, measures of wealth, and survival are all centrally important for any analysis of savings and labour supply. Past experience with similar analytical frameworks has also emphasised the importance of relationship status when seeking to capture transfer policy, labour supply and consumption decisions. The number of dependent children has an important bearing on the consumption needs of families, and education status is an important factor in enabling the model to capture financial inequality over extended time-frames.

Model projections start from a reference database of detailed microdata for a population cross-section; in the UK these data have been extracted from the Wealth and Assets Survey (WAS). The model generates and saves panel data for each adult in an evolving population cross-section, and can also be directed to produce a selected series of summary statistics output to Excel.

The remainder of this section provides a non-technical summary of the features of SIDD; further technical detail is reported for the UK variant of SIDD in van de Ven (2016a) and associated issues of calibration are discussed in van de Ven (2016b).

3.2 The Preference relation

Endogenous behaviour in the model is simulated based on an assumed utility function for each adult. The utility function takes a nested CES form that is assumed to be the same for all individuals, which is standard in the associated literature. The utility function depends on four key individual specific characteristics: equivalised consumption, leisure, salience costs, and prospective measures of liquid wealth.

Utility is assumed to be increasing in equivalised consumption. Consumption is adjusted for differences in family needs using the modified OECD scale, which assigns a value of 1.0 to the family reference person, 0.5 to each additional family member over age 13, and 0.3 to each child aged 13 and under. This adjustment is included for analysis to reflect the important influence that family size has been found to have on the timing of consumption.

Individuals who choose to work in the model are considered to trade reduced leisure time for an increase in private (labour) income. The default specification of the model is to allow for three employment alternatives (not-employed, part-time, and full-time employed) for each adult in a family. The number of employment options can, however, be increased to an arbitrarily high (discrete) number, so that the model is capable of capturing labour supply decisions over hours worked.

Salience costs are included in the preference relation to allow the model to reflect contexts where decisions appear to exhibit a systematic bias in favour of pre-existing default options. Salience costs are introduced to represent psychological rigidities, such as the difficulty and stress involved in researching alternative investment options. In contrast to financial transaction costs, which enter the budget constraint and may also act as a break on transactions, salience costs are assumed to be paid in the form of utility. Rigidities represented by salience costs have been much discussed in the savings literature, and can have a profound effect on projected temporal dynamics. Like all preference parameters, salience costs are unobservable, and need to be evaluated indirectly by matching simulated behaviour to behaviour that is observed in practice (see, e.g., van de Ven, 2013).

Prospective measures of liquid wealth enter the utility function directly to accommodate a so-called “warm glow” model of bequests. This model assumes that individuals derive satisfaction by the overall sum of money that they leave as a legacy, and do not care how that sum of money is subsequently used. This modelling approach helps to limit the computational costs of modelling bequest motives, relative to alternatives that are available in the literature. This feature, like almost all other features of the model, can be suppressed if desired.

One important feature of the preference relation that is assumed for SIDD is that it is capable of reflecting myopic decision making. That is, the parameters of the utility function can be specified to capture the shifting incentives faced by people who are prone to being disappointed by their past behaviour. This aspect of behaviour does not feature in the standard model of rational decision making, and can be important in reflecting the circumstances of the most deprived segments of society.

3.3 The wealth constraint

SIDD assumes that decisions are made as though people maximise the preference relation discussed in Section 3.2, subject to an age specific credit constraint imposed on non-pension wealth. Non-pension wealth is a net figure measured over all financial assets and liabilities of a family, excluding assets held in private pensions and rights to state benefits. An important asset class included in this measure is owner occupied housing. Non-pension wealth is assumed to evolve according to a simple accounting identity, increasing and decreasing over a given period depending upon whether consumption is less than or greater than disposable income.

It is possible to allow families to accrue unsecured debt in the model. In this case, the model assumes an upper bound to the available credit based on the net present value of the minimum potential future income stream (based on welfare payments). Furthermore, interest charges on non-pension wealth are assumed to increase as non-pension wealth declines. Hence, the model is capable of reflecting so-called “soft” and “hard” credit constraints.

Disposable income is equal to private (employment, investment and private pension) income plus welfare benefits less tax payments. The model comes prepackaged with a detailed description of transfer payments applicable in 2006, 2010, 2013, and 2016 for the UK. An important feature of the model is that the code that simulates taxes and benefits is accessible to any model user. This code can consequently be altered by the user, without the need to consult third-parties.

3.4 Pension savings

Three types of pension savings are explicitly represented in the model: (contributory) state pensions; private pensions, and a tax advantaged savings vehicle known as an Individual Savings Account (ISA) in the UK. All types of pension are modelled at the family level.

i. State pensions

SIDD includes a state pension, rights to which are accrued in respect of childcare, unemployment, or labour effort during the working lifetime. The model tracks the number of years that each family qualifies for accrual, up to the maximum required for the full state pension. Each family is assumed to draw on its state pension from a defined state pension age, which can be varied between alternative birth cohorts.

ii. Private pensions

Private pensions are Defined Contribution in the sense that every family is assigned an account into which their respective pension contributions are (notionally) deposited. Up to five private pension schemes can be run in parallel, which differ from one another in their required rates of personal pension contributions, employer pension contributions, management costs, and default options. In each period a family can be considered eligible for a single private pension scheme, choosing whether and how much it would like to contribute. Pension contributions are specified as a percentage of employment income, so that they can only be made by working families. The timing of pension receipt can also be specified as an endogenous decision.

The structure assumed for private pensions is designed to focus on key issues that have been raised in the contemporary debate regarding pension provisions. It is sufficiently flexible to reflect the multiplicity of pension arrangements that is a feature of private pensions in some countries. It can reflect the influence of management costs, and the risk-return trade-off on decision making. It can also reflect the influence that default options regarding pension participation and contribution rates may have, as a result of investor inertia (e.g. Carroll, 2009, McKay, 2006). It is consequently possible, for example, to use the model to consider the capacity for auto-enrolment to increase private sector retirement savings.

iii. ISAs

ISAs are an asset class that is designed to encourage savings for retirement in the UK. Participation in ISAs is encouraged by a ceiling on the aggregate contributions that can be made in any year, and by the fact that investment returns on ISA assets are tax free. Wealth held in an ISA can be withdrawn at any time, so that ISAs avoid the illiquidity problems associated with traditional pension savings schemes.

3.5 Education

Up to five alternative levels of education can be distinguished for each adult in the model. One of these levels is reserved for tertiary education, and the remainder distinguish alternative pre-tertiary states. Each adult is assigned an education level when they first enter the simulated population. This assignment is either drawn from the survey data from which model projections are made (for adults present in the reference population cross-section), or randomly allocated based on the relative prevalence of alternative education levels reported in survey data (for adults maturing into the simulated population). Adults who enter the simulated population with sub-tertiary education may also be identified as tertiary students. Tertiary students are associated with a probability of graduating to tertiary status at a later age. Otherwise, education is assumed to remain fixed throughout an individual's simulated lifetime.

The pre-tertiary levels of education can differ from one another in relation to the assumed probabilities of receiving a low wage offer, and associated wage premia. In addition, individuals with tertiary education can be distinguished from non-tertiary educated in relation to the age specific evolution of latent wages, and transition probabilities governing marriage and divorce.

3.6 Labour income dynamics

Wages are modelled at the family level, and depend upon four key elements, each of which is described in turn.

The first element is the family's latent wage, which can be thought of as the underlying value of the labour services that the family has to offer. Latent wages are assumed to evolve following a random process that depends upon age, year, education, and relationship status, and the preceding year's latent wage and labour supply. The specification adopted for latent wages has been very carefully chosen, and is closely related to specifications commonly considered in the wider economics literature (including the regression-toward-the-mean as studied by

Atkinson *et al.*, 1992, and the classical model of income dynamics advocated by Mincer, 1974; see Sefton and van de Ven, 2004, for discussion).

The second element applies a factor adjustment to reflect the impact on wages of alternative labour supply decisions. The factors associated with alternative labour supply decisions are the same for all families and are exogenously imposed.

The third element is an indicator variable that identifies whether each adult in a family receives a wage offer in a given period. If the individual fails to receive a wage offer, then they do not receive any employment income regardless of whether or not they choose to supply labour. Receipt of a low wage offer is assumed to be randomly allocated in each year, based upon age, year, and education specific probabilities. This feature of the model is designed to capture the incidence of involuntary unemployment in the model.

The fourth and final element is a factor adjustment to reflect the impact on wages of a previous decision to access retirement income. This feature can be useful for ensuring that the model captures rates of employment participation late in the simulated working lifetime reported by survey data.

3.7 Relationship status and fertility

Modelling relationship status

A 'relationship' is defined as a cohabitating partnership (including formal marriages and civil partnerships). The relationship status of each reference adult in each prospective year is considered to be uncertain. The transition probabilities that govern relationship transitions depend upon a reference adult's existing relationship status, their education, age, and birth year. These probabilities are stored in a series of 'transition matrices', each cell of which refers to a discrete relationship / education / age / birth year combination. Relationships are assumed to form between members within the simulated population, and newly married adults are organised into couples according to a sorting routine that refers to each adult's age and education.

Modelling children

The model takes explicit account of the number and age of dependent children in each family. The birth of dependent children is assumed to be uncertain in the model, and is described by transition probabilities that vary by the age, birth year, relationship status, and previously born children of a reference adult. These transition probabilities are stored in a series of transition matrices, in common with the approach used to model relationship status (described above). Having been born into a family, children are assumed to remain dependants until an exogenously defined age of maturity. A child may, however, depart the modelled family prior to attaining maturity, if the family experiences a relationship dissolution, in which case the children are divided equally between the separating spouses to the nearest integer.

3.8 International migration

As the review by O'Donoghue *et al.* (2010) makes clear, there are a wide range of alternative approaches used to simulate the effects of migration in the microsimulation literature. Key modelling decisions include whether to model net migration or immigration and emigration separately, the variables that describe the likelihood of emigration, the approach taken to generate the characteristics of immigrants, and whether to accommodate re-entry of emigrants. These decisions depend upon the reasons for the respective model's development, and the data that are available for parameterisation.

SIDD is fundamentally designed to reflect projections for the age distribution of the evolving population cross-section, and the bearing of migration on the distribution of income. These objectives motivated accommodation of both immigration and emigration in each simulated period. The pageant-algorithm described by Chénard (2000) is adopted, in which emigrant families are randomly selected from within the simulated sample to achieve exogenously defined targets that vary by age, year, education, marital status, dependent children, and past migrant status. Similarly, the characteristics of new immigrants are initialised in the model by cloning 'donor' families randomly drawn from targeted subgroups within the existing model population.

3.9 Health and mortality

SIDD can distinguish between 10 discrete health conditions for each adult in each simulated period. Simulated health conditions evolve through time, based on exogenously defined transition probabilities that vary by each adult's prevailing health condition, education, age, and year.

Health states can influence families in a variety of ways. The health condition of each adult can affect their likely health condition in the future, allowing the model to capture the persistence of conditions. The health condition of one adult in a couple can affect the carer responsibilities of their spouse (discussed below). The health state can also be defined to limit the discrete set of labour alternatives available to each adult, the probabilities of receiving a low wage offer and wages earned, welfare benefits, the likely evolution of relationship status in prospective years, and non-discretionary costs.

A "carer state" can be generated for each family in each simulated period, where carer families include one adult with carer responsibilities. The carer state evolves through time, based on exogenously defined transition probabilities that vary by the individual's prevailing carer state, the disability state of their spouse, age, and year. Carers can be limited to families with at least one adult who is sufficiently healthy, as defined by a pre-defined value of the health state. Carers can be distinguished from other adults in regards to the benefits that they are eligible for, their employment opportunities, and the time that they have available for leisure.

The model simulates survival for each adult based upon age and year specific mortality rates, which are commonly reported components of official life-tables.

3.10 Projecting a population through time

The model uses a two-step procedure to simulate a population through time. The first step solves for utility maximising decisions, given any potential combination of individual specific circumstances. The second step then projects a reference population through time, given the behaviour solutions derived in the first step.

Solving for optimising decisions

The allowance for stochastic income and asset returns implies that an analytical solution to the utility maximisation problem does not exist, and that numerical solution routines need to be employed. The model uses a standard backward induction technique for solving the lifetime decision problem, based on interpolation methods and a grid that overlays all conceivable combinations of individual specific characteristics. A full description of this approach is beyond the scope of the current report; see van de Ven (2016a) for details. In essence, however, this stage of the simulation can be thought of as generating a set of functions that describe endogenous decisions (e.g. consumption, labour supply, pension scheme participation) in terms of alternative combinations of individual specific characteristics.

Projecting individual circumstances through time

The model starts with a reference database that describes all relevant characteristics for each family, for a sample that is designed to reflect the population cross-section observed at a point in time. These data are used to identify endogenous decisions of each family, using the solutions to the utility maximisation problem identified in the preceding stage of the analysis. Given a family's characteristics and behaviour, its characteristics are projected through time following the processes that are considered to govern their intertemporal variation. Where these processes depend upon random terms, draws are taken from their defined distributions in a process that is common in the microsimulation literature (sometimes referred to as Monte Carlo simulation).

4. Directions for Further Research

At the time of writing, on-going analysis is organised under three principal themes.

In continuous development since 2002, SIDD is now a highly evolved framework for structural dynamic microsimulation. Nevertheless, work continues to *develop* the model, extending functionality along a number of dimensions. The most obvious focus of development is to expand the characteristics that the model is designed to reflect. In this respect, our interest is focussed upon allowing for the gender of adults, which is moderately challenging as it has an important influence on the way that many of the other characteristics that are currently included in the model are simulated. Housing is another key subject of interest, as it features prominently in the budgets of families in many countries. More generally, a longer term objective is to build macro-economic endogeneity into the model, so that the supply of capital and labour influence to investment returns and wages.

Less obvious, but no less important, developmental efforts focus on extending the functionality of SIDD. One important aspect of this effort concerns integrations for cloud computing, which is becoming increasingly important from a hard-ware perspective. Another focus of interest is to simplify use of the model, with alternative options including a SIDD-lite for new users, and a more intuitive user experience.

Our *academic focus* is to use the model in one of two key ways. The first focusses upon use of the model as a framework for empirical investigation of alternative assumptions concerning agent behaviour. The second stream of analysis is to use the model to explore the implications of policy alternatives of contemporary interest. Most of our work to date has focussed upon the second of these two subjects, but we anticipate this to change during the next five years.

Finally, we are now actively *seeking third parties to use SIDD*. We are pursuing this objective both for philosophical reasons, and because of the substantive economies of scale that exist. Our hope is to encourage a community of like-minded individuals to support one another in the development of cutting-edge economic research. To this end, we offer the product of a decade of research and development for free, thereby reducing the set-up costs associated with structural dynamic microsimulation. With SIDD, it is reasonable to expect a fully functional model to be up and running for a new economic context within months, rather than years. This is the purpose for which the model has been designed.

Acknowledgements

SIDD is written in Fortran, with parallelisations implemented using OpenMP and MPI. The model architecture was established by James Sefton. The following UK institutions have provided financial support essential to the model's development: the Joseph Rowntree Foundation, HM Treasury, the Economic and Social Research Council, HM Revenue and Customs, the Department for Work and Pensions, the Department for Business, Enterprise and Regulatory Reform, the Leverhulme Trust, and the National Association of Pension Funds. Funding has also been provided by the European Commission. The usual disclaimer applies.

References

Atkinson, A.B., Bourguignon, F. and Morrisson, C. (1992), *Empirical Studies of Earnings Mobility*. Reading: Harwood Academic Publishers.

Blundell, R., Brewer, M. and Shephard, A. (2005), "Evaluating the labour market impact of Working Families' Tax Credit using difference-in-differences", *HM Revenue and Customs*, ISBN 1-904983-12-X.

Blundell, R, Meghir, C, Symons, E, and Walker, I. (1984), "A Labour Supply Model for the Simulation of Tax and Benefit Reforms". In R. Blundell and I. Walker (eds.), *Unemployment, Search and Labour Supply*. Cambridge University Press: Cambridge.

Brewer, M., Browne, J., Emmerson, C., Goodman, A., Muriel, A. & Tetlow, G. (2007), "Pensioner poverty over the next decade: what role for tax and benefit reform?", *IFS Commentary* No. 103.

Browning, M. & Lusardi, A. (1996), "Household saving: micro theories and macro facts", *Journal of Economic Literature*, 34, pp. 1797-1855.

Capéau, B. & Decoster, A. (2016), "Getting tired of work or re-tiring in absence of decent job opportunities", *EUROMOD Working Papers* EM4/16.

Capéau, B., Decoster, A. & Dekkers, G. (2016), "Estimating and simulating with a random utility random opportunity model of job choice", *International Journal of Microsimulation*, 9, pp. 144-191.

Callan, T., Keane, C. and van de Ven, J. (2012), "A framework for pension policy analysis in Ireland: PENMOD, a dynamic simulation model". In T. Callan and J. O'Connell (eds.), *Analysing Pensions: Modelling and Policy Issues*. ESRI: Dublin.

Carrol, C. (1992), "The buffer-stock theory of saving: Some macroeconomic evidence", *Brookings Papers on Economic Activity*, 23, pp. 61-156.

Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. (2009), "Optimal defaults and active decisions", *Quarterly Journal of Economics*, 124, pp. 1639-1674.

Chénard, D. (2000), "Individual alignment and group processing: an application to migration processes in DYNACAN". In L. Mitton, H. Sutherland and M. Weeks (eds.), *Microsimulation modelling for policy analysis: challenges and innovations*. Cambridge: Cambridge University Press.

Creedy, J., Duncan, A.S., Harris, M. & Scutella, R. (2002). *Microsimulation Modelling of Taxation and The Labour Market: The Melbourne Institute Tax and Transfer Simulator*. Cheltenham: Edward Elgar.

Deaton, A. (1991), "Saving and liquidity constraints", *Econometrica*, 59, pp. 1221-1248.

Dekkers, G., Desmet, R., Fasquelle, N. & Weemaes, S. (2015), The social and budgetary impacts of recent social security reform in Belgium. In I. Salagean, C. Lomos & A. Hartung (eds.), *The young and the elderly at risk: Individual outcomes and contemporary policy challenges in European societies*. Intersentia.

Duncan, A. (2001), Microsimulation and Policy Setting in the United Kingdom, http://www.cerc.gouv.fr/meetings/seminaireoctobre2001/intervention1.doc

Giles, C. and McCrae, J. (1995), "TAXBEN: the IFS microsimulation tax and benefit model", *IFS Working Paper* W95/19.

Kimball, M.S. (1990), "Precautionary saving in the small and in the large", *Econometrica*, 58, pp. 53-73.

Li, J. & O'Donoghue, C. (2013), "A survey of dynamic microsimulation models: uses, model structure and methodology", *International Journal of Microsimulation*, 6, pp. 3-55.

Liegeois, P. & Islam, N. (2013), "Dealing with negative marginal utilities in the discrete choice modeling of labor Supply", *Economics Letters*, 118, pp. 16-18.

Lucas, R. (1976), "Econometric Policy Evaluation: A Critique". In K. Brunner and A. Meltzer (eds.), *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy 1. New York: American Elsevier.

McKay, S. (2006), "Employers' pension provision survey 2005", *Department for Work and Pensions Research Report*, 329.

Mincer, J. (1974), *Schooling, Experience, and Earnings*. New York: Columbia University Press.

O'Donoghue, C., Redway, H. & Lennon, J. (2010), "Simulating migration in the PENSIM2 dynamic microsimulation model", *International Journal of Microsimulation*, 3, pp. 65-79.

O'Donoghue, C., Lennon, J. & Hynes, S. (2009), "The Life-cycle Income Analysis Model (LIAM)", *International Journal of Microsimulation*, 2, pp. 16-31.

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

Pylkkänen, E. (2002), Modelling wages and hours of work. In E. Pylkkänen (ed.), *Studies on household labor supply and home production*. Goteborgs Universitet: Gothenburg.

Richiardi M. & Richardson R. (2016), "JAS-mine: A new platform for microsimulation and agent-based modelling", *Nuffield College Economic Working Papers* W04.

Rust, J. (2008), Dynamic Programming. In S. N. Durlauf & L. E. Blume (eds.), *New Palgrave Dictionary of Economics*. Palgrave Macmillan: New York.

Sefton, J. & van de Ven, J. (2004), "Simulating household savings and labour supply: An application of dynamic programming", *National Institute Economic Review*, 188, pp. 70-86.

Steventon, A., Sanchez, C. & Curry, C. (2007), "Increasing the value of saving in Personal Accounts: taking small pension pots as lump sums", *Equal Opportunities Commission Working Paper*, 58.

Sutherland, H. and Figari, F. (2013), "EUROMOD: the European Union tax-benefit microsimulation model", *International Journal of Microsimulation*, pp. 4-26.

van de Ven, J. (2016a), "LINDA: A dynamic microsimulation model for analysing policy effects on the evolving population cross-section", *NIESR Discussion Paper*, 459.

van de Ven, J. (2016b), "Parameterising the LINDA microsimulation model of benefit unit savings and labour supply", *NIESR Discussion Paper*, 464.

van de Ven, J. (2013), "The influence of decision costs on investments in Individual Savings Accounts", *National Institute Discussion Paper*, 407.

van de Ven, J. and Weale, M. (2009), "A structural dynamic micro-simulation model for policy analysis: Application to pension reform, income tax changes and rising life expectancy", *NIESR Discussion Paper*, 336.

Simulating the long-term labour market effects of economic transition in Hungary: the MIDAS_HU microsimulation model

KRISZTIÁN TÓTH

Abstract: Simulation of both labour market participation and future entitlements play a key role in the majority of dynamic microsimulation models, and the Hungarian MIDAS_HU model is no exception. These models (e.g. SESIM, MIDAS_BE, DESTINIE II.[1]) are often based on very recent (base year) employment states. While the Hungarian labour market is highly influenced by the major transformations of the early 90's, we used historical entitlement acquisition information from recent decades in our development process. This article gives insights into the methodology of using historical information to set up the labour market module of a dynamic microsimulation model.

In order to make use of historic information, the module's main explanatory variable is the retrospective labour market profile. This reflects the characteristics of a person's career over more recent decades. MIDAS_HU subsequently uses different logistic regression equations in combination with alignment as reduced-form behavioural equations to determine the probability of employment for individuals with different labour market profiles.

According to the validation results of the model, this method significantly improves the accuracy of the objective estimation of employment. This methodology also makes it possible to model more accurately than before the impact that the great transition (which took place in the first half of the 1990s) had on the pension system. We find that using labour market profiles is an effective tool for the forecasting of individual career diversities.

JEL classifications: H55, C15, C40, C53

---

[1] Detailed description of the mentioned models: SESIM – Flood et al. (2012); MIDAS_BE – Dekkers (2010); DESTINIE II – Blanchet et al. (2010) and Bachelet et al. (2014)

Introduction

The accurate simulation of labour market participation for individuals or groups is crucial for the creation of every pension model, and even more important in the case of microsimulation pension models. These models are usually used to project and analyse the prospective evolution of old-age poverty, which, in Hungary, greatly depends on the future distribution of service time (Dekkers et al., 2015). In order to make accurate projections of future distribution of service time we had to take into consideration that the Hungarian labour market is still highly influenced by the major transformations of the early 90's (Kézdi, 2002; Kertesi and Köllő 2002).

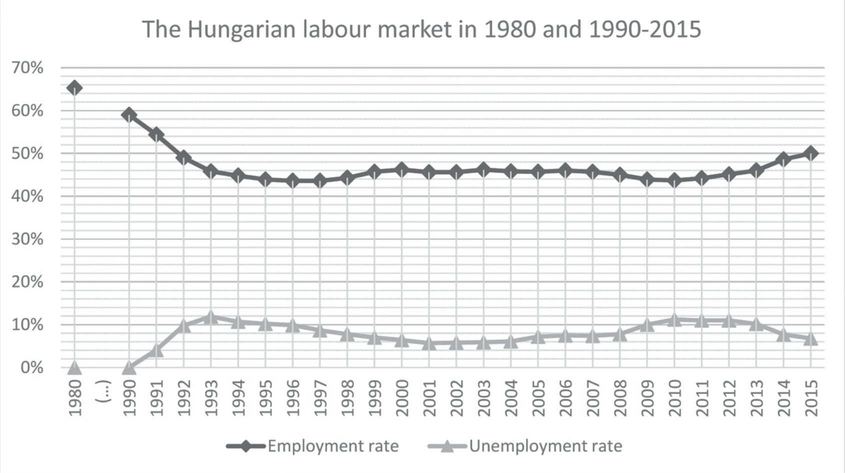


Figure 1: Employment and unemployment rate in Hungary in 1980 and 1990-2015

Source: MTA-KRTK databank: The Hungarian Labour Market Yearbook

As shown in the graph above, the employment rate declined by more than 13 percentage points between 1990 and 1993, and currently remains close to this level still. Nearly a million people lost their jobs during those three years. Many of them could not return to the labour market on a permanent basis even in later years, which resulted in more fragmented or even broken career paths. This means that there is a significant group of (permanently) unemployed persons whose employability is much lower than certain other (temporarily) unemployed persons[2]. This is why it would distort our findings too much if we didn't

---

[2] It is known from the literature that the duration of unemployment negatively affects the employability (Lindsay et al., 2003; Thomsen, 2009; McQuaid and Lindsay, 2002)

distinguish between persons who lost their job a few years ago, and those who lost their jobs decades ago.

Hence, we felt that if we used all available past employment data when setting up the estimation we would be able to predict an individual's following year's employment status more accurately.

Underlying data and the Data Warehouse

Long term data associated with the acquisition of entitlement

Entitlement to a pension in the Hungarian pension scheme is based on the service time accumulated during the individual's official working life, and – as a general rule – on the income acquired from 1988 onwards, and for which pension insurance contributions were typically made[3].

By contrast, the calculation of an individual's service time is based on the whole of the individual's working life, upon which the entitlement to and the amount of the pension benefit can be established. These entire (working life) periods comprise the period during which the individual was earning income and on the basis of which contributions had to be paid, together with other periods specified by the applicable statutory regulations.

Database of pension entitlements

The very first form of the database underlying our projection function forms the basis of the historic microsimulation pension projection model (NYIKA model) developed by the Pensions and Old-Age Roundtable, a forum that operated between 2007 and 2009 (Holtzer, 2010). The basic administrative database used for analysis presented in this paper is a more recent version of that original database developed further. A Data Warehouse was created in late 2014 for this purpose. This development enables the querying of continuously generated eligibility and benefit provision data, through the *data mart* system.

The Data Warehouse comprises all but only data relevant to the establishment of entitlements recorded between 1970 and 2012.

---

[3] See Dekkers et al. (2015) for a detailed description.

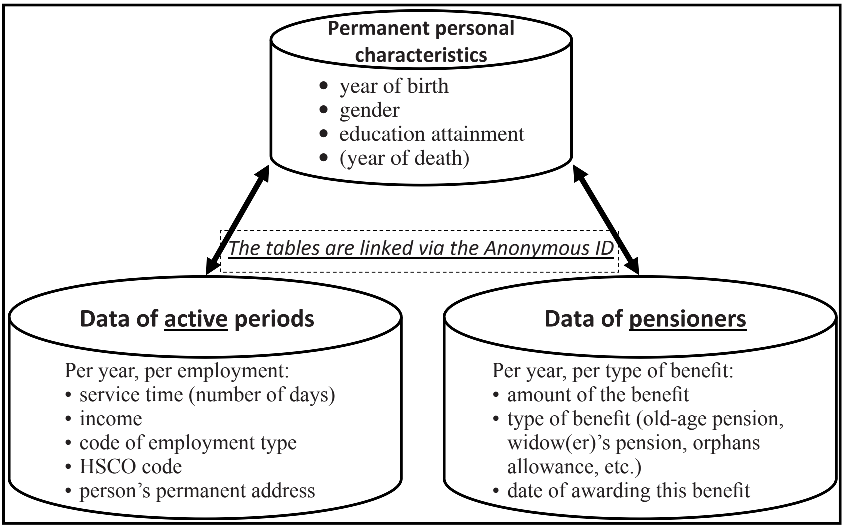


Figure 2: The structure of the Data Warehouse

For privacy reasons, the Data Warehouse contains only anonymized data. Each individual in the Data Warehouse has a unique anonymity identifier (ID) which establishes a connection between the individual's various data elements stored in different parts of the Data Warehouse.

The Data Warehouse can be divided into three parts as shown in Figure 2. The first part contains those personal characteristics which will not (or not often) change over time. These include, *inter alia*, demographical characteristics (like year of birth, gender), the termination of the accumulation of entitlement (so-called inactivation) and the imputed educational attainment[4].

---

[4] Due to a lack of a suitable reporting system, the period of full time tertiary studies are missing from our administrative database. Hence an estimated education attainment is "imputed", on the basis of the Hungarian Standard Classification of Occupations (HSCO) codes. This matching process is discussed in a study by Bálint, Köllő and Molnár (2010).

The second part contains all available data on the accrual of pension rights (the data of the active periods). This includes the variables detailed below, per year and per employment (or non-gainful activities):

- Service time, based on income generation or some other legal/official grounds (e.g. as a recipient of child care allowance);
- From 1988 onwards, **incomes** essentially constituting the basis of pension insurance contributions; distinguishing, also in this case, "real" incomes from "pseudo[5]" entitlement acquisitions;
- The code of employment type: employee, entrepreneur, public servant, parent on child care leave, etc.;
- The Hungarian Standard Classification of Occupations[6] (HSCO) code
- The postal code relating to the person's permanent address of residence

The third part of the Data Warehouse contains data from pensioners. This pension disbursement data provides information about the insured individual's active statuses or retirement. Anonymous pension disbursement micro-data from January each year (since 2014) indicates the given insured individual's active beneficiary status, i.e. information on whether pension benefit(s) have been established as due and whether pension benefit(s) or any other regular allowances are being disbursed to the individual. Parameters relating to these benefits or allowances are also available, such as:

- the amount of the benefit or allowance for which the given income-earning working life was sufficient;
- the type of benefit or allowance;
- the date of awarding this benefit

---

[5] Pseudo entitlement acquisition includes legal relationships during which the income earned is not (as a general rule) included in the pension (e.g. maternity allowances).

[6] For a detailed description of HSCO codes, see KSH (2011)

Completeness of the data

Figure 3 shows what percentage of the Hungarian population by age and gender are historically covered by the data of the current Data Warehouse.

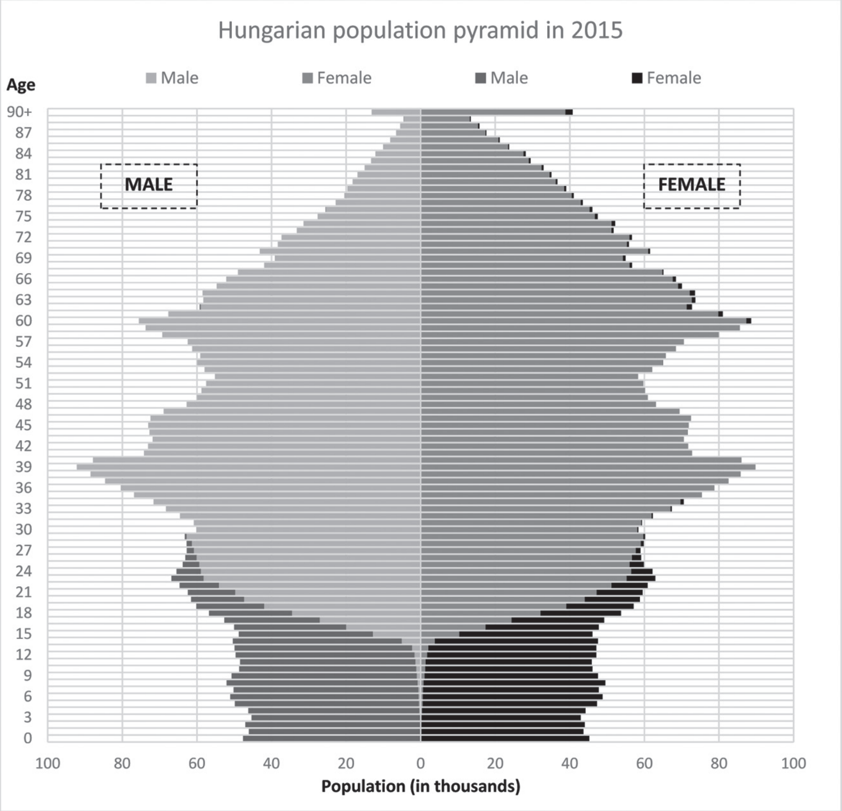


Figure 3: Cross sectional coverage

Source: KSH. (2015). Demographic Yearbook, 2014

As shown in this graph, the Data Warehouse includes the data of almost every person currently over the age of 29, while the younger generations' data is almost completely missing. The reason for this steady decline is that we only have data for persons who have been insured for at least one day in his/her life or who ever received any kind of official pension benefits such as orphans allowance. This latter situation explains why we actually do have data for a few individuals below the age of 15. Between the age of 16 and 28 we can see how these generations progressively enter the labour market, which is the school-to-work transition of the young.

Beside the youngest generations' missing data, the Data Warehouse has other kinds of data gaps and errors. A small portion of our administrative data may be incorrect owing to errors in the entering of data or as a consequence of other factors, while some other data items – not strictly linked to income generating activities – are sporadically incomplete. These may be corrected by a process called data reconciliation, or when certain benefit packages are later taken up.

The one major category in which data is significantly lacking is the entitlement acquisition before 1997, since a considerable proportion of that data is still not available in an electronic format.

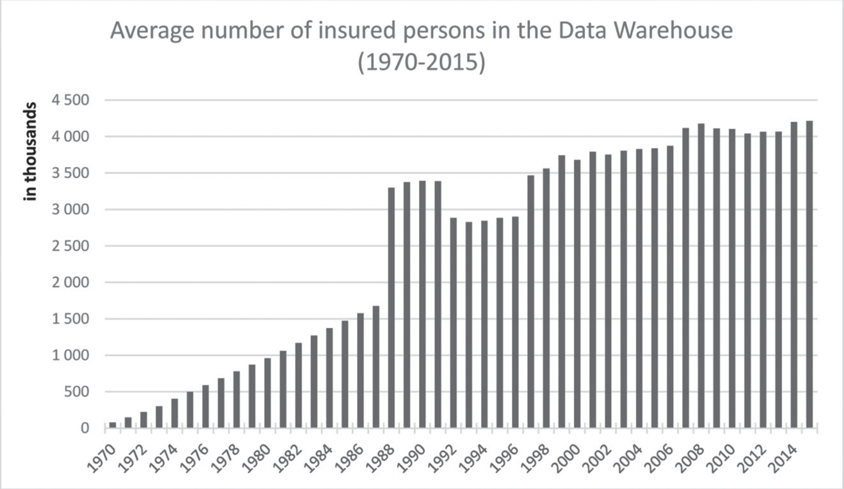


Figure 4: Longitudinal coverage

Source: own calculation

As shown in the Figure 4. before 1988 the yearly average number of insured persons in the Data Warehouse was very low. From 1970 to 1987 this average number increased almost linearly by 80,000 to 100,000 persons per year. The reason for this is that the employment data of those born before 1955 was neither processed nor electronically available, so their employment data is almost completely missing from the Data Warehouse, while those persons who were born in 1955 had just started to enter the labour market in 1970. So this increase is in line with that what time the successive generations entered the labour market.

Between 1992 and 1996 there is another data gap, which we also had to handle. But it is not just the average number of insured persons per year which showed that the database had significant data gaps. As Figure 5 shows, the average years of entitlement – based on the data we had in 2014 (light grey line) – was also distorted in the case of certain cohorts. For example in the case of the cohort born in 1960, the average years of entitlement was 20.2. This is almost six years less than the average of the cohort born in 1959.

In order to fill these data gaps, we used a multiple imputation process, based on entitlement acquisition data pertaining to the age group of people born between 1955 and 1959 (hereafter referred to as basis age group). This was because in regard to that age group, data on entitlements acquired in the past is available electronically, therefore their entitlement acquisition data may be regarded as well-nigh complete. In regard to people born before 1955 or after 1959, i.e. those who were of economically active age even before 1997 but who were not yet recipients according to the pension benefit disbursement data of January 2013[7], the missing data on service periods has been generated at random, from the distribution of the service time data observed in regard to the basis age group.

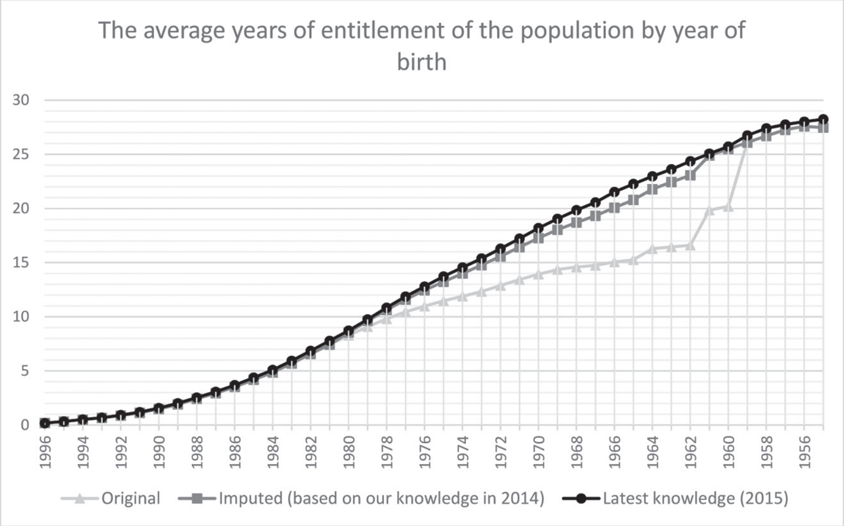


Figure 5: The result of the imputation

---

[7] The imputation was performed done in the first half of 2014 using employment data from 1970 to 2012 and pension benefit disbursement data of January 2013

Our imputation (dark grey line) resulted in less distorted and more plausible average years of entitlement in the case of every cohort. We used this imputed database for the creation of the current labour market profiles.

At the time of our analysis (second half of 2014) it was still not possible to confirm whether the results of our imputation were correct or not. But the new entitlement data which has become available since then (black line) show that the imputation produced reliable data and that the previous real averages had been underestimated only slightly.

The creation of the labour market profiles

The analysis of longitudinal life career data has primarily revealed the dominant impact of two variables on individual chances of employment. Individuals' labour market profiles are comprised of possible combinations of these two variables. One of them is a binary figure indicating employment during the preceding period. The other variable integrates employment characteristics of the active life career so far into a single data item. To enable more accurate differentiation, individuals were separated into two groups in advance, because regarding the differentiation of earlier careers in the labour market, the data did not account for those who had been recipients of some long term (e.g. disability) benefits earlier on, or those with pseudo-legal relationships[8] (e.g. maternity allowances) during the basis year. These two groups were assigned a separate individual labour market profile (either No. 1 or No. 3, respectively).

---

[8] Pseudo legal relationships include legal relationships during which the income earned is not (as a general rule) included in the pension (e.g. maternity allowances).

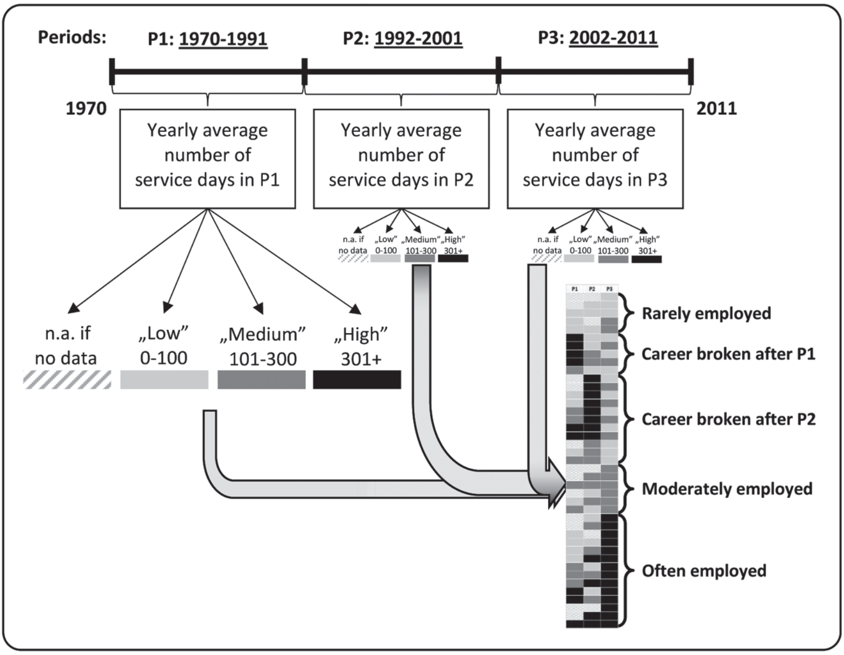


Figure 6: Creation of the labour market profiles

To determine further labour market profiles, the 1970-2012 period was divided into the following three parts: 1970-1991 (P1), 1992-2001 (P2) and 2002-2011 (P3). Then for each employee we calculated the average number of days covered by pensionable earnings that can be included in the calculation of the pension benefits for the various periods. Of course, when an employee had no pensionable earning days in a given calendar year because he or she had not at that time started the economically active period of his or her life, the missing data was ignored.

Next, in terms of the total number of days covered by pensionable earnings, the following three categories of levels of employment were set up in each of the three periods concerned; i) where the average annual number of accountable days was below 100, we call this "low" ii) between 100 and 300 – "medium" and iii) over 301 – "high". In view of the possibility of the occurrence of "empty" values as well, a total of 41 different simplified historic employment careers were created. After an assessment of various categorisation possibilities on the basis of professional and statistical criteria (e.g. significant differences between existing average employment levels) these employment working-life cases were assigned to the following 3 main groups:

- **Weak labour market history:** This group is comprised of three clearly distinguished sub-groups.
    - *Rarely employed:* This sub-group is made up of persons who have hardly ever been present in the labour market, having obtained sporadic entitlements. They were never grouped into the "high" category (although at least in two periods they got a "low" rating).
    - *Career path was broken after the 1st period:* The second sub-group is made up of those who lost their jobs during the transformation of the labour market during the early 1990s and who have never managed to get permanently reintegrated in the labour market since. Their labour market participation decreased in the second period and remained at this level (or even decreased) in the third one.
    - *Career path was broken after the 2nd period:* The third sub-group is made up of people who managed to successfully adapt to the labour changes in the early 1990s but who have typically been absent from the labour market since the beginning of the 21st century. Their labour market participation decreased only in the third period compared to the second.
- **Medium labour market history:** This group is comprised of two distinguishable sub-groups.
    - *Moderately employed:* The first sub-group is made up of those who acquired medium levels of entitlements during the last period, by increasing or at least stagnating employment frequencies, relative to the preceding periods.
    - *Often employed with break(s):* The next group is made up of individuals who had high employment frequency during the last period but for whom there was then a break in their employment, e.g. in one year they had more than 100 days, while in the next year they had fewer than 100 days of insurance relationship.
- **Strong labour market history:** This, the largest and main group, comprises people with high(er) degrees of employment without breaks during the last period (*Often employed without breaks*). Belonging to this group is independent of an individual's preceding employment patterns.

Accordingly, the following profiles (with group ID code) were put together by way of the procedure described above:

- 10: recipient, inactive in 2011,
- 11: recipient, active in 2011,
- 20: low employment, inactive in 2011,
- 21: low employment, active in 2011,
- 30: special legal relationship, inactive in 2011,
- 31: special legal relationship, active in 2011,

- 40: medium employment, inactive in 2011,
- 41: medium employment, active in 2011,
- 50: high employment, inactive in 2011,
- 51: high employment, active in 2011.

Model selection and validation[9]

In order to correctly measure the mathematical explanatory power, we used a cross validation approach, using 90% of the data (randomly selected) as the training set and 10% as the testing dataset. We have estimated the parameters of the logit model on the learning dataset, and tested the goodness of fit across the testing dataset. Owing to the large sample size and the random selection, both sub-samples remained representative in the case of each categorical variable (Vékás, 2015).

In order to find the optimal complexity of the logit model we have estimated and evaluated the following five nested models.

- 0th model: one equation which contains only the intercept
- 1st model: one equation which contains the intercept and the person's employment status in 2011 as an explanatory variable
- 2nd model: one equation which contains the intercept and the person's employment status in 2011 plus the labour market profiles
- 3rd model: contains one separate equation for each combination of the 5 labour market profiles with the 2 possible employment statuses in 2011. Each one of the 10 equations contains the gender (2 categories), region of residence (NUTS II, 7 categories), type of settlement (4 cat.), and type of occupation (10 cat.), age (continuous variable), squared age and constant, as explanatory variables.
- 4th model: contains one separate equation for each combination of the labour market profiles with the individual's employment statuses in 2011. The equations contain every explanatory variable of the 3rd model plus all possible pairwise interactions of the 3rd model's categorical explanatory variables.

Several goodness of fit indicators have been calculated in each model using the testing dataset. Table 1 and Figure 7 show the values of these indicators.

---

[9] This chapter is partly based on Vékás (2015)

Table 1: Creation of the labour market profiles

Model	**Log-likelihood**	**McFadden's Pseudo-R2 (%)**	**Cox-Snell's Pseudo-R2 (%)**	**Nagelkerke's Pseudo-R2 (%)**
0th model	-358 640	0.0	0.0	0.0
1st model	-234 614	34.6	36.9	50.1
2nd model	-210 400	41.3	42.3	57.5
3rd model	**-198 727**	**44.6**	**44.7**	**60.8**
4th model	-77 876	22.5	25.9	35.2
	Tjur's Pseudo-R2 (%)	**Area under the ROC**	**Gini coefficients (%)**	**p-value of the LR-test (%)**
0th model	0.0	0.500	0.0	0.0
1st model	43.2	0.833	66.6	0.0
2nd model	48.7	0.893	78.5	0.0
3rd model	**51.6**	**0.908**	**81.7**	**0.0**
4th model	44.6	0.837	67.3	100.0

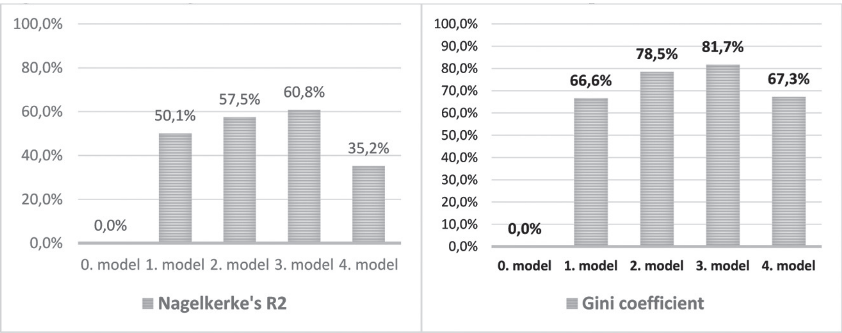


Figure 7: Values of goodness of fit indicators for various model specifications

As we expected, the previous year's employment status has the greatest explanatory power. The labour market profiles also have significant explanatory power, i.e. more than twice as reliable as the explanatory power of all other explanatory variables together (like age, gender, type of occupation, etc.).

Our results firmly support the use of model 3, since other models were not so good in terms of measures of fit. Nested likelihood ratio tests also confirmed that the complexity of model 3 is sufficient to explain the Hungarian labour market and the involvement of further interaction terms were disadvantageous in terms of bias-variance trade-off.

As can be seen, all examined goodness of fit indicators reach their maximum level in the case of the 3rd model which means increasing the complexity of the logit model would only be worthwhile for this model. In the case of the 4th model, the decreasing indicators indicate multicollinearity.

Table 2 contains the estimated coefficients of the largest segment's (high employment and active in 2011) equation. It also contains the odds ratio, and the p value, of the estimated coefficients.

Table 2: Estimated coefficients of the largest segment's (high employment & active in 2011)

Variable		**Coefficients**	**p value**	**Odds ratio**
Constant		–0.693	0.0	0.500
Age		0.159	0.0	1.172
Age x Age		–0.001	0.0	0.999
Dummy of gender	Female	–0.309	0.0	0.734
Dummy variables of HSCO groups[1]	1. Managers	0.239	0.0	1.270
	2. Professionals	0.547	0.0	1.728
	3. Technicians and associate professionals	0.289	0.0	1.336
	4. Office and management (customer services) occupations	0.147	0.0	1.158
	5. Commercial and services occupations	–0.243	0.0	0.785
	6. Agricultural and forestry occupations	–0.077	14.7	0.926
	7. Industry and construction industry occupations	–0.111	0.0	0.895
	8. Machine operators, assembly workers, drivers of vehicles	–0.021	42.6	0.979
	9. (Elementary) occupations not requiring qualifications	–0.366	0.0	0.693
Dummy variables of regions	Southern Great Plain	0.177	0.0	1.194
	Southern Transdanubia	0.003	89.2	1.003
	Northern Great Plain	0.109	0.0	1.115
	Northern Hungary	0.053	0.5	1.055
	Central Transdanubia	0.188	0.0	1.207
	Western Transdanubia	0.271	0.0	1.312

Variable		**Coefficients**	**p value**	**Odds ratio**
Dummy variables of type of settlement	Towns with county right, major provincial towns	–0.005	80.0	0.995
	Other towns	0.109	0.0	1.115
	Villages	0.069	0.0	1.071

* Reference category: men living in Budapest (Central Hungary) and having another (not listed) occupation

[1] For a detailed description of HSCO codes, see KSH (2011)

As this table shows, women were, ceteris paribus, 0.734 times less likely to be employed in 2012 than men, which is a significant difference between labour market chances of women and men. Based on the odds ratios of the HSCO groups, we can find significant differences in employment chances among highly educated and low skilled persons e.g. persons whose occupation was categorised as "professionals" were, ceteris paribus, 1.728 times more likely to be employed in 2012 than those whose occupation was categorised as "other". While the odds ratio of the low skilled ("occupations not requiring qualifications") workers means that their employment chances were, at the time, ceteris paribus, 0.693 times lower than those persons' belonging to the reference category.

In the case of almost every settlement type and region – see Table 2 – the odds ratio of employment was higher than in the reference category (capital, Central Hungary) despite the fact that the employment rate is higher in Budapest and Central Hungary than in the rest of the country. This may seem odd, but it can be attributed to Budapest's different composition in terms of variables.

Within the examined segment logit, i.e. log(odds) of market activity as a function of age can be modelled, ceteris paribus, by a concave parabola which reaches its global maximum at the age of 48. This is consistent with the fact that the employment chances rise dynamically after the person enters the labour market, and start to decrease as the person gets closer to retirement age. Results regarding gender, age and main employment category, in every segment, were in line with existing experiences, and deviations from the model were generally not significant.

Application of the labour market profiles in the MIDAS_HU model

MIDAS_HU belongs to the MIDAS dynamic microsimulation model family developed by the Federal Planning Bureau of Belgium. It is a cross-sectional model of the whole population that simulates processes at the level of the individual and household, and then at the annual period level.

The starting data for the model consists of a 20% random sample of the 2012 population stratified by age, gender, work status (employed, unemployed) and type of provision (old-age pension, widow's pension and orphan's allowance). The simulated data were aligned according to the Hungarian AWG macroeconomic data (e.g. employment rates, average gross income, mortality rates, etc.) and the corresponding data of the Central Statistical Office in Hungary (e.g. marriage or divorce). The adjustment was always based solely on proportions.

The simulation of the labour market activity in the model is based on logistic regression equations in combination with alignment. The model uses the logistic regression equations presented (as the 3rd model) in the previous chapter to determine the probability of employment for individuals of different labour market profiles. This is followed by picking the *n* individuals having the highest estimated probability from each age and gender group, making sure that the ratio of the number of those so selected (*n*), to the total number of individuals in that particular group, equals or is as close as possible to the macro data specified in the alignment table.

In continuing to work with the labour market profiles the model recalculates the employment frequency of the preceding 10 years each year, identifying the ratio of the days spent in insurance relationship during the period to the total number of the same. (Of course, if one entered the labour market only 3 years ago, in his or her case the ratio is calculated only for the past 3 years.) Based on the employment frequency so recalculated, the model revises the labour market profiles.

Meanwhile, if this ratio is below 40 % among men, the individual concerned is given the No. "2" (low employment) profile, while if it is between 40 % and 60 %, (medium employment) or if it is over 60 %, the individual is assigned a "4" or "5" (high employment) profile. In the case of women, the corresponding values are 0-30, 30-50, 50+, respectively. The key consideration in establishing the cutting values was the temporal stability requirement of those belonging to the various profiles. (This is the reason for breaking down the cutting values into men and women as well.)

It should be noted that the persons who entered the labour market at some period of the projection horizon, are provided with their first labour market profiles randomly, in accordance with ratios measured relying on basis data.

Conclusions

In this paper we have attempted to show that using individuals' past entitlement acquisition data in addition to the persons' existing attributes can improve the accuracy of related estimations. By using the earlier introduced labour market profiles we have been able to correctly predict the next year's employment status in 80% of the cases, while both type I and type II errors were 10%. The improvement in accuracy means better projections in the short run, which is important when the model is used for budgetary planning. Besides the short term advantages, the use of these profiles is favourable in the long run too. They can help when considering the value of individual career diversity and diversity in the lengths of service, both of which plays a key role in the long term projection of ”service year sensitive” benefits. Furthermore, the use of these profiles can be considered as the first step to modelling and projecting labour market exclusion, which is very important when we try to project how the value of various poverty indicators will change over the coming decades.

It should be noted that the use of labour market profiles was only one of the possible choices open to us in terms of extracting historic data and making predictions.

It was clearly an advantage that our analysis was based on integrated databases covering a broad range of subject areas. The data available was far more detailed than that published previously. With the use of these extended databases, we were able to assess the structural changes at the beginning of the 1990's in the Hungarian labour market more precisely and in more detail. Working lives and careers broken following the Hungarian system change around 1990 "became visible", providing important information for the model. Now it can be determined how and to what extent the members of various social groups were affected by the labour market transformation that took place two and a half decades ago. Accordingly, it is now possible to project, for instance, who, by social status/category, may face higher or lower risks of old age poverty during the coming decades. In general, the data available as it is now, and the established analytical methodology used, may provide a comprehensive picture of income and employment differences and factors among people making up Hungarian society and their impact on current and future pension schemes.

Literature cited

Bálint, M., Köllő, J. & Molnár, G. [2010]: Pension entitlement acquisition and the full active life career (*Nyugdíjjogszerzés és a teljes aktív életpálya*). *Statisztikai Szemle, vol. 88(6), pp 623-647.*

Bachelet, M., Leduc, A., Marino, A. [2014]: The Destinie II biographies: new sample and projection (*Les biographies du modèle Destinie II: rebasage et projection*). *Working Papers of the DESE G2014-01*, Institut National de la Statistique et des Etudes Economiques.
http://www.insee.fr/fr/publications-et-services/docs_doc_travail/G2014-01bis.pdf

Blanchet, D., Buffeteau, S., Crenner, E., Le Minez, S. [2010]: The new Destinie 2 microsimulation model: main characteristics and illustrative results (*Le modèle de microsimulation Destinie 2: principales caractéristiques et premiers résultats*). *Working Papers of the DESE g2010-13*, Institut National de la Statistique et des Etudes Economiques.
https://www.insee.fr/fr/statistiques/fichier/1380833/G2010-13.pdf

Dekkers, G. [2010]: The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model. Working Paper, No. 10-10. Federal Planning Bureau, Brussels, http://www.plan.be/admin/uploaded/201005110837500.wp201010.pdf

Dekkers, G., Desmet, R., Rézmovits, Á., Sundberg, O., Tóth, K. [2015]: *On using dynamic microsimulation models to assess the consequences of the AWG projections and hypotheses on pension adequacy: Simulation results for Belgium, Sweden and Hungary.* Paper presented to the SPC Working Group on Ageing Issues (SPC-WG-AGE), Brussels, March 27th, 2015.
http://www.plan.be/admin/uploaded/201506121351500.REP_SIMUBESEHU0515_11026.pdf

Flood, L., Jansson, F., Pettersson, T., Pettersson, T., Sundberg, O., Westerberg, A., [2012]: SESIM III - A Swedish dynamic micro simulation model.
http://sesim.org/Documents/Handbook%20of%20SESIM.pdf

Holtzer, P. ed., [2010]: *Jelentés a Nyugdíj és Időskor Kerekasztal tevékenységéről.* Miniszterelnöki Hivatal, Budapest, ISBN 978-963-9682-04-7

Kézdi, G. [2002]: Two Phases of Labor Market Transition in Hungary: Inter-Sectorial Reallocation and Skill-Biased Technological Change. *Budapest Working Papers on the Labour Market, 2002/3.*

Kézdi, G. & Köllő, J. [2002]: Economic Transformation and the Revaluation of Human Capital – Hungary, 1986–1999. *The Economics of Skills Obsolescence, pp 235-273.*

KSH [2011]: Hungarian standard classification of occupations HSCO-08/FEOR-08. Budapest, http://www.ksh.hu/docs/szolgaltatasok/eng/feor08/efeor08.pdf

KSH [2015]: Demographic yearbook 2014. Budapest,
http://www.ksh.hu/docs/hun/xftp/idoszaki/evkonyv/demografiai_evkonyv_2014.pdf

Lindsay, C., McCracken, M. & McQuaid, R. W., [2003]: Unemployment duration and employability in remote rural labour markets. *Journal of Rural Studies, Volume 19, Issue 2, April 2003, pp 187-200,* ISSN 0743-0167, doi:10.1016/S0743-0167(02)00067-0.

McQuaid, R. W. & Lindsay, C. [2002]: The 'employability gap': long-term unemployment and barriers to work in buoyant labour markets. *Environment and Planning C: Government and Policy. 20 (4)*, pp *613-629.* doi:10.1068/c22m. ISSN 0263-774X

Thomsen, S. L. [2009]: Explaining the Employability Gap of Short-Term and Long-Term Unemployed Persons. Kyklos, 62: 448–478. doi:10.1111/j.1467-6435.2009.00445.x

Vékás, P. [2015]: Estimating individual labor market activity in Hungary in a microsimulation framework (*Az egyéni munkaerő-piaci aktivitás becslése mikroszimulációs modellkeretben*). *Közgazdasági Szemle, vol. 62, pp 1291-1308.*