CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economics Studies

Bachelor thesis

2017 Veronika Marková

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economics Studies

Veronika Marková

How Smart will be Europe 2020? A Panel Data Analysis.

Bachelor thesis

Prague 2017

Author: Veronika Marková
Supervisor: doc. Ing. Tomáš Cahlík CSc.

Academic Year: 2016/2017

Bibliographic note

MARKOVÁ, Veronika *How Smart will be Europe 2020? A Panel Data Analysis.* Prague 2017. 73 pp. Bachelor thesis (Bc.). Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor doc. Ing. Tomáš Cahlík CSc.

Range of thesis

Number of character with spaces: 55 217

Abstract

This thesis analyses the targets for education under Europe 2020 strategy - early leavers from education and tertiary education attainment. The main objective is to predict the future development for these indicators up to 2020 and answer the question whether these targets will be fulfilled. From the empirical perspective, it applies econometric analysis of panel data based on human capital theory, that relates the education level of an individual with the parental education and the labour market conditions.

Keywords Europe 2020, European Union, Early leavers from education, Tertiary education, Human capital theory, Panel data analysis

Author's e-mail **veronik.markova@gmail.com**
Supervisor's e-mail **cahlik@fsv.cuni.cz**

Bibliografický záznam

MARKOVÁ, Veronika *How Smart will be Europe 2020? A Panel Data Analysis.* Praha 2017. 73 pp. Bakalářská práce (Bc.) Univerzita Karlova, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí práce: doc. Ing. Tomáš Cahlík CSc.

Abstrakt

Tato práce analyzuje cíle strategie Evropa 2020 zaměřené na vzdělávání - míru předčasného ukončování školní docházky a podíl osob s dokončeným terciárním vzděláním. Hlavní cíl této práce je předpovědět budoucí vývoj těchto indikátorů až do roku 2020 a odpovědět na otázku, jestli budou tyto cíle splněny. Z empirického hlediska je použita ekonometrická analýza panelových dat založená na teorii lidského kapitálu, která dává do souvislosti úroveň dosaženého vzdělání jedince s úrovní vzdělání jeho rodičů a s podmínkami na pracovním trhu.

Klíčová slova Evropa 2020, Evropská unie, Předčasné ukončování školní docházky, Terciální vzdělávání, Teorie lidského kapitálu, Analýza panelových dat

E-mail autora **veronik.markova@gmail.com**
E-mail vedoucího **cahlik@fsv.cuni.cz**

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 31, 2017

Signature

Acknowledgment

I would like to express my sincere gratitute to my supervisor, doc. Ing. Tomáš Cahlík CSc., for his remarks and patient guidance, without which it would not be possible to complete this thesis.

Bachelor Thesis Proposal

Author	Veronika Marková
Supervisor	doc. Ing. Tomáš Cahlík CSc.
Proposed topic	How Smart will be Europe 2020? A Panel Data Analysis.

Topic specification:

Europe 2020 Strategy was launched in 2010 with aim to create conditions for smart, sustainable and inclusive growth. There are five headline targets that were agreed to be achieved by 2020 including employment, research and development, climate and energy, education, social inclusion and poverty reduction. My thesis will focus on the targets for education, I will analyse the current development and I will construct an econometric model to forecast the future progress. The main objective is to answer the question, whether the Europe 2020 targets will be fulfilled. Following the approach of Dragomirescu-Gaina & Weber (2013), the empirical model will be set on the human capital theory, presented by Becker (1964), that describes the incentives of individuals for entering or leaving the education processes based on their socio-economic background. As an empirical method, I will use the econometric analysis of panel data.

Research question:

Will be the targets for education under Europe 2020 strategy fulfilled?

Outline

1. Introduction
2. Europe 2020
- overview of Europe 2020 targets
- actual fulfilment of education targets
3.Theoretical background
- human capital theory
4.Methodology
- overview of panel data theory and methods
5.Empirical Model
6.Forecast Method
7.Results & Discussion
8.Conclusion

Core bibliography

1. BECKER, G. (1993) [1964]. *Human capital: a theoretical and empirical analysis, with special reference to education (3rd ed.).* Chicago: The University of Chicago Press. ISBN 0226041204.
2. DRAGOMIRESCU-GAINA, C., WEBER, A. (2013). *Forecasting the Europe 2020 headline target on education and training: A panel data approach.* Luxembourg: Publications Office. ISBN 9789279325090.
3. EUROPEAN COMMISSION. (2010) *Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth.* Communication from the Commission: Publications Office. Brussels. Available at: http://ec.europa.eu/research/era/docs/en/ec-understanding-era-13.pdf [05-12-2015]
4. GREENE, W. H. (2012) *Econometric analysis. 7th ed.* Boston: Prentice Hall. ISBN 9780131395381.
5. WOOLDRIDGE, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Contents

List of Figures

List of Tables

List of Country Abbreviations

AT	Austria
BE	Belgium
BG	Bulgaria
CY	Cyprus
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
EL	Greece
ES	Spain
FI	Finland
FR	France
HR	Croatia
HU	Hungary
IE	Ireland
IT	Italy
LT	Lithuania
LU	Luxembourg
LV	Latvia
MT	Malta
NL	Netherlands
PL	Poland
PT	Portugal
RO	Romania
SE	Sweden
SI	Slovenia
SK	Slovakia
UK	United Kingdom

Appendix

Introduction

How smart will be Europe 2020? How smart will be the Europe by year 2020? With this headline, I am referring to the smart pillar of the strategy of European Union - the Europe 2020 presented by European Commission (2010). Europe 2020 was launched in 2010 as a successor of Lisbon Strategy, with similar goals - to enhance growth, promote jobs and to make European Union stronger and better. As the Lisbon Strategy, Europe 2020 is a ten-year plan. Its subheading is "A strategy for smart, sustainable and inclusive growth" reflecting the three perspectives or pillars from which the growth is perceived, as well as the three right paths to achieve it. With its headline targets and initiatives Europe 2020 focuses on five areas: Employment, Research & Development, Climate Change & Energy, Education and Poverty & Social Exclusion.

Among the areas of Europe 2020 strategy, I have decided to focus on education as the importance of knowledgeable society is undoubtable. The targets for Education under Europe 2020 strategy are the following: rate of *early leavers from education* below 10% and percentage of *tertiary education attainment* within population aged 30-34 at least 40%. The key question of my thesis is: How Smart will be Europe 2020? in other words: Will these targets be fulfilled?

To answer, I will set the empirical model based on human capital theory presented by Becker (1964, 1993), that describes the incentives of individuals for entering or leaving the education process based on their socio-economic background. I will follow the approach of Dragomirescu-Gaina & Weber (2013) and Dragomirescu-Gaina et al. (2015) with more current data or/and other variables used in the model. As an empirical method, I will use the econometric analysis of panel data. My final objective is to forecast the values for year 2020 and to evaluate the results on

the member states and on the European Union level.

In the first chapter I will go further with the Europe 2020 strategy settlement, I will explain the education targets in more detail and I will evaluate the actual status; in the next, I will discuss the theory of human capital - the basis and argumentation for my empirical model. Then I will shortly review the econometrics background and the methodology concerning the panel data methods - the method used in my models. Furthermore, I will describe the data source and assumptions used while selecting the indicators, the main features of our empirical model and the best method discussion. The last part of my thesis covers discussion over the results and the conclusion.

1 Europe 2020

The Europe 2020 Strategy was prepared by the European Commission in March 2010 and the Council adopted it in June 2010 as a ten-years growth plan following the path of former Lisbon Strategy, that was on run from 2000 to 2010. The new European strategic framework considered both non-achievements of some of the goals of Lisbon Strategy and the recent worldwide financial and economic crisis still resonating in 2010. Considering the lessons learned from the crisis, 2010 was seen as a new beginning. And the complex new strategy was built in a way to make Europe stronger, stable and sustainable in the future. While setting the new goals, the previous targets from the Lisbon Strategy, as well as a competitive comparison with other Western developed countries (such as USA and Japan) were considered. As stated by J. M. Barroso in the Communication from the Commission (European Commission, 2010) these targets are ambitious, but attainable and their attainment is crucial for sustainable future of Europe.

As was already mentioned, Europe 2020 is a strategy for smart, sustainable and inclusive growth. But what are these characteristics supposed to mean? The Smart Growth is perceived as developing an economy towards knowledge based economy; in other words, more (and more effective) investment in education, research and innovation; the Sustainable Growth focuses on greener, more resource efficient and more competitive economy, and it also covers the agenda for environmental policies; the Inclusive Growth encourages high employment and enhances the economic, social and territorial cohesion, it emphasis job creation and poverty reduction. Putting all these three growth perspectives together, it results in the following headline targets:

Table 1: EUROPE 2020 targets

1. Employment	- 75% of the 20-64-year-olds to be employed
2. Research & Development	- 3% of the EU's GDP to be invested in R&D
3. Climate Change & Energy (compared to 1990 levels):	- Limiting greenhouse gas emissions by 20% or even 30% - Providing 20% of energy from renewables - 20% increase in energy efficiency
4. Education	- At most 10% of early leavers from education and training - At least 40% of 30-34 years-old completing tertiary education
5. Poverty & Social Exclusion (compared to 1990 levels):	- Ensuring at least 20 million fewer people in or at risk of poverty and social exclusion

(source: author's elaboration; based on Eurostat)

1.1 Education targets: Overview

Even though the future vision of European Union is (and always was) to reduce dramatic territorial differences between member states, the situation is not in all areas close to reality. To ensure the targets of Europe 2020 will be reasonable for every of the 28 member states, and having in mind the different starting levels in 2010, the global European targets of Europe 2020 were split and cascaded down into a national level (as a result of discussions between EU and Member States). The National Targets (set out in National Reform Programmes in April 2014)[1] for the Education area are as listed in Table 2 below:

[1]National Reform Programme (NRP) is a part of check & balance mechanism regarding every Member State in so called European Semester. European Semester (first kicked of in 2015) is a yearly cycle of economic policy coordination that evaluates the achievements and risks in fulfilment of targets and also provides recommendations for every Member State

Table 2: National targets for Education indicators

Member States		ELE	TEA	Member States		ELE	TEA
Austria	AT	9.5%	38.0%[1]	Ireland	IE	8.0%	60.0%
Belgium	BE	9.5%	47.0%	Italy	IT	16.0%	26-27%
Bulgaria	BG	11.0%	36.0%	Lithuania	LT	9.0%	48.7%
Cyprus	CY	10.0%	46.0%	Luxembourg	LU	10.0%	66.0%
Czech Republic	CZ	5.5%	32.0%	Latvia	LV	10.0%	34-36%
Germany	DE	10.0%	42.0%[2]	Malta	MT	10.0%	33.0%
Denmark	DK	10.0%	40.0%	Netherlands	NL	8.0%	40.0%
Estonia	EE	9.5%	40.0%	Poland	PL	4.5%	45.0%
Greece	EL	9.7%	32.0%	Portugal	PT	10.0%	40.0%
Spain	ES	15.0%	44.0%	Romania	RO	11.3%	26.7%
Finland	FI	8.0%	42.0%	Sweden	SE	10.0%	40-45%
France	FR	9.5%	50.0%[3]	Slovenia	SI	5.0%	40.0%
Croatia	HR	4.0%	35.0%	Slovakia	SK	6.0%	40.0%
Hungary	HU	10.0%	30.3%	United Kingdom[4]	UK	NA	NA

(source: author's elaboration; based on Eurostat)

ELE = Early Leavers from Education; TEA = Tertiary Education Attainment

Comments: 1. including ISCED 4/4a; 2. including ISCED 4;3. 17-33-year-olds;

4. UK has no target in National Reform Programme

At this point, it is necessary to specify how exactly are the educational indicators calculated and to which level of education they refer. *Tertiary education attainment* is defined as the percentage of the population aged 30-34 who have successfully completed tertiary studies, *early leavers from education* are defined as the percentage of the population aged 18-24 with at most lower secondary education and who were not in further education or training during the last four weeks preceding the survey. The levels of education are defined by International Standard Classification of Education (ISCED), classification maintained by the United Nations, that is dividing education attainment and training into several levels. It is important to note that there was the change in a classification: former ISCED 1997 was valid until 2013 and from year 2014 the data are classified by the new ISCED 2011. Basically, the new classification extends the former scale. For Tertiary Education Attainment, the relevant educa-

tion levels are *First stage of tertiary education (5)* and *Second stage of tertiary education (6)* under ISCED 1997, under ISCED 2011 the levels are differentiated into *Short-cycle tertiary education (5)*, *Bachelor's or equivalent level (6)*, *Master's or equivalent level (7)* and *Doctoral or equivalent level (8)*. For Early Leavers from Education and Training the relevant levels are *Pre-primary education (0)*, *Primary education or first stage of basic education (1)* and *Lower secondary or second stage of basic education (2)* under both classifications even though under ISCED 1997 also some sub-category of Level 3 (3C short) was included and in the ISCED 2011 the levels are renamed. Following table represents the changes. For more details, see OECD (2015).

Table 3: ISCED 1997 vs ISCED 2011

	ISCED 1997		->	**ISCED 2011**
	Level	Description	Level	Description
ELE	0	Pre-primary education	0	Less than primary education
	1	Primary education or first stage of basic education	1	Primary education
	2	Lower secondary or second stage of basic education	2	Lower secondary education
	3	(Upper) secondary education	3	Upper secondary education
	4	Post-secondary non-tertiary education	4	Post-secondary non-tertiary education
TEA	5	First stage of tertiary education	5	Short-cycle tertiary education
	6	Second stage of tertiary education	6	Bachelor's or equivalent level
			7	Master's or equivalent level
			8	Doctoral or equivalent level

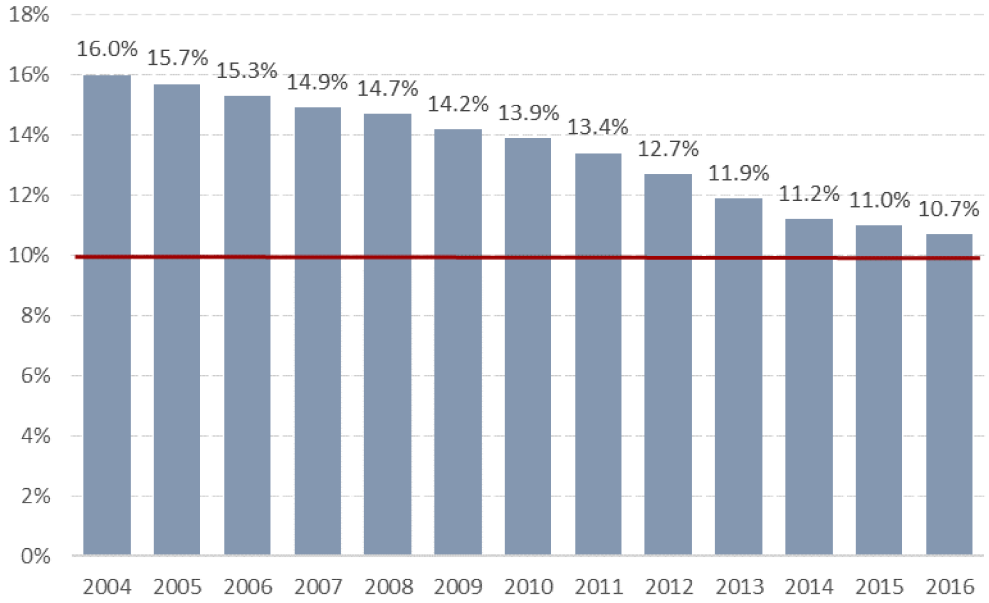
(source: author's elaboration; based on Eurostat)

1.2 Education targets: Fulfilments

The status of the education targets fulfilment is monitored on yearly basis, together with the recommendations and possible fulfilment exposures in *Education and training monitor*, produced by the European Commission, for the latest one see European

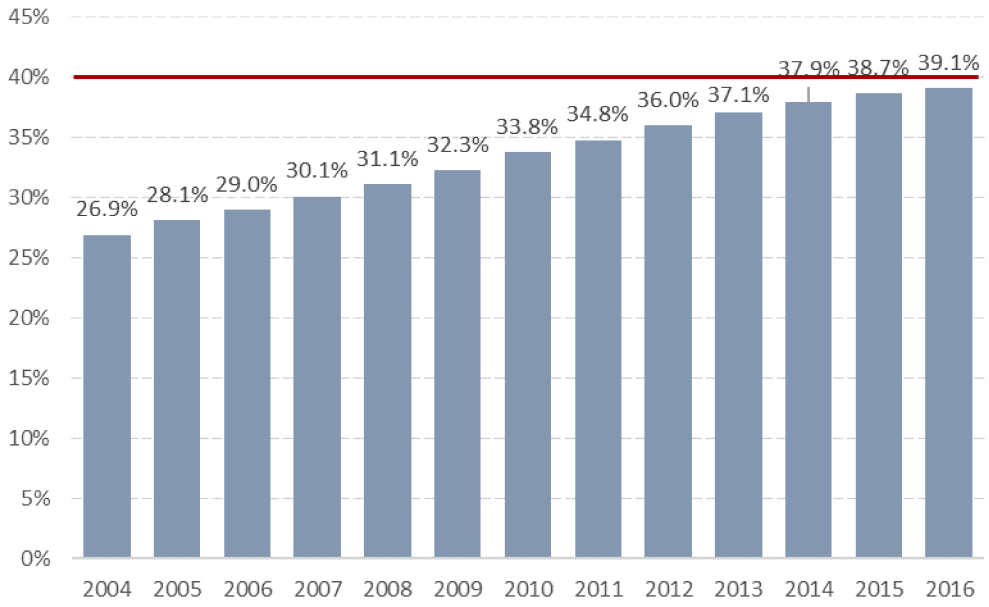
Commission (2016). The current development of the education targets has a positive trend, for both indicators, and in 2016 the total European Union values (for 28 member states) are very close to the targeted value. Number of early school drop-outs decreased from the original level of 13.9% in 2010 to 10.7% in 2016 and the Tertiary education levels increased from 33.8% in 2010 to 39.1% in 2016. The detailed development of both indicators can be find below in Figure 1 (for ELE) and Figure 2 (for TEA).

Figure 1: Early Leavers from Education (EU28 level): Development



Europe 2020 target is 10%, indicated by red line (source: author's elaboration; based on Eurostat)

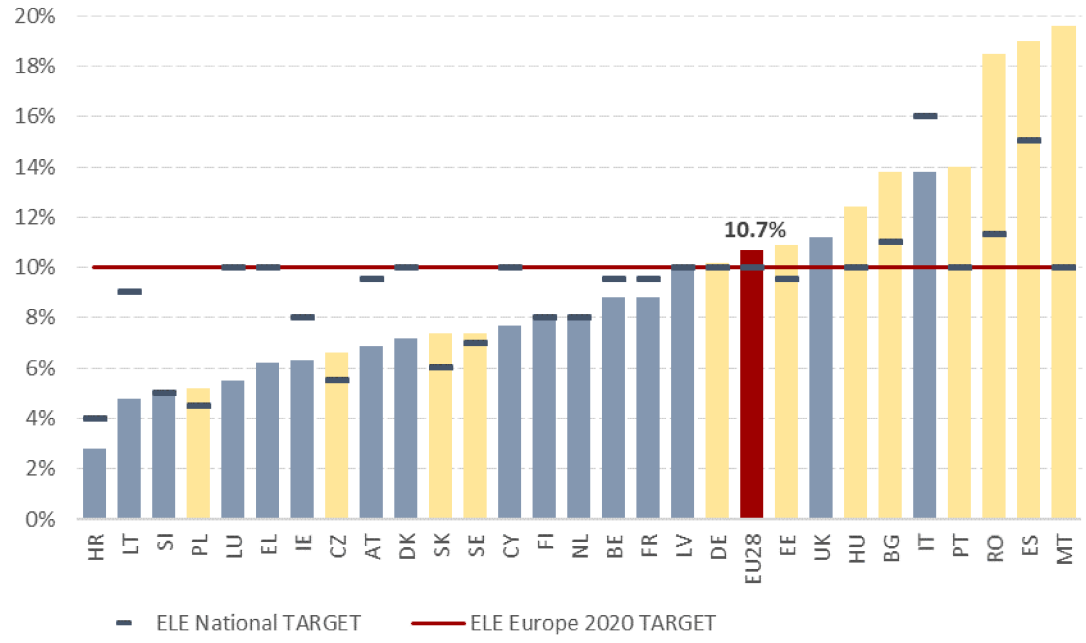
Figure 2: Tertiary Education Attainment (EU28 level): Development



Europe 2020 target is 40%, indicated by red line (source: author's elaboration; based on Eurostat)

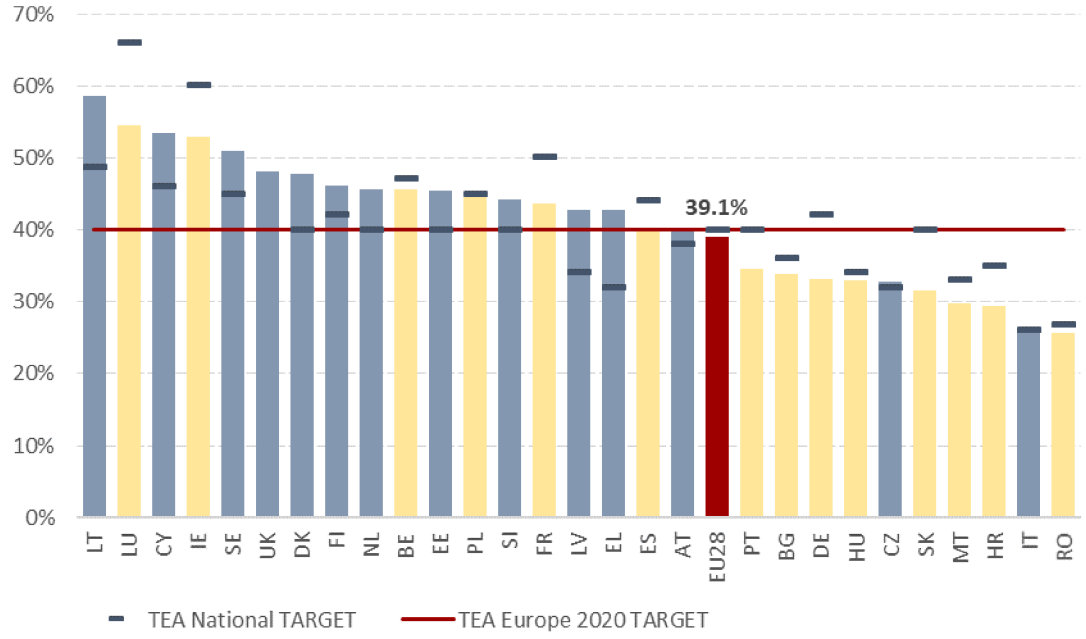
Nevertheless, examining the Member States National targets, the 2016 fulfilment of some Member States is still far from the targeted level and these targets might have been set too ambitiously. Of course, these countries have four more years to grow and reach their targets. The detailed fulfilment for every Member State can be found below in Figure 3 (for ELE) and Figure 4 (for TEA). To conclude, by simply observing the data, it seems likely that both Europe 2020 education targets will be achieved by 2020 but plausibly not all the individual National targets.

Figure 3: Early Leavers from Education in 2016 by Member state



Member States not fulfilling National target are indicated in yellow
(source: author's elaboration; based on Eurostat)

Figure 4: Tertiary Education Attainment in 2016 by Member state



Member States not fulfilling National target are indicated in yellow
(source: author's elaboration; based on Eurostat)

2 Human capital theory

In the theoretical human capital framework, pioneered by Becker (1964, 1993), Ben Porath (1967) and Mincer (1958), the term human capital refers to the stock of knowledge, skills, experience, talent of an individual or in general everything that contributes to his work performance or productivity. In the theory, Becker (1964, 1993) explores the investment in human capital, where the education takes a major part. While taking the investment decision, as for any other type of investment, the individual considers several factors such as rate of return or opportunity cost. Then it is reasonable to assume that while taking the decision about investment in human capital, in education, the individual also considers these factors - in this case the individual values earning prospects against opportunity cost of not entering immediately the labour market. And, following the approach of Dragomirescu-Gaina & Weber (2013) and the later Dragomirescu-Gaina et al. (2015), this was the starting point for explaining the dynamics behind the development of both indicators - Early leavers from education and Tertiary education attainment.

According to the human capital theory, the incentives for an education of a young individual hugely depend on his socioeconomic background. That means mix of several factors where his family plays the most important role. Therefore, the education level of an individual is highly correlated with the educational level of his parents. For instance, the study of Chevalier et al. (2013) presents the relationship between the parental education and income and the schooling of their children, resulting with significant effects.

The next important factor is the incentive towards entering or leaving the educational process itself, the motivation for education. As was already mentioned, there is the trade-off between

the short-term benefit from immediate entrance to employment (probably lower-skilled, not as well payed) and the longer-term perspective of expectation of higher income, where the higher level of education is needed as it is connected to more specialized and expertized job position.

These two perspectives are crucial when observing the two dimensions of education attainment relevant for the Europe 2020 indicators. Following the terminology used by Dragomirescu-Gaina & Weber (2013), it can be defined as an *Income* and *Employment prospect*.

2.1 Early leavers from education

The common characteristic for the group of early school leavers is the lack of skills and therefore poorer employment prospects, or stated differently, lower salary expectations. Due to these conditions, this group is more short-term oriented and its discount rate is higher, therefore these individuals more likely tend to choose to work instead of continuing to study. Especially when chances of employment are higher due to situation on the labour market. On the contrary, according to the study of Petrongolo & Segundo (2002) based on drop-out rates and unemployment in Spain, high levels of youth unemployment can drive the need for education and consequently the stay-on rates among young people can increase. The level of unemployment is a driver in decision-making process about longer engagement in education and training, in Dragomirescu-Gaina & Weber (2013) approach it is named as *employment prospect*.

Following the logic that less educated individuals are not skilled enough for better payed jobs, it is likely that they also create less favourable socioeconomic family background - and plausibly some financial constraints for a young individual for continuing to study

in comparison to their wealthier peers. Therefore, choosing the parents' education as a proxy for family background can cover more factors influencing drop-out rates.

Hence putting these arguments together, the simple model can be specified by following equation where t stands for time and the *parents education* stands as a proxy for socio-economic background of a family. *Employment prospect* is denoted by *t-k* as it is referring to decision made in past:

$$\begin{aligned} &Early\ leavers\ from\ education_t \\ &= \beta \cdot Parent's\ education_t + \gamma \cdot Employment\ prospect_{t-k} \end{aligned}$$

where β and γ are model coefficients and k is the time lag lenght.

2.2 Tertiary education attainment

Compared to Early school leavers, individuals that achieved tertiary education are mostly highly skilled with significant proficiency. They also have higher earning profiles together with higher stability (and probability) of employment as their employment rate is less volatile and depends less on the business cycle than in less educated groups.

Individuals attaining tertiary education have longer time horizons. When deciding about entering tertiary education the important decision factor is the expected wage from the future employment that can be defined as *skill premium*. This is important especially considering the current changes in the labour market - with the technology innovation in a digitalized world where many former tasks performed by men are now robotized, the skilled workers are favoured. In other words, these individuals are not influenced by current employment situation as their situation is more stable than for the low educated workers, therefore the *employment prospect* used for early leavers from education is not

valid for them. On the other hand, the important driver for tertiary education attainment can be defined as *income prospect* as these individuals consider more the added value of education in terms of expected wage earned.

Then, similarly to Early school leavers indicator, also Tertiary education attainment is hugely connected to family background. Following the assumption that tertiary education guarantees better wage (and therefore better economic conditions for a family), it is likely that with parents achieving tertiary education level, their children would achieve it also.

Putting these arguments together, the simple model can be specified by following equation where t stands for time and the *parents education* stands as a proxy for socio-economic background of a family. *Income prospect* is denoted by *t-k* as it is referring to decision made in past:

$$\begin{aligned} &Tertiary\ education\, attainment_t \\ &= \beta \cdot Parent's\, education_t + \gamma \cdot Income\, prospect_{t-k} \end{aligned}$$

where β and γ are model coefficients and k is the time lag lenght.

3 Panel data approach

Panel or longitudinal data combine both cross-sectional and time series dimension, the two primary types of data used in econometrics. Panel data allow to monitor the same cross-sectional units over certain period of time even though the total number of observation is not large. The main advantage of this method is the flexibility for modelling as it is possible to control the unobserved time constant characteristics of each cross-sectional unit - the correlation between unobserved heterogeneity and explanatory variables is allowed. Consequently, the omitted variable problem does

not need to be an issue here which can be very beneficial for simple models.

For explaining the Europe 2020 education indicators it is natural to use panel data method from the characteristics of the data itself as I observe development of member states (= cross-sectional units) over certain period of years. Also, I can use very simple empirical model with only a few explanatory variables as the unobserved heterogeneity is controlled.

In the panel data analysis, we can have either panel that are *balanced* or *unbalanced*. When the panel consists from n set of observations on individual on T time periods, therefore the total number of observations is nT, the panel is called *balanced*. On the other hand, if some of the observations for any individual in any time are missing, i.e. the total number of observations is less than nT, the data set is *unbalanced*. With *unbalanced* panel data, some additional estimation issues and computations are arising.

In this section, I follow the theoretical framework presented by Greene (2012) and Wooldridge (2002). The basic framework for panel data regression can be represented by the following equation:

$$y_{it} = x'_{it}\beta + z'_i\alpha + \varepsilon_{it} = x'_{it}\beta + c_i + \varepsilon_{it}$$

where i denotes the cross-sectional unit and t is a time period. K regressors in x_{it} do not include constant term, $z_i\alpha$ is the heterogeneity (or individual effect) where z_i contains a constant term and observed or unobserved group specific variables that affect the dependent variable, all set to be constant over time. The unobserved time-constant heterogeneity c_i and the idiosyncratic error ε_{it}, that contains time varying factor influencing the dependent variable, are often together referred to as a composite error.

There are several approaches and estimation methods to be

used, depending on the observability (or unobservability) of this set of variables. If z_i is observed for all cross-section units, then the model can be treated by ordinary least square estimation method. When it is not observed, which is in most of the applications, there are other methods to proceed with. The basic panel data models can be defined followingly:

1. **Pooled OLS regression:** If z_i contains only a constant term, the OLS provides consistent and efficient estimates. The model can be represented by the general panel data equation:

$$y_{it} = x'_{it}\beta + z_i + \varepsilon_{it}$$

2. **First differencing estimation** simply transforms the latent heterogeneity out of the model. As a basis, it takes the general panel data equation:

$$y_{it} = x'_{it}\beta + c_i + \varepsilon_{it}$$

and by first differencing, the first differences equation is

$$\Delta y_{it} = y_{it} - y_{it-1} = (\Delta x_{it})'\beta + \Delta c_i + \Delta\varepsilon_{it}$$

The advantage is that is removes the latent heterogeneity, that can be caused by both fixed or random effects, from the model. Nonetheless at the same time, by differencing, it leads to the loss of first observation and more importantly, it removes all the time-invariant variables from the model.

3. **Fixed effects:** If z_i is unobserved, but correlated with x_{it}, the least squares estimator of β is biased and inconsistent which is the consequence of an omitted variable as $E(c_i|X_i) \neq 0$. The model, in this instance, can be rewritten as

$$y_{it} = x'_{it}\beta + c_i + \varepsilon_{it}$$

where $c_i = z_i'\alpha$ contains all the observable effects - group-specific constant term defined as **fixed effects**. The term fixed signifies the *fixed* correlation of c_i and x_{it} , c_i is still stochastic.

4. **Random effects:** When the individual heterogeneity z_i is unobserved, but can be assumed to be uncorrelated with x_{it}, i.e. $E(c_i|X_i) = 0$, the model can be formulated as

$$\begin{aligned} y_{it} &= x_{it}'\beta + E[z_i'\alpha] + z_i'\alpha - E[z_i'\alpha] + \varepsilon_{it} \\ &= x_{it}'\beta + \alpha + u_i + \varepsilon_{it} \end{aligned}$$

where u_i is a group-specific **random** element, identical in each period. This model may by consistently but inefficiently estimated by least squares, therefore other efficient methods are needed.

For the best model spelection, I applied the procedure presented by Park (2011). The general approach in modelling starts with the pooled OLS model and proceeds to more complicated methods. In most of the instances, the pooled OLS are inconsistent and/or inefficient due to the unobserved heterogeneity caused by fixed or random effects. To formally demonstrate the presence of either of them, method of random effects can be tested over pooled OLS by Lagrange multiplier test by Breusch-Pagan (1980) and the fixed effects over pooled OLS can be tested by F-test. To determine the best method between fixed effects and random effects, the Hausman specification test is conducted. For the empirical model, I used the tests specifications as they are described in the package for panel data models in R (*plm* package by Croissant & Millo 2008).

Fixed effects and the method of first differences yields the same results for $T = 2$, for $T > 2$, the fixed effects are usually more efficient, unless there is serial correlation of regression disturbances

present. Compared to fixed effects, the disturbances in the first differences method follow a random walk and therefore the autocorrelation can be lowered by this method. One of the common tests for serial correlation is presented by Wooldridge (2002), that is also part of the *plm* package by Croissant & Millo (2008).

4 Data specification

All data used later in the empirical models have been taken from European Statistical Office (Eurostat)[2]. Data for EUROPE 2020 headline indicators and for the socio-economics proxies are available on country level on yearly basis. Data for education attainment are available among various age groups and genders; and are calculated as annual averages of quarterly EU Labour Force Survey data (EU-LFS).

For the empirical analysis, the data are limited due to missing values and breaks in time series(mostly due to methodological change)[3], which can cause problems in the modelling process. The overview for the main indicators can be found in the table 4.

The missing values mostly occur for more historical data, especially in case of *early school leavers* for many countries in years 2000 and 2001. The big break in the time series is in 2014 for all educational indicators in 2014 where the new ISCED classification started to be valid (for more details about ISCED, see chapter 1.1).

[2]downloaded on 17th April 2017

[3]Only breaks with label "b" were considered, more details can be found in explanatory explanation see Eurostat metadata

Table 4: Overview of missing values and breaks in the time series

headline indicators	period	missing values & breaks in the data
Early school leavers (18-24)	2000 -2016	- for years 2000 and 2001, missing values for 29% and 18% of member states respectively - breaks in the data around 2003 and 2006 - break in 2014 for all member states
Tertiary education attainment (30-34)	2000 -2016	- missing values for years 2000-2003 for 4-7% member states - some minor breaks in time series around 2003 and 2006 - break in 2014 for all member states
supporting indicators	**period**	**missing values & breaks in the data**
Parent's education (low education levels)	2000 -2016	- missing values for years 2000-2003 for 4-7% member states - some minor breaks in time series around 2003 and 2006 - break in 2014 for all member states
Parent's education (tertiary education level)	2000 -2016	- missing values for years 2000-2003 for 4-10% member states - some minor breaks in time series around 2003 and 2006 - break in 2014 for all member states
Unemployment rate	2000 -2016	- missing values for years 2000-2002 for 4-14% member states - some minor breaks in time series in the data set for this period
Real labour productivity	1995 -2016	- missing values for years 1995-1999 for 10-17% member states - very few breaks in time series

(source: author's elaboration; based on Eurostat)

To avoid drawbacks, that the missing values resulting in unbalanced panel data could cause in the empirical models, I respecified my panel set to obtain a *balanced panel*. In general, there are several methods to deal with missing values, the main question to start with is whether the data are missing completely at random or whether there is some pattern (or cross-dependence) of the missingness. In the first case, eliminating these cases does

not bias the inference, in the latter it does. As most of the data in this data set are missing for the more historical years, it is unlikely, that these data are missing completely at random. To avoid complicated data imputation, I sacrificed some years in the beginning of the panel and a few member states, that were lacking the data the most. I imput the remaining missing values using one of the simple approaches presented by Gelman & Hill (2007) - carrying over the last value (or using the consequent one). As I limited the number of missing value cases to a very few instances, I believe it is a sustainable approach.

In the last part of this chapter I will specify the selection of indicators for the *parent's education* level. The Eurostat data are limited by specific age-groups brackets that are available and it is not possible to choose every age-group. While choosing the proper age group, that would match the best the relationship between individual and parents, I used the average age of women at childbirth for the two education attainment levels combined with an age of a mother when first child and consequently I estimated the relevant age group of parents. Several possibilities were used when specifying the empirical model. Tabel 5 represents the result.

Table 5: Education attainment age groups

Europe 2020 indicator	Proxy for Parents' education	age group:
Early leavers from education (age group: 18-24)	Females with at most lower secondary education attainment	35-44
	Males with at most lower secondary education attainment	45-54
Tertiary education attainment (age group: 30-34)	Adults with tertiary education attainment	55-64

(source: author's elaboration)

5 Empirical model & Best model selection

This chapter specifies the empirical side of my models together with the path for attaining the best one. Both models are built on the theoretical framework described in chapter 2 - i.e. the level of *early school leavers* (or *tertiary education attainment*) is explained by the relationships with the level of parental education (which figures as a proxy for a family social-economical background) and by the decision about staying or leaving the education process made in past, motivated by the income or the employment prospect (for *tertiary education attainment* and *early school leavers* respectively). During the modelling process, I considered several specifications.

From the empirical perspective, I followed the panel data approach presented by Dragomirescu-Gaina & Weber (2013), which is the model of first differences, and I used the logarithmic transformation. All specifications were estimated with the R software, using the plm package described by Croissant & Millo (2008).

The argumentation for the use of first differences is that this method (i) transforms the latent heterogeneity out of the model; (ii) reduces the residual autocorrelation and (iii) eliminates the country specific differences.

5.1 Early leavers from education

Using the theoretical framework described in chapter 2 transformed to the logarithmic values, the equation for a regression by first differences can be written as:

$$\begin{aligned}&\Delta log(early\ leavers\ from\ education)_t \\ &= \beta \cdot \Delta log(parent's\ education)_t + \gamma \cdot \Delta log(employment\ prospect)_{t-k} + \varepsilon_t\end{aligned}$$

where β and γ are model coefficients, Δ represents the first differences, k is the time lag lenght and ε is the error term.

As proxies for *parental education* I use the age-groups by gender specified in chapter 4, that are capturing an "average family", i.e. *share of females with low education levels, age 35-44* and *share of males with low education levels, age 45-54*. For the *employment prospect*, the preferred specification is *unemployment rate within the low educated group*[4]. Alternative specifications were used during the modelling process and I will discuss them later in this section.

According to the human capital theory (Becker 1964, 1993), the hypothesis for the model outcomes are the following:

- increase in share of low educated parents generates higher percentage of early school leavers
- higher unemployment rate generates less young people leaving the education process, i.e. decreases the share of early school leavers

I arrived at the specification presented in Table 6. Observing the results, the hypothesis about the characteristics of the explanatory variables were correct, from the results it can be said that the unemployment has greater effect that the parental education. The time lag for the unemployment rate was selected empirically, the lag length might be interpreted as the time when the decision about enrolment in higher education take place.

Table 6: Regression Results - Early leavers from education

Δ log(females with low education levels, age 35-44)	0.258***
	(0.087)
Δ log(males with low education levels, age 45-54)	0.320***
	(0.082)
Δ log(unemployment rate within the low educated group), lag 7	−0.858***
	(0.244)
Observations	182
Time period	2002-2016
Number of countries	26*
Balanced panel	yes
R^2	0.159
Adjusted R^2	0.149
F Statistic	10.832*** (df = 3; 179)

Note: *p<0.1; **p<0.05; ***p<0.01

*Austria and Croatia were excluded from the data set due to missing values for more historical years

[4] *low education level* corresponds to ISCED levels 0-2

The following table presents the overview of alternative specifications for the *early school leavers*, the preferred model is marked as (1).

Model (1) differs from (2) by constant, that turn out not to be significant in the specification. Model (3) uses shorter time lag resulting in insignificant unemployment rate. In model (4) I used the *total unemployment rate* instead of the *unemployment rate within low educated individuals*, as was presented in the original model used by Dragomirescu-Gaina & Weber (2013). In the end, I opted for the model (1) as the *total unemployment rate* was less significant. It can be interpreted in a way that *unemployment rate within low educated individuals* has higher influence during the decision-making process as it more reflects the *employment prospects* within this group which is more affected by the seasonality on labour market (in contrast with the more educated and/or trained groups that are included in the *total unemployment rate*).

Table 7: Alternative specifications - Early leavers from education

	(1)	(2)	(3)	(4)
Δ log(elef)	0.258***	0.238***	0.387***	0.345***
	(0.087)	(0.089)	(0.080)	(0.086)
Δ log(elem)	0.320***	0.297***	0.373***	0.309***
	(0.082)	(0.085)	(0.078)	(0.084)
Δ log(unempl), lag 7	−0.858***	−0.738***		
	(0.244)	(0.270)		
Δ log(unempl), lag 6			−0.307	
			(0.237)	
Δ log(unempl2), lag 7				−0.329**
				(0.159)
Constant		−0.009		
		(0.008)		
Observations	182	182	208	182
R^2	0.159	0.159	0.200	0.138
Adjusted R^2	0.149	0.145	0.192	0.129
F Statistic	10.832***	11.197***	16.240***	7.863***
	(df = 3; 179)	(df = 3; 178)	(df = 3; 205)	(df = 3; 179)

Notes: *p<0.1; **p<0.05; ***p<0.01
elef = females with low education levels, age 35-44; elem = males with low education level, 45-54; unempl = unemployment rate within the low educated group; unempl2 = total unemployment rate

5.2 Tertiary education attainment

Based on the theoretical framework described in chapter 2 and using the logarithmic values, the equation for the first differences regression is:

$$\Delta log(tertiary\ education\ attainment)_t = \beta \cdot \Delta log(parent's\ education)_t + \gamma \cdot \Delta log(income\ prospect)_{t-k} + \varepsilon_t$$

where β and γ are model coefficients, Δ represents the first differences, k is the time lag lenght and ε is the error term.

As proxy for *parental education* I use *share of adults with tertiary education attainment, age 55-64* (as specified in chapter 4) and for the *Income prospec*t I use the *real labour productivity*, as proposed by Dragomirescu-Gaina & Weber (2013).

Following the human capital theory (Becker 1964, 1993), the hypothesis for the model outcomes are:

- increase in share of parents with tertiary education generates higher percentage of tertiary educated individuals in the age group 30-34
- higher real labour productivity increases the incentive for tertiary education attainment and therefore has also a positive effect

I arrived at the specification presented in Table 8. Examining the results, the estimated effects of the explanatory variables are as expected. The estimated *real labour productivity*, representing the *income prospect*, has twice as big effect than the *parental education*. Together with the results from the regression for *early school leavers* it can drive a hypothesis that the labour market conditions have greater effect for education decisions than the level of parental education. The time lag for the *real labour productivity* was selected empirically, the lag length equal to 13 might correspond to the decision about enrolment to university, as Dragomirescu-Gaina & Weber (2013) comments.

Table 8: Regression Results - Tertiary education attainment

Δ log(adults with tertiary education, age 55-64)	0.286***
	(0.061)
Δ log(real labour productivity), lag 13	0.676***
	(0.096)
Observations	184
Time period	1995-2016
Number of countries	23*
Balanced panel	yes
R^2	0.062
Adjusted R^2	0.056
F Statistic	-6.347 (df = 2; 182)

Note: *p<0.1; **p<0.05; ***p<0.01

*Austria, Croatia, Ireland, Malta and Poland were excluded from the data set due to missing values for more historical years

Table 9 presents the overview of alternative specifications for the *tertiary education attainment*, the preferred model is marked as (1). Model (1) differs from (2) by constant, but for the less significant variables and higher F statistic, model (1) was preferred. Model (3) uses detailed specification of the parental education, splitting the *share of adults with tertiary education* into *share of females with tertiary education* and *share of males with tertiary education*, but as the results are not significant for the males but otherwise similar, I opt for model (1) instead. Model (4) uses shorter time lag for *real labour productivity*, resulting in overall less significant explanatory variables and lower R^2.

Table 9: Alternative specifications - Tertiary education attainment

	(1)	(2)	(3)	(4)
Δ log(teat)	0.286***	0.158**		0.108*
	(0.061)	(0.063)		(0.061)
Δ log(teaf)			0.260***	
			(0.060)	
Δ log(team)			0.003	
			(0.067)	
Δ log(rlp), lag 13	0.676***	0.286**	0.632***	
	(0.096)	(0.117)	(0.097)	
Δ log(rlp), lag 12				0.240**
				(0.114)
Constant		0.024**		0.028**
		(0.005)		(0.005)
Observations	184	184	184	207
R^2	0.062	0.063	0.068	0.035
Adjusted R^2	0.056	0.052	0.058	0.030
F Statistic	-6.347	6.046***	-2.552	3.699**
	(df = 2; 182)	(df = 2; 181)	(df = 3; 181)	(df = 2; 204)

Note: *p<0.1; **p<0.05; ***p<0.01
teat = adults with tertiary education, 55-64; teaf = females with tertiary education, 55-64; team = males with tertiary education, 55-64; rpl = real labour productivity

6 Forecast method

As the last step of the modelling procedure, to be finally able to answer the question "How smart will be Europe (in) 2020?", I forecasted the values of the two-fold educational indicators, *Early leavers from education* and *Tertiary education attainment*, up to 2020 based on the econometric models derived in the previous chapter.

There are, indubitably, many approaches how to design a forecast, some of them very complex and some of them very simple. Due to the limitation of this thesis, the complex approach, that would mostly require further modelling, is out of the scope. Therefore, I have opted for the simple method, on the ground of the approach of Dragomirescu-Gaina & Weber (2013). The steps for this procedure are the following:

1. **Construction of country specific projections for all explanatory variables** used in the econometrics models up to 2020. Parental education levels for both indicators are projected from 2016 values up to 2020 following the population cohort method - the details can

be found later in this chapter. The other explanatory variables, unemployment rate in *Early leavers from education* model and labour productivity in the model for *Tertiary education attainment*, are used with a time lags long enough (7 and 13 years respectively) that the data for 2020 forecast are already available and there is no need for any further assumptions.

2. **Final calculation of the country-specific forecast** for the two-fold Europe 2020 education target, including the uncertainty inherent in the forecast expressed by confidence intervals[5].

3. **Calculation of the probability of reaching both EU and national targets** for the two educational indicators - comparison of the forecasted and targeted value.

The assumptions behind this forecast is that there is *no policy change* neither from European union, nor from any of the member states in the following years.

6.1 Projecting the parental education levels

The projections of the parental education levels up to 2020 were forecasted by a simple dynamic equation describing the aging process of a particular age-group with specific education level attainment. The assumptions for this approach are the following: (i) all the decisions about education for these age groups were taken in the past, in other worlds, I presume that the education level attainment for these individuals is not changing [6] as usually the educational process is a matter of a younger age, i.e. no life-long learning or retraining assumed; (ii) it ignores all inward and outward migration effects also as the different mortality risk between groups with different educational attainment level. The projection is illustrated by following equation:

$$Pop(g, t+10) = Pop(g-10, t) + \lambda \cdot gap$$

[5]Confidence intervals were calculated under normality assumptions on 90% probability level.

[6]Of course, I am aware that there are some individuals that are increasing their education level in later age but I assume: (i) it is not a majority of population and (ii) as the education attainment level is only a proxy for a social-economic background of a family in my models (including financial budget), the further education levels achieved would have, nevertheless, only a longer term realization and would not affect considerably the social-economic background of an individual when he was taking his education decision.

where *Pop* is a share of individuals with a given education level at time t in a group with age index g. The age groups g and $g-10$ are referring to two different population cohorts, separated by 10 years. The 10 years difference between cohort groups was chosen due to the available of the data for the education attainment level, I was working with the following age groups: 25-34, 35-44, 45-54 and 55-64.

However, as the projection model is carrying considerable simplifications, the last part of the equation was added to smooth the prediction and to narrow the gap in education levels between two consecutive cohorts, which could be, in some cases, very significant. The *gap* was calculated as a difference in the education attainment levels in a one cohort group over time on a country specific level, to illustrate, the difference between education level attainment of an age group 25-34 in 2006 and an age group 35-45 in 2016 (which is, simply by aging process, considered to be a same group of individuals), λ as a smoothing adjustment coefficient was set at 0.5, following the Dragomirescu-Gaina & Weber (2013) approach. For the instances, where the differences of the two consecutive cohorts were huge, such as for Baltic countries or for Luxemburg, I used slightly higher smoothing adjustments coefficient to offset it. The big differences are mainly caused by the methodological breaks in the data.

The result of the data extrapolation for all the parental education age groups are in the table 10. For a cross check of the reasonability and validity of the extrapolation, detailed development of all three indicators can be found in Appendix A-1, A-2 and A-3 for the projections of *females with low education levels, age 35-44*, *males with low education levels, age 45-54* and *adults with tertiary education, age 55-64* respectively.

Table 10: Projection of the Parental education levels up to 2020

	Low education level				**Tertiary education**	
	Females, age 35-44		**Males, age 45-54**		**Adults, age 55-64**	
	2016	2020^f	2016	2020^f	2016	2020^f
EU 28	17.9	16.6	24.9	23.8	22.3	23.7
AT	14.9	13.0	11.5	10.4	22.5	22.7
BE	17.2	16.7	25.7	22.4	27.0	31.0
BG	15.5	17.5	16.6	16.9	23.7	23.0
CY	14.4	14.6	22.3	17.1	25.9	30.4
CZ	4.3	4.8	4.1	3.8	15.2	17.1
DE	14.6	12.5	10.9	10.4	26.2	27.4
DK	13.1	14.8	22.1	18.6	29.7	29.5
EE	6.3	6.0	5.6	6.0	36.1	39.2
EL	17.7	17.8	32.7	31.9	20.7	22.7
ES	28.7	28.6	46.2	42.7	23.2	26.0
FI	6.3	6.2	14.0	13.3	36.8	40.3
FR	14.7	14.3	22.6	21.2	22.1	23.0
HR	11.5	9.0	16.7	17.6	15.5	15.0
HU	13.5	13.1	15.5	15.0	17.2	18.4
IE	11.0	11.2	28.9	22.6	27.6	31.8
IT	30.4	26.7	47.3	43.4	12.3	12.7
LT	7.8	5.8	3.9	3.4	30.3	33.0
LU	13.8	12.4	25.2	14.0	27.8	34.1
LV	9.0	7.9	7.2	8.4	26.2	27.9
MT	49.9	43.6	63.6	58.9	9.0	11.2
NL	16.9	15.5	23.4	23.7	27.3	28.7
PL	5.7	4.4	8.5	8.2	13.7	14.5
PT	36.8	34.2	66.0	66.6	12.9	12.4
RO	21.4	24.8	15.0	17.6	9.7	10.2
SE	11.2	10.9	14.2	13.3	30.5	30.7
SI	7.7	5.3	13.9	14.1	19.5	20.1
SK	5.7	5.2	6.7	6.0	13.4	14.2
UK	15.7	13.7	23.9	21.2	34.1	35.3

(source: author's elaboration; based on Eurostat and own calculation)

7 Results & Discussion

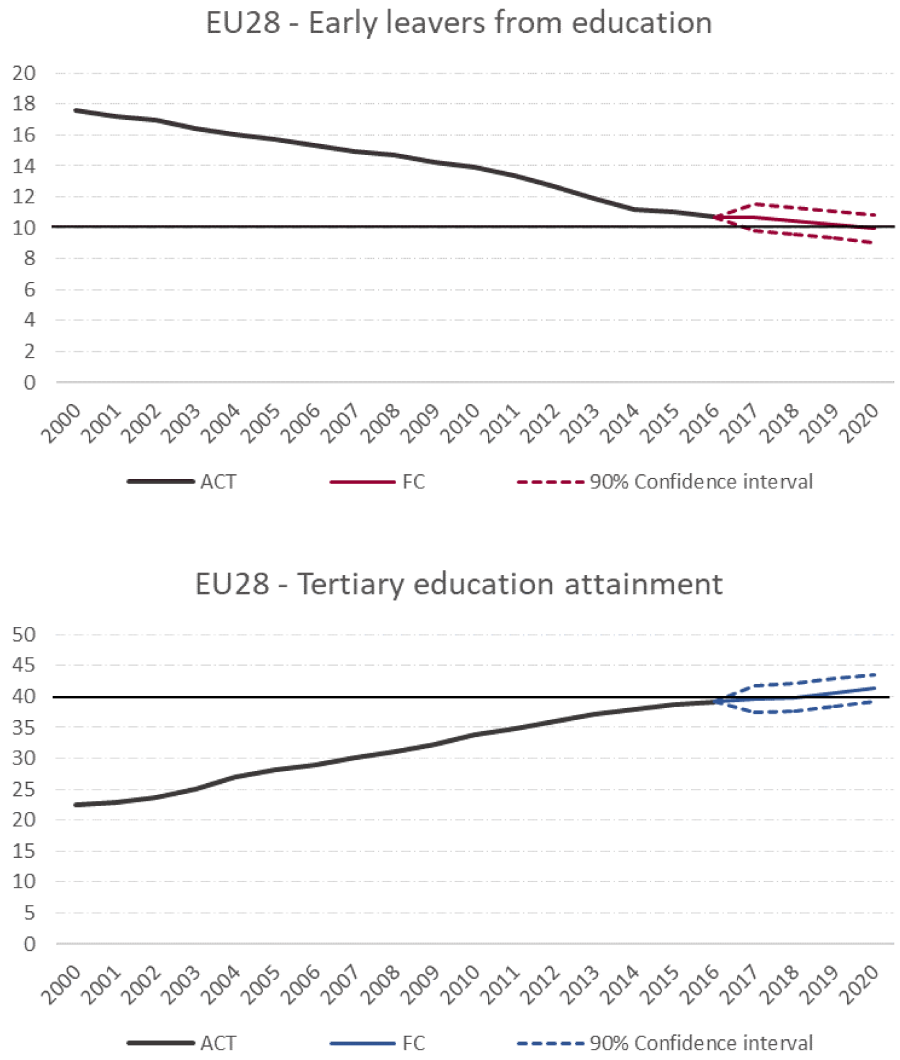
After calculation of the country specific forecast for 2020 and the confidence intervals[7], the probability of reaching both Europe 2020 and individual member state targets was set based on the following criteria:

- *very high* probability: the forecasted value and its confidence interval is above target (for *tertiary education attainment*) or below the targeted value (for *early leavers from education*)
- *high* probability of fulfilment: the forecasted value and its upper bound of the confidence interval is above 2020 target (for *tertiary education attainment*) or the forecasted value and its lower bound of the confidence interval is below the target (for *early leavers from education*)
- *fair* probability corresponds to only the upper bound of the confidence interval above the target, not the forecasted value itself (for *tertiary education attainment*) or only the lower bound of the confidence interval below (for *early leavers from education*)
- *low* probability was assigned when both forecasted value and its confidence interval is below the target (for *tertiary education attainment*) or above the targeted value (for *early leavers from education*)

According to this classification, the results seem very optimistic - the overall probability of reaching the Europe 2020 education targets on the aggregated 28-member states level is assessed to be *high* for both indicators, with the forecasted development path drawn in the charts below.

[7]All the country specific forcasts with the confidence intervals can be found in the Appendix B-1 for *Early leavers from education* and Appendix B-2 for *Tertiary education attainment*

Figure 5: Forecasted path for the two-fold Europe 2020 Education target on the EU28 aggregated level



(source: author's elaboration)
Notes: ACT represents the avalaible data up to 2016; FC represents the forecasted values up to 2020

The probabilities of targets fulfilment on a member state level are listed in tables 11 (for *early school leavers*) and 12 (for *tertiary education attainment*).

In case of *Early leavers from education*, the overall Europe 2020 target 10% is very highly probable to be achieved by 17 member states, and for another 3 it is highly probable to be fulfilled - most of these countries were reaching the target in 2016 already. Nevertheless, there are 6 countries with low probability of reaching the 10% level. These countries are still dealing with quite high share of *early school leavers*, but in majority the percentage of young people out of education system is decreasing, except for the case of Romania, where the share of *early school leavers* is likely to increase during

the following years up to forecasted 20%.

From the national targets fulfilment perspective, 15 countries are very highly or highly probable to meet their national targets whereas 8 member states will most likely not. While evaluating the performance on the national target level, it is important to note that some countries have set less ambitious targets than the Europe 2020 one and will very likely fulfil them (for instance Italy or Spain), but on the other hand, some countries with low percentage of *early school leavers* in the population in general have set their national targets even more challenging that the 10% in Europe 2020 and will most probably fail to fulfil them - case of Czech Republic or Poland. The reasonability of the national targets is further discussed by Břízová (2013).

Table 11: Probability of fulfilling targets for Early school leavers

Early leavers from education

			Probability of fulfiling:					*Probability of fulfiling:*	
	2016 status	**MS target**	**EU2020 target**	**MS target**		**2016 status**	**MS target**	**EU2020 target**	**MS target**
AT	7.0	9.5	*very high*	*very high*	IE	6.5	8.0	*very high*	*very high*
BE	9.2	9.5	*very high*	*very high*	IT	14.1	16.0	*low*	*very high*
BG	13.4	11.0	*low*	*low*	LT	5.0	9.0	*very high*	*very high*
CY	6.7	10.0	*very high*	*very high*	LU	6.5	10.0	*very high*	*very high*
CZ	6.7	5.5	*very high*	*low*	LV	10.7	10.0	*fair*	*fair*
DE	10.3	10.0	*high*	*high*	MT	19.8	10.0	*low*	*low*
DK	7.6	10.0	*very high*	*very high*	NL	8.2	8.0	*very high*	*high*
EE	9.7	9.5	*high*	*fair*	PL	5.2	4.5	*very high*	*fair*
EL	6.5	10.0	*very high*	*very high*	PT	13.6	10.0	*high*	*high*
ES	19.4	15.0	*low*	*low*	RO	18.5	11.3	*low*	*low*
FI	8.7	8.0	*very high*	*low*	SE	7.5	7.0	*very high*	*fair*
FR	8.9	9.5	*very high*	*very high*	SI	4.7	5.0	*very high*	*very high*
HR	2.8	4.0	*very high*	*very high*	SK	6.8	6.0	*very high*	*low*
HU	12.5	10.0	*low*	*low*	UK	11.2	-	*fair*	-

(source: author's elaboration; based on Eurostat and own computation)
Notes: all data are in %. MS target stands for Member state or national target
Global EU2020 target for *Early leavers from education* is maximum 10%;
UK has not set any national target

Evaluating the forecasted values for *tertiary education attainment*, achieving the target 40% is very highly (or highly) probable for 18 countries whereas for 9 member states the probability is very low. Compared to *early school leavers*, this indicator has positive trend for all countries in the European union.

With the national target, the discussion is the same as for *early school leavers* - some countries have set their national target below 40%, i.e. less ambitious, and some countries committed to more challenging goals. From

the 8 countries that are unlikely to reach their national target, some are over-achieving the Europe 2020 one, for instance Luxemburg, Ireland or France.

Table 12: Probability of fulfilling targets for Tertiary education attainment

Tertiary education attainment

			Probability of fulfiling:					*Probability of fulfiling:*	
	2016 status	**MS target**	**EU2020 target**	**MS target**		**2016 status**	**MS target**	**EU2020 target**	**MS target**
AT	39.7	38.0	*high*	*very high*	IE	52.7	60.0	*very high*	*low*
BE	44.6	47.0	*very high*	*very high*	IT	26.3	26.0	*low*	*high*
BG	33.4	36.0	*low*	*high*	LT	58.3	48.7	*very high*	*very high*
CY	53.8	46.0	*very high*	*very high*	LU	54.1	66.0	*very high*	*low*
CZ	31.9	32.0	*fair*	*very high*	LV	41.0	34.0	*very high*	*very high*
DE	33.3	42.0	*low*	*low*	MT	29.8	33.0	*low*	*fair*
DK	47.9	40.0	*very high*	*very high*	NL	45.7	40.0	*very high*	*very high*
EE	46.0	40.0	*very high*	*very high*	PL	44.3	45.0	*very high*	*high*
EL	42.1	32.0	*very high*	*very high*	PT	34.3	40.0	*low*	*low*
ES	40.2	44.0	*high*	*low*	RO	25.5	26.7	*low*	*very high*
FI	45.6	42.0	*very high*	*very high*	SE	50.8	45.0	*very high*	*very high*
FR	44.0	50.0	*very high*	*low*	SI	43.5	40.0	*very high*	*very high*
HR	30.5	35.0	*low*	*low*	SK	30.5	40.0	*low*	*low*
HU	32.8	34.0	*low*	*high*	UK	47.9	-	*very high*	-

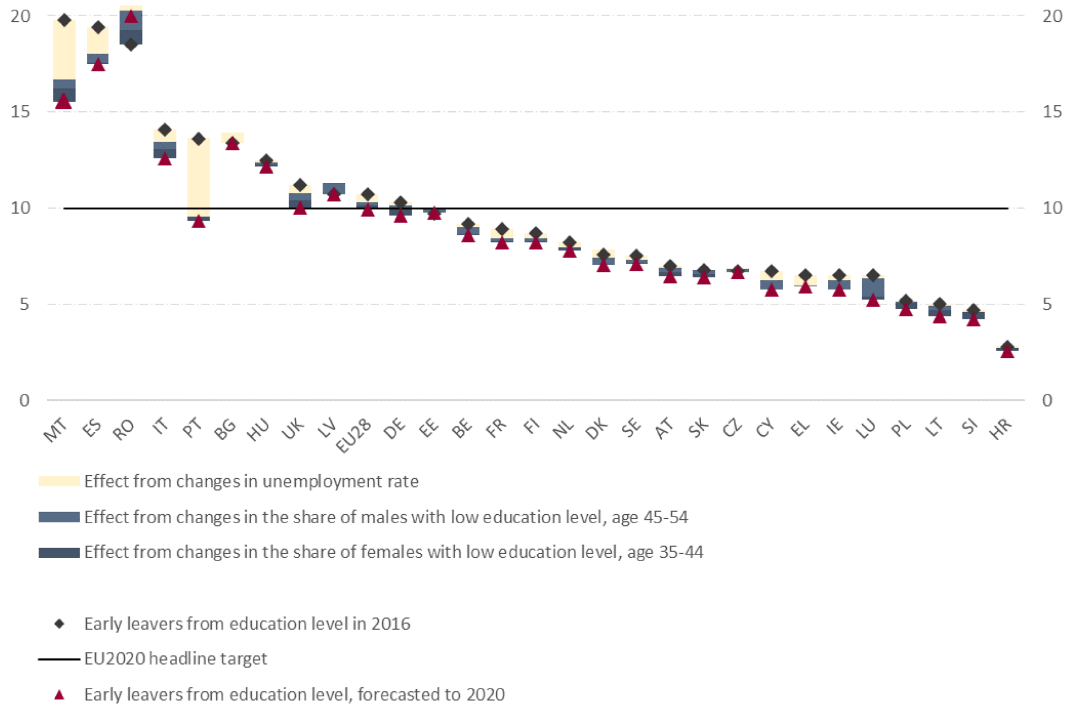
(source: author's elaboration; based on Eurostat and own computation)
Notes: all data are in %. MS target stands for Member state or national target;
Global EU2020 target for *Tertiary education attainment* is at least 40%;
UK has not set any national target

In general my results correspond to the conclusions derived by Dragomirescu-Gaina & Weber (2013). For some countries (for instance Austria or Greece), the forecasted results are even more optimistic as the development in previous years was better than anticipated in year 2012, where the analysis of Dragomirescu-Gaina & Weber (2013) was constructed.

Examining the forecasted results in more detail, the two following charts Figure 6 and 7 present the graphical overview of country specific fulfilment in 2016 and the forecasted levels for 2020 together with the decomposition to the two main drivers - the parental education and the income employment prospect, for *tertiary education attainment* and *early school leavers* respectively.

The member states are ordered based on their initial fulfilment level in 2016. According to the figures, around two-thirds of the countries are reaching the targets, which correspond to the conclusion made in previous paragraphs.

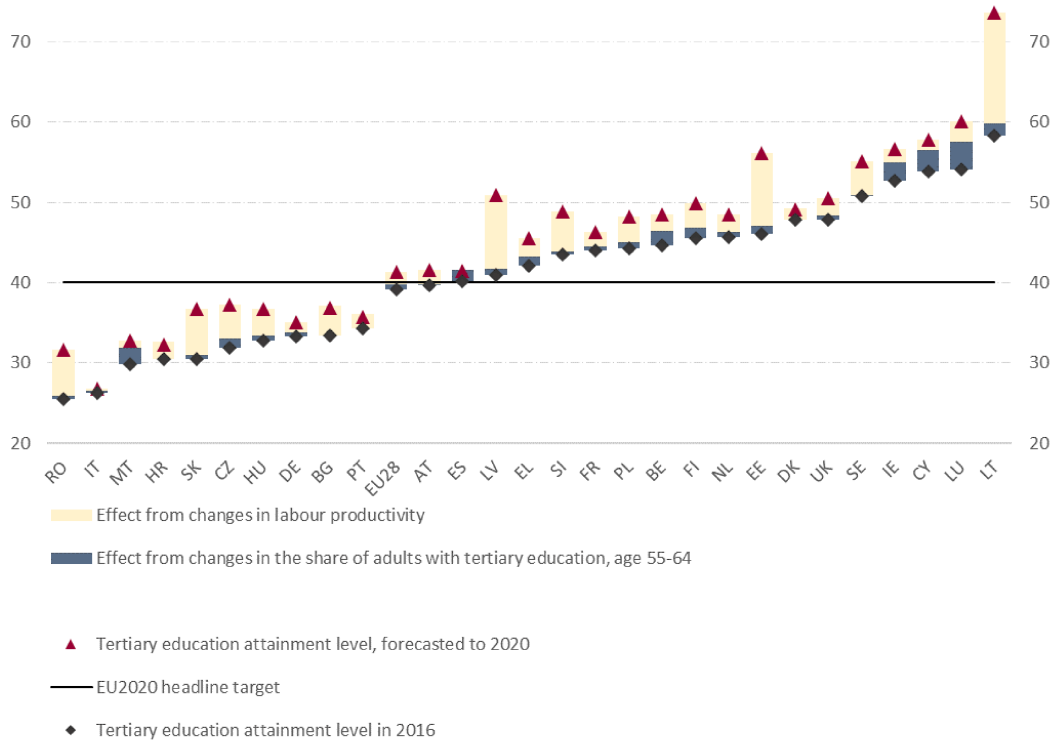
Figure 6: Decomposition of the country specific forecast for Early leavers from education



(source: author's elaboration)

For most of the countries, the changes in the *early leavers from education* indicator are rather minor and mostly are similarly related to changes in unemployment and in the level of parental education. Nonetheless, the effects from changes in unemployment rate of the low educated group of people are very high in Portugal, Malta or Spain. Furthermore, the contribution from changes in the parental level education is high for Luxemburg or for Romania, where the increase in share of low educated individuals in the age-groups for parents causes the increase in the overall forecasted share of *early school leavers*.

Figure 7: Decomposition of the country specific forecast for Tertiary education attainment



(source: author's elaboration)

In general, substantial improvements in *tertiary education attainment* are due to vast changes in the labour productivity, influencing the incentives for tertiary education attainment, rather than due to bigger share of tertiary educated parents with exceptions of Luxemburg, Cyprus or Ireland. The biggest effects from labour market conditions are in Baltic countries, the considerable improvement on the labour market can be a consequence of structural reforms and/or institutional changes implemented during the catching-up phase as Dragomirescu-Gaina et al. (2015) arguments.

Conclusion

My thesis started with two questions, I asked How smart will be Europe 2020? and How smart will be the Europe by year 2020? The main objective of my thesis was to predict the future path for the two educational indicators under the strategy Europe 2020 - *early leavers from education* and *tertiary education attainment* - and to answer the probability, that these targets will be fulfilled by year 2020. The overall educational target under Europe 2020 are at most 10% of *early leavers from education* (aged 18-24) and at least 40% of *tertiary educated* individuals aged 30-34 within the population. As the educational situation is not homogenous across European Union, besides the global European target, the individual national targets were set up as well. Some of these targets are more challenging than the Europe 2020 strategy, whereas others are not.

To be able to answer my research question, I established econometrics models of panel data, using the method of first differences. The theoretical basis for these specifications was the theory of human capital, pioneered by Becker (1964, 1993), describing the incentives of an individual to leave or enter the education process based on his socio-economic background and labour market conditions, defined as *income* and *employment prospect*, for *tertiary education attainment* and *early leavers from education* respectively. Following the approach of Dragomirescu-Gaina & Weber (2013) and Dragomirescu-Gaina et al. (2015), the selected proxies for this theoretical ground were *parental education level* as a proxy for the socio-economic family background, *unemployment rate within the low educated group* as a proxy for *employment prospect* and *real labour productivity* as a proxy for *income prospect*. From the results of the empirical modelling, I established a simple forecast method and I constructed the country specific predictions, together with the uncertainty intervals, for both educational targets. As the last step, I calculated the probability level of reaching the targeted level.

Overall the forecasted results paint rather a favourable future. Looking at the *early leavers from education*, the target is likely to be fulfilled on the aggregated EU28 level, as two thirds of the member states are predicted to be below the 10%. Examining the probability of reaching the national targets, around 50% of the countries are very likely to fulfil their national goals. Nevertheless some countries were plausibly too optimistic with their future development and around 4% of countries will most likely fail in their national targets fulfilment and in Romania, the share of *early leavers from education* is predicted to increase by 2020.

For *tertiary education attainment* the overall aggregated EU28 perspec-

tive is also very positive, again, about two thirds of the countries are expected to have higher share of tertiary educated individuals aged 30-34 than 40%. Nonetheless some countries, such as Romania or Italy are still a lot behind. In general, the majority of the countries is also expected to reach their national targets, with 8 countries that are likely to fail fulfilling them.

While decomposing the contribution to the changes in the educational indicators (*early leavers from education* and *tertiary education attainment*), it is valid for both indicators that higher effects are expected from the changes in the labour market conditions rather than from parental education levels.

To conclude, based on the results, it is very probable that the Europe 2020 will be fulfilled on the aggregated EU28 level and that the future development path for education is positive.

References

[1] BECKER, G. (1993). [1964]. *Human capital: A theoretical and empirical analysis, with special reference to education.* Chicago: The University of Chicago Press. ISBN 0226041204.

[2] BEN PORATH, Y.(1967). *The Production of Human Capital and the Life Cycle of Earnings.*, JP E 75 (1967), pp. 352-365.

[3] BREUSCH, T. S., PAGAN, A. R. (1980). *The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics.* Review of Economic Studies, 47(1):239-253.

[4] BŘÍZOVÁ, P. (2016). [2013]. *Europe 2020 Strategy: Are National Goals Reasonable?* LAP LAMBERT Academic Publishing.

[5] CHEVALIER et al. (2013).*The impact of parental income and education on the schooling of their children.* IZA Journal of Labor Economics 2013, 2:8.

[6] CROISSANT, Y., MILLO, G. (2008). *Panel Data Econometrics in R: The plm Package.* Journal of Statistical Software.

[7] DRAGOMIRESCU-GAINA, C., WEBER, A. (2013). *Forecasting the Europe 2020 headline target on education and training: A panel data approach.*

[8] DRAGOMIRESCU-GAINA, C., ELIA, L. and WEBER, A. (2015). *A fast-forward look at tertiary education attainment in Europe 2020.* Journal of Policy Modeling, Elsevier, vol. 37(5), pages 804-819.

[9] EUROPEAN COMMISSION. (2010). *Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth.* Communication from the Commission: Publications Office. Brussels. Available at: http://ec.europa.eu/research/era/docs/en/ec-understanding-era-13.pdf [05-10-2016]

[10] EUROPEAN COMMISSION. (2016). *Education and training monitor 2016.* Luxembourg: Publications Office of the European Union. Available at: https://ec.europa.eu/education/sites/education/files/ monitor2016_en.pdf [04-06-2017]

[11] GELMAN, A., HILL, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

[12] GREENE, W. H. (2012). *Econometric analysis.* 7th ed. Boston: Prentice Hall. ISBN 978-0131395381.

[13] MINCER, J. (1958) *Investment in Human Capital and Personal Income Distribution.* Journal of Political Economy,University of Chicago Press, vol. 66, No. 4, pp. 281-302.

[14] OECD, European Union, UNESCO Institute for Statistics. (2015). *ISCED 2011 Operational Manual: Guidelines for Classifying National Education Programmes and Related Qualifications.* OECD Publishing. http://dx.doi.org/10.1787/9789264228368-en.

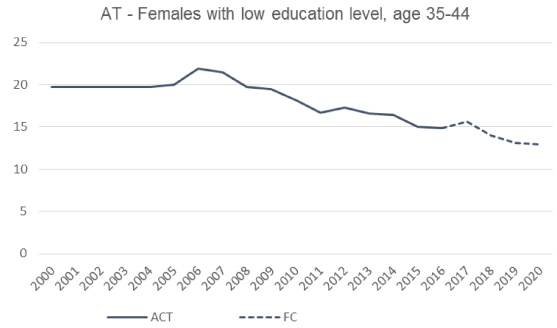
[15] PARK, H. M. (2011). *Practical Guides To Panel Data Modeling: A Step-by-step Analysis Using Stata.* Tutorial Working Paper. Graduate School of International Relations, International University of Japan.

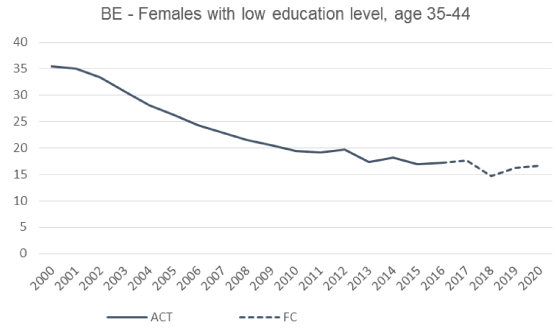
[16] PETRONGOLO, B., SEGUNDO, M. J. S. (2002). *Staying-on at school at 16: the impact of labor market conditions in Spain.* Economics of Education Review, Volume 21, Issue 4, Pages 353-365

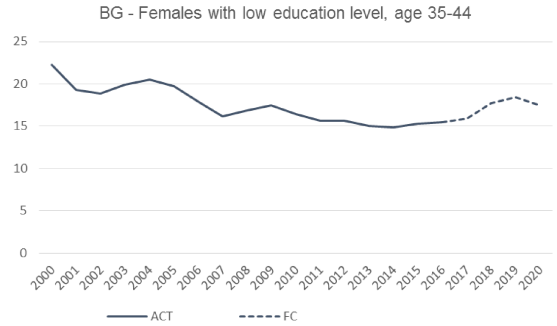
[17] WOOLDRIDGE, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

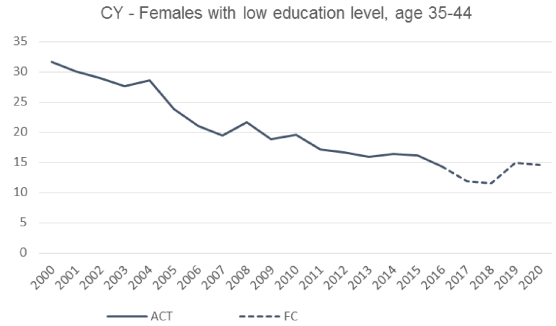
Appendix A-1

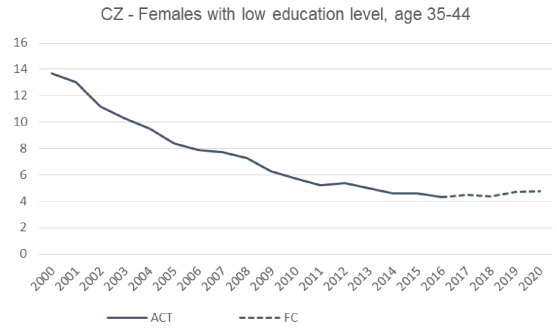
Projection of females with low education levels, age 35-44

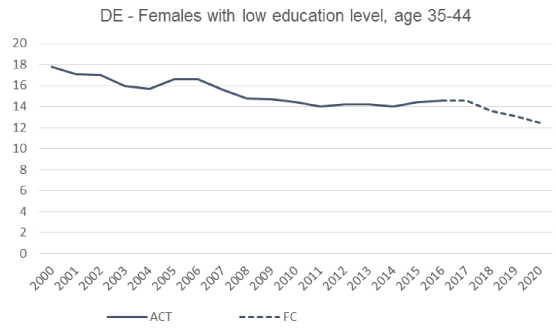


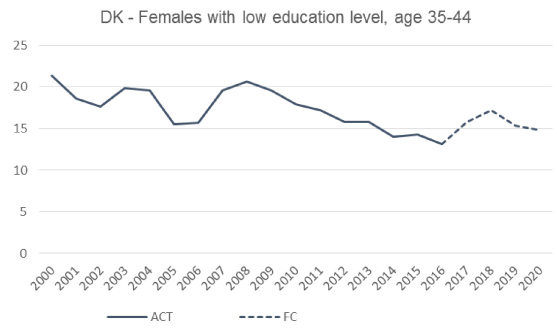


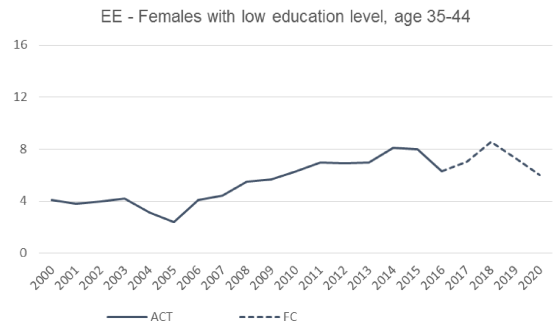






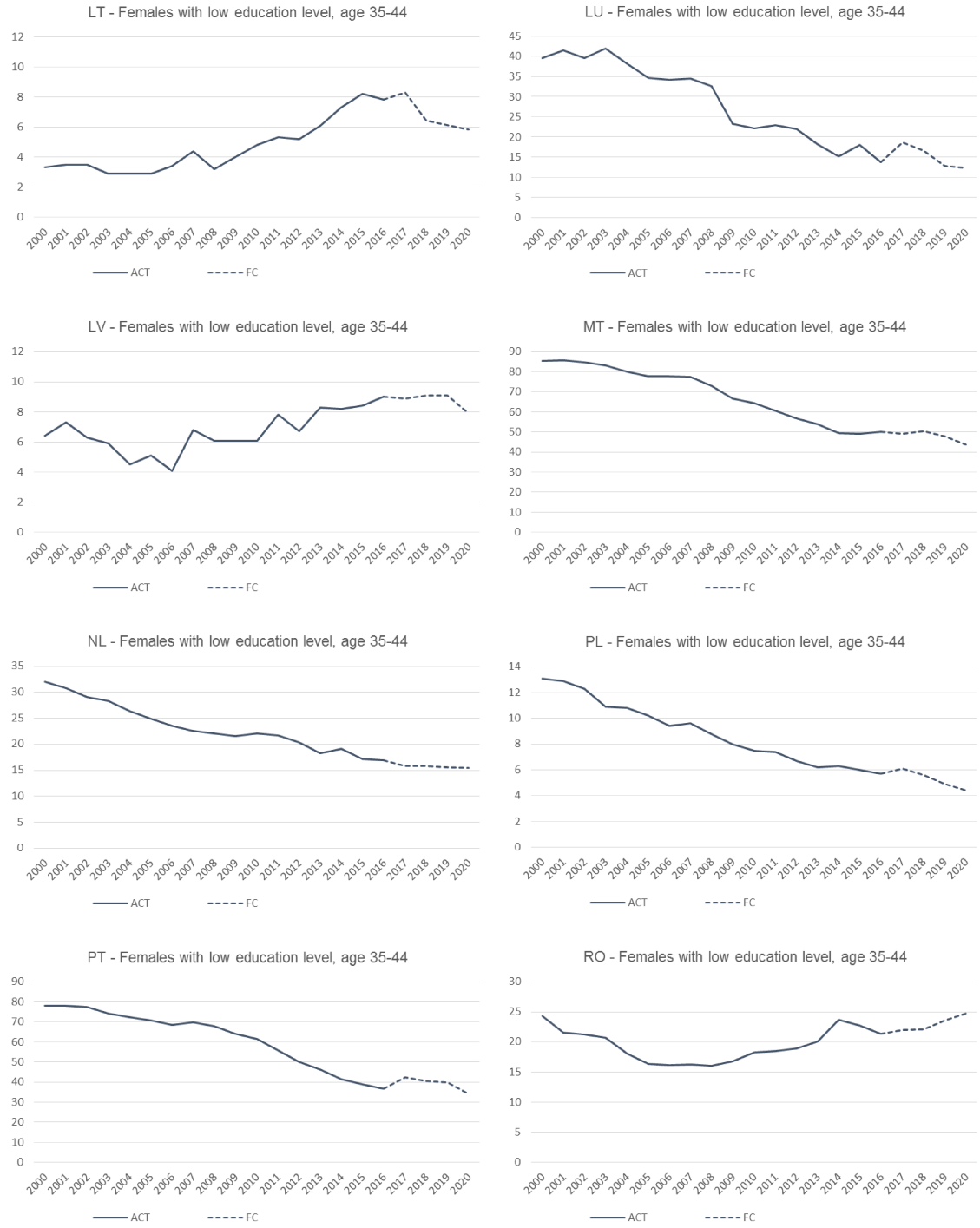




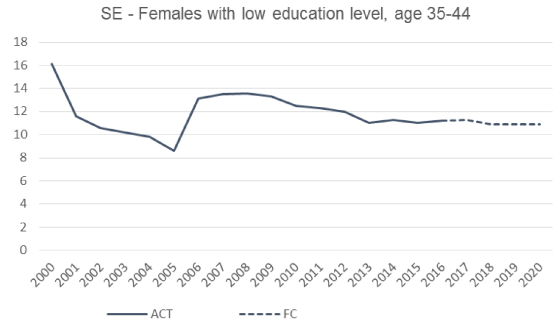


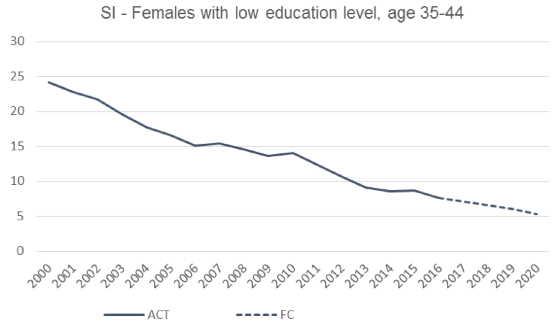
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

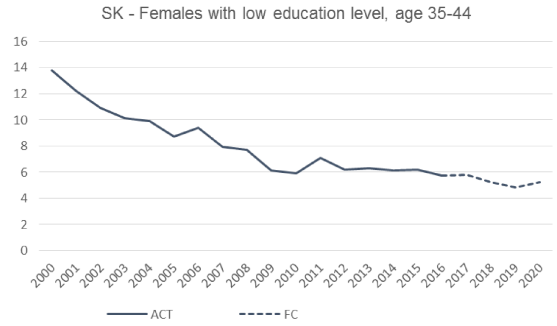
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

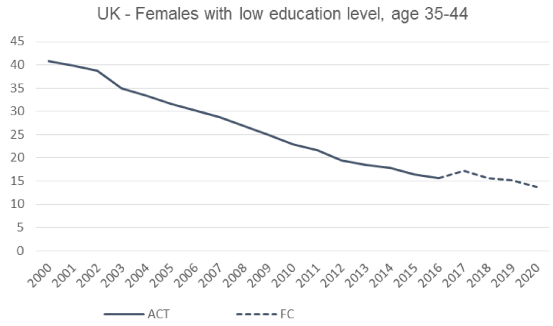


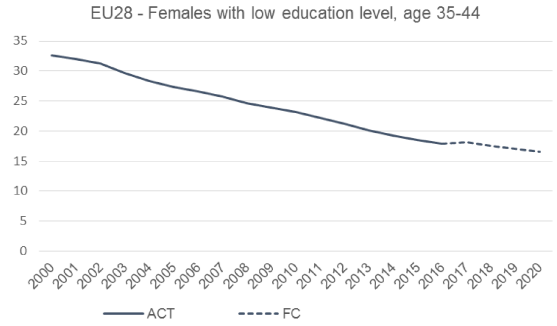
Notes: ACT = development up to 2016; FC = forecasted values up to 2020







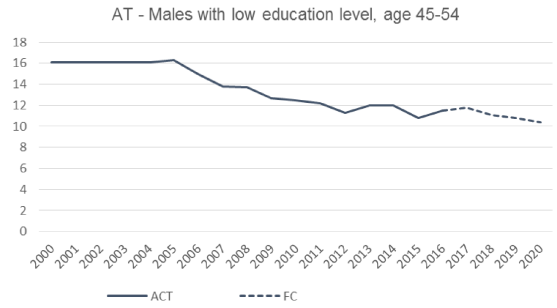


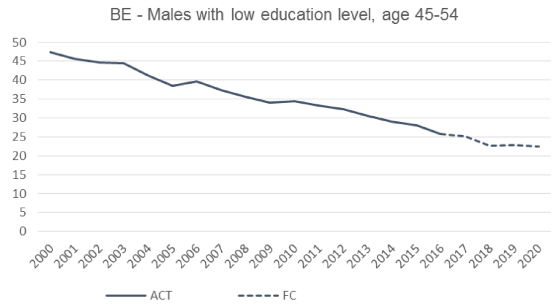


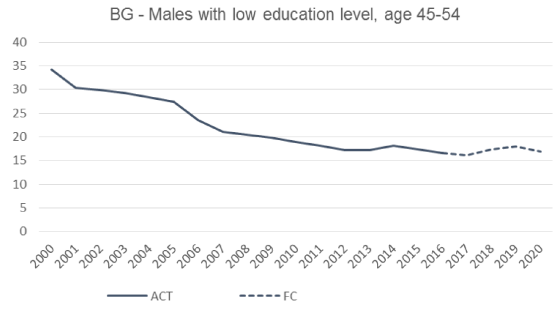
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

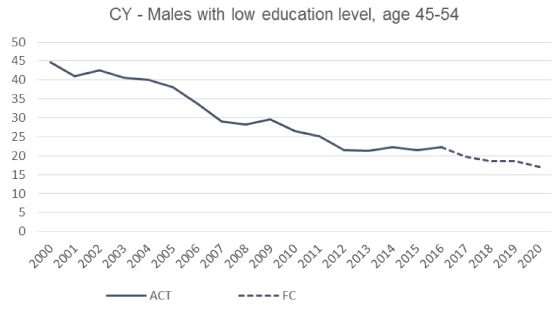
Appendix A-2

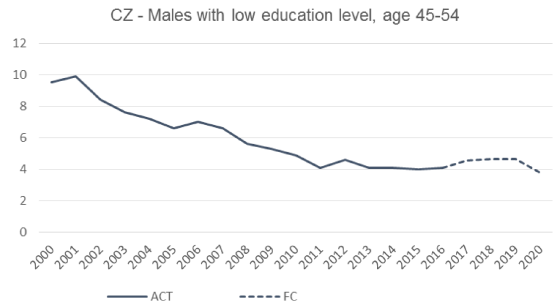
Projection of males with low education levels, age 45-54

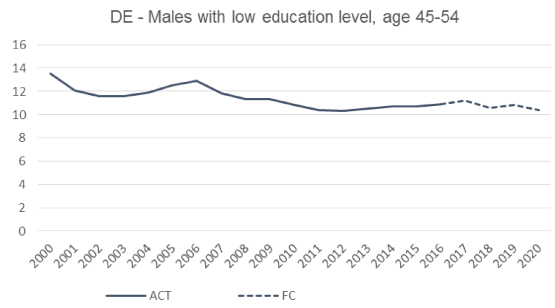


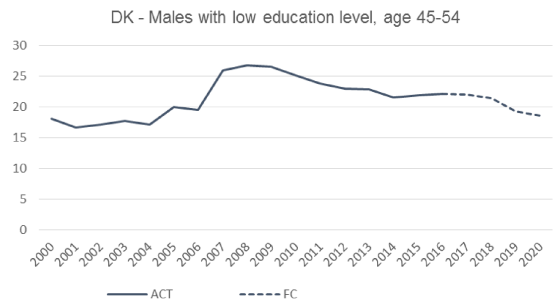


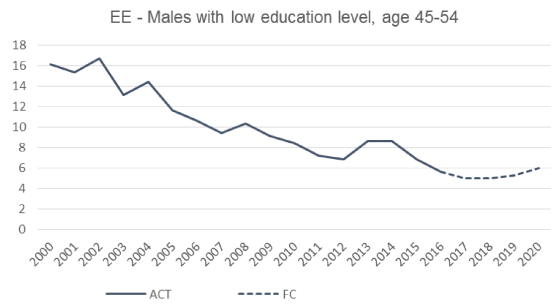








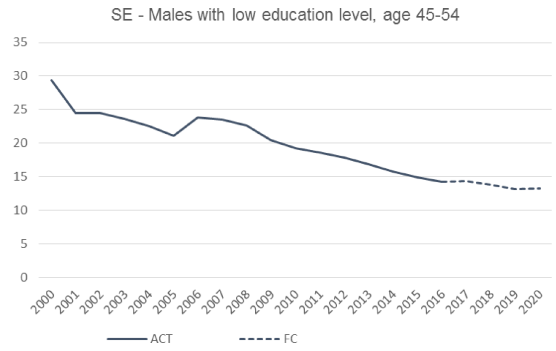


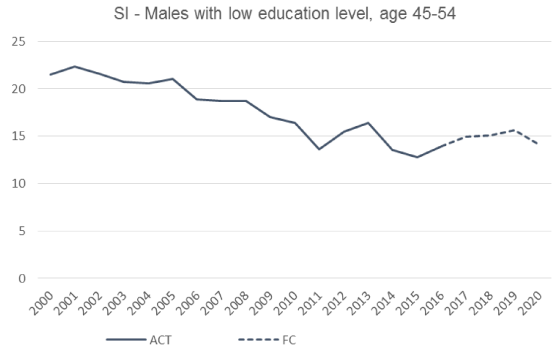


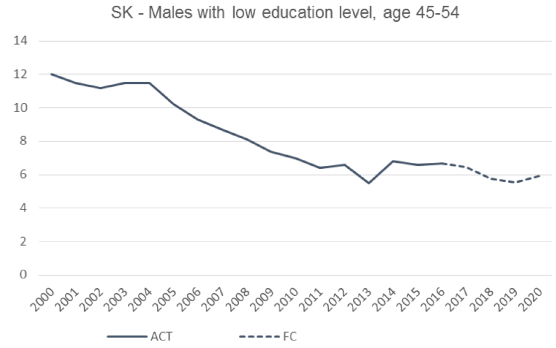
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

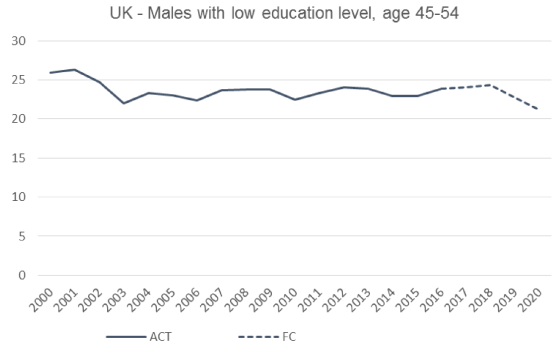
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

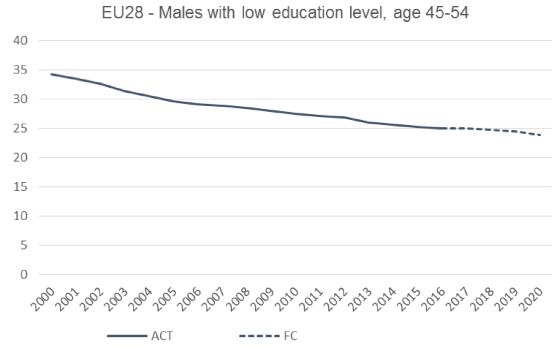
Notes: ACT = development up to 2016; FC = forecasted values up to 2020











Notes: ACT = development up to 2016; FC = forecasted values up to 2020

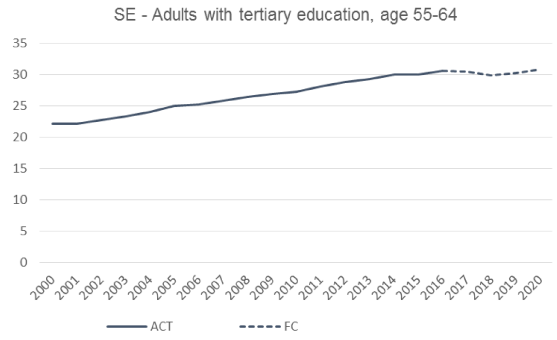
Appendix A-3

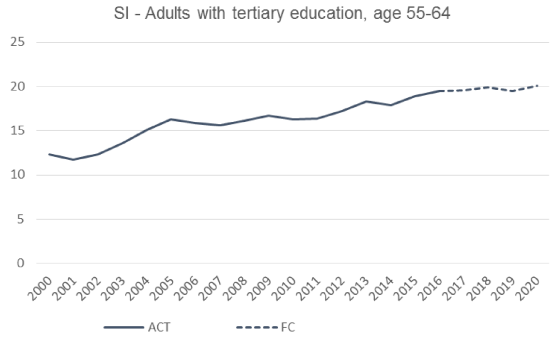
Projection of adults with tertiary education, age 55-64

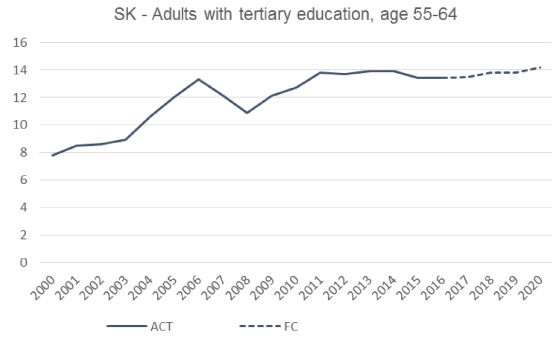
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

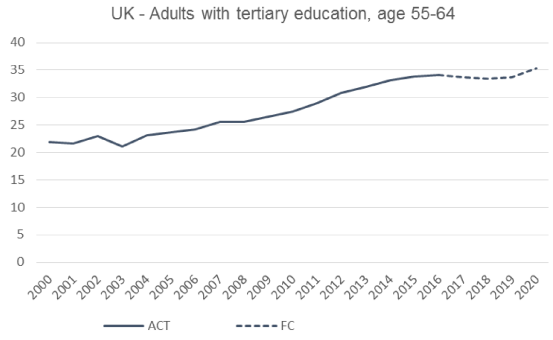
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

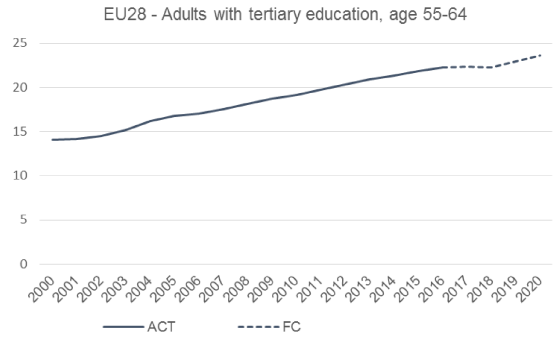
Notes: ACT = development up to 2016; FC = forecasted values up to 2020







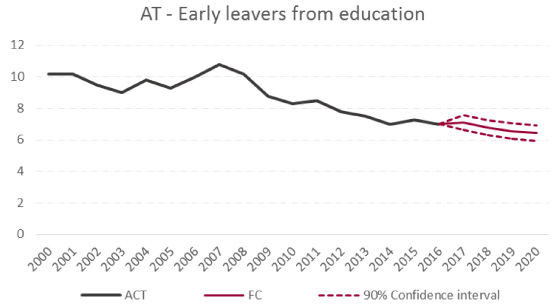


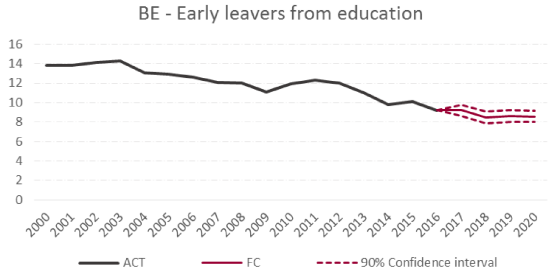


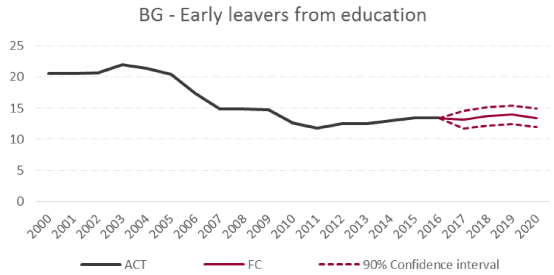
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

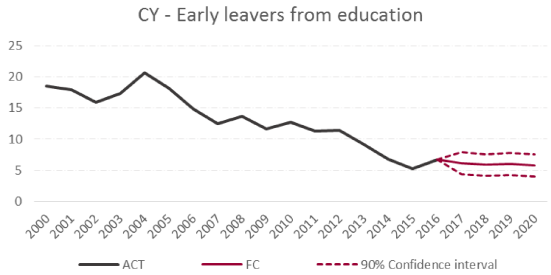
Appendix B-1

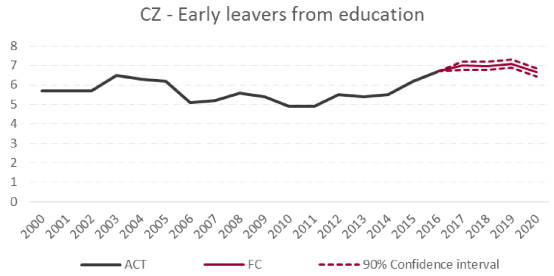
Country specific forecast for Early leavers from education

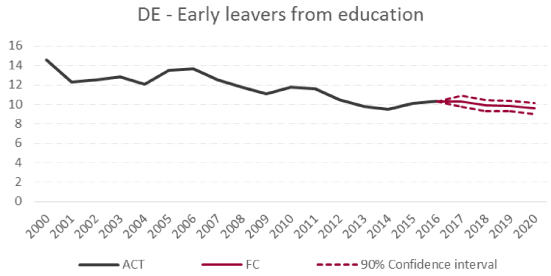


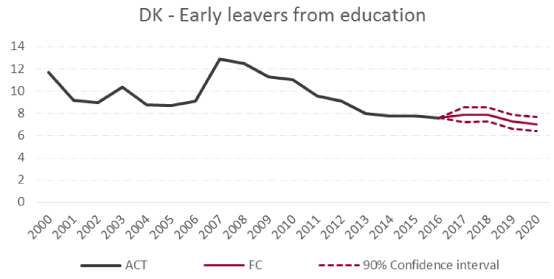


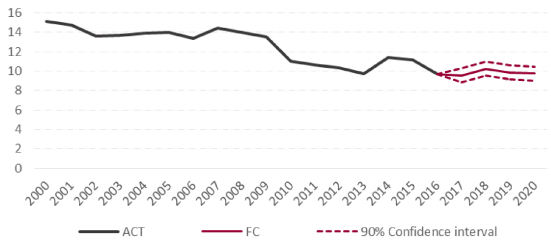




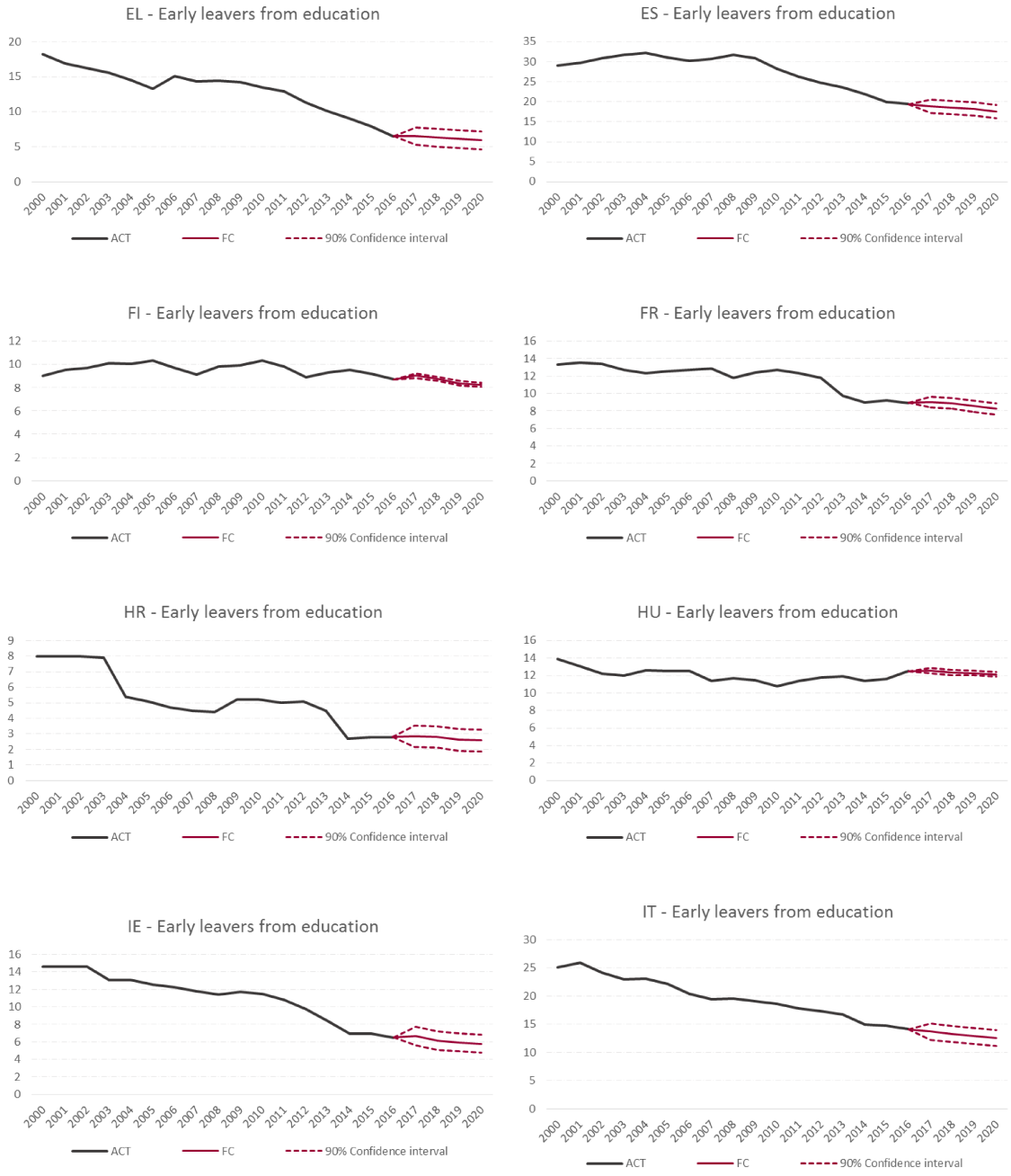




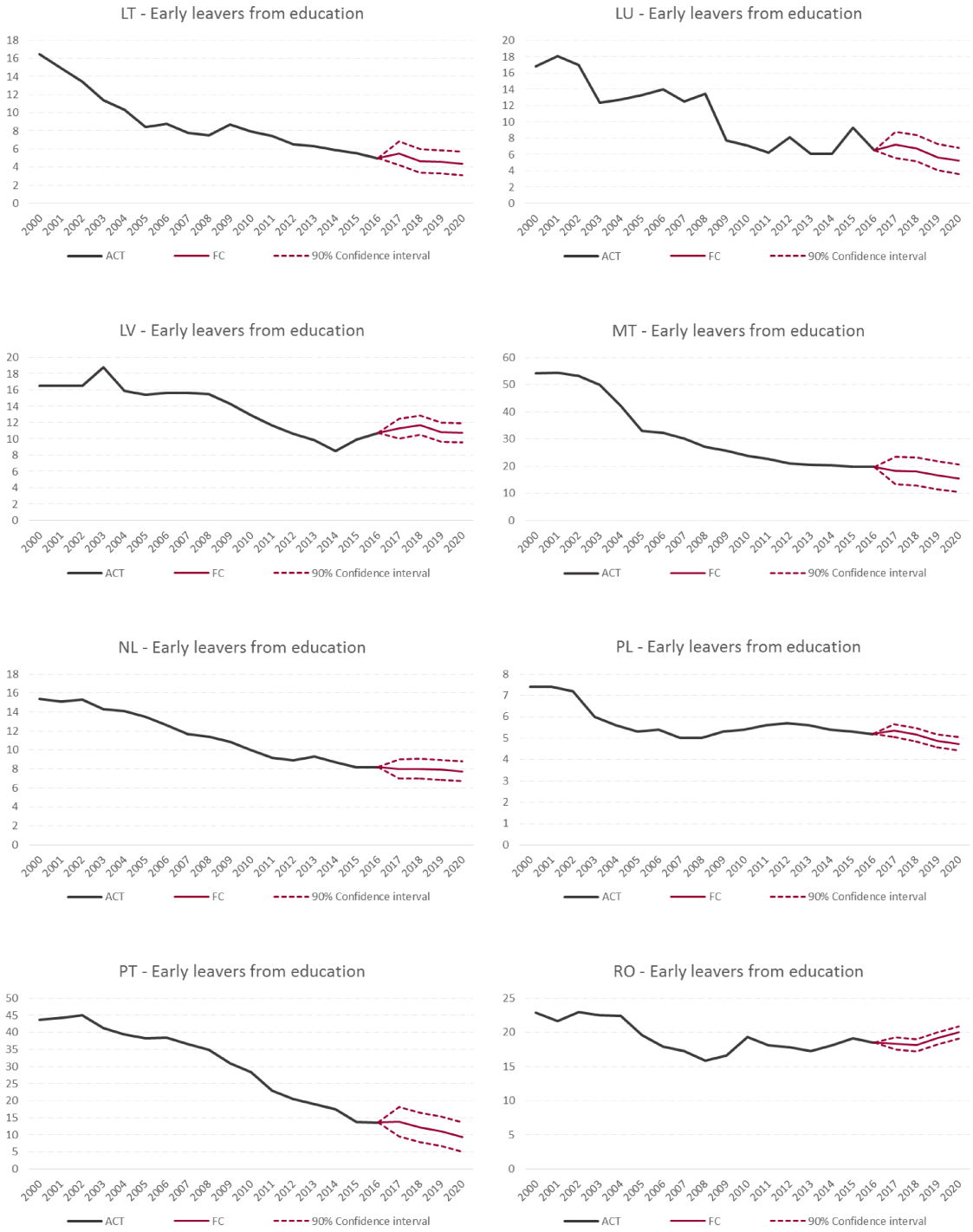




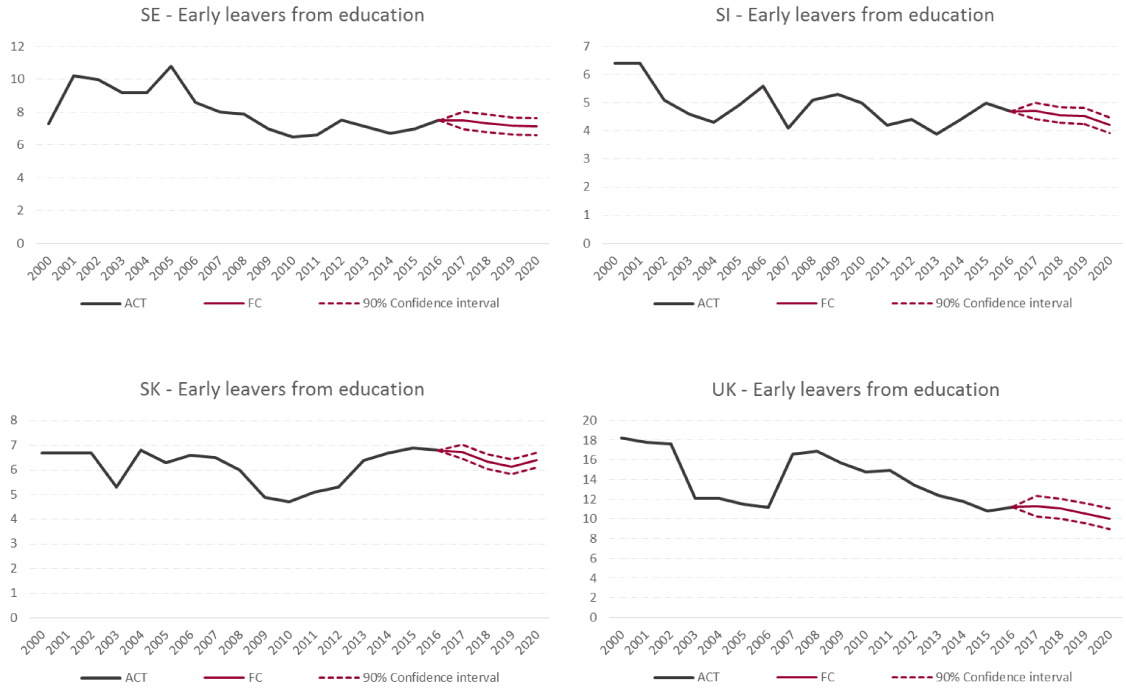
Notes: ACT = development up to 2016; FC = forecasted values up to 2020



Notes: ACT = development up to 2016; FC = forecasted values up to 2020



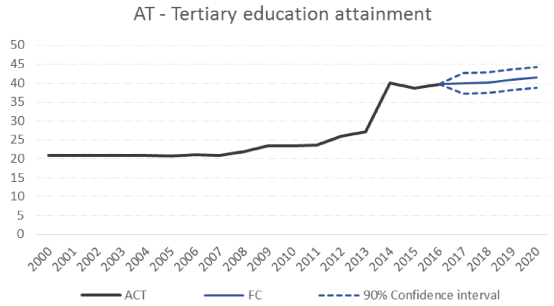
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

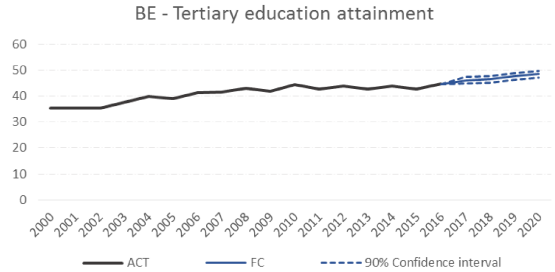


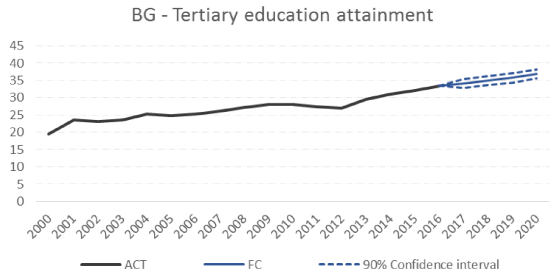
Notes: ACT = development up to 2016; FC = forecasted values up to 2020

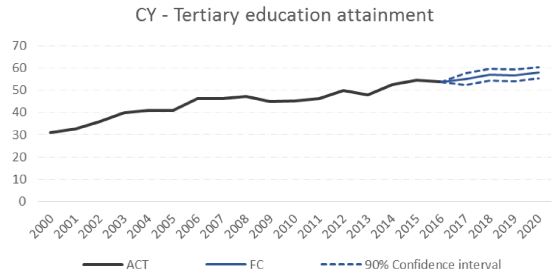
Appendix B-2

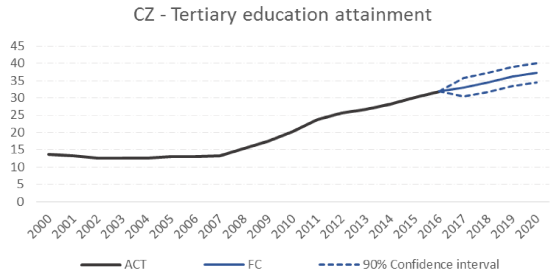
Country specific forecast for Tertiary education attainment

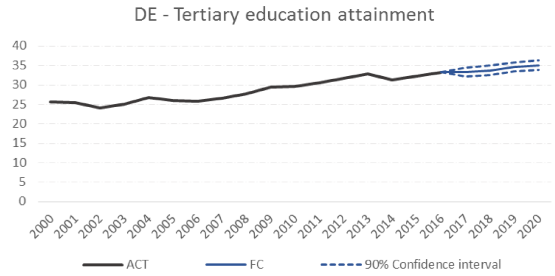


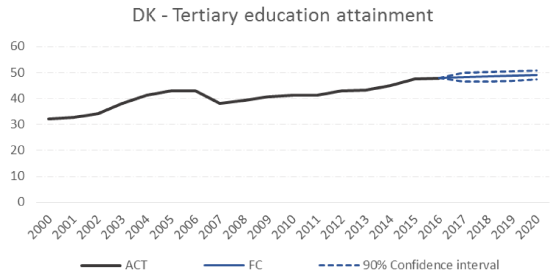


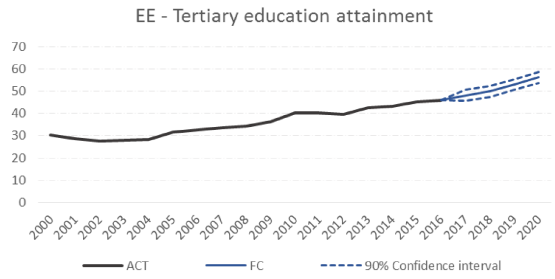




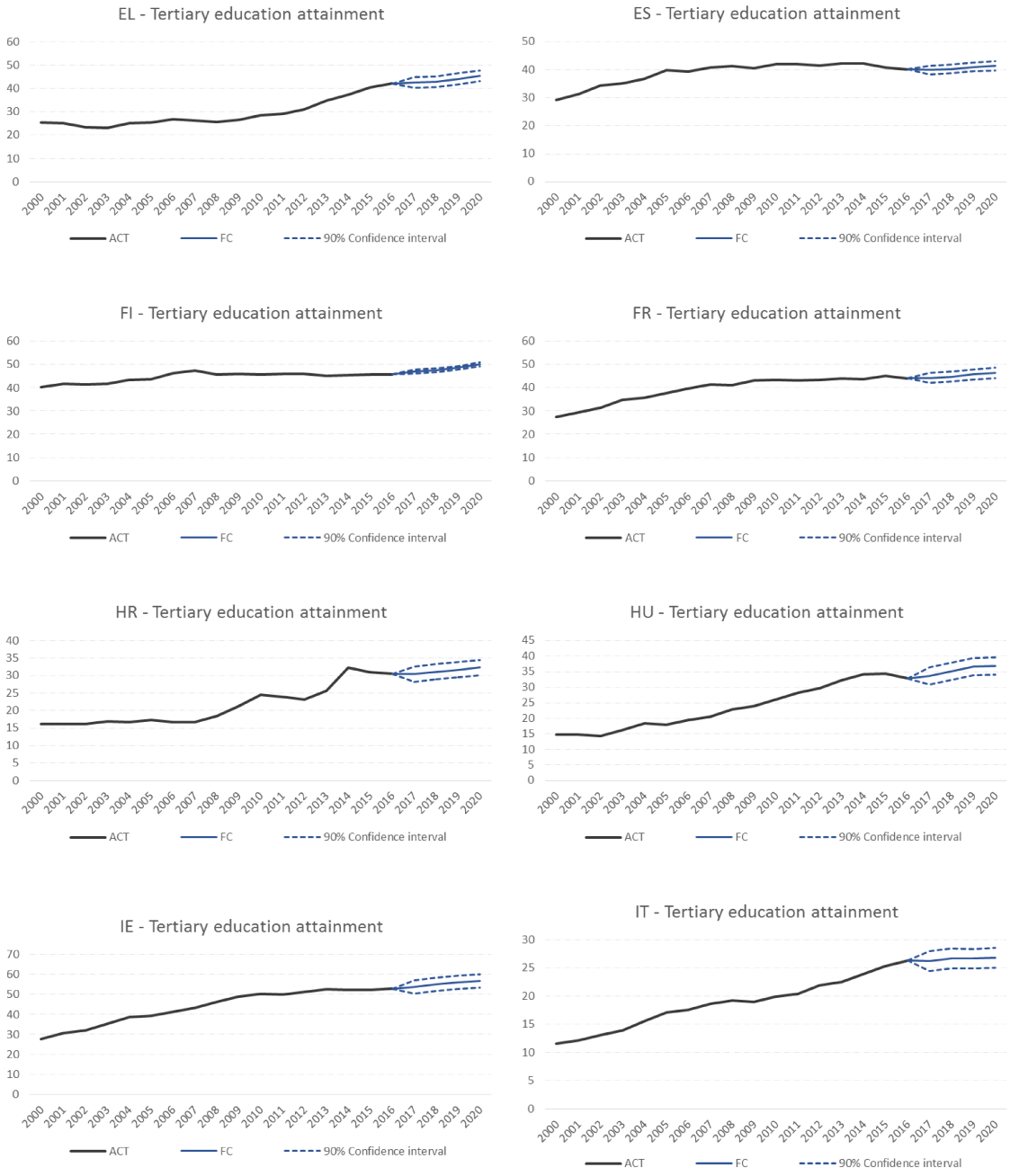




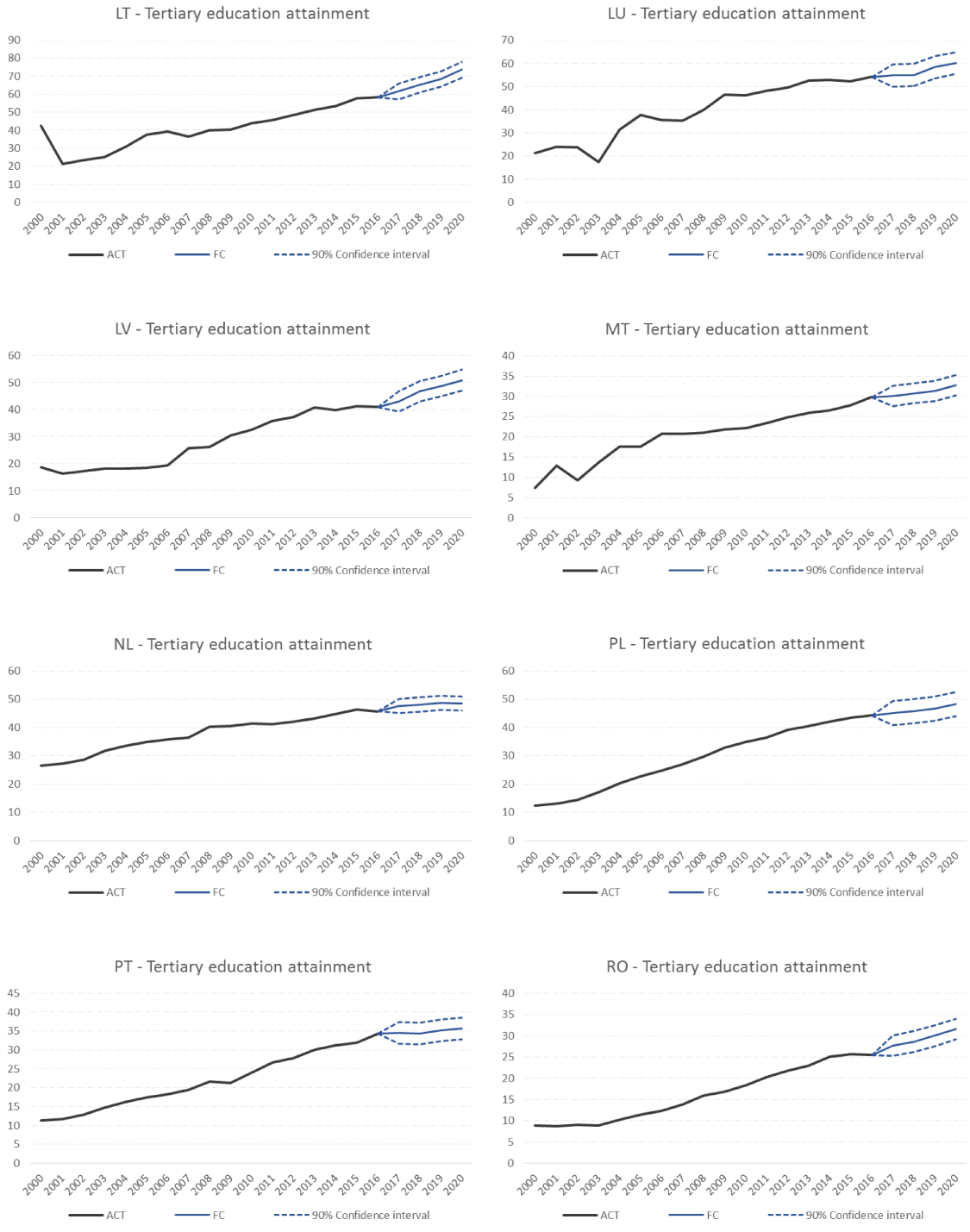




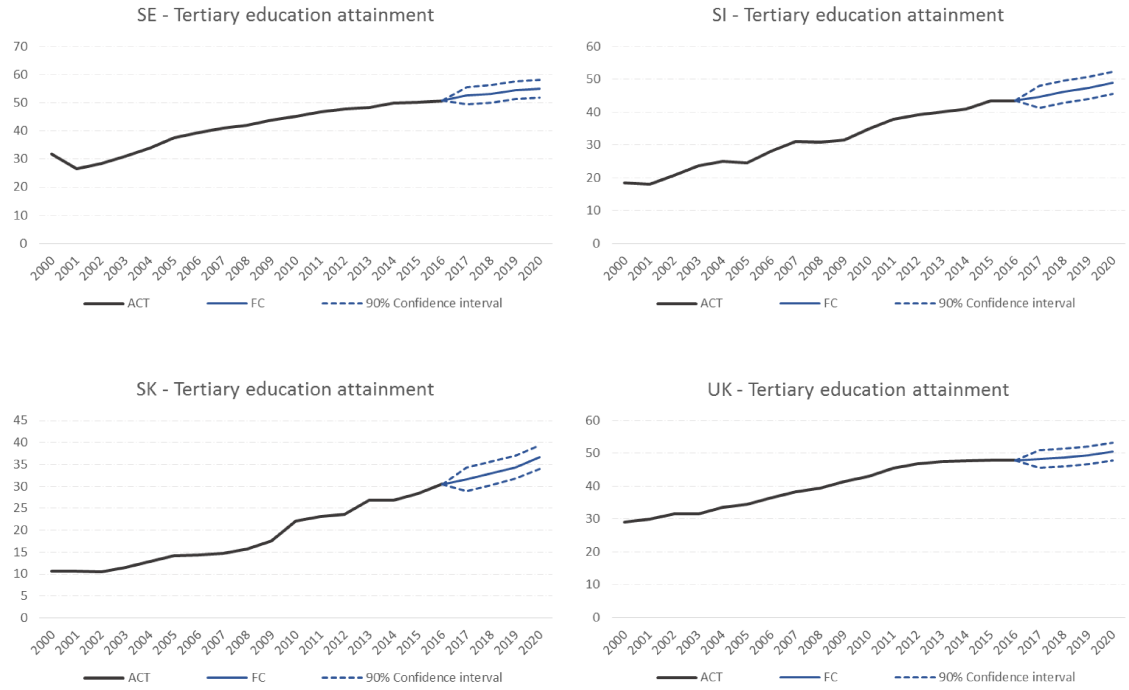
Notes: ACT = development up to 2016; FC = forecasted values up to 2020



Notes: ACT = development up to 2016; FC = forecasted values up to 2020



Notes: ACT = development up to 2016; FC = forecasted values up to 2020



Notes: ACT = development up to 2016; FC = forecasted values up to 2020