A fast-forward look at tertiary education attainment in Europe 2020

Dragomirescu-Gaina, Catalin and Elia, Leandro and Weber, Anke

Econometrics and Applied Statistics, Joint Research Centre - European Commission, Via E. Fermi, 2749, Ispra, Italy, Econometrics and Applied Statistics, Joint Research Centre - European Commission, Via E. Fermi, 2749, Ispra, Italy, Econometrics and Applied Statistics, Joint Research Centre - European Commission, Via E. Fermi, 2749, Ispra, Italy

1 June 2014

A fast-forward look at tertiary education attainment in Europe 2020

Catalin DRAGOMIRESCU-GAINA[†] *

Leandro ELIA[†]

Anke WEBER[†]

Abstract

This paper provides an answer to the question of whether Europe will be able to reach its tertiary education target by 2020. Insights into the dynamics of future education attainment and areas for effective policy interventions in the long-run are highlighted. We model the dynamics behind education decisions as a balance between investment and consumption motivations. We use a panel approach and a wide range of statistical tests to insure that model specifications are stable and robust. We find that while Europe is likely to achieve its target, there is a growing divide between best performing countries and low performers.

Keywords: human capital investment; tertiary education; panel data; forecasting; Europe 2020 strategy

JEL classifications: C23; C52; C53; J24

[†] Econometrics and Applied Statistics, European Commission, Via E. Fermi, 2749, Ispra, Italy
* Corresponding author. Catalin Dragomirescu-Gaina, email: catalin.dragomirescu-gaina@jrc.ec.europa.eu
Leandro Elia, email: leandro.elia@jrc.ec.europa.eu, Anke Weber, email: anke.weber@ec.europa.eu

1 Introduction

In Europe, the share of highly-educated individuals has steadily increased over the past decade. However, most European countries fall short of having figures comparable to the U.S. or other high income countries. This unsatisfactory outcome has led the European Commission (EC), within the strategic framework for European cooperation in education and training, to put forward a political agenda emphasising the need to increase tertiary education attainment. Specifically, EC is committed to increase the proportion of 30-34 year-olds in Europe having completed tertiary education to at least 40% in 2020. This overall goal or benchmark is then translated into national targets by taking into account country-specific situations. In 2012 for example, the share of tertiary graduates remains below the EU target, namely at 35.7%, with some countries such as Italy, Malta and Romania scoring as low as 23%, far below their national targets.

This paper poses the question whether the Europe 2020 target on tertiary education can be achieved. In particular, we estimate a model of tertiary education attainment based on a standard theoretical framework (Becker, 1964; Becker and Tomes, 1986; Keane and Wolpin, 1997; Cameron and Heckman, 1998; Cameron and Taber, 2004; Acemoglu and Pischke, 2001; Behrman and Rosenzweig, 2002; Todd and Wolpin, 2006), which considers family background characteristics and expected returns to schooling.

To see how tertiary education attainment might look like by 2020, some possible approaches would be to build a utility-maximizing dynamic model (see amongst others, Keane and Wolpin, 1997; Todd and Wolpin, 2006) or to estimate a statistical model of schooling attainment using microdata (see amongst others, Cameron and Heckman, 1998; Kaufmann, 2010). This paper adopts a different approach. It presents estimates of schooling attainment derived using a time-series/cross-section data for 27 European countries.

The estimated model is then used to construct long-term forecasts for Europe as well as for individual countries. The major advantage of this panel data model versus other approaches based on model simulations or scenarios analysis, is that we use a minimum number of realistic assumptions that make all our empirical model determinants exogenous in the process. This forecasting exercise shows that Europe as a whole is likely to reach its target of 40% on tertiary education attainment by 2020, but the pace of improvement will be slower than in the past. In addition, our analysis shows that there exists a diverging pattern among members, with some countries showing faster progress than others. The different trajectories of best and worst performers point to important policy challenges which need to be addressed sooner rather than later.

This paper contributes to the relevant empirical literature in at least three aspect. First, using a panel regression analysis, we investigate the dynamics of tertiary education attainment over time, disregarding any constant country-specific factors such as those related to traditions, cultural identity or institutional arrangements. Instead, we focus on the *within* changes in tertiary education attainment, which might be triggered by the changing attitude towards education of young cohorts or the introduction of new learning methods. Second, under theoretically grounded assumptions, we offer an econometric model capable of delivering long-term forecasts on tertiary education attainment in the presence of limited data availability. We show that the empirical model is in line with existing theoretical and empirical evidence, while having a parsimonious specification and stable coefficients over time. Third, our analysis offers suggestions for effective policy interventions to increase tertiary education attainment. In particular, we decompose the expected increase in tertiary education until 2020 into its main determinants and point to areas where policy actions might have longer-term consequences on education attainment.

The analysis ramps up the debate on tertiary education attainment in Europe by offering a baseline on which policy interventions might be advanced. We discuss two main areas of policy intervention. The first area addresses family background and intergenerational mobility, with possible actions directed towards broadening access to education, reducing financial constraints etc. The second one refers to labour market functioning, with a focus on the expectation formation process and the consequences of technological progress on wage distribution and skill premium. With regard to policy effectiveness, 2020 might be too tight of a deadline to achieve a meaningful impact on the benchmark. Still, strong policies might be needed today in order to reduce the widening gap between the best and worst performers in terms of education, by acting along the mechanisms suggested by our model.

The paper is structured as follows. Section 2 provides the theoretical framework for the econometric model. Section 3 discusses the specification of the econometric model and the estimation approach. Section 4 presents the empirical results and various robustness checks. Section 5 presents forecast results and section 6 concludes.

2 Theoretical Framework

2.1 Literature Review

There is a rich theoretical literature addressing the economics of human capital investment, education and schooling decisions. A first major strand of research embarks on the earlier perspective provided by Ben-Porath (1967), where education is a pure *investment* good with foregone earnings being the only cost. The focus is on the individual, who chooses education to maximize his expected discounted life-time utility (see Becker, 1964; Mincer, 1974; Heckman, 1976; Keane and Wolpin, 1997; Card, 2001).

In deciding on the amount of education, individuals formulate expectations of future earnings based on information available at the moment when the decisions are made, and choose the option with the highest expected return. The dynamics of the process driving these expectations is still subject of ongoing research. After the seminal papers by Freeman (1971, 1975), only very recently the literature has shown a growing interest in investigating the effect of subjective expectations on education decisions. Dominitz and Mansky (1996) showed that students are capable of making realistic estimates of future incomes and these are consistent with their performance on the labour market (although their study does not analyze the relationship between earnings expectations and investment into schooling). Oosterbeek and Webbink (1995) study the influence of income expectations on higher education enrolment in the Netherlands. Their results suggest that potential future earnings matter for education decisions. Recently, Buchinsky and Leslie (2010), Kaufmann (2010), Jensen (2010) and Attanasio and Kauffman (2012) provide important evidence on the role of individuals' expectations on labour market outcomes in determining schooling decisions.

A measure of subjective expectations is not always available in cross-section data, but might become less of a problem in time-series analysis. With time passing, the distance between expectations and actual realizations gets blurred. We take advantage of this idea and use long time-lags to reflect labour market dynamics associated with schooling choices as a proxy to income expectations. There is a wide range of available theoretical studies in labour economics discussing the interplay between income and (higher) education. It is commonly assumed that a good approximation of the return to education is the difference in earnings from undertaking and not undertaking education. Acemoglu and Pischke (2001) approximate labour market payoff with a measure of skill wage premium. Bils and Klenow (2000) question the standard explanation for the strong empirical relation between education and economic growth and provide a discussion about the causality running in the opposite direction. Autor et al. (1998) and Acemoglu (2002) amongst others find that technology shifts over recent decades have favoured skilled or highly educated workers. Their evidence suggests the presence of a positive feedback loop between labour incomes and skills on the back of productivity advancements.

A second major theoretical strand dates back to Becker and Tomes (1979, 1986) who consider the role of family socio-economic[1] background in the education investment process. Utility is maximized across generations, with family acting as the central decision maker in this process. Within this framework, both theoretical and empirical literature reveals a positive relationship between family income and children's education attainment (see Kodde 1986; Jacoby 1994; Benabou 1996; Jacoby and Skoufias 1997; Cameron and Heckman 1998; Cameron and Taber 2004; Acemoglu and Pischke 2001; Black and Devereux, 2001). This positive relationship is generally explained by credit constraints, where less wealthy families are not able to borrow to finance the optimal investment in their children´s education. Education in this case can be thought of as a consumption good, denoting that wealthier families can *consume* more of it. Hence, controlling for intergenerational transmission is crucial when explaining education levels. We do so by using parents' education instead of income. Unlike earnings, education has some advantages in terms of estimating intergenerational transmission (see Black and Devereux, 2001). First, measurement issues are less of a worry given that, once completed, education is not subject to transitory shocks or life-cycle movements. Second, parents' education might better reflect family permanent income. Indeed, Cameron and Heckman (1998) find that permanent instead of current family income has the leading role in explaining the impact of financial constraints on children education attainment.

2.2 The Workhorse Model

In this section, we provide a stylized model formulation of investment in education consistent with the above-mentioned literature. In our setup, education attainment, which we denote by H is a function of both family socio-economic background, F and expected labour market payoff, E. All three model variables H, F and E can carry two indexes: an age index g and a time index t. Both indexes are measured in years meaning they grow at the same pace, but are completely independent otherwise. One can fix the age and let the time vary to see, for example, how the education of individuals with a given age evolves over time (i.e. time-series analysis). Likewise, one can fix the time and vary the age to see how the education attainment differs by age for different population cohorts (i.e. cross-section analysis). Given our objective and our benchmark definition, we are clearly in the first situation.

[1] The socio-economic background includes education and income along some other dimensions.

Before writing any equation, we need to introduce some additional notation, to ease the understanding of the model. Let g^0 be the age of the child at the time schooling decision is taken and g the age when his education attainment is measured. Assuming a fix distance between the age of the child and the age of his parents, we can easily characterize family variable using the same age index, g. The observed education attainment is the result of a decision process that occurred in the past, at an age $g^0 < g$. As such, the information set available when formulating expectations for future income should correspond to the decision relevant age, g^0. Therefore, we can write:

$$H(g,t) = f(F(g,t), E(g^0,t)) \tag{1}$$

Equation (1) represents the basis for our empirical model outlined in the next section.

3 Empirical Strategy

3.1 Model Specification

We are interested in modelling the dynamics of tertiary education attainment over time, for the reference population aged 30 to 34. Keeping the age constant and allowing for a time subscript, we can write the empirical counterpart of equation (1) as:

$$H(g)_{c,t} = \alpha_c + \beta_c * t + \gamma * F(g)_{c,t} + \lambda * E(g^0)_{c,t} + \varepsilon_{c,t} \tag{2.1}$$

where $H(g)_{c,t}$ represents the population share of young individuals in the age bracket g having tertiary education, as measured in country c and at time t. The term $F(g)_{c,t}$ represents the proxy for family socio-economic background and $E(g^0)_{c,t}$ is the expected labour market payoff. The coefficient α_c is the country fixed effect meant to capture stable institutional factors affecting education attainment over time while β_c is the country-specific time trend, summarizing consistent and gradual institutional improvements. $\varepsilon_{c,t}$ is a country- and time-specific error term.

The age g in equation (2.1) is observable, but the age g^0 relevant for the decision-making process is unobservable. We are assuming that education choices are based only on the set of information and expectations available at the decision-making age, g^0. Within a time-series approach, the unobservable nature of g^0 can be overcome by using lags that can empirically approximate the time gap between education decision and education measurement. Dropping g from equation (2.1), we can write an empirical equation that suits the benchmark definition:

$$H'_{c,t} = \alpha_c + \beta_c * t + \gamma * F'_{c,t} + \lambda * E'_{c,t-s} + \varepsilon'_{c,t} \tag{2.2}$$

where $H'_{c,t}$ is the tertiary education benchmark, $F'_{c,t}$ is the proxy for family socio-economic background and $E'_{c,t-s}$ is the labour market payoff relevant for individuals aged 30-34. As such, all variables are measured at time t, except $E'_{c,t-s}$ which is the s time-lag of the proxy for expected return. Once the empirical estimate of the s time-lag is available, one can easily infer the unobserved age g^0 using the relation that $g - g^0 = s$.

When trying to estimate (2.2) directly, one needs to address the potential time dependence of the dependent variable $H'_{c,t}$. This is necessary since the tertiary education attainment benchmark spans over 5 consecutive cohorts, covering all individuals aged between 30 and 34. In this context, an alternative would have been to estimate a dynamic panel, i.e. using difference or system GMM as proposed by Arellano-Bond (1991), Blundell and Bond (1998). However, we cannot adopt such estimation strategy given the nature of the benchmark variable and the short time-series available. In particular, given the 5 consecutive cohorts included in the benchmark, one would need to use at least a 5-year lag as instrument to remove correlation between (differenced) error term and the (differenced) lagged dependent variable. This would severely limit sample size and thus weaken the estimation with negative consequences on the forecasting exercise.

In addition, equation (2.2) cannot be rigorously estimated if the data are non-stationary. Unfortunately, univariate and multivariate unit root test cannot be applied in our context because of the short time series, the presence of structural breaks due to methodological changes in collecting the data and the limited number of countries.[2] To tackle the potential non-stationarity in the data, we prefer to specify the model in first-differences:[3]

$$\Delta H'_{c,t} = a + \beta_c + \gamma' * \Delta F'_{c,t} + \lambda' * \Delta E'_{c,t-s} + e_{c,t} \qquad (3)$$

where β_c is now interpreted as the country-specific fixed effect, a is the regression constant and Δ is the first-difference operator.

Equation (3) sets out the dynamics of tertiary education attainment as a combination of consumption and investment motivations. As extensively pointed out by the literature and in line with our discussion in section 2.1, a decision about getting (higher) education is partly explained by better-educated and/or wealthier parents and partly by more favourable earnings expectations. The first determinant encompasses the consumption motive while the second one would reflect the investment motive.

Our dataset lacks specific information on households and individuals, making it impossible for us to identify family bonds. In our setup, individuals belong to a *synthetic* family composed of adults and young individuals matching as closely as possible a typical parental relation.[4] We use the (share of) tertiary educated adults aged 55-64 in lieu of parent's education attainment as a proxy for family background, i.e. the F' variable in equation (3). Obviously, the word *family*[5] would lose its literal meaning in this context and become completely exogenous to the model.

[2] Cross sectional dimension N equals at most 27 and time-dimension T is 13 annual observations. Asymptotic properties of panel unit root tests require N/T→0, which is not met by our dataset. The Im-Pesaran-Shin-type test would be appropriate for fixed N and fixed T but does not allow for serial correlation which is a major concern in this context, as we will see later. Moreover, multivariate unit root tests require the assumption of independence of the units, which cannot be held in our case.

[3] First-differentiating implies that some of the information contained in the original data would be lost if models' variables were cointegrated. However, this assumption is hardly testable in our data. See Asteriou and Agiomirgianakis (2001) for a model using cointegration techniques on education attainment data.

[4] According to 2012 data from Eurostat, for EU27 the mean age of women at childbirth was 30 as of 2011, with a minimum of 27 for Bulgaria and Romania and a maximum of 31 for Ireland and Spain.

[5] Throughout the paper, we use the word *family* to describe both concepts.

For the expected labour payoff, we use an array of proxies in the empirical application. Our choices are based on the discussion provided in section 2.1, but severely limited by data availability. We proxy expected labour payoff using a set of different but related variables: (i) labour productivity, (ii) total factor productivity (TFP), (iii) employment rate and (iv) real compensation per employee.[6] The expected relationship between these indicators and tertiary education attainment is in all cases positive, except for employment rate for which we expect a negative sign.[7]

Heterogeneity issues are more challenging in models estimated from panel data. Here, we consider a more general structure of the model by allowing for unobserved common factors influencing all parameters. In so doing, the estimation approach is based on the Common Correlated Effects (henceforth CCE) method advanced by Pesaran (2006). The method consists of approximating the linear combinations of the unobserved factors by cross section averages of the dependent and explanatory variables, and then running our regressions augmented with these cross section averages. An advantage of this method is that it yields consistent estimates under a variety of situations such as serial correlation in errors and contemporaneous dependence of the observed regressors with the unobserved factors (Pesaran and Tossetti, 2009). We can then estimate a model of the form:

$$\Delta H'_{ct} = a + \beta_c + \gamma' * \Delta F'_{c,t} + \lambda' * \Delta E'_{c,t-s} + \varphi * \overline{\Delta H'}_{ct} + \delta * \overline{\Delta F'}_{c,t} + \theta * \overline{\Delta E'}_{c,t-s} + e'_{ct} \qquad (4)$$

where the last three terms on the right-hand side are the cross section averages of the variables included in equation (3). As pointed out in Chudik and Pesaran (2013) cross correlations of errors could be due to omitted common effects, spatial effects, or could arise as a result of interactions within socioeconomic networks. In our model, cross section dependence may stem from Europe-wide policy recommendations and/or reforms, such as those included in the Bologna process.

3.2 Selecting the best model specification for forecasting purposes

In this section we discuss the various tools employed to identify the best model in terms of statistical and forecasting properties. In particular, we adhere to the recommendations advanced in Zellner (1992), namely avoid over-fitting the model and *keep it sophisticatedly simple*, but *select appropriate empirical proxies* to maintain model consistency. Since we aim at providing *long-term* forecasts for the Europe 2020 benchmark, it is crucial to keep the model simple. In particular, we need to reduce the forecasting uncertainty by relying on clearly defined assumptions for our exogenous variables. The use of adults' education attainment as a regressor is particularly suitable for such a purpose. It allows us to *anchor* the forecast of tertiary education attainment to an exogenous path derived using a formula based on demographic

[6] We also tried using different proxies for labour market payoff (e.g., per capita GDP, overall and skill-specific unemployment rate, skill-specific employment rate) but we found neither significant coefficients nor robust specifications.

[7] The negative relation between employment and wages or productivity derives from a standard labour demand framework. Please note that our model could be interpreted as a model of labour supply, where labour demand has been exogenized.

dynamics (i.e. a replication of an aging process) and thus keep the number of assumptions at minimum.

We follow a general-to-specific approach and consider a series of specification tests when selecting the final model specification. In particular we go through the following steps: first, we choose the appropriate lag structure of the determinants of tertiary education attainment by using Akaike (AIC) and Bayesian (BIC) information criteria; second, we test for lack of residual autocorrelation using a number of tests as proposed by the literature (Arellano-Bond, 1991; Baltagi-Wu, 1999; Wooldridge, 2002 p.282–283); third, we check coefficients stability by varying the estimation time-sample; fourth, we rely on o*ut-of sample* root mean square errors (RMSE) computed over 1 to 4 years ahead horizons to evaluate forecasting accuracy.[8]

Heterogeneity is an important concern for our modelling approach due to the fact that some countries might have the potential of driving the estimated coefficients. We follow Li (1985) and perform a robust regression analysis each time we change the model specification so as to exclude those countries indicated as gross outliers by the procedure. Data availability and our slight preference for working with balanced panels might further reduce the number of countries included in some estimation samples.

Residuals autocorrelation and coefficients stability are of substantial significance for our empirical exercise and we discuss them carefully below. Residuals autocorrelation may originate from the nature of the education attainment variable being measured over consecutive population cohorts. First-differentiating the data mitigates to some extent the presence of lower order autocorrelation, but it might not affect higher order autocorrelation. Assume, for example, no inward/outward migration flows that might alter the composition of the population cohorts except aging. Taking first-difference of tertiary educated individuals aged 30-34 might produce residuals autocorrelated at a higher order. In particular, every 5th lag of $\Delta H'$ would produce almost perfect negative autocorrelation in this setup.[9] Other sources of potential residual autocorrelation might, for example, arise in the presence of siblings. To dismiss all these potential possibilities, we provide a number of autocorrelation tests as follows. We first check for first-order autocorrelation using Baltagi-Wu (1999) and Wooldridge (2002) tests which are shown to be robust even in small samples. We then move from 1st-order up to a maximum of 7th-order autocorrelation using Arellano-Bond (1991) test[10].

Coefficient stability over time is a particularly important feature in forecasting models. This is extremely relevant when most recent observations become available and forecasts need to be updated. Model misspecification for instance as a result of short time-series, omitted relevant variables, or breaks in the series would generally undermine such a property. We duly

---

[8] Given that different model specifications might include different sub-set of countries, for comparability's sake we compute equally weighted RMSE statistics taking into account forecast accuracy in two separate cases: (i) for the maximum number of countries (27 annual observations or less) for which the RMSE statistics could be computed and (ii) only for the biggest 5 member states in terms of (2012 year) population shares, i.e. Germany, U.K., France, Italy and Spain. All together, these 5 countries comprise around 63% of EU27 total population. Adding more countries to the list, would face a challenge: the lack of long enough time-series for some economic indicators that would not allow the computation of *comparable* RMSE for all periods and all model specifications.

[9] Recall the definition of the variable H' measuring the education attainment of five consecutive age cohorts, i.e. from 30 to 34. Assuming no inward or outward migration flows, we can simply approximate $H'_{age+1,t}$ by $H'_{age,t-1}$ where age is any value between 29 and 34. The first difference of the benchmark, $\Delta H'_{30-34,t} = H'_{30-34,t} - H'_{30-34,t-1}$ can be computed as $H'_{34,t} - H'_{29,t}$, where we have assumed that all the terms in between cancel out. This formulation of the $\Delta H'_{30-34,t}$ might produce fifth-order autocorrelation in the residuals of equation (3).

[10] Implemented in Roodman (2009).

check for this by estimating equation (3) over different time samples, by gradually expanding a short initial sample with new observations until we reach the full available sample. In this way we obtain recursive coefficients of the same model specification. For each estimation we go through all tests mentioned above so as to guarantee that our criteria are always met.

The results of this forecasting exercise are discussed and illustrated in section 5.

3.3 Data

The data on tertiary education attainment are available from Eurostat and also come disaggregated by gender and age. We use data for the 2000-2012 period at country level. Data on economic indicators used in this paper, such as labour productivity, total factor productivity (TFP), real compensation per employee are drawn from AMECO database of the European Commission's Directorate General for Economic and Financial Affairs. However, the employment rate of individuals aged 15-64 is taken from Eurostat, were longer time series were available.

Figure 1. Tertiary education attainment by country between 2000 and 2012

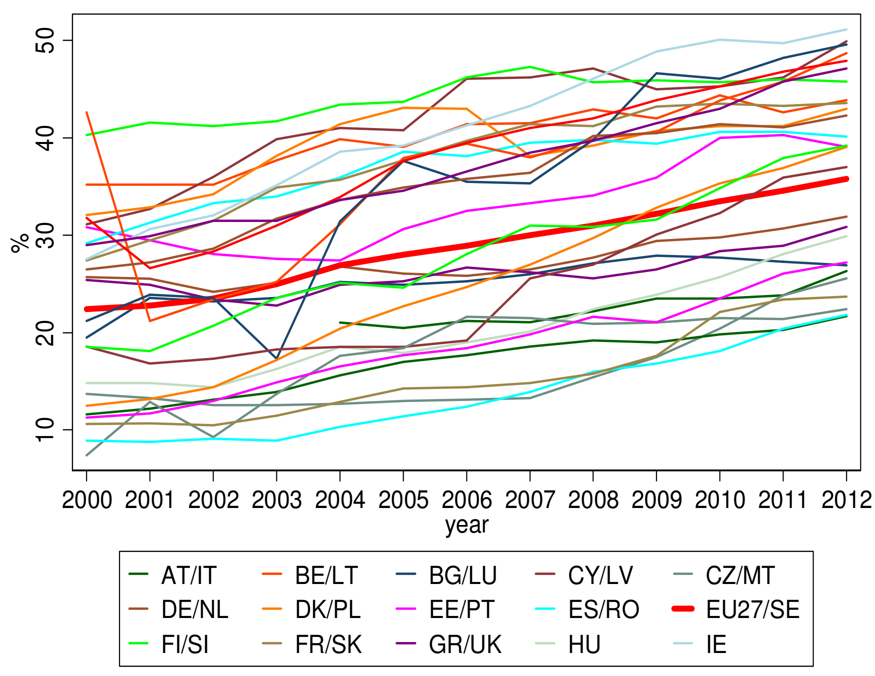
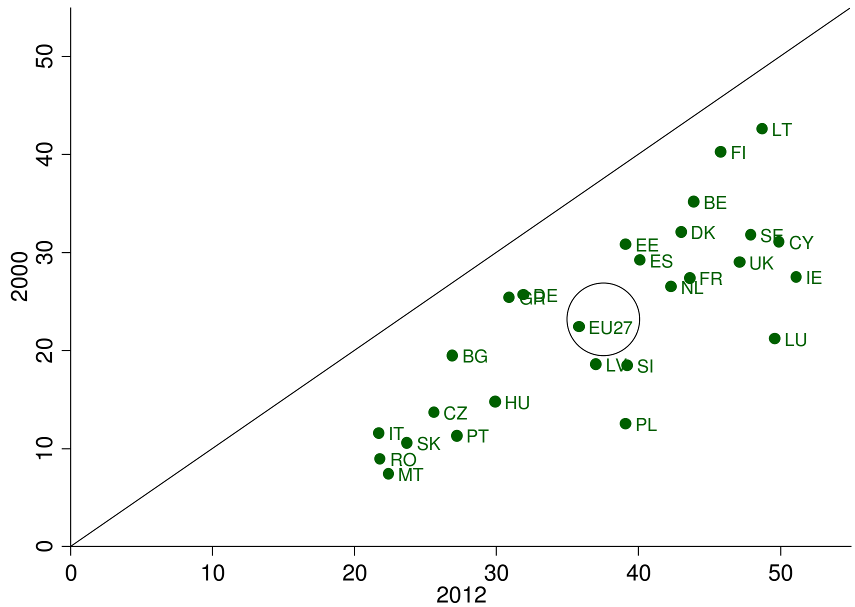


Figure 2. Country-specific tertiary education attainment between the 2000 and 2012



Figures 1 and 2 above provide an initial perspective on the Europe 2020 tertiary education attainment dynamics.[11] Three important findings could be noticed from these graphs. Firstly, the gradual increase in tertiary education attainment across all EU27 members has been stark and continuous over time. This suggests that younger cohorts have a more open attitude towards (higher) education attainment. Secondly, all countries seem to share a common trend, yet with no evident sign of convergence. Figure 2 especially, shows no clear evidence of convergence based on the initial education attainment level. Countries starting from a low level are not improving faster than countries with high initial levels. Thirdly, it appears that two main groups of countries become visible, one in the bottom left and another in the upper right corner. The distinction between the two groups is not very clear, with the EU27 average standing right between the two. We will come back to this in section 5.

4. Results

4.1 Preferred model specification

Following the testing procedure indicated in section 3.2, we arrive at a preferred model estimated on a sample of 12 European countries, namely Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the U.K. As of 2012, these 12 countries represent 76% of total EU27 population and generate 88% of its GDP.

[11] Please note that many member states experienced methodological breaks in the education attainment time-series around the year 2003.

Table 1 reports the estimated coefficients from equation (3) in its best specification. Each column in the table represents a regression of the fraction of tertiary educated people aged 30-34 on the share of adults with high education, meant to approximate the family background, and labour productivity[12], our proxy for expected returns in this case. In the first column, model (1) includes dummies to control for major breaks in tertiary education attainment time-series, occurred around 2003 in most countries. Next, model (1.1) adds to the initial specification a vector of country dummies that controls for mean differences in tertiary education attainment across countries. In model (1.2), we include a vector of year dummies that takes into account shifts in higher education attainment common to all countries. Throughout the paper we use Huber-White standard errors to allow for arbitrary correlation of residuals within each country.

Last, model (1.3) in Table 1 shows the estimated CCE specification. In line with our discussion from section 3.1, we tackle the potential cross correlations of errors due to omitted common effects by adding to model (1) the cross-section averages of the dependent and explanatory variables (coefficients not reported). Controlling for cross-section dependence does not significantly change the magnitude of the estimates, nor the sign or the statistical significance.

Table 1. OLS estimates of the share of tertiary educated individuals aged 30-34.

Model	(1)	(1.1)	(1.2)	(1.3)
Δ log(adults' share high education, 55-64) t	0.34***	0.29***	0.30***	0.34***
	(0.09)	(0.09)	(0.09)	(0.08)
Δ log(labour productivity), t-13	0.58**	0.54**	0.53*	0.57**
	(0.23)	(0.28)	(0.30)	(0.23)
Constant	0.48	-0.67	-1.37	-
	(0.60)	(1.21)	(2.35)	
Observations	144	144	144	144
R^2	0.32	0.42	0.44	0.62
Adj. R^2	0.30	0.35	0.33	0.61
# of countries	12	12	12	12
Estimation period	2001-2012	2001-2012	2001-2012	2001-2012
Control for breaks in series	yes	yes	-	-
Year dummies	-	-	yes	-
Country dummies	-	yes	yes	-
Recursive estimation summary:				
Wooldridge test for AR(1)	Ok	Ok	Ok	Ok
Baltagi-Wu LBI test	-	[2.12-2.28]	-	-
Arellano-Bond test for all AR(1)-(7)	Ok	No	No	Ok

Note. - Ordinary least squares estimates given. Model (1) includes dummies for 2003 and 2004 to control for the presence of structural breaks in the series. Model (1.1) adds country dummies, while model (1.2) includes all time dummies. Model (1.3) includes cross section averages of the dependent and explanatory variables. Recursive estimations summary refers to the estimation of the same model specification over expanding time samples, starting with 2001-2006 and ending with 2001-2012. The "ok" label means that the specification passes the respective test for all samples considered, while the "no" label indicates at least one failure. For the Baltagi-Wu test, we indicate the min-max range for LBI statistics (i.e. values around 2 denote lack of significant residual correlation). For the Arellano-Bond autocorrelation test, all lags up to 7-years were considered; the label in the table summarizes the results from all seven test statistics and all samples. Huber-White robust standard errors in parentheses allow for arbitrary correlation of residuals within each country.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[12] Restuccia and Vandenbroucke (2014) have recently proposed a model that explains education attainment over time and across countries by relying on productivity, along with life expectancy.

The results of the model are economically and statistically significant. In particular, the coefficient of 0.34 in the first column indicates that the share of tertiary educated young people might grow by 0.34 percentage points as a results of a 1% increase in the share of adults with a university degree or higher. After removing country fixed effects and year effects the impact of our proxy for family background slightly reduces the point estimates to 0.29-0.30. These results are consistent with the wealth of empirical investigations that typically find an intergenerational education correlation of the order of 0.3-0.5 for Western Europe (see for instance, Hertz et al., 2007, Chevalier et al., 2003, 2013).

Higher labour productivity leads to more schooling in our model, as it would be associated with an increase in the expected payoff. More specifically, a 1% rise in productivity causes an increase in tertiary education attainment by an amount of 0.6%. According to the skill-biased technological change theory, there is a positive relation between education choices and skill premium on the back of increases in technological progress. The 13th lag of labour productivity has been selected empirically but it also overlaps with the decision time on enrolment and graduation, representing therefore the proxy used by individuals in our model for expected returns to higher education. Note that using a long lag allays any concerns about potential reverse causality between productivity and higher education attainment.

Overall, Table 1 shows no significant change in the size of estimated coefficients across the four model specifications. Despite the introduction of some additional controls, the two main model coefficients were not altered, except for the change in the constant term and the gradual increase in the R^2 which were both to be expected. Less parsimonious specifications clearly provide better explanatory power, but with a dataset dominated by methodological breaks this is not very useful for our forecasting purposes.

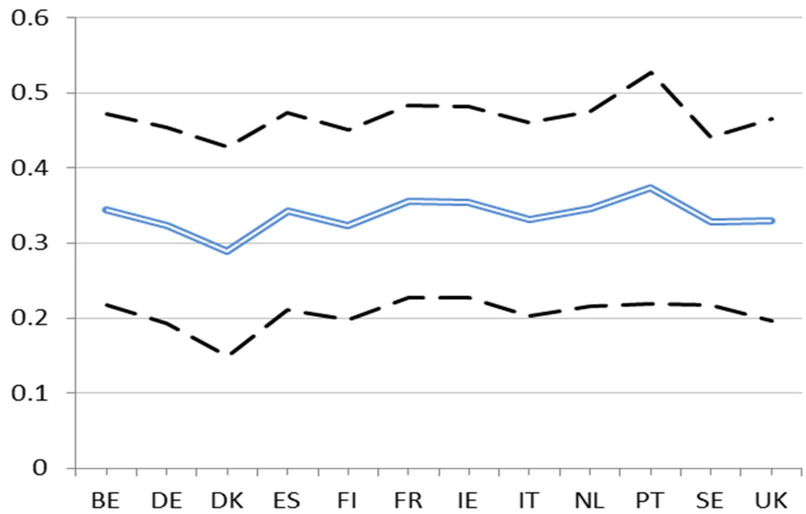
Our findings seem to suggest that higher education attainment is more responsive to investment-like motivations that better reflect expected income than to intergeneration mobility. With rising inequalities over the past decades in many countries, it is not surprising to see that the decision to invest in higher education today is driven to a large extent by income motivations.

4.2 Further robustness checks

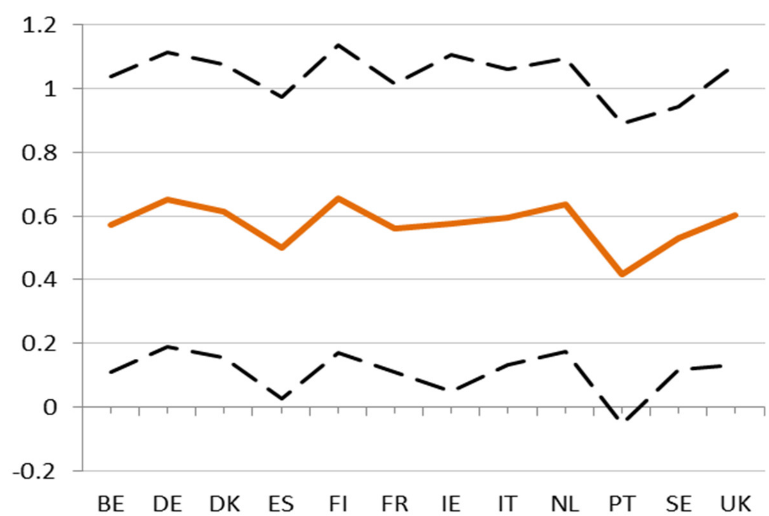
This section discusses some further robustness checks on the preferred specification.

To exclude the possibility that our results are driven by some specific countries, we re-estimate model specification (1) by taking out one country at the time. Results are displayed in Figure 3, where the horizontal axis indicates the excluded country and the vertical axis reports the magnitude of the estimated coefficients along with 95% confidence intervals of the coefficients. Figure 3 clearly suggests that the sample used in our preferred specification is quite homogenous and excluding one country at a turn does neither alter the size nor the statistical significance of the coefficients.

Figure 3. Check of coefficients' stability in the preferred model specification, excluding one country at a time.



a) Coefficient of the share of adults with high education, age 55-64



b) Coefficient of labour productivity

Note. – Ordinary least squares regression coefficients are given. 95% confidence intervals refer to coefficients shown in Table 1, column 1. Each coefficient is from separate regressions taking out one country at a time. Country removed from the sample is indicated in the horizontal axis.

Our model selection strategy was based on the range of statistical tests and rules as described in section 3.2. However, we provide estimates of equation (3) under a different selection of empirical proxies in order to demonstrate that the economic interpretation of the model is not sensitive to the chosen proxies of parental background and expected payoff. Tables 2 and 3 show the results of such an exercise. All model specifications provide a consistent story in line with the predictions of theoretical models that perceive education both as a consumption and investment good. For comparability's sake, our preferred specification is reported in the first column of each table. We consider separately the share of women and men aged 55-64 and the share of females aged 45-55 as proxies for parental background[13] along with TFP, employment rate and real compensation per employee as proxies for expected return. In Table 3 models are in general estimated by considering different sets of countries. This difference is the consequence of closely following the approach described in section 3.2 that only excludes countries identified as gross outliers in a robust regression procedure.

As Tables 2 and 3 clearly reveal, we find statistically significant coefficients of similar magnitude, with only a couple of exceptions. In Table 3, models (5) and (6), the coefficient of the share of tertiary educated adults almost doubles compared to best model specification in the first column; however these models do not pass autocorrelation tests. Note that the sign of employment rate is correctly reversed: the higher the employment rate, the lower the expected labour income and, therefore, the lesser the incentives to get higher education.

[13] We also use different education attainment levels and different gender/cohort combinations (e.g. the share of males with high education aged 45-54). Results not reported here, did not improve compared to those presented in Table 2 and 3.

Table 2. OLS estimates of the share of tertiary educated people aged 30-34.

Model	(1)	(2)	(3)
Δ log(adults' share high education, 55-64) t	0.34***		
	(0.09)		
Δ log(females' share high education, 55-64) t		0.25***	
		(0.06)	
Δ log(females' share high education, 45-54) t			0.15*
			(0.09)
Δ log(males' share high education, 55-64) t			0.23**
			(0.11)
Δ log(labour productivity) t-13	0.58**	0.65***	0.39*
	(0.23)	(0.21)	(0.24)
Constant	0.48	0.25	1.32*
	(0.60)	(0.63)	(0.73)
Observations	144	144	144
R^2	0.32	0.29	0.24
Adj. R^2	0.30	0.27	0.22
AIC	776.3	782.5	792.5
BIC	791.1	797.3	807.3
Control for breaks in series	yes	yes	yes
Panel	balanced	balanced	balanced
# of countries	12	12	12
Recursive estimation summary			
Wooldridge test for AR(1)	ok	ok	ok
Arellano-Bond test for AR(1)-(7)	ok	ok	ok
Out-of sample RMSE statistics (≤27) †			
1 year ahead	1.38	1.38	1.43
2 years ahead	2.23	2.23	2.22
3 years ahead	2.94	2.98	3.00
4 years ahead	3.54	3.61	3.66
Out-of sample RMSE statistics (=5) ‡			
1 year ahead	0.66	0.71	0.86
2 years ahead	1.08	1.15	1.42
3 years ahead	1.49	1.59	2.04
4 years ahead	1.87	2.02	2.72

Note. - Ordinary least squares estimates given. All models include dummies for 2003 and 2004 to control for the presence of structural breaks in series. Recursive estimations summary refers to the estimation of the same model specification over expanding time samples, starting with 2001-2006 and ending with 2001-2012. The "ok" label means that the specification passes the respective test for all samples, while the "no" label indicates at least one failure. For the Arellano-Bond autocorrelation test, all lags up to 7-years were considered; the label in the table summarizes the results from all seven test statistics and all samples. Huber-White robust standard errors in parentheses allow for arbitrary correlation of residuals within each country. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

† The RMSE is calculated as an equally weighted average of all available countries (at most 27). The 1 year ahead RMSE is computed using 6 observations, 2 years ahead RMSE - using 5 observations and so on.

‡ The RMSE is calculated as an equally weighted average of the biggest 5 countries in terms of population (as of 2012). These countries are: Germany, U.K., France, Italy and Spain; together, these 5 countries comprise around 63% of EU27 total population.

Table 3. OLS estimates of the share of tertiary educated people aged 30-34.

Model	(1)	(4)	(5)	(6)
Δ log(adults' share high education, 55-64) t	0.34***	0.35***	0.57***	0.65***
	(0.09)	(0.09)	(0.17)	(0.15)
Δ log(labour productivity) t-13	0.58**			
	(0.23)			
Δ log(TFP) t-12		0.41*		
		(0.25)		
Δ log(employment rate) t-11			-0.58***	
			(0.20)	
Δ log(real compensation per employee) t-11				0.42*
				(0.23)
Constant	0.48	1.23*	2.34***	-0.11
	(0.60)	(0.65)	(0.68)	(0.75)
Observations	144	132	201	168
R^2	0.32	0.31	0.38	0.51
Adj. R^2	0.30	0.29	0.38	0.50
AIC	776.3	717.7	1232	1013
BIC	791.1	732.1	1242	1028
Panel	balanced	balanced	unbalanced	balanced
Control for breaks in series	yes	yes	-	yes
# of countries	12	11	27	14
Recursive estimation summary:				
Wooldridge test for AR(1)	ok	ok	no	no
Arellano-Bond test for AR(1)-(7)	ok	ok	no	no
Out-of sample RMSE statistics (≤27):				
1 year ahead [+)]	1.38	1.36	1.32	1.86
2 years ahead	2.23	2.32	2.26	3.00
3 years ahead	2.94	3.04	2.91	3.89
4 years ahead	3.54	3.53	3.49	4.92
Out-of sample RMSE statistics (=5):				
1 year ahead	0.66	0.67	0.79	0.72
2 years ahead	1.08	1.10	1.48	1.13
3 years ahead	1.49	1.56	2.41	1.54
4 years ahead	1.87	1.99	3.37	1.90

Note. - Ordinary least squares estimates given. All models include dummies for 2003 and 2004 to control for the presence of structural breaks in series. Recursive estimations summary refers to the estimation of the same model specification over expanding time samples, starting with 2001-2006 and ending with 2001-2012. The “ok” label means that the specification passes the respective test for all samples, while the “no” label indicates at least one failure. For the Arellano-Bond autocorrelation test, all lags up to 7-years were considered; the label in the table summarizes the results from all seven test statistics and all samples. Huber-White robust standard errors in parentheses allow for arbitrary correlation of residuals within each country. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

† The RMSE is calculated as an equally weighted average of all available countries (at most 27). The 1 year ahead RMSE is computed using 6 observations, 2 years ahead RMSE - using 5 observations and so on.

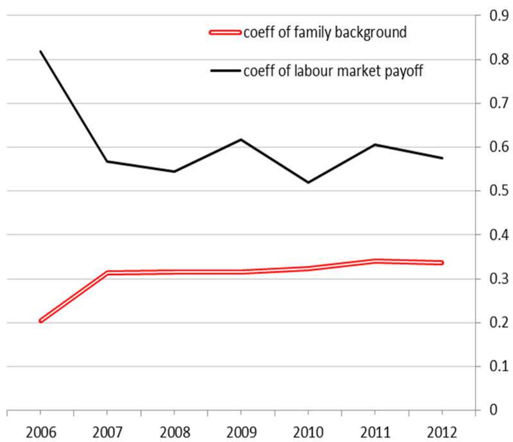
‡ The RMSE is calculated as an equally weighted average of the biggest 5 countries in terms of population (as of 2012). These countries are: Germany, U.K., France, Italy and Spain; together, these 5 countries comprise around 63% of EU27 total population.

In line with our discussion from section 3.2, we recursively estimate all models presented in Tables 2 and 3. We start with an initial sample 2001-2006 and expand it until the full sample 2001-2012 is reached. Figure 4 reports the results of the recursive estimation of models of Tables 2 and 3. For the readers’ convenience we provide a plot of the recursive coefficients estimated by each model specification (Figure 4).

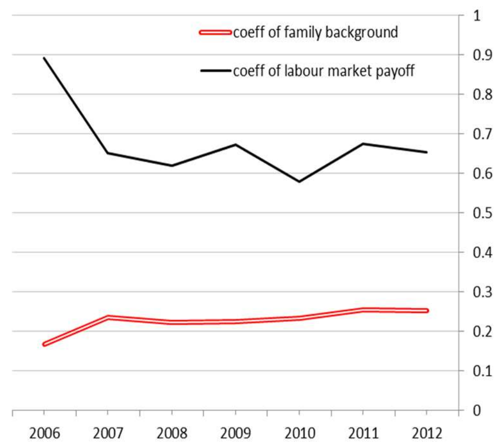
Model specifications (1) and (2) show very high coefficient stability and have very similar properties. However, the latter specification has smaller RMSE at long horizons, which is one of the most common criterion in assessing forecast accuracy. Model specification (6) also shows quite stable recursive coefficients over time, but it does not pass the autocorrelation tests. For all other model specifications, the recursive coefficients are either trending or unstable.

Note that the ranking of the models (1)-(6) based on forecasting accuracy does not change between the two sets of RMSE indicators computed on different country sets. Overall, we conclude that our preferred model specification has superior forecast accuracy and also successfully passes the coefficients' stability check.

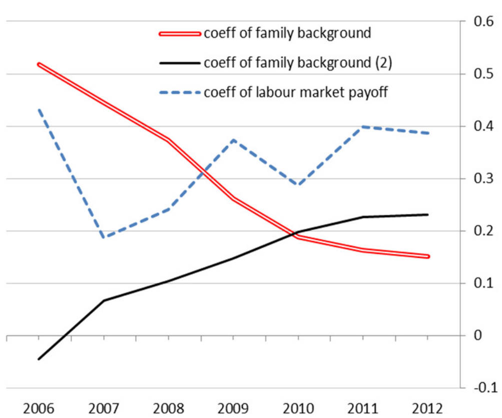
Figure 4. Coefficients stability check for models of Tables 2-3.



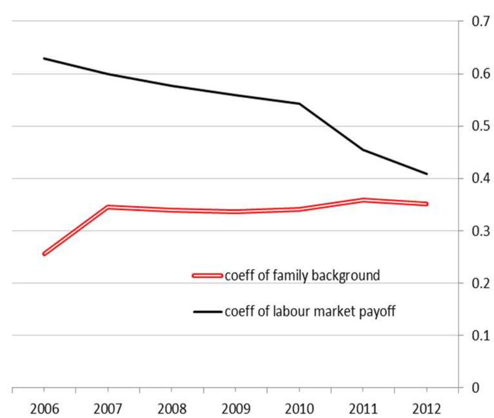
a) Model (1)



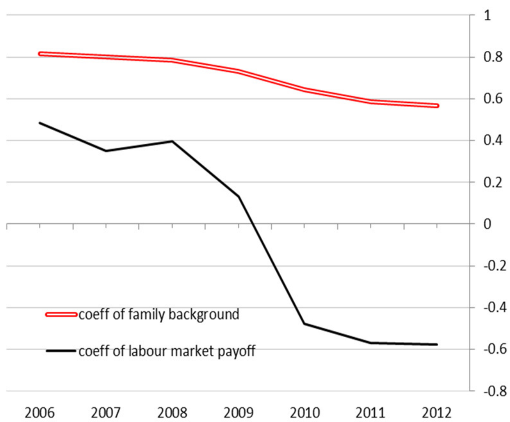
b) Model (2)

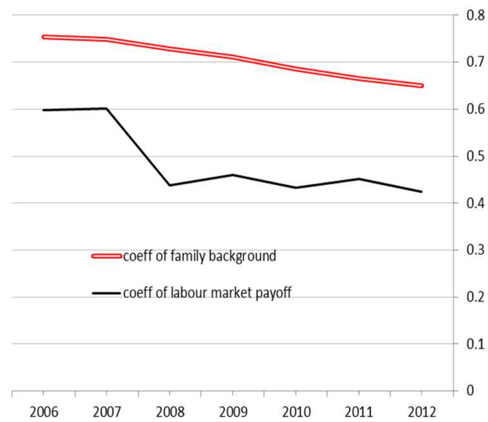


c) Model (3)



d) Model (4)





e) Model (5)

f) Model (6)

Note. – Recursive ordinary least squares regression coefficients of models in Table 2 and 3 are given. The shortest time-sample is 2001-2006, while the longest is 2001-2012.

5 Forecasting tertiary education attainment up to 2020

To provide an answer to our initial question whether EU27 will be able to reach the target of 40% tertiary education attainment by 2020, we need two additional assumptions. First, we assume that, at aggregate level, EU27 behaves according to the preferred model specification discussed in section 4.1. With no country-specific terms included, this assumption should be straightforward. Second, we assume that the 12 countries included in the final estimation sample provide a good approximation of the EU27 aggregate mainly because they represent a significant share in terms of both population and GDP. Including more than 12 countries was not supported by our empirical strategy as discussed in sections 3.2 and 4.

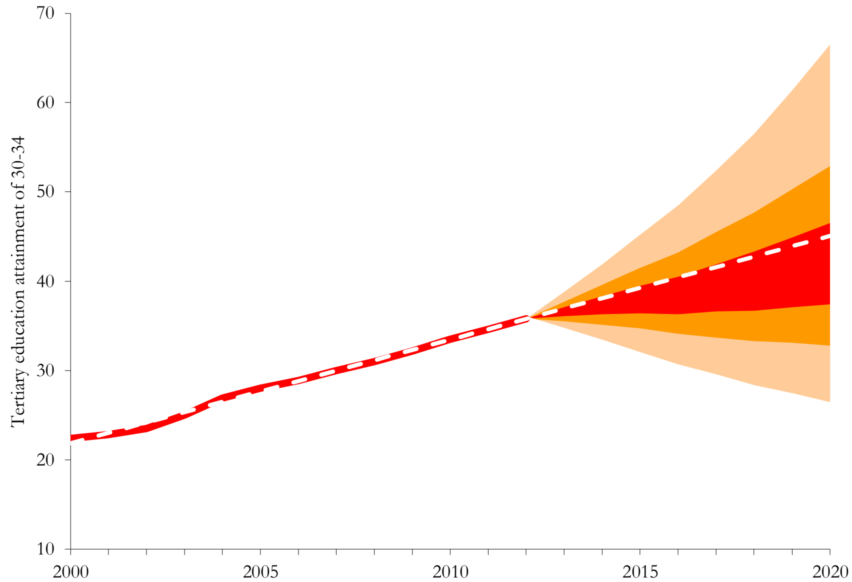
The EU27-aggregate forecast displayed in Figure 5 is built in two steps. The first step is concerned with obtaining EU27 projections for the two model's determinants, i.e. education level of the adult population and labour productivity. For the first determinant, we adopt a simple extrapolation method based on aging for projecting the shares of 55-64 year-olds with tertiary education over the period 2013-2020. Appendix A describes in greater detail the extrapolation strategy and discusses potential caveats. For the second determinant, the lagged labour productivity allows generating forecasts up to 2020 using readily available data.[14] As a consequence, our forecasts will be *conditional* on the expected developments of the model's regressors.

In the second step, we use the preferred model coefficients to compute the expected change in the share of tertiary educated 30-34 year-olds up to 2020. Based on the estimated variance and assuming normal distribution, we can derive upper and lower bounds associated

[14] In fact, the 13-year lag might, in principle, offer estimates at least up to 2025.

with standard probability levels. Figure 5 depicts the forecasts for EU27 aggregate as a fan chart, using the 30%, 60% and 90% probability intervals.

Figure 5. Forecasts of tertiary education attainment of 30-34 years olds.



Note. – The shaded areas in the figure above plot the 30% (dark red), 60% (red) and 90% (light red) confidence intervals for EU27 *tertiary education attainment* forecast constructed according to model (1) of table 1.

The superimposed dashed white line in Figure 5 is the time trend and represents the naïve forecast based on historical data. Both the naïve and the conditional forecasts show a similar trend. However, the linear trend suggests a slightly faster growth in the share of tertiary educated individuals compared to the central 30% distribution of the conditional forecast. This however can be explained by the expected dynamics of the two model's determinants. Firstly, the slow growth in the share of tertiary educated adults aged 45-54 registered in the period 2000-2010 translates into a slower expected improvement for the 55-64 year-olds over the period 2010-2020. Secondly, aggregate labour productivity has slowed down over the period 2000-2007 with respect to the 1990s', producing a flatter trajectory for the period 2013-2020.[15]

The contribution of the two determinants to the forecast could be assessed by decomposing the forecast of the EU27 tertiary education attainment. More precisely, in Table 4, we present the percentage point contributions of adult´s tertiary education and labour productivity to the change of tertiary education over the period 2013-2020. As can be seen, labour productivity plays a larger role than adults' education in explaining the improvement in tertiary education (7.3 pp versus 3.6 pp).

[15] Despite the economic boom occurred in this period, most European countries have registered lower economic growth and productivity dynamics compared to the period before 1999 in the run-up to the introduction of the common currency.

Table 4. Forecasts decomposition of the EU27 tertiary education attainment of 30-34 year-olds.

2013-2020 cumulated change in:	
Tertiary education attainment, 30-34 year-olds,	15.2
contributions in p.p.:	
Adults' share high education, 55-64 year-olds	3.6
Labour productivity	7.3
Constant	4.3

Note. – All contributions in the table are in log terms.

We construct country-specific forecasts by following the same two steps described above.[16] One caveat though relates to the importance of country heterogeneity in the context of a *future probable convergence process* in higher education. If this were to be the case, a country starting from a low level but with significantly faster than average improvement in education attainment would have an underestimated forecast based on a model that omits the linear time trend from equations (2.1)-(2.2) or, equivalently, the country-specific constant from equation (3). The first argument that stands against this interpretation is that our estimated fixed-effects specification did not pass the required autocorrelation tests. By not including country fixed effects in the preferred model specification, extending the model to other countries becomes more straightforward. As a second argument, we provide additional evidence to substantiate the use of our econometric model to construct country-specific forecasts. In particular, we compare our conditional forecasts to an alternative set of forecasts derived from a cohort-based approach. As described in greater detail in Appendix B, this cohort-based model uses country-specific data on tertiary enrolment, duration of studies and completion rates. Being built on two different sets of modelling assumptions the two approaches are completely independent. Due to data limitations, we are able to provide country-specific forecasts of the tertiary education attainment of the 30-34 years old for a limited number of years and countries. By comparing the two sets of forecasts, we find no evidence that the forecasts from the econometric model are systematically biased due to the omission of country-specific characteristics.

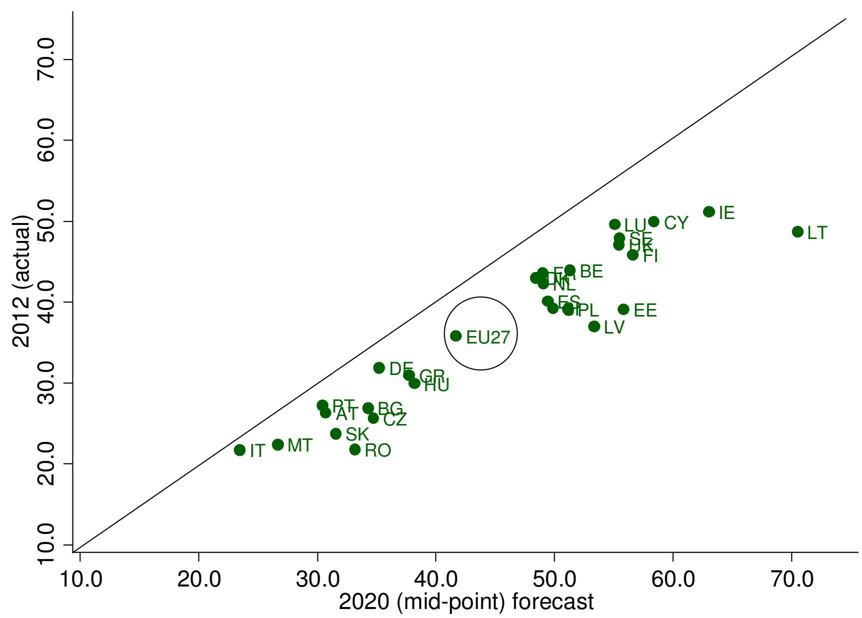
The country-specific forecasts for tertiary attainment are suitable to understand how fast countries are expected to improve their level of tertiary education until 2020. Graphical evidence of this exercise is shown in Figure 6. In particular, tertiary education attainment estimates in 2020 are contrasted with 2012 values for each European member state. The farther away a country is situated from the diagonal, the larger its expected improvement. The average distance from the main diagonal corresponds to the constant term of the model, while variations around this level reflect the country-specific contributions of model the determinants. If a country has better a than average projection, it will have better than average outcome in terms of education attainment and will lie farther away from the main diagonal.

Figure 6 clearly identifies two major groups of countries: one in the upper right corner and another one in the lower left corner. These findings can be interpreted as showing a divided Europe, with some countries experiencing fast improvements (upper right) and other countries

[16] A common practice in forecasting assumes "no policy change" over the entire horizon. At country level, one can relax this assumption and use different scenarios for the model determinants, especially for adults' education attainment. A different exogenous path than the one discussed in Appendix A might include some factors omitted in the empirical model, such as migration flows.

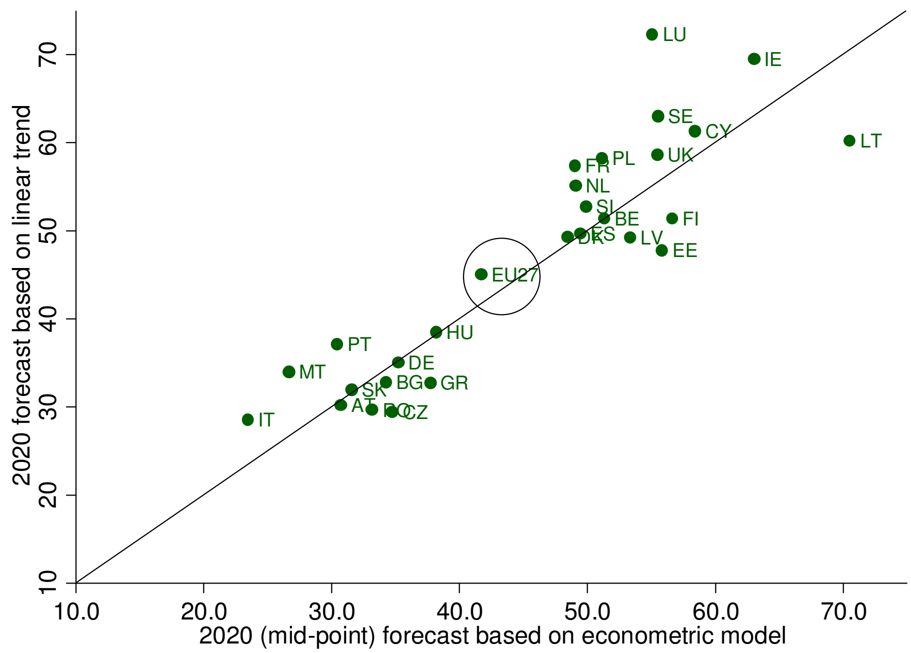
lagging behind (lower left). This development poses a major policy challenge for the EU as a whole, given that human capital is an important driver of innovation that supports the real convergence process. By comparing the evidence provided in Figure 2 and Figure 6, the gap between the two groups seems to be widening. In fact, the best performers group seems to lie on average farther away from the main diagonal compared to the worse performers group, suggesting a much bigger improvement in percentage points for the former.

Figure 6. Country-specific tertiary education attainment between the 2012 (actual) and 2020 (forecast)



Next, we compare country-specific forecasts based on the econometric model with the naïve country-specific forecasts (i.e. derived using linear trends). We plot the results in Figure 7. The most interesting aspect is that, within the *leaders* group (upper right), there are more countries with a naïve forecast higher than the one generated by econometric model determinants, while the opposite is true for the *laggards* group (lower left). This means that the worst performers might have a chance in the future and beat a naïve forecast, but only if they are able to turn the odds in their favour, by improving the along the mechanisms suggested by the econometric model presented.

Figure 7. Comparison between country-specific forecasts: linear trend versus econometric model



6 Concluding remarks

The European Commission is committed to increase tertiary education attainment to at least 40 percent of the relevant population (30-34 year-olds) by 2020. This paper aims to address this major policy concern based on a standard theoretical framework and using a rigorous empirical approach. In so doing, we had to deal with a very limited data environment such as the short time span and the presence of breaks, gaps in data and outliers that make it possible only to focus on a restricted sample of countries. We show that our empirical model meets important features such as coefficient stability, robustness, small out-of-sample RMSE, and is thus capable to provide insightful long-term forecasts.

Our forecasting exercise shows an overall optimist picture for Europe as a whole and up to 2020, despite some remaining challenges. We point to two main groups of countries with diverging dynamics in terms of tertiary education attainment. Our forecasts are mainly driven by demographics and past labour market dynamics, leaving a narrow space for effective policy adjustment before 2020. When compared against the naïve forecasts, our model-driven projections show that the future does not necessarily have to look like the past. However, due to the long lags involved, strong policies might be needed today to improve the odds of those countries that lag behind. The widening gap between the two groups has better chances of being reversed after 2020, but strong policy actions might be needed today in line with the transmission mechanisms outlined by the model.

All estimated model specifications provide a consistent story in line with the predictions of theoretical models that perceive education both as a consumption and investment good.

However, schooling decisions show a higher dependence on expected income than on family background, which is not surprising in a context dominated by rising returns to education and income inequalities.

The paper also indicates relevant policy areas where improvements could be most efficient. A policy of improving labour market efficiency could reinforce monetary incentives in getting educated. A frictionless labour market may simplify the way individuals anticipate income distribution and skill premium, which feeds back into their education decisions. Labour market reforms, thus, could have important indirect effects on education attainment by increasing the positive spill-overs associated with skilled workers, education decisions and technological progress. Our empirical findings clearly expose this dynamic link, highlighting the reverse causality channel that runs from economic developments to education decisions.

Increasing education participation can have positive externalities to future cohorts through intergenerational transmission of educational choice. However, this might raise a discussion on inequality and the role of education for intergenerational mobility in the long-run. Well-designed retraining programmes with a focus on life-long learning might prove effective in improving the education level of adult cohorts and thus affect youth attainment in the end.

Appendix A

Projecting the education attainment of adult population

In this Appendix we provide details about the projections of adults' education attainment up to 2020. Ignoring (i) inward and outward migration flows, (ii) mortality risk differential between individuals with different education attainment and (iii) assuming no change or improvement in education attainment for adult population cohorts, we can write the evolution of education attainment of adult population according to the following dynamic equation:

$$Z(g, t + 10) = Z(g - 10, t) + \lambda * gap$$

where $Z(g, t)$ is the share of individuals with a given education level measured at time t and having age g. Note that, if the time index is fix, the two age groups g and $g - 10$ would be interpreted as referring to different population cohorts separated by 10 years. However, in this case, the equation above is the formalization of a simple aging process viewed over time from the perspective of a single cohort. We chose to work with 10 years knots due to data availability.

However, since our simplifying assumptions above might not hold in the data, a gap will emerge between the two indicators. This gap between $Z(g, t + 10)$ and $Z(g - 10, t)$ could be significant in some cases, reflecting a departure from our assumptions set. Inward/outward migration seems to be the strongest assumption in this case, especially since migration flows have increased across Europe during the last decade, mainly as a result of EU enlargement. Thus, we allow for its gradual phasing out at the rate given by λ, arbitrarily set at 0.5. These gaps might be calculated for 2010, 2011 and 2012 time points only, using the available data for adults' education attainment between 2000-2012 in the age brackets: 25-34, 35-44, 45-54, 55-64. However, we only need the 2012 value of the gap in order to construct the exogenous path for adults' education attainment according to equation above.

Appendix B

Checking the unbiasedness of the country-specific forecasts

This appendix introduces the cohort model as an alternative method for forecasting tertiary education attainment of 30-34 year-olds. Given its methodological independence from previous modelling assumptions of section 5, it provides a critical test for the unbiasedness of the country-specific forecasts.

Cohort methods are widely employed to analyse a targeted population with certain characteristics born in a specific period (see Yang and Land, 2013; Myrskylä et al., 2013). Here, we use birth cohorts to assess the education level pertaining to individuals aged 30-34 years. We calculate the number of tertiary graduates by employing country-specific information on (i) new entrants to tertiary education by age group, (ii) the average duration of studies and (iii) the average completion rate. Data on the number of individuals that entered tertiary education by age group and population by age group are drawn from UNESCO-UIS/OECD/Eurostat data collection on statistics of education, while tertiary education completion rate and average duration of studies are taken from OECD.[17] Due to data availability, we are only able to calculate the tertiary education attainment for the years 2014-2017.[18]

For each year during the 2014-2017 period and for each birth cohort, we calculate the number of individuals aged 30-34 that enter tertiary education, i.e. ISCED 5A and 5B levels. In addition, by using the average duration of studies, we ensure that new entrants had enough time to finish their studies before counting them in the benchmark. Next, we divide the number of new entrants of each cohort by the total birth cohort population and then multiply this share by a country-specific completion rate.[19]

The results from the birth cohort method can be used to check whether country-specific forecast using the panel approach are systematically biased. Figure 8 shows the correlations between the difference of panel- and cohort-based forecasts and the average improvement of the benchmark value until 2020 as a measure of the omitted country fixed-effects. We consider two groups of countries: the round points denote countries included in our favourite model specification, while the square points indicate countries not included in the sample but for which the cohort forecasts were available.

If country-specific forecasts are biased due to the omission of fixed effects or due to omission of the country from the estimated sample, the scatter plot should display a strong positive correlation: the higher the past improvement of the country, the more underestimated the conditional forecasts versus the cohort-based forecasts and the higher the difference between the two. However, looking at the scatter plot in Figure 8 we do not observe any correlation for the two groups considered. For information, we plot the estimated R^2 of a best linear fit for both

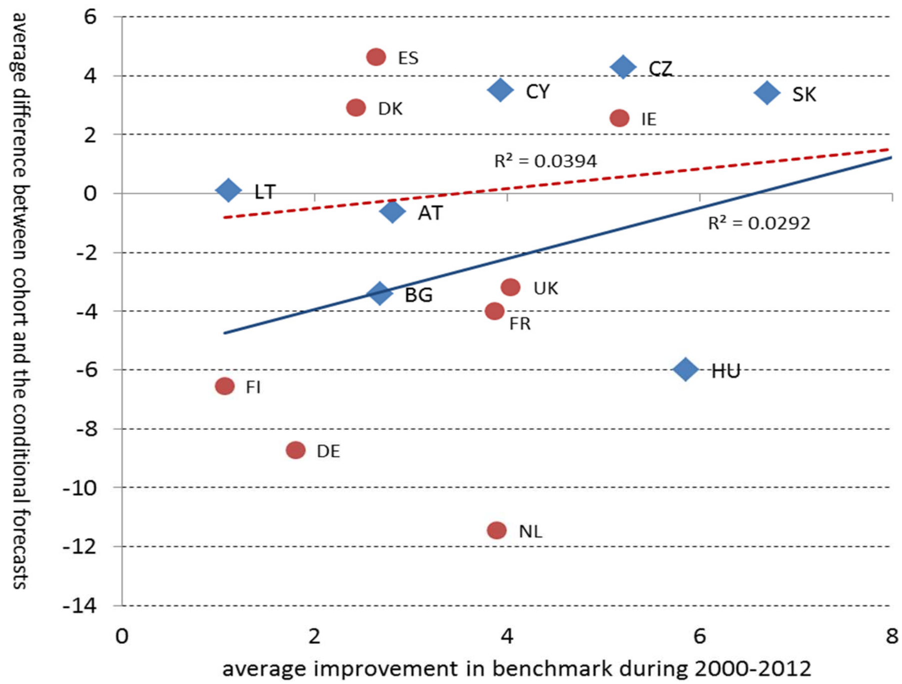
[17] See OECD (2008, 2010), indicators published in Education at a Glance reports.

[18] For the years 2013 and 2018-2020 some of the data on new entrants is unavailable. More precisely, for the year 2013 data on new entrants in 1997 is required. However, new data on new entrants to tertiary education is only available from 1998 onwards. In addition, for the years 2018-2020 data on new entrants from 2004 onwards would be required.

[19] Note that the completion rate used for the birth cohort model is the completion rate for the tertiary education level 5A only. This implies that the calculated values of the benchmark crucially depend on having a constant tertiary completion rate across years and types of programs (5A and 5B).

groups. Hence, we can conclude that extending our forecasting exercise to all EU27 countries is not subject to substantial biases induced by our empirical strategy.

Figure 8. Comparison of conditional forecasts with cohort-based forecasts.



References

Acemoglu, D. (2002). Technical Change, Inequality, and the Labour Market. *Journal of Economic Literature*, *40*, 7-72.

Acemoglu, D. & Pischke, J. S. (2001). Changes in the wage structure, family income, and children's education. *European economic review*, 45 (4-6), 890-904.

Arellano, M., & S. Bond. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Asteriou, D., & Agiomirgianakis, G. M. (2001). Human capital and economic growth: time series evidence from Greece. *Journal of Policy Modeling*, 23(5), 481-489.

Attanasio O. & Kaufmann K. (2012). "Education Choices and Returns on the Labour and Marriage Markets: Evidence from Data on Subjective Expectations", Working Paper, Bocconi Univesity, Milan.

Autor D., Katz, L. F. & Krueger, A. B. (1998). Computing Inequality: Have Computers Changed the Labour Market? *The Quarterly Journal of Economics 113 (4): 1169-1213.*

Baltagi, B. H. & Wu, P. X. (1999). Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory*, 15: 814–823.

Becker, G. (1964). *Human Capital. A Theoretical and Empirical Analysis, with Special Reference to Education*. NBER: Chicago.

Becker, G. S. & Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, 87, 1153–89.

Becker, G.S. & Tomes, N. (1986). Human Capital and the Rise and Fall of Families, *Journal of Labour Economics* 4, no. 3, pt. 2, S1-S39.

Behrman, J. R. & Rosenzweig, M. R. (2002). Does Increasing Women's Schooling Raise the Schooling of the Next Generation? *American Economic Review*, 91(1), 323–334.

Benabou, R. (1996). Heterogeneity, stratification and growth: Macroeconomic implications of community structure and school finance". *American Economic Review*, 86, 584-609.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4), pp. 352-365.

Bils M. & Klenow, P. J. (2000). Does Schooling Cause Growth?, *American Economic Review*, 90(5), 1160-1183.

Black, S. E., & Devereux, P. J. (2011). Recent developments in intergenerational mobility, in Ashenfelter, O. & Card, D. (eds), *Handbook of labour economics*, 4, 1487-1541.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data model*, Journal of Econometrics,* 87(1): 115-143.

Buchinsky, M. & Leslie, P. (2010). Educational Attainment and the Changing U.S. Wage Structure: Dynamic Implications on Young Individuals' Choices. *Journal of Labour Economics* 28(3), 541-594.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica*, 69(5), 1127-1160.

Cameron, S. V., & Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts (No. w6385). *National Bureau of Economic Research*.

Cameron, S. V., & Taber, C. (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy*, *112*(1), 132-182.

Chevalier, A., Denny, K. & McMahon, D. (2003). A Multi-Country Study of Inter-Generational Educational Mobility. In Dolton, P., Asplundh, R. & Erling, B. (eds). Education and Inequality Across Europe. Edward Elgar, London.

Chevalier, A., Harmon, C., O'Sullivan, V. & Walker, I. (2013). The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics*, 2:8.

Chudik, A., & Pesaran, M. H. (2013). Large panel data models with cross-sectional dependence: a survey, *CESifo Working Paper*, 4371.

Dominitz, J. & Manski C. F. (1996). Eliciting student expectations of the return to schooling. *Journal of Human Resources*, 31(1), 1-26.

Freeman, R. B. (1971). The market for College-trained Manpower. Cambridge, Massachusetts: Harvard University Press.

Freeman, R. B. (1975). Legal Cobwebs: a recursive model of the market for new layers. *Review of Economics and Statistics*, 57, 171-179.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. & Verashchagina, A. (2007). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends *The B.E. Journal of Economic Analysis & Policy*, 7(2): 1935-1682.

Heckman, J. J. (1976). A Life-Cycle Model of Earnings, Learning, and Consumption. *Journal of Political Economy,* 84(4) part 2: S11–S44.

Jacoby, H.G. (1994). Borrowing Constraints and Progress Through School: Evidence from Peru. *The Review of Economics and Statistics*, 76(1), 151-160.

Jacoby, H.G. & Skoufias, E. (1997). Risk, financial markets and human capital in a developing country. *Review of Economic Studies*, 64, 311-335.

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 125(2), 515–548.

Kaufmann, K. (2010). Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns, Working Papers 362, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, *105*(3), 473-522.

Kodde, D. (1986). Uncertainty and the Demand for Education, *Review of Economics and Statistics*, 68 (3), 460-467.

Li, G. (1985). Robust regression. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey, (eds), Exploring Data Tables, Trends, and Shapes, 281–340. New York: Wiley.

Mincer, J. (1974). Schooling Experience and Earnings NY: NBER Columbia U. Press.

Myrskylä, M., Goldstein, J. R., & Cheng, Y. H. A. (2013). New cohort fertility forecasts for the developed world: rises, falls, and reversals. *Population and Development Review*, 39(1), 31-56.

OECD (2008). Education at a Glance. OECD Indicators. Paris: Organisation for Economic Co-operation and Development.

OECD (2010). Education at a Glance. OECD Indicators. Paris: Organisation for Economic Co-operation and Development.

Restuccia, D. & Vandenbroucke, G. (2014). Explaining educational attainment across countries and over time, *Review of Economic Dynamics*, Available online, ISSN 1094-2025, http://dx.doi.org/10.1016/j.red.2014.03.002

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86-136.

Todd, P. E., & Wolpin, K. I. (2006): "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 96(5), 1384–1417.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA, MIT Press.

Yang, D.Y, & Land, K. C. (2013). Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications, Chapman and Hall/CRC; Taylor & Francis Group, Boca Raton.

Zellner, A. (1992). Statistics, science and public policy. *Journal of the American Statistical Association*, 87, 1-6.