Projection of the Educational Attainment of the Hungarian Population from 2001 to 2020

Modelling Education with a Dynamic Microsimulation Model - ISMIK

Zoltán Hermann and Júlia Varga

The paper presents a dynamic microsimulation model (ISMIK) which was developed to project individuals' educational participation and qualification. The microsimulation model simulates education events at the individual level. The model projects educational participation and attainment of the population over a period of 20 years (2001-2020) by age, gender, region in 6 qualification categories using a 50 per cent sample of the Hungarian Census of 2001 as the initial model population. We present the results of the base scenario when all parameters (transition probabilities) are assumed to be time invariant, that is, this scenario projects education level of the population on the condition that everything continues as in the early 2000s. In addition to the base scenario we present the results of alternative scenarios. The first one investigates what would have been educational attainment of the population had Roma and Non-Roma students had the same opportunities in their schooling career. The further scenarios measure the effects of educational policy changes: reducing the duration of compulsory schooling; increasing the share of students in secondary vocational training schools; and reducing the number of state-funded places in higher education.

1. Microsimulation modelling

Microsimulation models have been increasingly applied in recent years in quantitative analyses of economic and social policy problems. The main idea of microsimulation is that processes resulting from the actions and interactions of a large number of micro-units can be explained best by analysing the behaviour of these micro-units, that is events can be best explained at the level on which decisions are made. Behavioural relationships seem to be more stable on micro-level than in aggregated data and changes in aggregated data may be the result of changes in the composition of the micro-units even if the behaviour of the individual micro-units and their individual characteristics do not change.

Microsimulation models use simulation techniques and take micro-level units – individuals, families or firms - as the basic units of analysis. These micro units are identified by characteristics such as age of an individual, educational attainment, children in a family, income etc. These characteristics are then modified in the microsimulation modelling depending on their individual behaviour and the institutional relationships in which they operate. Using micro data allows for the widest range of heterogeneity in the population to be captured in the model.

Microsimulation models are used for different purposes: projections, evaluations of public policy or designing policy reform etc.

In recent years a lot of countries developed and applied microsimulation models for projecting educational attainment of the population in the future. *Payne et al. (2008)* give a detailed summary of the education modules of different microsimulation models. Most of these education modules were developed as part of a broader microsimulation model. For example the DYNAMOD microsimulation model *(King et al, 1999, Robinson - Baekgaard, 2002.*) in Australia was developed to project characteristics of the Australian population over a period of up to 50 years or the LifePath model in Canada (*Statistics Canada, 2011*) projects the characteristics of the Canadian population. One of the modules of these models is the education module. Some microsimulation models which contain education modules were developed for other purpose. For example the MOSART model in Norway (*Fredriksen, 1998*) is used to assess the public pension system and it contains a separate education module. There are some separate education microsimulation models also as the SESIM model in Sweden which was developed to assess the Swedish education finance system (*Flood J. et al, 2005*).

1.1. DIFFERENT TYPES OF MICRO SIMULATION

The principal characteristics of microsimulation models can be classified in different ways. Here we summarize some of the main differences between different models.

Base dataset selection – ***Cross-Sectional or Synthetic:***

Base dataset selection is important in a microsimulation model as the quality of the input data determines the quality of the output. Cross-sectional base data we get from a cross-sectional dataset. The source of cross sectional data may be administrative data, survey datasets or

Census data. Synthetic datasets are artificially created; they often contain all the variables required.

***Cohort or Longitudinal Models*:** cohort models model a single cohort over their lifetime, population/cross-section models model a population cross-section over a period of time.

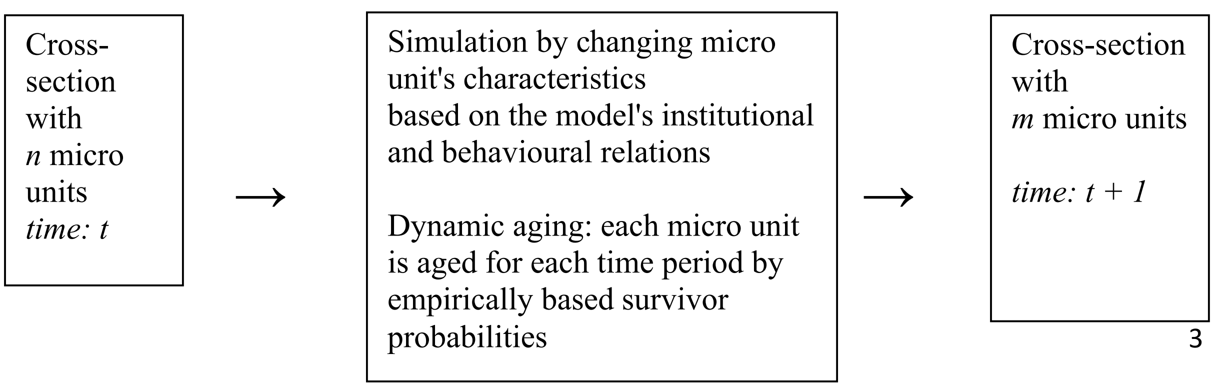
Behavioural Equations or Probabilistic Based Mode*lling:* Behavioural models are based on economic theory, in the sense that changes to characteristics result in a change in the behaviour of agents within the model. A probabilistic model on the other hand attempts to reproduce observed distributional characteristics in sample surveys without necessarily a theoretical underpinning

***Static or Dynamic Microsimulation**:* “Ageing” within microsimulation refers to the process of changing characteristics of micro units over time. *Static ageing* means that the weights of the observations are adjusted so that the simulated population distribution matches the macro aggregates. For example, in order to simulate an ageing society, the weighting of young people gradually decreases over time while the weighting of elderly people would increase; however, there is no change to the attributes of these individuals. *Dynamic ageing* changes the attributes of the individuals instead of altering their weights. In a dynamic microsimulation models each microunit is aged individually by empirically based survival probabilities.

In *dynamic cross-section microsimulation* individually based aging by survival probabilities is applied for each microunit of an entire cross sectional representative sample. (*See Figure 1)*

Figure 1

Dynamic cross-section microsimulation



Discrete Time or Continuous Time: Discrete time models simulate which individuals experience particular events in given time intervals. Continuous time models treat time as a continuous variable and determine the exact time that an event occurs.

Events: a dynamic microsimulation model with discrete time is based on the assumption that each person in each period has certain (transition) probabilities of experiencing transitions from one state to another. Each of these transition probabilities will normally constitute an event, and depends on each person's characteristics.

Drawing in Microsimulation**:** Dynamic microsimulation models are generally solved by stochastic drawing, often called Monte Carlo technique. In the *random number* method the computer generates so-called random numbers with uniform distribution (0,1), and if the random number is smaller than the probability, the event occurs. This stochastic drawing generates an extra uncertainty in the projections often called Monte Carlo Variability (MCV). The simplest method to reduce Monte-Carlo variability is to increase the number of simulated units or to repeat the simulation several times for the same sample. Another solution for the problem is to constrain the simulated number of events to its expectation value without changing each person's probability of experiencing the event. With exogenous probabilities the expectation value of the model is the sum of the transition probabilities. This constraining is easy when probabilities are (group-wise) homogeneous, and for example used when a stratified sample comprises a given number of persons from each stratum.

2. THE ISMIK MODEL

The ISMIK model is a dynamic microsimulation model with a cross section of the Hungarian population. The model simulates the basic demographic and educational events of the Hungarian population. The simulation horizon of the model was 2001-2020. The simulation is carried out by drawing if each person each year makes certain transitions from one state to another with transition probabilities depending on each person's characteristics. Transition probabilities are estimated from observed transitions in a recent period.

Events included in the simulation are (1) demographic events: births, deaths, (2) educational activities and (3) migration.

The ISMIK model is a discrete time model with the calendar year as time unit. It's a cross-sectional model, it simulates the whole population each year before entering the next year. The ISMIK model 'ages' the population, individuals born and are added to the population, while others die and leave the simulation. With each passing year each individual passes through a series of events that predict changes in characteristics what will happen to them in that year. The model simulate one event at the time in a fixed order, it's a recursive simulation

Table 1 summarizes the main characteristics of the model.

Table 1

Main characteristics of the ISMIK model

Microsimulation model – simulates education events at the individual level
Dynamic microsimulation – population is aged by drawing of events
Real initial population based on Census data with corrections
Simulation units - persons
Cross sectional simulation - simulates the whole population year-by-year
Discrete time model - calendar year as time unit
Recursive simulation – simulates events in fixed order with conditional transition probabilities
Transition probabilities estimated with observed rates
Drawing method : mean-constrained stochastic drawing
National and regional models

Initial population of the ISMIK model is a 50 per cent household sample of the 2001 Hungarian Census supplemented by a 50 per cent sample of boarding in inhabitants. The initial population of the model consisted of 5 096 323 individual records. In the initial population some characteristics of the individuals were determined by simulation.

We also added a *„Roma status"* variable to individuals in the initial population of the model with the help of microsimulation. The reason for this was that the demography of the Roma population in Hungary seems to have differed sharply from the country average for quite a long time, but the size of the Hungarian Roma population cannot be established directly on the basis of ethnicity and language use as recorded in the population censuses. As the population considered being Roma is far larger than that of Roma minority as reported in the Census - while both populations are strongly disadvantaged concerning their socio-economic situation and education career. Using the microsimulation technique we assigned Roma status to a subset of the non-Roma population. The probability of being Roma was calculated from a probit model of the reported Roma status. We pre-defined the number of people with assigned Roma status in order to get a Roma population (including people with both reported and assigned Roma status) equal to the population being considered as Roma according to representative sociological surveys.

Transition probabilities in the model

Transition probabilities in the ISMIK model were mostly estimated with econometric models using individual level data. In case of some events in the model macro level data were used. The following datasets were used to estimate transition probabilities:

- 2006-2009 waves of *Hungarian Life Course Survey* conducted by *TÁRKI- Educatio Kht*. The survey is an individual panel survey administered yearly which follows education history of 10 000 students from final year of lower secondary school (grade 8). We know from the survey family background variables of students also.
- *Public Education Statistics of Ministry of Education (KIRTAT*). Data for years 2001-2010
- Individual level *Higher Education Application-Admission Statistics* of Ministry of Education (*FELVI)* , Data for years 2001-2009

- *Higher Education Statistics of Ministry of Education*
- *2001 Census*
- *2005 Micro Census*
- *Follow-up Survey of Hungarian Higher Education Graduates (DPR)* 2010

Drawing method in the ISMIK model

In the ISMIK model - following the Norwegian MOSART model - we use "means constrained stochastic drawing method". That is we constrain the simulated number of events to the expected number of events, which is in general the sum of the transition probabilities without changing each person's probability of experiencing the event. With exogenous probabilities the expectation value of the model is the sum of the transition probabilities.

The events in the ISMIK model are summarized in *Table 2,* including covariates and estimation methods of transition probabilities

Table 2

Events in the ISMIK models and estimation of transition probabilities

Events in sequence	*Who is „at risk" commencing*	*Outcomes*	*Covariates*	*Estimation technique*
I. DEMOGRAPHIC EVENTS				
I. 1. FERTILITY				
I. 1.1. Fertility	women aged 15- 50	1. Yes 2. No	Age, educational attainment, Roma, type of settlement, county of settlement	logit
I.1.2. Determining date of birth	newborns	1. Until September 2. After 1st of September	–	ratios
I.1.3. Sex of newborn	newborns	1. Girl 2. Boy	–	ratios
I.1.4. Educational attainment of the father	newborns	1. Less than lower secondary 2. Lower secondary 3. Vocational attaining school 4. Upper secondary school 5. Higher education	Educational attainment of mother, Roma, type of settlement, county of settlement	multinomial logit
I.2. MORTALITY	newborns	1. Yes 2. No	Gender, educational attainment of mother	ratios
	aged 1-25		Age, gender	mortality probability ratios
	aged more than 25		Age, gender, educational attainment	

II. EDUCATION EVENTS				
II.1. PRIMARY AND LOWER SECONDARY EDUCATION				
Events in sequence	***Who is „at risk" commencing***	***Outcomes***	***Covariates***	***Estimation technique***
II.1.1. Completion of lower secondary education in full time education	aged 14-17	1. At age 14 2. At age 15 3. At age 16 4. At age 17 5. Does not complete lower secondary education in full time education	Gender, educational attainment of mother and father, Roma, type of settlement, county	multinomial logit
II.1.2. Completion of lower secondary school in part type education	aged 20, doesn't have lower secondary school qualification	1. Yes 2. No	–	ratios
II.2. SECONDARY EDUCATION				
II.2.1. Following studies at upper secondary level	14-17 year olds, finished lower secondary education in the given year of	1. Following studies in gymnasium 2. Following studies in secondary vocational school 3. Following studies in vocational training school 4. Does not follow further studies	Gender, educational attainment of father and mother, Roma, type pf settlement, county	logit

Events in sequence	***Who is „at risk" commencing***	***Outcomes***	***Covariates***	***Estimation technique***
II.2.2. Graduates at upper secondary school in full time education?	Who have followed studies at upper secondary level 4 or 5 year before	1. Upper secondary qualification in gymnasium in 4 years 2. Upper secondary qualification in vocational secondary schools in 4 years 3. Vocational training qualification in 4 years 4. Upper secondary qualification in gymnasium in 5 years 5. Upper secondary qualification in vocational secondary schools in 5 years 6. Vocational training qualification in 5 years 7. Does not get qualification	Gender, educational attainment of father and mother, Roma, type pf settlement, county	multinomial logit
II.2.3. Graduates from upper secondary school in part time education	22-29 year olds with lower secondary or vocational training school qualification	1. Upper secondary qualification in gymnasium 2. Upper secondary qualification in vocational secondary school 3. Does not get qualification	Gender, age, qualification	ratios

Events in sequence	*Who is „at risk" commencing*	*Outcomes*	*Covariates*	*Estimation technique*
II.3. HIGHER EDUCATION				
I.3.1. Following studies in higher education Tertiary B type education, full time education?	Graduated from upper secondary education 0-3 years before and haven't followed studies at tertiary level before	1. Yes 2. No	Gender, age, type of secondary school, year of graduation in secondary education, type of settlement, region	logit
II.3.2. Following studies in higher education Tertiary A type education, full time education	Graduated from upper secondary education 0-3 years before and haven't followed studies in full time education at tertiary level	1. Yes 2. No	Gender, age, type of secondary school, year of graduation in secondary education, type of settlement, region	logit
II.3.3. Graduates from tertiary B type education, full time education	Enrolled in full time tertiary B type education 1-3 years before	1. Yes, in 1 year 2.Yes, in 2 years 3. Yes,in 3 years 4. No	Gender	ratios
II.3.4. Graduates from tertiary A type education, full time education	Enrolled in full time tertiary A type education 3-8 years before	1. Yes, in 3 years 2. Yes, in 4 years 3. Yes, in 5 years 4. Yes, in 6 years 5. Yes, in 7 years 6. Yes, in 8 years 7. No	Gender	ratios
II.3.5. Graduates from part time tertiary A type education?	18-39 year olds with upper secondary or tertiary B type qualification who are not enrolled in full time tertiary education	1. Yes 2. No	Gender, age	ratios

II.3.6. Level of tertiary degree BA/college level or MA/university level?	Graduated from tertiary A type education in 0-1 year before	1. BA/college level degree 2. MA/university level degree	Gender	ratios

III. MIGRATION BETWEEN REGIONS				
Events in sequence	***Who is „at risk” commencing***	***Outcomes***	***Covariates***	***Estimation technique***
1. Region of habitation	aged 0-23 whose educational attainment is at most lower secondary school and who are living with parents	1. Central Hungary 2. Central Transdanubia 3. Western Transdanubia 4. Southern Transdanubia 5. Northern Hungary 6. Northern Great Plan 7. Southern Great Plan	Age, educational attainment of mother (or father), region of habitation in the previous year	multinomial logit
	aged 23-74 and aged 0-23 whose educational attainment is higher than lower secondary school an/or who are not living with parents		Gender, age, educational attainment, region of habitation in the previous year	multinomial logit

3. SIMULATION RESULTS FOR THE EDUCATIONAL ATTAINMENT OF THE POPULATION 2001-2020

The following table and figures present the results of the microsimulation. Results for the base scenario are summarized in Section 3.2. The base scenario assumes that all transition probabilities remain the same as were observed in the early 2000s, that is, this scenario projects education level of the population on the condition that everything continues as in the early 2000s.

Results for Scenario 2 are summarized in Section 3.2. Scenario 2 investigates what would have been educational attainment of the population had Roma and Non-Roma students had the same opportunities in their schooling career.

Results for Scenario 3 and Scenario 4 are summarized in Section 3.3. These scenarios measure the effects of educational policy changes: reducing the duration of compulsory schooling; increasing the share of students in secondary vocational training schools; and reducing the number of state-funded places in higher education.

3.1. RESULTS OF BASE SCENARIO –"EVERYTHING GOES ON AS BEFORE"

Results of the base scenario for the 25-64 year old population are summarised in Table 3.1 and Fugure 3.1. The overall results show that the long term trend of increasing educational attainment continues, though its pace is slowing down regarding the share of the low skilled. Note that changes in educational attainment are driven by two factors. First, as a young cohort replaces an old one in each year attainment rises "mechanically" as old cohorts were in school 50-40 years ago, when the average attainment was much lower than these days. Second, changes in educational attainment are also affected by the difference between current young cohorts in one year and the next year. Although the impact of this second factor on the average attainment of the population is much smaller, it is more interesting, as it reveals future trends of attainment. Moreover, this is what the current educational policy can change. For this reason we present the results for the young cohorts in more detail.

Table 3.1

Distribution of 25-64 year olds by highest educational attainment per cent

Year	Less than lower secondary	Lower secondary	Vocational training school	Upper secondary	Higher education	Total
2001	3.7	29.6	24.1	27.9	14.6	100
2010	2.4	21.7	27.5	31.1	17.3	100
2015	2.2	19.1	27.3	32.1	19.3	100
2020	2.1	16.5	26.8	32.8	21.3	100

Figure 3.1

Distribution of the 25-64 year olds by highest educational attainment

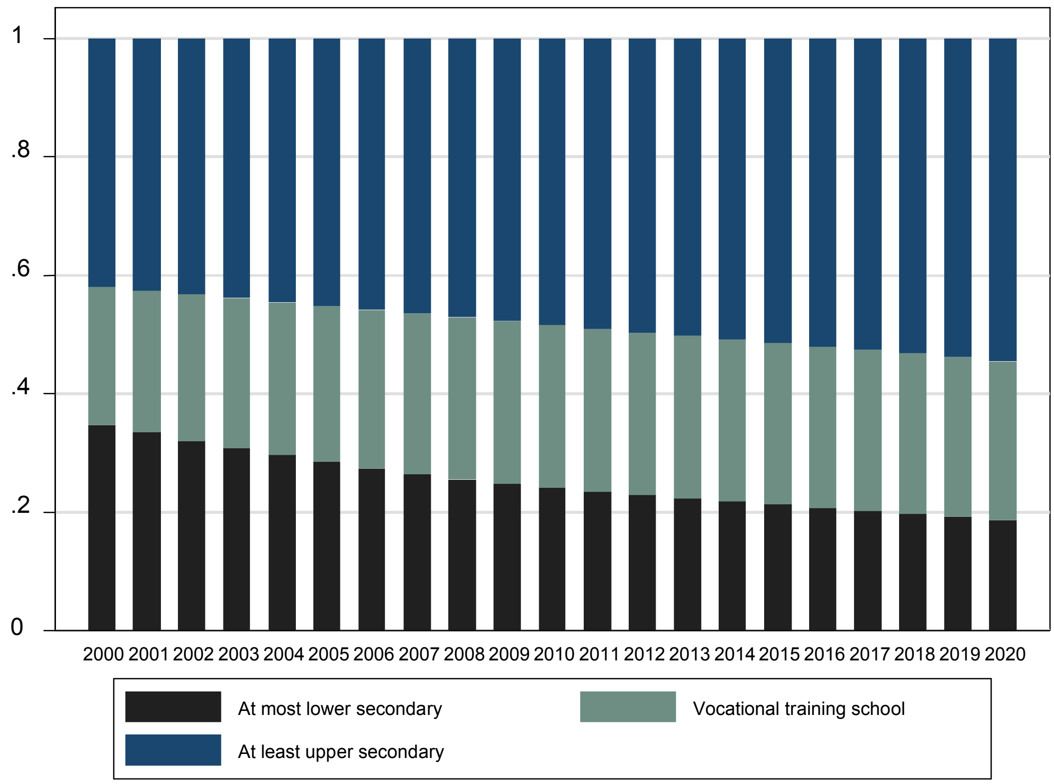


Figure 3.2-3.6 displays changes in educational attainment for three groups of young cohorts by level of education. Overall the results suggest the increasing trend can be expected to slow down or even halt.

Regarding the share of people with no education, i.e. drop outs before grade 8, the end of lower secondary education, the trend even shifts from decreasing to increasing, though the share of this attainment group remains small even at the end of the period (Figure 3.2). The share of people with lower secondary education appears to stagnate among the young in the next decade, but can not be ruled out to increase later, as indicated by the trend for the 20-24 cohorts (Figure 3.3). The main reason of these changes is the increasing share of Roma population, with a non-negligible propensity to drop out from school at the end of the compulsory education period. Note that these changes suggest increasing poverty, as employment rates for the low skilled are especially low in Hungary.

Figure 3.2

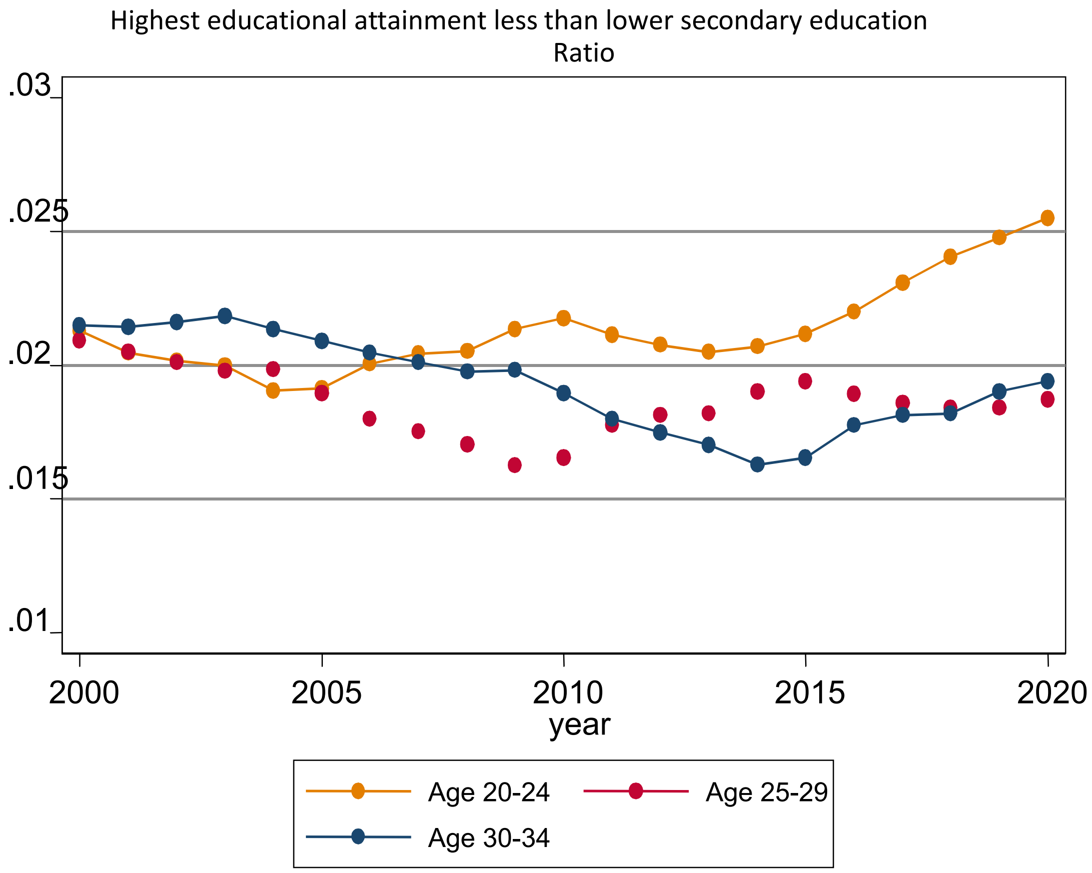
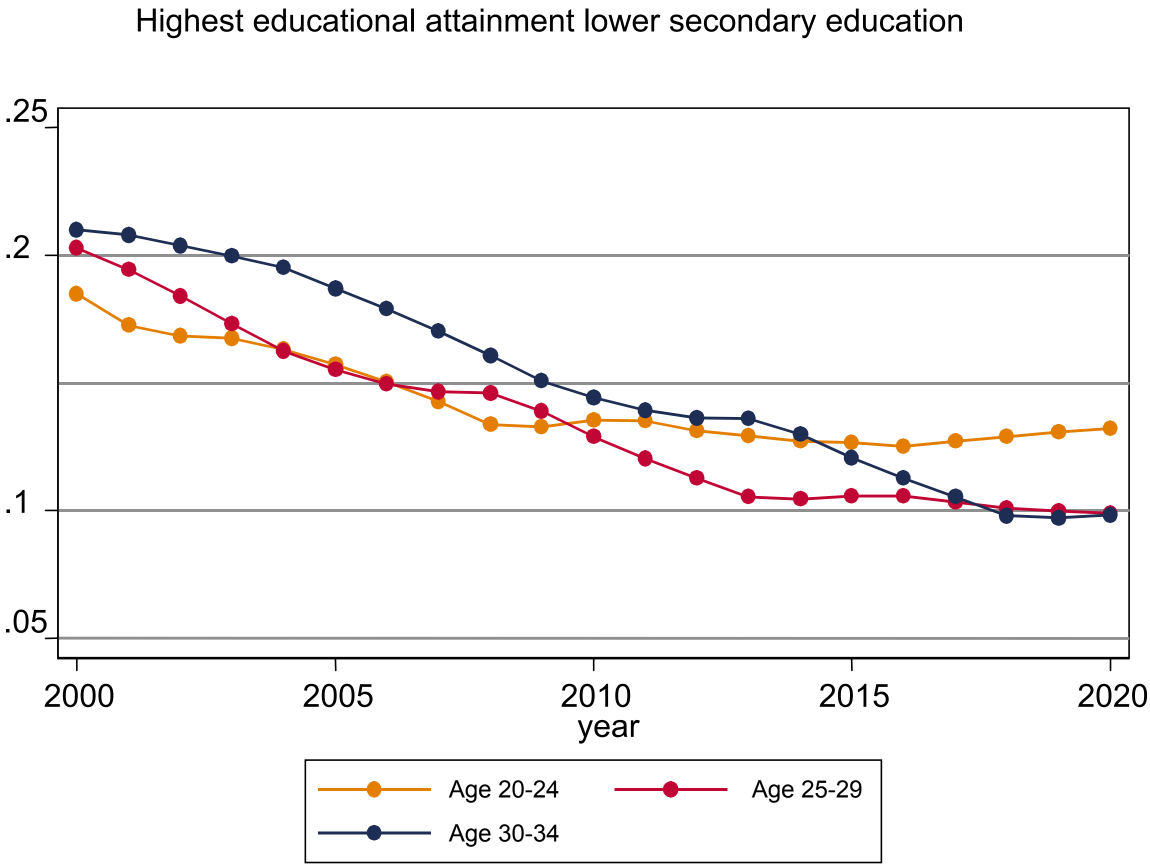


Figure 3.3



Upper secondary education is divided into three tracks in Hungary, but the most important distinction is between the vocational track and the other two; the academic track and a mixed one with a curriculum very similar to the academic track complemented by some vocational education. The vocational track follows a less demanding curriculum and does not qualify for higher education.

The share of people with upper secondary education in the vocational track has been rising until the mid nineties, as the share of early school leavers has decreased. However, vocational schools became less popular in the past two decades, and the share of this attainment group started to drop (Figure 3.4). This can be expected to turn into stagnation in the next few years. However, note that the share in the working age population is still increasing in the entire period, due to old cohorts with an even smaller share, thus a shortage in the labour supply with a vocational school degree is not likely to emerge on the short and medium run.

The share of people with an upper secondary degree from the academic or mixed track displays a mirror image of that of vocational degree (Figure 3.5). Following a period of rapid expansion it slows down and turns into stagnation.

The share of graduates shows a similar trend (Figure 3.6).

Figure 3. 4.

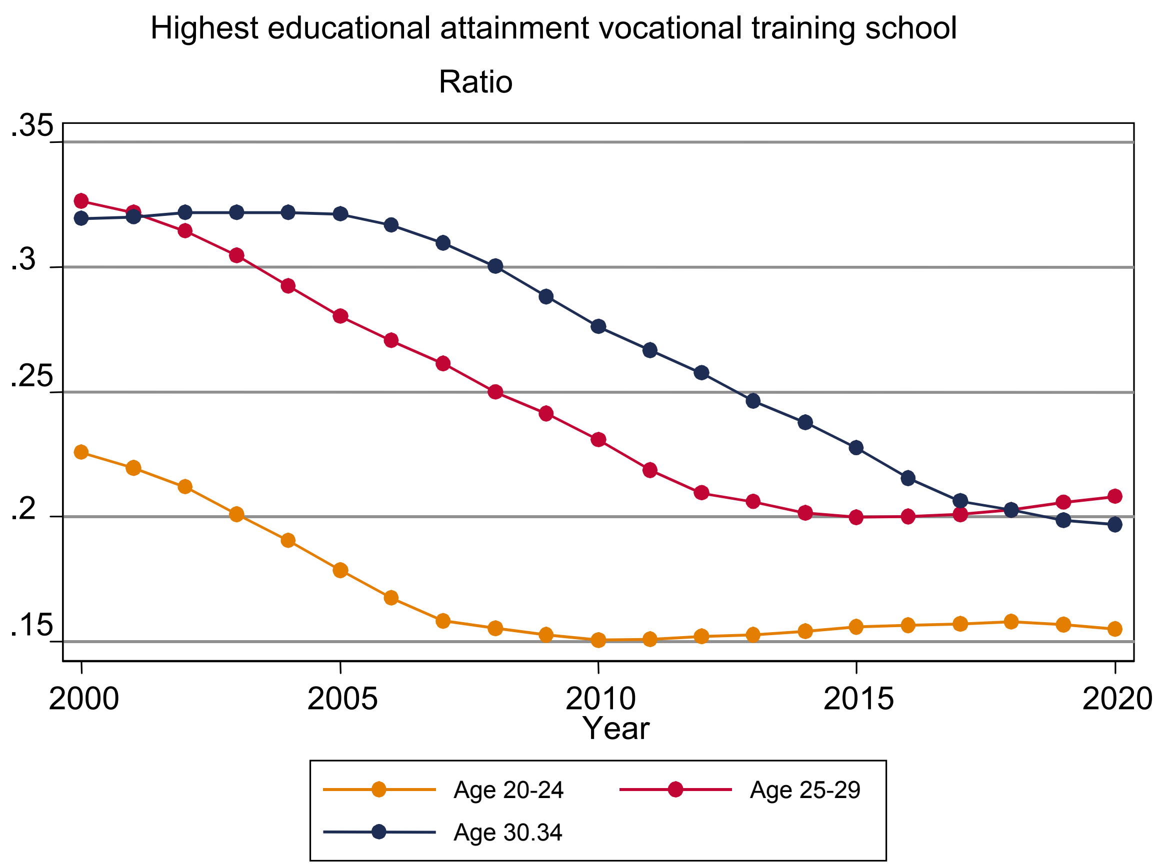


Figure 3.5

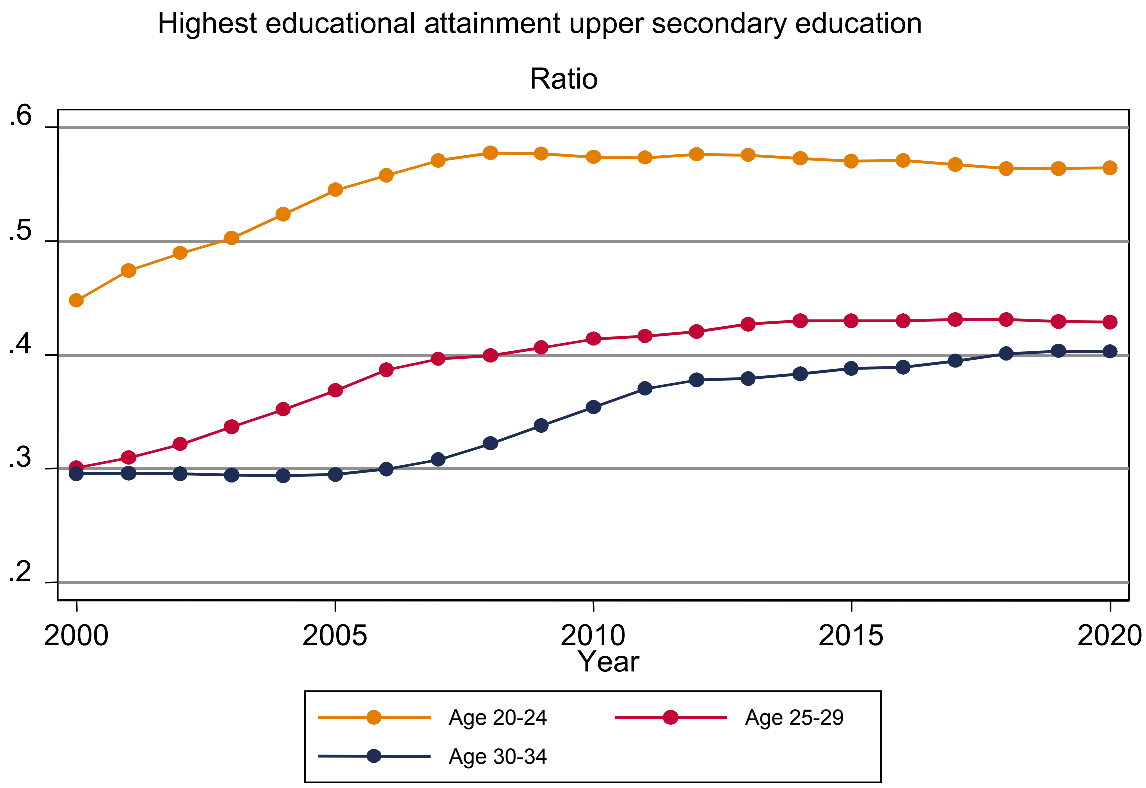
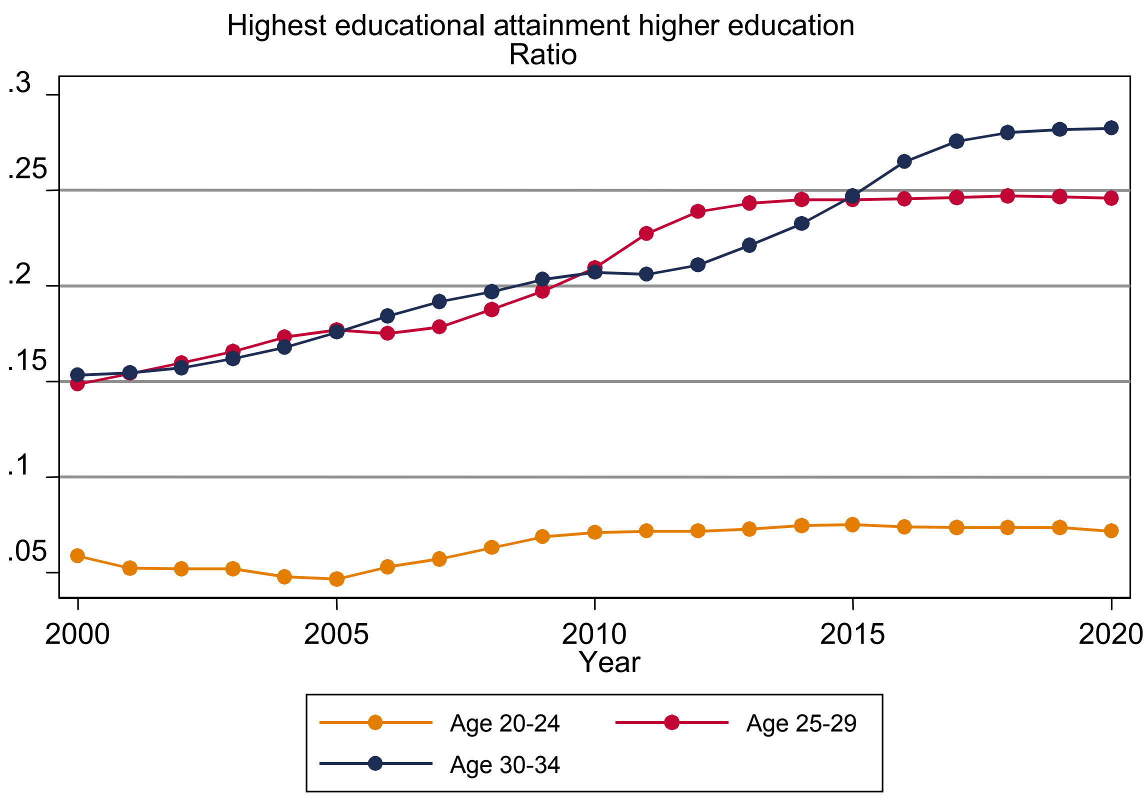


Figure 3.6



Beside the overall attainment shares it is worth to look at these by gender. Historically men outperformed women in educational attainment, but this has turned into the opposite in the past decades. The share of low skilled in the entire population of women has decreased more, and in the young cohorts it is already 2-3 percentage points below that of men (Figure 3.7 and 3.8).

Figure 3.7.

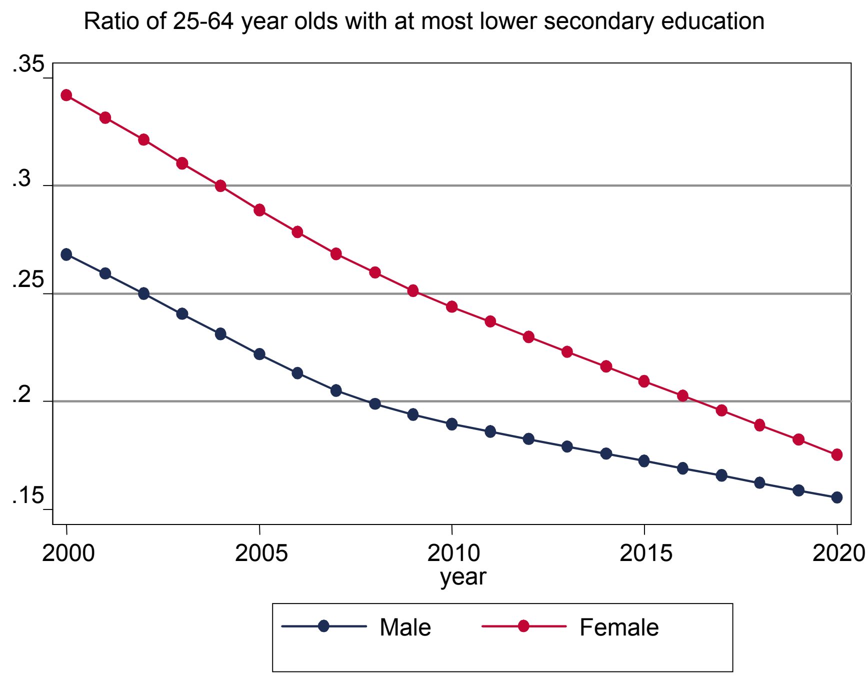
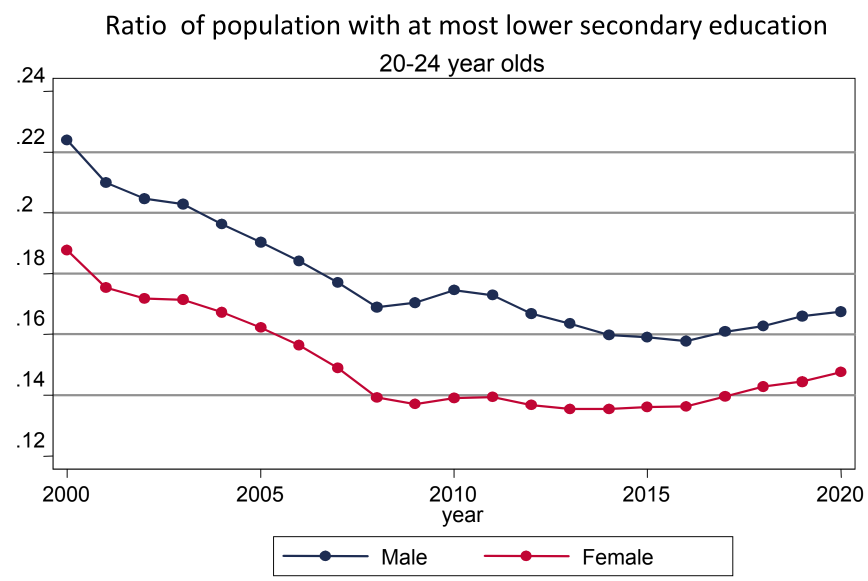


Figure 3.8.



Changes are even more marked regarding the share of graduates, where women do account for the major part of the overall increase, both for the 25-64 year old population and young cohorts (Figure 3.9 and 3.10).

Figure 3.9

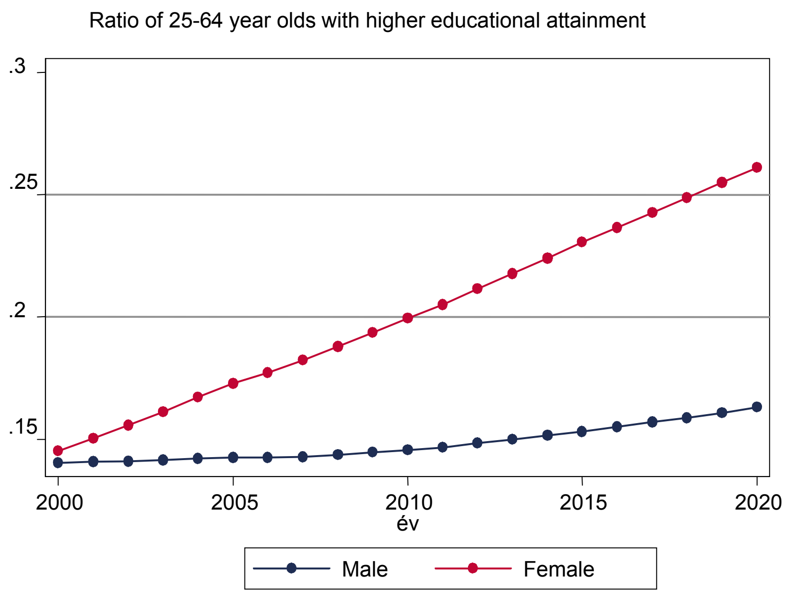
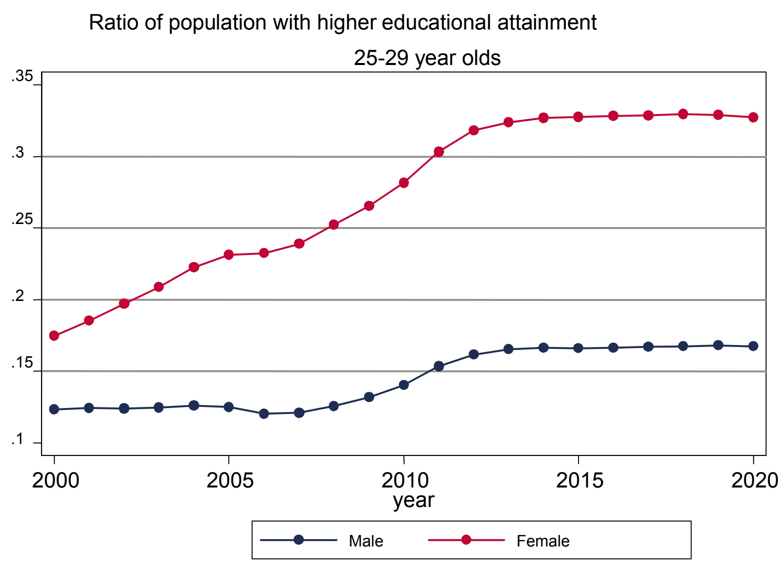


Figure 3.10.



3.2. RESULTS SCENARIO 2 - ROMA AND NON-ROMA STUDENTS HAVE THE SAME OPPORTUNITIES IN THEIR SCHOOLING CAREER

Providing an effective education for the Roma children in order to compensate at least in part for the effect of family background is one of the major challenges faced by education policy in Hungary. In this section we present predictions assuming that education policy is more successful in this respect than nowadays. In this scenario we assume that Roma children follow identical educational careers to the non-Roma with a similar family background, i.e. we discard the disadvantages face specifically by the Roma in school. However, the effect of the low education of parents is assumed to remain at work. Technically speaking the effects captured by the Roma dummy variable in the transition probability estimates are assumed to be zero in this scenario.

Figure 3.11-3.13 compares the educational attainment in this scenario to the base model. The results suggest that a significant improvement could be achieved by eliminating the disadvantages of Roma children in schools. Note that this is not about compensating the family background effect, which any socially responsible educational policy programme should target, but a less ambitious goal of merely providing similar educational quality than the non-Roma with similar social status face in schools.

Finally Figure 3.14 represents regional disparities in the share of low skilled in young cohorts at the end of our forecasting period in the base and the more effective Roma education scenarios. The results reveal that the massive (and according to the base scenario increasing) regional disparities could be mitigated substantially if Roma students were face better opportunities in education.

Figure 3.11.

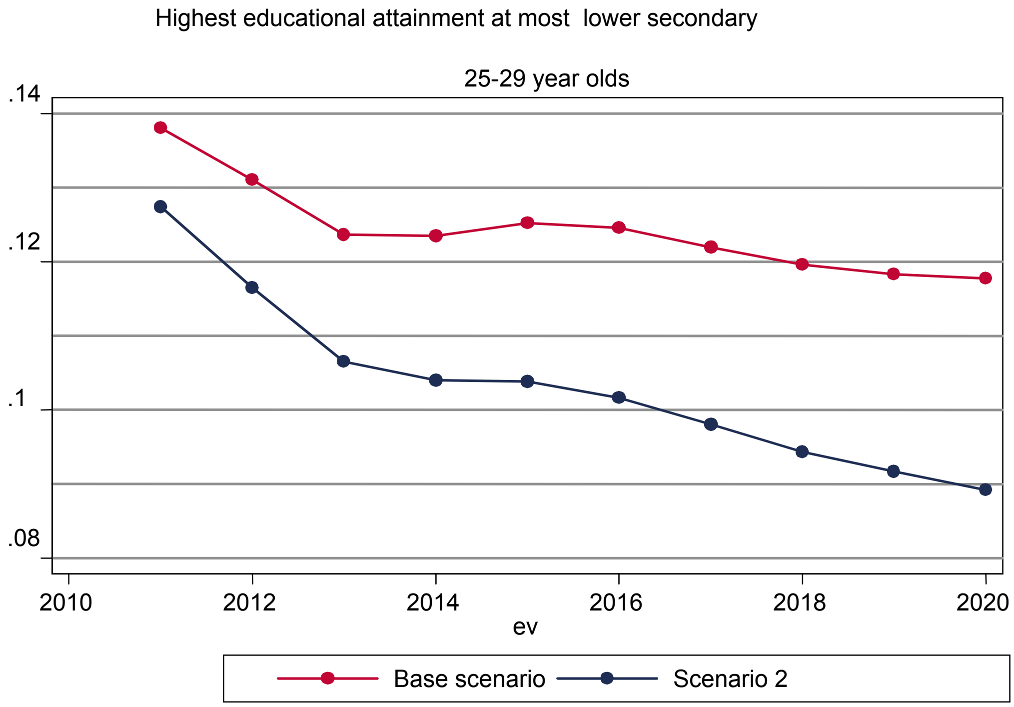


Figure 3.12.

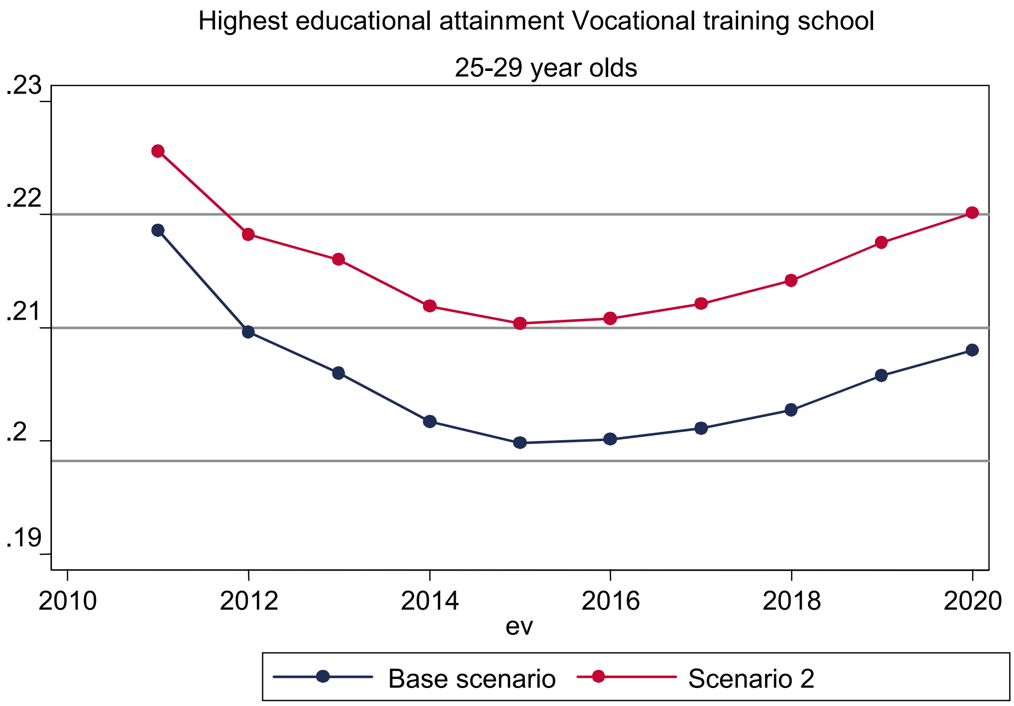


Figure 3.13.

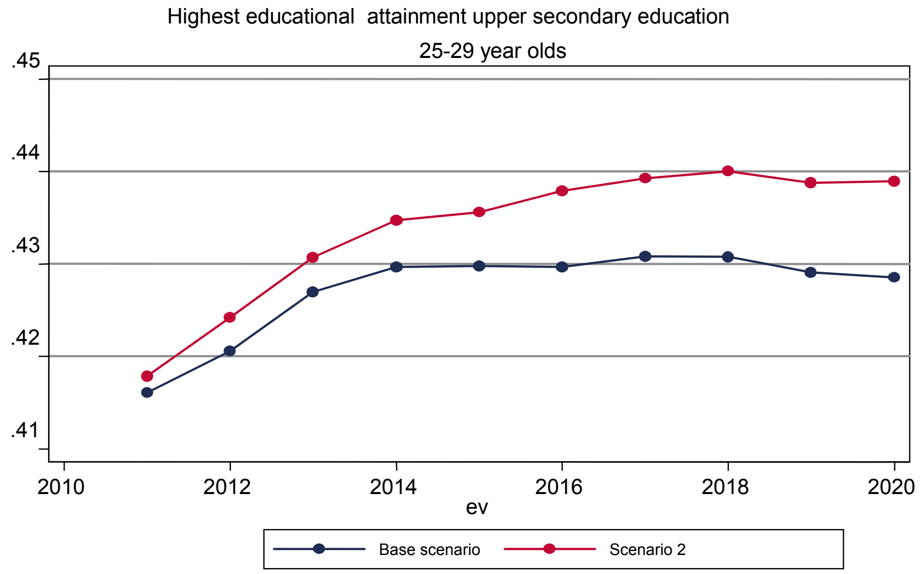
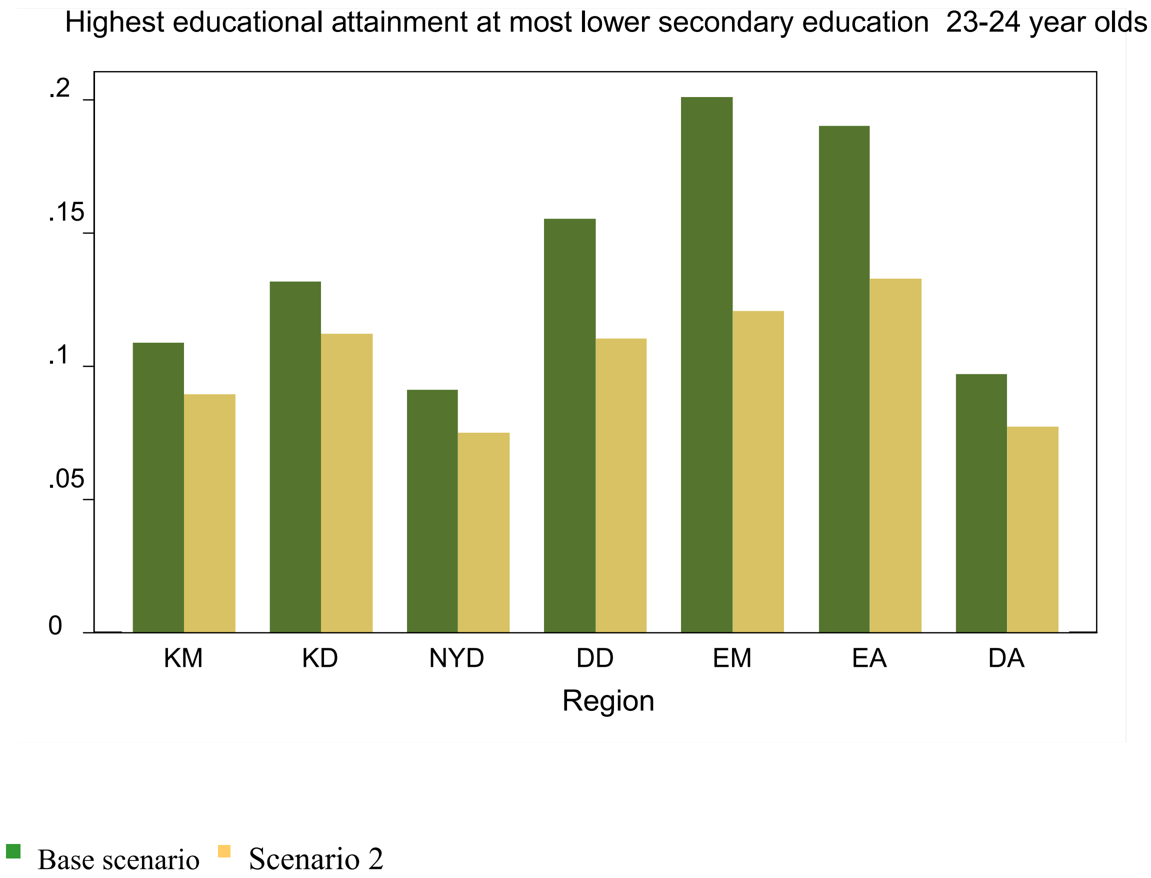


Figure 3.14.

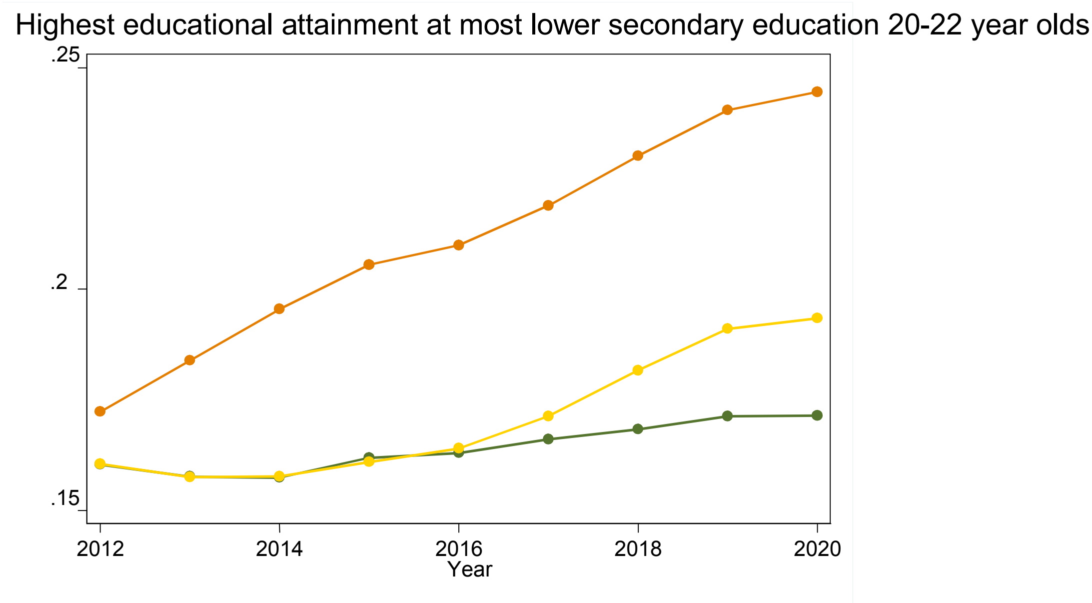


3.3. RESULTS SCENARIO 3 AND SCENARIO 4 - EFFECTS OF EDUCATIONAL POLICY CHANGES

In the 2010-2014 period the government initiated some major reforms in education. In this section we try to disclose the likely effects of these reforms by comparing forecasts taking into account these to our no-reform base forecast. Three policy reforms are considered here: decreasing the compulsory education age from 18 to 16 years, increasing the share of vocational track enrolment to 35 percent and decreasing the number of students in higher education with no tuition. The second and third policies are relatively easy to cope with in the microsimulation model as external constraints, while the first one raises questions more directly about behavioural responses. Compulsory education age has been raised from 16 to 18 years recently. Thus it is a natural to assume that this policy pushes drop out rates in upper secondary education to the level prevailing before the 18 year compulsory age episode, as measured in the Census of 2001. The first variant of the third scenario is built on this assumption. However, it can be argued that the effects of changes in compulsory education are not symmetrical. As most students and families have grown accustomed to education until the end of upper secondary education, drop out rates might not return to the previous level when compulsory education age is lowered again. The second variant of the third scenario reflects this argument, providing a lower bound estimate of the effect of current policies.

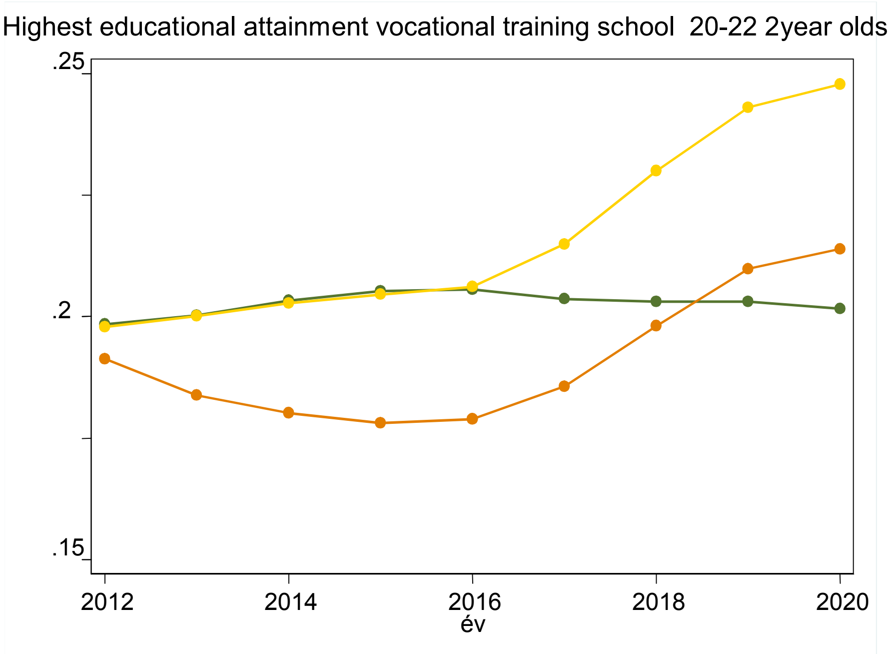
Our microsimulation model suggests that these educational reforms are likely to increase the share of low skilled, while decreasing the graduation rate and also the share of students leaving education with an upper secondary degree from the academic or mixed track (Figure 3.15, 3.17 and 3.18). The effect on the share of students with a vocational school degree is somewhat ambiguous (Figure 3.16). If drop out rate does not increase, the share of students finishing vocational school goes up, as intended by policy makers. However, if the drop out rate returns to previous levels (i.e. that before the compulsory education age was raised to 18 years), this may in part offset increasing enrolment in the vocational track, since dropping out of upper secondary education affects mostly students of this track.

Figure 3.15.



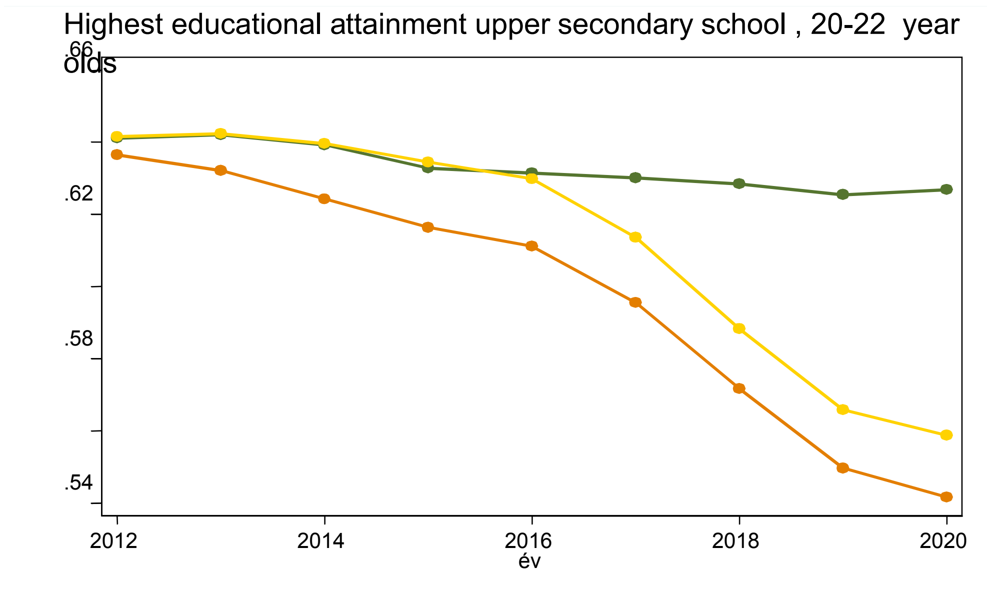
— Base scenario — Scenario 3 with unchanged drop out rates — Scenario 3 with changing drop out rates

Figure 3.16.



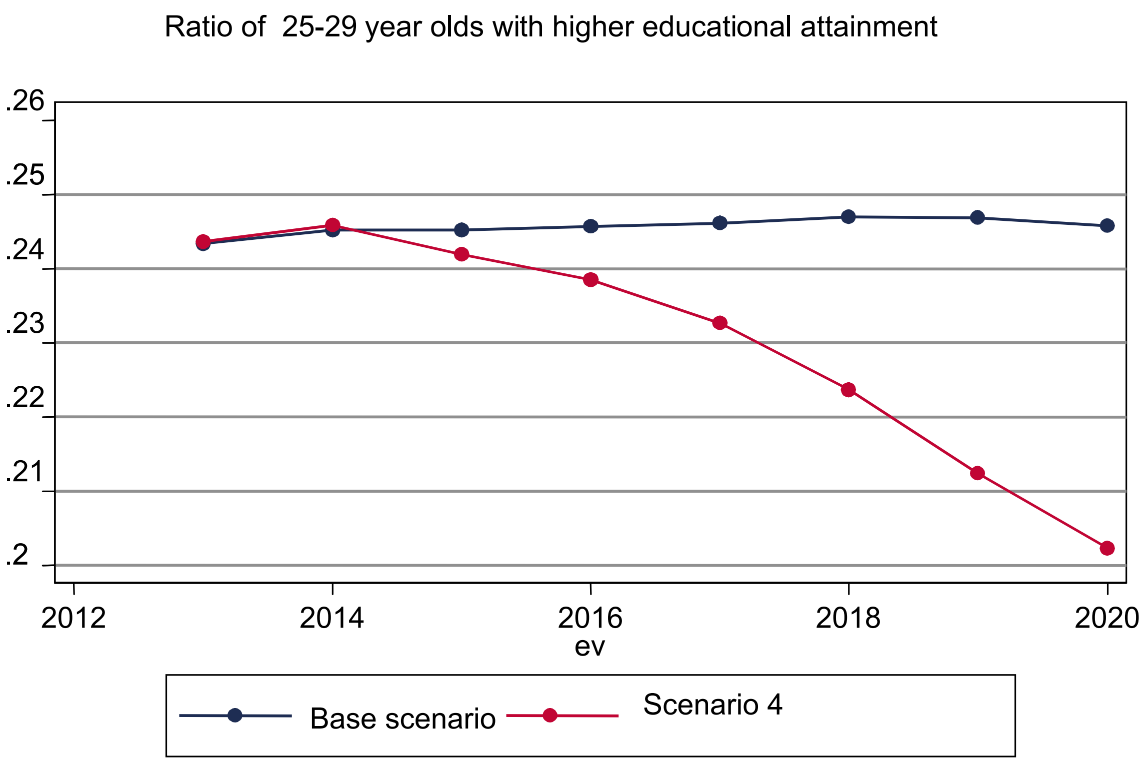
— Base scenario — Scenario 3 with unchanged drop out rates — Scenario 3 with changing drop out rates

Figure 3.17.



— Base scenario — Scenario 3 with unchanged drop out rates — Scenario 3 with changing drop out rates

Figure 3.18.



References

Courtioux,P. – Gregoir, S. – Houe, D. (2009): The Simulation of the Educational Output over the Life Course: The GAMEO Model Paper presented at the 2nd general conference of the International Microsimulation Association "Microsimulation: Bridging Data and Policy" 2009, Ottawa (Ontario), Kanada

Flood, J., Jansson, F, Petterson T, Sundber, O., Westerberg, A (2005): SESIM III.A Swedish dynamic microsimulation modell. http://www.sesim.org/

Fredriksen, D. (1998): Projections of Population, Education, Labour Supply and Public Pension Benfits. Analysis with the Dynamic Microsimulation model MOSART. Statistics Norway

King, A - Bækgaard , H. - Robinson, M. (1999): DYNAMOD2 – an Overview. National Centre for Social and Economic Modelling. University of Canberra Technical Paper No. 19

Liskó Ilona (2002): Cigány tanulók a középfokú iskolákban. Kutatás közben sorozat, 234. kötet, Oktatáskutató Intézet, Budapest.

Martin Spielauer (2010): Persistence and change of the relative difference in educational attainment by ethnocultural group and gender in Canada. Vienna Yearbook of Population Research 2010 (Vol.8), pp. 261-282

Merz, J.(1994): Microsimulation – A Survey of Methods and Applications for Analyzing Economic and Social Policy. FFB Discussion Paper N.9. Universität Lünebug

Payne, A. - Precival R .- Harding A.: (2008): APPSIM – Modelling Education: NATSEM University of Canberra Working Paper No. 11

Robinson M., - Baekgaard, H. (2002): Modelling Students in DYNAMOD2. NATSEM. University of Canberra . Technical Paper No. 23.

Statistics Canada (2011):The LifePath Microsimulation Model. An Overview. http://www.statcan.gc.ca/microsimulation/lifepaths/lifepaths-eng.htm

Zaidi A., Rake, K. (2001): Dynmic Microsimultion Models. Review and Some Lessons from SAGE. SAGE Discussion Papers No. 2.