Available online at www.sciencedirect.com

ScienceDirect

Journal of Policy Modeling 37 (2015) 804–819

Journal of Policy Modeling

www.elsevier.com/locate/jpm

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

Catalin Dragomirescu-Gaina *, Leandro Elia, Anke Weber

*European Commission, Joint Research Centre (JRC), Econometrics and Applied Statistics, via Enrico Fermi 2749, 21027 Ispra, VA, Italy*

Received 14 December 2014; received in revised form 22 April 2015; accepted 10 May 2015
Available online 10 June 2015

**Abstract**

This paper strives to answer the question whether Europe will reach its tertiary education attainment target by 2020. 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. Insights into the dynamics of future education attainment and remaining policy challenges are highlighted. While Europe is likely to achieve its target, there is a growing divide between best and low performing countries that raises doubts with respect to real economic convergence prospects.


*JEL classification:* C23; C52; C53; J24

*Keywords:* Human capital investment; Tertiary education; Panel data; Forecasting; Europe 2020 strategy

## 1. Introduction

In Europe, the share of highly-educated individuals has steadily increased over the past decade. However, most European Union (EU) countries fall short of having figures comparable to the U.S. or other high-income countries. This unsatisfactory outcome has led the European Commission

* Corresponding author.
*E-mail addresses:* catalingaina@gmail.com (C. Dragomirescu-Gaina), leandro.elia@jrc.ec.europa.eu (L. Elia), anke.weber@ec.europa.eu (A. Weber).

http://dx.doi.org/10.1016/j.jpolmod.2015.05.002

(EC), within the strategic framework for European cooperation in education and training, to put forward a political agenda emphasizing 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 is then translated into national targets by taking into account country-specific situations. In 2012, for example, the share of tertiary graduates (henceforth the benchmark) remains below the EU target, namely at 35.9%, with some countries such as Italy, Malta and Romania scoring as low as 22%, far below their national targets.

This paper poses the question whether the Europe 2020 target on tertiary education can be achieved. To provide an answer, we estimate a model of education attainment based on a standard theoretical framework (Acemoglu & Pischke, 2001; Becker, 1964; Becker & Tomes, 1986; Behrman & Rosenzweig, 2002), 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 & Wolpin, 1997; Todd & Wolpin, 2006) or to estimate a statistical model of schooling attainment using microdata (as in Cameron & Heckman, 1998; Kaufmann, 2010). This paper adopts a different approach. It presents estimates of schooling attainment derived using 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 our approach is the use of very few realistic assumptions that make the determinants of the model become exogenous with respect to schooling decision. Our 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, we expose a diverging pattern among Member States, with some countries improving faster than others. Given the complex causality links between economic dynamics and education attainment, the different trajectories of best and worst performers point to important policy challenges with respect to real economic convergence prospects within the EU.

This paper contributes to the relevant empirical literature in at least three aspects. 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 (higher) education of young cohorts. 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 consistent with existing theoretical and empirical evidence, while having a parsimonious specification and stable coefficients over time. Third, by decomposing the expected increase in tertiary education until 2020 into its main determinants–that is labour productivity and adults' education, we point to policy areas that might have longer-term consequences on education attainment.

The analysis ramps up the debate on tertiary education attainment in Europe by offering a baseline scenario 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 and reducing financial constraints. The second refers to labour market functioning, with a focus on the formation process of wage expectations and the consequences of technological progress on wage distribution and skill premium. With regard to policy effectiveness, the year 2020 might be too tight of a deadline to achieve a meaningful impact on the benchmark. Still, strong policies might be needed today to reverse 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 the forecast results and their policy implications. Last, Section 6 concludes.

## 2. Education decisions, income expectations and family background

There is a rich theoretical literature addressing the economics of human capital investment. 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 her expected discounted life-time utility (see Becker, 1964; Heckman, 1976; Keane & Wolpin, 1997; Mincer, 1974).

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. Dominitz and Manski (1996) showed that students are capable of making realistic estimates of future incomes and these are consistent with their performance on the labour market. Oosterbeek and Webbink (1995) highlight the influence of income expectations on higher education enrolment in the Netherlands. Recently, Buchinsky and Leslie (2010), Kaufmann (2010), Jensen (2010) and Attanasio and Kaufmann (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-sectional data, but might become less of a problem in time-series analysis. We take advantage of this fact and use long time-lags to reflect labour market dynamics associated with schooling choices. With time passing, the distance between expectations and actual realizations gets blurred, allowing us to pick up empirical proxies easier. It is commonly assumed that a good approximation of the expected return to education is the difference in earnings from undertaking and not undertaking education (see Acemoglu & Pischke, 2001). Autor, Katz, and Krueger (1998) and Acemoglu (2002) among many others suggest the existence of a positive feedback loop between labour incomes and skills on the back of productivity advancements. Their evidence shows that technology shifts over recent decades have favoured skilled or highly educated workers.

A second major theoretical strand dates back to Becker and Tomes (1979), Becker and Tomes (1986) who consider the role of family socio-economic 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 (Acemoglu & Pischke, 2001; Black & Devereux, 2011; Cameron & Heckman, 1998). 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 thus can be thought of as a consumption good, where wealthier families can *consume* more of it. Therefore, controlling for intergenerational transmission is crucial when explaining education levels (Pascual, 2009). We do so by using adults' education instead of income. Unlike earnings, education has some advantages in terms of estimating intergenerational transmission (see Black & Devereux, 2011). 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 key role in explaining the impact of financial constraints on children education attainment.

Consistent with the above-mentioned literature, we consider education attainment, which we denote by $H$, to be a function of both family socio-economic background, $F$, and expected labour market payoff, $E$. All three variables $H$, $F$ and $E$ can carry two indexes: an age and a time index. To see how the education level of individuals in a given age bracket evolves over time, one needs to fix the age index, $g$, and let the time index, $t$, vary. Assuming a constant distance between child's and parents' age, we can characterize both $H$ and $F$ using the same age index, $g$. Yet, the observed education attainment $H$ is the result of a decision process that occurred in the past, at an age $g^0 < g$, the age when the education attainment is measured. As such, the information set available when formulating expectations for future income should correspond to the decision relevant age, i.e. $g^0$. The following equation represents the basis of our empirical strategy outlined in the next section:

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

where $f$ is a function capturing the stylized model formulation above.

## 3. Empirical strategy

### 3.1. Model specification

Since we are interested in modelling the dynamics of education attainment for a reference population cohort, we can replace the time index by a time subscript and write the empirical counterpart of (1) as:

$$H(g)_{c,t} = \alpha_c + \beta_c * t + \gamma * F(g)_{c,t} + \lambda * E\left(g^0\right)_{c,t} + \varepsilon_{c,t} \tag{2}$$

where $H(g)_{c,t}$ represents the population share of young individuals in the age bracket $g$ having completed 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 $E\left(g^0\right)_{c,t}$ is the expected labour market payoff. The coefficient $\alpha_c$ is a country-specific constant term meant to capture stable institutional factors affecting education attainment over time while $\beta_c$ is the country-specific time trend, summarizing any consistent and gradual institutional improvements. $\varepsilon_{c,t}$ is a country- and time-specific error term.

Our reference population includes all individuals aged 30–34, thus effectively identifying age $g$ in Eq. (2). However, $g^0$ being relevant for the decision-making process remains unobservable. We assume that education choices are based only on the set of information available at the decision-making age, $g^0$. Within our 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 Eq. (2), we get:

$$H'_{c,t} = \alpha_c + \beta_c * t + \gamma * F'_{c,t} + \lambda * E'_{c,t-s} + \varepsilon'_{c,t} \tag{3}$$

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. All variables are measured at time $t$, except $E'_{c,t-s}$ which is the $s$ time-lag of the proxy for expected return (where $s = g - g^0$).

When trying to estimate (3) directly, one needs to address the potential time dependence of the endogenous variable $H'_{c,t}$. This is necessary since the tertiary education attainment benchmark spans over 5 consecutive cohorts, covering all individuals aged 30–34. In this context, an alternative would have been to estimate a dynamic panel, i.e. using difference or system GMM as proposed by Arellano and Bond (1991), Blundell and Bond (1998). Unfortunately, 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 adverse consequences on the forecasting exercise.

In addition, Eq. (3) cannot be rigorously estimated if the data are non-stationary. Unfortunately, univariate and multivariate unit root tests 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.[1] To tackle the potential non-stationarity in the data, we prefer to specify the model in first-differences:[2]

$$\Delta H'_{c,t} = a + \beta_c + \gamma' * \Delta F'_{c,t} + \lambda' * \Delta E'_{c,t-s} + e_{c,t} \tag{4}$$

where $\beta_c$ is now interpreted as a country-specific constant term, $a$ is a regression constant and $\Delta$ is the first-difference operator. Eq. (4) sets out the dynamics of tertiary education attainment as a combination of consumption and investment motivations, in line with our discussion in Section 2.

Our dataset lacks specific information on households and individuals, making it impossible to identify family bonds. In our setup, individuals belong to a *synthetic* family composed of both adults and young individuals, matching as closely as possible a typical parental relation.[3] 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 Eq. (4). We then proxy expected labour payoff using a set of different but related variables, i.e. (i) labour productivity, (ii) total factor productivity (TFP), (iii) employment rate and (iv) real compensation per employee.[4] All these indicators are proxies for the skill premium, so their relationship with tertiary education attainment is expected to be positive in all cases, except for employment rate.[5]

[1] Cross sectional dimension N equals at most 27 and time dimension T is 13 annual observations. Asymptotic properties of panel unit root tests require $N/T \to 0$, 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.

[2] 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.

[3] 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.

[4] Additional robustness checks also employed per capita GDP, overall and skill-specific unemployment rate, skill-specific employment rate, but none of them were found significant and robust to different model specifications.

[5] Both labour productivity and TFP are good proxies to reflect changes in the wage distribution and the skill premium (see Autor et al., 1998; Acemoglu, 2002). In addition, employment rate and real compensation per employee relate to labour market developments observed over the last decades: changes in real compensation per employee are a proxy the expected payoff in the higher-skilled part of the employment distribution, while changes in employment rate are a proxy for the expected payoff in the lower-skilled part of the distribution.

Heterogeneity issues are 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 doing so, the estimation approach is based on the Common Correlated Effects (henceforth CCE) method advanced by Pesaran (2006). This method 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 & Tossetti, 2011). We can then estimate a model of the form:

$$\Delta H'_{ct} = a + \beta_c + \gamma' * \Delta F'_{c,t} + \lambda' * \Delta E'_{c,t-s} + \varphi * \overline{\Delta H'}_{ct} + \delta * \overline{\Delta F'}_{c,t} + \theta * \overline{\Delta E'}_{c,t-s} + e'_{ct} \tag{5}$$

where the last three terms on the right-hand side are the cross section averages of the variables included in Eq. (4). Thereby, we are able to account for cross-correlations, which might stem from Europe-wide policy recommendations and/or reforms, such as those included in the Bologna process.

### 3.2. Model selection and specification tests

We use data on tertiary education attainment for the 2000–2012 period from Eurostat, disaggregated by country, gender and age groups. Most economic indicators, including labour productivity, total factor productivity (TFP), real compensation per employee are drawn from AMECO database of the European Commission, except for the employment rate which is taken from Eurostat, due to better data availability.

To provide long-term forecasts for the Europe 2020 benchmark on tertiary education attainment, it is important to rely on simple but robust model specifications, supported by a clear theoretical framework. We consider a battery of model selection and specification tests. 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 proposed by the literature (Arellano & Bond, 1991; Baltagi & Wu, 1999; Wooldridge, 2002); third, we check coefficients stability by varying the estimation time-sample; fourth, we evaluate forecasting accuracy using out-of-sample root mean square errors (RMSE) 1–4 years ahead.[6]

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 that excludes gross outliers from the sample.

Residuals autocorrelation tests are extremely important and we discuss them carefully below. Autocorrelation may originate from the nature of the education attainment variable being measured over consecutive population cohorts. First-differentiating can mitigate to some extent the presence of lower order autocorrelation, but it might not affect higher order autocorrelation. For example,

[6] 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 population shares (as of 2012), 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 would not allow the computation of comparable RMSEs for all periods and all model specifications.

assume no inward/outward migration flows that might alter the composition of the population cohorts, except ageing. Taking first-difference of tertiary educated individuals aged 30–34 might produce autocorrelation at a higher order. In particular, every 5th lag of $\Delta H'$ could produce almost perfect negative autocorrelation in this setup. Other sources of residual autocorrelation might, for example, arise in the presence of siblings. To dismiss all these potential possibilities we provide a number of autocorrelation tests. We check for first-order autocorrelation using Baltagi and Wu (1999) and Wooldridge (2002) tests which are shown to be robust in small samples; we then move from 1st-order up to a maximum of 7$^{\text{th}}$-order autocorrelation using the Arellano and Bond (1991) test, making sure all tests provide consistent results for the 1st-order case.

Coefficients' stability is relevant when forecasting models need to be updated to include new data. Model misspecification, for instance as a result of omitted variables or breaks in the series, would generally undermine such a property. We recursively estimate Eq. (4) over different time periods, gradually expanding the initial sample with new observations until we reach the full available sample. For each model we carry out all the tests mentioned above so as to guarantee that our criteria are met.

## 4. Econometric results

### *4.1. The preferred model specification*

Following the testing procedure indicated in Section 3.2, our preferred model specification is a model estimated on a sample of 12 EU countries, namely Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the U.K. Together, these countries represent 76% of total EU27 population and generate 88% of its GDP (as of 2012).

Table 1 reports the model's estimated coefficients along the specification given by Eq. (4). The best specification includes the share of adults with high education and labour productivity[7] as regressors. Model (1) also controls for the major breaks in tertiary education attainment time-series that occurred around 2003 in most EU countries. Model (2) includes a vector of country dummies that controls for mean differences in tertiary education attainment. In model (3), we add a vector of year dummies that takes into account shifts in higher education attainment common to all countries. Last, model (4) in Table 1 shows the estimated CCE specification according to Eq. (5).

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 controlling for country and year specific factors, 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, Chevalier, Denny, & McMahon, 2003; Chevalier, Harmon, O'Sullivan, & Walker, 2013; Hertz et al., 2007).

Higher labour productivity, our proxy for expected returns, leads to more schooling. A 1% rise in productivity causes an increase in tertiary education attainment by an amount of 0.6%. The 13th lag of labour productivity when subtracted from the age of the reference population, i.e. 30–34,

[7] See also Restuccia and Vandenbroucke (2014), who recently proposed a model that explains education attainment over time and across countries by relying mainly on labour productivity (along with life expectancy).

Table 1
Estimates of the log share of tertiary educated individuals aged 30–34.

| Model | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 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 |
| *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 coefficient 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*: OLS estimates. Huber-White robust standard errors in parentheses allow for arbitrary correlation of residuals within each country. Model (1) includes dummies for 2003 and 2004 to control for the presence of breaks in time-series. Model (2) adds country dummies that capture country-specific trends; model (3) includes year dummies. Model (4) factors in cross-section averages of the dependent and explanatory variables. Recursive coefficient estimation summary info 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 indicated test for all sub-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 were considered; the "Ok/No" labels summarize the results from all seven test statistics on all sub-samples.

* $p<0.10$.

** $p<0.05$.

*** $p<0.01$.

hints at an interesting overlap with the decision time on enrolment (and/or graduation). The long lag should allay any concerns about potential reverse causality between productivity and higher education attainment in the model.

Overall, Table 1 shows no significant change in the size of estimated coefficients across the four model specifications. Our findings seem to suggest that higher education attainment is more responsive to 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 larger extent by income motivations (Winchester & Greenaway, 2007).

### 4.2. Further robustness checks

To exclude the possibility that our results are driven by some specific countries, we re-estimate model (1) from Table 1 by leaving out one country at the time. Results are displayed in Fig. 1, 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 (dashed lines).

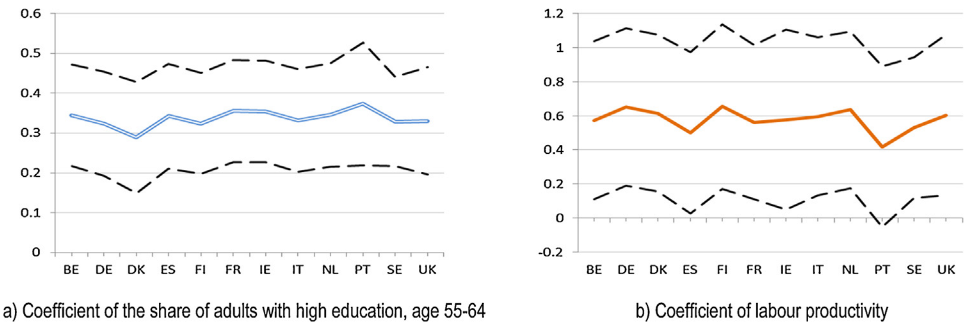


Fig. 1. Check of coefficients' stability in the preferred model specification, excluding one country at a time. *Note*: OLS regression coefficients. 95% confidence intervals refer to coefficients shown in Table 1, column 1. Each coefficient is from a separate regression taking out one country (indicated on the *x*-axis) at a time. For Fig. 1 (b), the 90% confidence interval excludes the zero line in case Portugal is removed.

Fig. 1 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.

To demonstrate that the interpretation of the results is not sensitive to the chosen proxies of parental background and expected payoff, we use a different set of explanatory variables. Table 2 shows the results of such an exercise. For readers' convenience, our preferred specification is reported under the first column. 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[8] along with TFP, employment rate and real compensation per employee as proxies for expected return. Note that models (7), (8) and (9) are estimated by considering different sets of countries as a consequence of closely following the approach described in Section 3.2.

As Table 2 clearly reveals, we find statistically significant coefficients of similar magnitude, with only few exceptions. For models (8–9), the coefficient of the share of tertiary educated adults almost doubles compared to the best model specification showed in the first column; however these models do not pass autocorrelation tests. Note the negative coefficient for employment rate in model (8): individuals choose more education due to fears of unemployment risks associated with the low-skill part of the employment distribution. In model (9), individuals prefer the higher real wages that accrue to the high-skill group.

For the readers' convenience, we plot the recursive coefficients estimated for each model specification in Fig. 2. Models (1) and (5) show high coefficient stability and have very similar properties; however, the former specification has smaller RMSE at long horizons – one of the most common criterion in assessing forecast accuracy. Overall, we conclude that our preferred model specification (1) has superior forecast accuracy and also successfully passes the coefficients' stability check.

## 5. Forecasts for tertiary education attainment and policy implications

To provide an answer to our initial question whether the EU27 as a whole will be able to reach the target of 40% tertiary education attainment by 2020, we need two additional assumptions. First,

[8] 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.

Table 2
Estimates for the log share of tertiary educated individuals aged 30–34.

| Model | (1) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|
| Δ log (adults' share high education, 55–64) $t$ | 0.34*** (0.09) | | | 0.35*** (0.09) | 0.57*** (0.17) | 0.65*** (0.15) |
| 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** (0.23) | 0.65*** (0.21) | 0.39* (0.24) | | | |
| 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 (0.60) | 0.25 (0.63) | 1.32* (0.73) | 1.23* (0.65) | 2.34*** (0.68) | −0.11 (0.75) |
| Observations | 144 | 144 | 144 | 132 | 201 | 168 |
| *Adj.* $R^2$ | 0.30 | 0.27 | 0.22 | 0.29 | 0.38 | 0.50 |
| *AIC* | 776.3 | 782.5 | 792.5 | 717.7 | 1232 | 1013 |
| *BIC* | 791.1 | 797.3 | 807.3 | 732.1 | 1242 | 1028 |
| Control for breaks in series | Yes | Yes | Yes | Yes | – | Yes |
| Panel | Balanced | Balanced | Balanced | Balanced | Unbalanced | Balanced |
| # of countries | 12 | 12 | 12 | 11 | 27 | 14 |
| Recursive estimation summary | | | | | | |
| Wooldridge test for AR(1) | Ok | Ok | Ok | Ok | No | No |
| Arellano–Bond test for AR(1)–(7) | Ok | Ok | Ok | Ok | No | No |
| Out-of sample RMSE statistics (≤27)[a] | | | | | | |
| 1 year ahead | 1.38 | 1.38 | 1.43 | 1.36 | 1.32 | 1.86 |
| 2 years ahead | 2.23 | 2.23 | 2.22 | 2.32 | 2.26 | 3.00 |
| 3 years ahead | 2.94 | 2.98 | 3.00 | 3.04 | 2.91 | 3.89 |
| 4 years ahead | 3.54 | 3.61 | 3.66 | 3.53 | 3.49 | 4.92 |
| Out-of sample RMSE statistics (=5)[b] | | | | | | |
| 1 year ahead | 0.66 | 0.71 | 0.86 | 0.67 | 0.79 | 0.72 |
| 2 years ahead | 1.08 | 1.15 | 1.42 | 1.10 | 1.48 | 1.13 |
| 3 years ahead | 1.49 | 1.59 | 2.04 | 1.56 | 2.41 | 1.54 |
| 4 years ahead | 1.87 | 2.02 | 2.72 | 1.99 | 3.37 | 1.90 |

*Note*: OLS estimates. Huber-White robust standard errors (in parentheses) allow for arbitrary correlation of residuals within each country. All models include dummies for 2003 and 2004 to control for the presence of breaks in time-series. Recursive estimations summary info refers to the estimation of the same model specification over expanding sub-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.

[a] The RMSE is the equally weighted average of country-specific RMSE. The 1 year ahead RMSE is computed using 6 observations, 2 years ahead RMSE – using 5 observations and so on.

[b] The RMSE is the equally weighted average of the top 5 most populated countries' RMSEs (as of 2012). These countries are: Germany, UK, France, Italy and Spain; together, these 5 countries comprise around 63% of EU27 total population.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

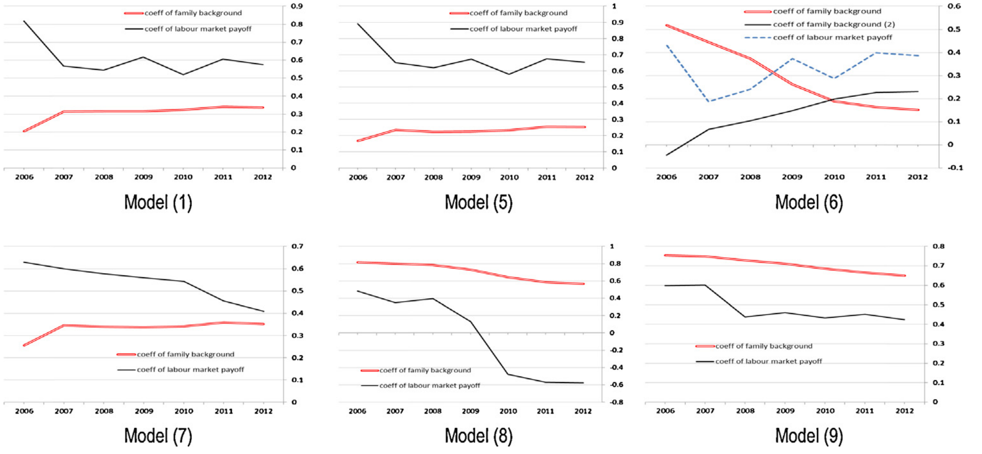


Fig. 2. Coefficient stability check for models of Table 2. *Note*: Recursive OLS coefficients of models in Table 2 are given. The *x*-axis displays the end-point of each estimation sub-sample. The recursive estimation starts with the first sub-sample i.e. 2001–2006, and ends with the 2001–2012 sub-sample.

we assume that the EU27-aggregate behaves according to the best model specification discussed in Section 4.1; with no country-specific terms included, this assumption should be straightforward. Second, the 12 countries in the best model specification provide a good approximation of the EU27-aggregate mainly because they represent a major 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.

We build the EU27-aggregate forecast in two steps. In the first step, we need EU27 projections for the two determinants, i.e. adult education attainment and labour productivity. For the first determinant, we adopt a simple extrapolation method based on replicating an ageing process to obtain projections for the share of 55–64 year-olds with tertiary education over the period 2013–2020 (see Appendix A for details on the extrapolation method). For the second determinant, the long lag used in the estimation allows us to generate forecasts up to 2020 using readily available data.[9] In the second step, we use the model coefficients to compute the expected change in the share of tertiary educated 30–34 year-olds up to 2020.

Country-specific forecasts can be constructed following the same two steps described above. One major 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 improvements in education attainment would have an underestimated forecast based on a model that omits the country-specific trends from Eqs. (2) and (3) or, equivalently, the country-specific terms from Eq. (4). The first argument against this interpretation is that such model specification did not pass the required auto-correlation tests. By not including country-specific factors in the preferred model specification, extending the model to other countries becomes straightforward. A second argument relies on alternative forecasts derived from a cohort-based approach, which uses country-specific data on tertiary enrolment, duration of studies and completion rates (for more details, see Dragomirescu-Gaina & Weber, 2013). Built on a different set of assumptions, this approach is completely

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

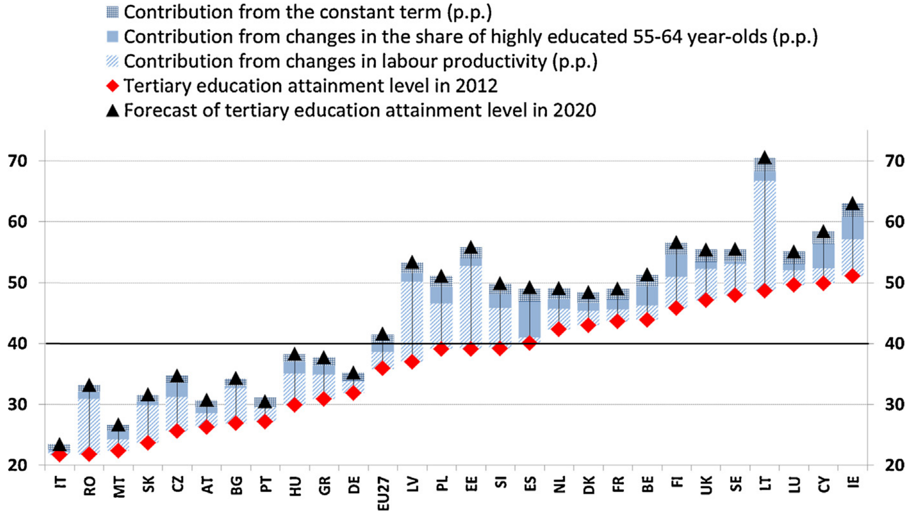


Fig. 3. Decomposition of the country-specific forecasts. *Note*: Countries are ordered according to their 2012 benchmark value on the *x*-axis. Benchmark's value is shown on the *y*-axis. The red tilted square denotes the share of tertiary educated 30–34 year-olds in 2012, i.e. the baseline year. The black arrow shows the expected value in 2020. The blue shaded areas denote the cumulated contributions of the three model components, i.e. constant term, changes in labour productivity and changes in adults' education attainment level. The horizontal black line displays the 40% EU target set for 2020. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

independent from the forecasting approach described in this paper. When comparing the two sets of forecasts we find no evidence that the forecasts based on the econometric model are systematically biased due to the omission of country-specific characteristics.

Fig. 3 summarizes the outcome of our forecasting exercise, including country-specific results. While the EU as a whole is indeed likely to reach the 40% target in 2020, this progress masks a vast heterogeneity of countries' progress. As of 2012, the group of countries on the right side of the chart are already beyond the 40% threshold, but the group of countries on the left side substantially lag behind. Moreover, when looking at the expected progress by the two country groups, it becomes clear that the divergence we observe with respect to 2012 values is also likely to persist in the future. Our calculations show a slower expected progress for those countries currently lagging behind, and a faster expected progress for the high performing countries. In particular, 11 countries are not expected to reach the 40% threshold, namely Italy, Romania, Malta, Slovakia, Czech Republic, Austria, Bulgaria, Portugal, Hungary, Greece and Germany; some are not even expected to reach their often less ambitious national targets.[10]

The graph additionally decomposes the predicted progress until 2020 in its main components: a contribution from constant (or structural time-invariant factors), a contribution from changes in labour productivity and a contribution from changes in adults' education level. In particular, across the EU, the contribution of adults' education level is substantially lower than the contribution from labour productivity. This means that progress on tertiary education is mainly driven by increased incentives arising from the labour market rather than by improved educational background of the parents. The exceptions are Spain, as well as Cyprus and Belgium. For those countries most of

[10] Italy (national target: 26–27% versus forecast: 23%), Malta (33% versus 27%), Slovakia (40% versus 24%), Bulgaria (36% versus 34%).

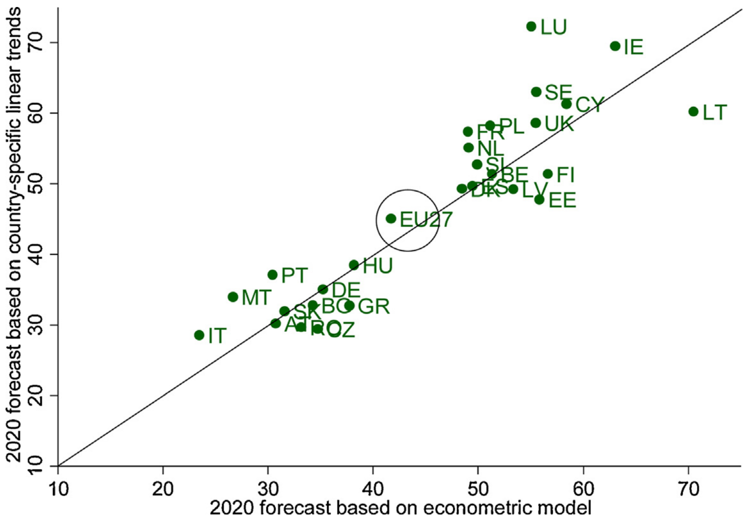


Fig. 4. Forecast comparison: linear trend versus econometric model.

the progress is driven by family background, and thus these countries seem better positioned to benefit from past investments in the education system. On the other hand, for Germany, parents' education does not contribute to any extent to the improvement in tertiary education, while the impact is even negative for Portugal.

It is noteworthy that most of the countries with particularly fast predicted progress, such as Romania, Latvia, Estonia and Lithuania have also a high percentage of this progress explained by labour productivity. In those countries, soaring productivity, mostly driven by improved institutions and wide structural reforms implemented during their catching-up process, has provided a strong push for the young generations to aspire tertiary education and to reap some of the benefits offered by these fast developing labour markets.

However, given the complex and bi-directional causality links between economic dynamics and education attainment (see Asteriou & Agiomirgianakis, 2001; Bils & Klenow, 2000), our country-specific forecasts also provide a weak support for real convergence prospects over the coming years. The current stall in the economic convergence process within the EU raises additional challenges, because it also exposes a divide between countries in terms of expected education attainment. This should pose a major policy challenge for the EU as a whole, given that human capital is the most important driver of innovation, the long-term engine of real economic convergence. The divergence in productivity dynamics within the EU has been one of the main roots of the current economic crisis. However, our model shows that past economic dynamics are likely to snowball due to their future consequences on education attainment that might nourish renewed economic divergence, as in a vicious spiralling process.

On a more positive note, Fig. 4 compares the forecasts built on our econometric model with the naïve forecasts derived using country-specific linear trends. Interestingly, within the group of 'leaders', there seem to be more countries with a naïve forecast higher than the one generated by the econometric model, while the opposite holds true for the group of laggards. While this is merely a comparison between two forecasts, the model-based forecasts seem to entail less divergent patterns between the two country groups. We speculate that a better outcome for EU might be achieved by improving along the mechanisms suggested by the econometric model, instead of simply replicating past performances.

## 6. Concluding remarks and discussion

The European Commission is committed to increase tertiary education attainment to at least 40% 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. We show that our empirical model meets important features such as robustness, stability, good out-of-sample forecasting properties, and is thus capable to provide insightful long-term forecasts. All estimated model specifications provide a consistent story in line with the predictions of theoretical models that perceive education both as 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 (see Winchester & Greenaway, 2007).

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. Due to the complex causality links between economic dynamics and education attainment, our forecasts provide a weak support for real economic convergence prospects over the coming years. Therefore, the current stall in the real convergence process within the EU raises additional challenges. Our forecasts are mainly driven by demographics and past economic dynamics, leaving a narrow space for effective policy adjustment before the year 2020. The widening gap between the two groups has better chances of being reversed only after 2020, but strong policy actions might be needed today along the main transmission mechanisms outlined by the model.

The paper also indicates relevant policy areas where improvements could be efficient. A frictionless labour market may simplify the way individuals anticipate income distribution and skill premium (Buchinsky & Leslie, 2010), which feeds back into their education decisions. Labour market reforms, for example, might have important indirect effects, by strengthening the positive spill-overs associated with skilled workers, education decisions, skill premium dynamics and technological progress. Our empirical findings clearly expose this dynamic link, highlighting the less noticeable reverse causality channel that runs from economic developments to education decisions. In fact, most of the new member states, which experienced major institutional reforms in the run-up to EU accession, illustrate the broad benefits of a process where better economic outcomes and higher education attainment strengthen each other, as in a virtuous cycle.

Increasing participation and broadening access to higher education can also have positive externalities to future cohorts through intergenerational transmission of education choice. However, strengthening this channel might raise a discussion on the role of education as an opportunity ‘equalizer’. Instead, well-designed retraining and life-long learning programmes, especially for low-skilled, might prove effective in raising the average education level of adult cohorts and thus affect youth education attainment in the end.

## Acknowlegements

We are thankful to the editor (A. M. Costa) and the anonymous referees of this journal for their comments that have significantly improved our paper. We are also grateful to Andrey Vasnev and Stelios Karagiannis for their useful comments. Part of this research project was financed under

CRELL VI AA. The responsibility for any remaining errors or omissions is our own. The view expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

## Appendix A. Projecting the education attainment of adult population cohorts

This appendix provides details on the country-specific projections of adults' education attainment up to 2020. Disregarding (i) inward and outward migration flows, (ii) mortality risk differential between individuals with different education attainment and (iii) assuming no change in the education attainment of adult population cohorts (e.g. lack of life-long learning, no re-training, etc.), we can write the evolution of education attainment for the adult population cohort 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 indexed by $g$. The two age groups $g$ and $g-10$ can be interpreted as referring to different population cohorts, separated by 10 years. The equation above is the formalization of a simple ageing process viewed over time, from the perspective of a single population 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 (inward/outward migration seems to be the strongest one). We allow for its gradual phasing out at the rate given by $\lambda$, arbitrarily set at 0.5. Using available data for adults' education attainment between 2000 and 2012 in the age brackets: 25–34, 35–44, 45–54, 55–64, gaps might be calculated for 2010, 2011 and 2012; however, we only need the 2012 gap value to construct an exogenous path for adults' education attainment according to equation above.

## 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., & Bond, S. (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). Milan: Bocconi University.

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*. Chicago: NBER.

Becker, G. S., & Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, *87*, 1153–1189.

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labour Economics*, *4*(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.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, *75*(4), 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 O. Ashenfelter, & D. Card (Eds.), *Handbook of labour economics* (Vol. 4) (pp. 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.

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*.

Chevalier, A., Denny, K., & McMahon, D. (2003). A multi-country study of inter-generational educational mobility. In P. Dolton, R. Asplundh, & B. Erling (Eds.), *Education and inequality across Europe*. London: Edward Elgar.

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.

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

Dragomirescu-Gaina, C., & Weber, A. (2013). *Forecasting the Europe 2020 headline target on education and training. A panel data approach. EUR 26077 EN*. Luxembourg: Publications Office of the European Union. JRC83346.

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 Pt. 2), S11–S44.

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.

Li, G. (1985). Robust regression. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes* (pp. 281–340). New York: Wiley.

Mincer, J. (1974). *Schooling experience and earnings*. NY: NBER Columbia U. Press.

Oosterbeek, H., & Webbink, D. (1995). Enrolment in higher education in The Netherlands. *De Economist*, *143*(3), 367–380.

Pascual, M. (2009). Intergenerational income mobility: The transmission of socio-economic status in Spain. *Journal of Policy Modeling*, *31*(6), 835–846.

Pesaran, M. Hashem. (2006). Estimation and inference in large heterogenous panels with a multifactor error structure. *Econometrica*, *74*(4), 967–1012.

Pesaran, M. H., & Tossetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, *161*(2), 182–202.

Restuccia, D., & Vandenbroucke, G. (2014). Explaining educational attainment across countries and over time. *Review of Economic Dynamics*, *17*(4), 824–841.

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.

Winchester, N., & Greenaway, D. (2007). Rising wage inequality and capital-skill complementarity. *Journal of Policy Modeling*, *29*(1), 41–54.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.