McGuinness, Seamus; Bergin, Adele; Whelan, Adele

Working Paper

Overeducation in Europe: Trends, Convergence and Drivers

IZA Discussion Papers, No. 10678

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: McGuinness, Seamus; Bergin, Adele; Whelan, Adele (2017) : Overeducation in Europe: Trends, Convergence and Drivers, IZA Discussion Papers, No. 10678, Institute of Labor Economics (IZA), Bonn

This Version is available at:
http://hdl.handle.net/10419/161301

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10678

Overeducation in Europe: Trends, Convergence and Drivers

Seamus McGuinness
Adele Bergin
Adele Whelan

MARCH 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10678

Overeducation in Europe: Trends, Convergence and Drivers

Seamus McGuinness
ESRI, Trinity College Dublin and IZA

Adele Bergin
ESRI and Trinity College Dublin

Adele Whelan
ESRI and Trinity College Dublin

MARCH 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Overeducation in Europe: Trends, Convergence and Drivers*

This paper examines patterns in overeducation between countries using a specifically designed panel dataset constructed from the quarterly Labour Force Surveys of 28 EU countries over a twelve to fifteen year period. It is not the case that overeducation has been rising rapidly over time in all countries and where overeducation has grown the trend has been very gradual. Furthermore, overeducation rates were found to be static or falling in approximately fifty percent of the 28 EU countries. The evidence points towards convergence in overeducation at a rate of 3.3 percent per annum. In terms of the determinants of overeducation we find evidence to support policies aimed at improving effective female participation, labour market flexibility and the practical aspects of educational provision as a means of reducing the incidence of overeducation within countries.

JEL Classification: I2, C23
Keywords: overeducation, dynamic panel data

Corresponding author:
Seamus McGuinness
Economic and Social Research Institute
Whitaker Square
Sir John Rogerson's Quay
Dublin 2
Ireland
E-mail: Seamus.McGuinness@esri.ie

* The research leading to these results has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613256.

Introduction

Overeducation describes the situation whereby individuals are employed in jobs within which the level of education required to either get, or do, the job in question is below the level of schooling held by the worker. Since the 1990s the literature on overeducation has ballooned with a number of stylized facts emerging including the association of overeducation with lower earnings and job satisfaction, relative to matched workers with equivalent education, and higher rates of job mobility although not necessarily to improved job matches (see Quintini, 2011; McGuinness, 2006 for reviews). It is these negative impacts coupled with growing evidence that overeducation tends to be persistent over time (Dolton and Vignoles, 2000; Rubb, 2003; Frenette, 2004; and McGuinness and Wooden, 2009) that makes the issue of concern from a policy perspective. However, while there are now many studies that examine both the determinants and impacts of overeducation at a national level, there is very limited research aimed at understanding overeducation from a cross-national perspective. The lack of comparative international research on the issue of overeducation is, arguably, due to a lack of available datasets that measure overeducation across countries over time. This paper aims to address this substantial gap in the literature by developing a time-series dataset for 28 European countries in order to assess variations in both levels and trends in overeducation within European countries, the degree of convergence in overeducation rates over time and the key factors explaining country level variations in overeducation rates in Europe.

The results show that it is not the case that overeducation is consistently rising across all European countries, in fact we observe positive trends in less than half the countries in the sample with overeducation remaining either constant or falling in the majority cases. Consistent with country level studies we find that overeducation is higher among females in the vast majority of countries. Looking at the direction of trends at a more aggregated level we find that while the average trend in overeducation across all 28 countries appears to be relatively stable over the period 2003 to 2013, substantial differences do exist depending on the geographical country block. Overeducation rates tend to be highest and most volatile over time in peripheral European countries, while overeducation in central European countries tends to be lower and appears to follow a somewhat cyclical pattern. Overeducation is consistently lowest and stable over time in eastern European countries. The study finds evidence of ongoing convergence in overeducation, whereby countries with the lowest initial values of overeducation tend to experience the highest growth rates over time. Finally, in terms of the factors driving cross country differentials, factors relating to both the composition and level of labour demand, labour supply and the structure of educational provision all appear important.

Existing Comparative Literature

Due mainly to data constraints there has been relatively little assessment of overeducation from an aggregate country level perspective; however, some exceptions do exist. The most comprehensive assessment to date comes from Verhaest & van der Velden (2012) who use a

multi-level model to explain cross-country variations in the incidence of graduate overeducation at a single point in time. Explanatory variables in the Verhaest & van der Velden (2012) study include measures for the composition of higher education supply in terms of both vocational versus academic orientation and field of study, proxies for educational quality, [i] measures of the output and unemployment gaps, [ii] indicators of employment protection legislation within each country and the level of educational over-supply. The model also includes controls for the share of graduates in the over 25 population and gross expenditure on R&D. Graduate over-supply is then taken as the difference in the standardized values of these two variables. Verhaest & van der Velden (2012) find that cross-country differences in overeducation were related to their measures which, they argue, capture variations in quality and orientation (general versus specific) of the educational system, business cycle effects and the relative oversupply of highly skilled labor. Croce and Ghignoni (2012) use annual data from the European Community Household Panel (ECHP) to examine differences in graduate overeducation in 26 European countries between 1998 and 2003. Based on a samples of between 80 and 100 country level observations, Croce and Ghignoni (2012) report that overeducation tends also to be influenced by business cycle variables and tends to be higher in countries with a lower wage gap between graduates and workers with upper secondary education. Davia, McGuinness and O' Connell (2017) attempt a similar exercise using EU-SILC data in order to explain regional variations in overeducation rates across 28 European countries between 2004 and 2009. Similar to Verhaest & van der Velden (2012), Davia et al. (2017) find evidence to support the notion that overeducation is higher in areas where the level of educated labor supply exceeds demand and where university enrolment levels are greatest. Davia et al. (2017) also report that the overeducation rate is positively related to the share of migrants in the labor market and is lower for females in regions with strong employment protection.

Thus, while some concerns may be raised regarding the quality of some of the indicator variables derived in studies relying on cross-sectional international data or very short panels, the existing literature points to potential importance of aggregate level variables in explaining international differences in overeducation. Studies to date indicate that factors such as the composition and level of labour demand, educational provision and business cycle effects may play some role in explaining spatial differences. Nevertheless, the limited data available to researchers to date has resulted in international variations in overeducation being assessed over a relatively narrow period of time; furthermore, no study has been able to effectively deal with issues of country level unobserved heterogeneity or potential endogeneity in attempting to identify causal relationships. This study aims to address these deficits in the literature in order to provide the first robust combined assessment of both the evolution of overeducation within Europe and the factors determining observed international differences.

Data and Methods

There are no reliable time-series data in existence on European overeducation to allow a systematic cross-country comparison across time and, consequently, the data development aspect is a key contribution of the current study. The data used in this study is the quarterly anonymized country level files of the European Union Labour Force Study (EU-LFS) for the period Q1 1998 up to Q4 2012. In terms of the key metric, there are essentially three standard methods of measuring overeducation, the most commonly used is a subjective measure based on individual responses comparing their attained education levels with their perceived job entry requirements. Second in popularity is an objective approach, termed the empirical method, that compares individual levels of schooling with either the mean or mode level of schooling of their respective occupation usually measured at a 2 digit level. The third, least commonly applied, approach is an occupational dictionary method that compares individual level education with the required level of schooling detailed for specific occupations in the documentation accompanying occupational classification systems. Existing studies indicate that while the correlation between the various definitional approaches tends not to be particularly high, they generate very similar results with respect to both the incidence and impacts of overeducation (Sloane 2003, McGuinness, 2006).

As there exists no subjective overeducation question within the EU-LFS overeducation is measured objectively using the empirical method. For each country, in each quarter, overeducation is defined as the proportion of employees in employment whose ISCED level of schooling lies one level or more above the occupational mode. The occupational modal level of education is the most common qualification possessed by workers in each two-digit occupation group. Overeducation is calculated within two-digit occupational codes and using five ISCED categories of <2, 3, 4, 5B and 5A+6. Thus, if the modal level of schooling in a particular two-digit occupation was measured at ISCED 3, then all individuals educated to ISCED levels 4 and above would be deemed to be overeducated in our approach. We then use the individual level metrics to calculate the overall rate of overeducation in each country for each quarter and we also calculate rates separately for males and females. Given that we are dealing with a large number of countries, for the purposes of our analysis we group these into three categories on the basis of common linkages in terms of geographical proximity, levels of economic development and access to the single market. The first category comprises the countries that acceded to the EU from 2004 which include Bulgaria, Czech Republic, Estonia, Hungary, Lithuania, Latvia, Poland, Romania, Slovenia, and Slovak Republic and are referred to as the 'Eastern' states. The second category refers to Portugal, Ireland, Italy, Greece and Spain, the traditional 'Periphery' of the EU. The third group ('Central') comprises the remaining countries located in central and northern Europe and includes Austria, Belgium, Denmark, Finland, France, Iceland, Luxemburg, Netherlands, Norway, Sweden and the UK.[iii] Generally, we found that the average rate of overeducation is lowest in the Eastern European countries, highest in the Periphery and the average rate lies somewhere in between in the Central European countries (see descriptive evidence in the next section).

In addition to describing long-run trends in overeducation rates both within and between countries, we also examine the extent to which overeducation rates in Europe have been converging or diverging over time by estimating a Barro regression (Equation 1) (Barro, 1997). The Barro model examines the relationship between the growth rate of overeducation and the initial level of overeducation using a regression model. If countries with lower initial levels of overeducation tended to have a higher growth in overeducation over time, then the estimate of the coefficient of interest (β_1 in equation 1) would be negative and significant implying convergence. In contrast, a positive coefficient would point towards divergence in overeducation rates across countries. In addition to the Barro regression, we also check for convergence by plotting the cross-country variance in overeducation rates for specific groups of countries. Arguably, we might expect overeducation rates to converge over time as workers from saturated graduate labor markets relocate to areas with greater levels of job opportunity and lower levels of overeducation.

$$\frac{\ln Ov(t) - \ln Ov(0)}{t} = \beta_0 + \beta_1 \ln Ov(0) + \varepsilon \qquad (1)$$

Finally, we examine the determinants of cross-country variations in overeducation for countries within a panel estimation framework and the general relationship is:

$$y_{it} = \beta_0 + \beta_j X_{ijt} + \alpha_i + \varepsilon_{it} \qquad (2)$$

where y_{it} is the dependent variable observed for country *i* in time *t*, β_0 is a constant term, X_{ijt} represents a number of j independent variables with β_j the associated coefficients, α_i is the unobserved time-invariant country effect and ε_{it} the error term. In terms of the specific panel modeling approach adopted, we opt for a fixed effect estimator that allows us to model the determinants of overeducation while controlling for time invariant country level fixed effects. One complication that does arise with our data is that our dependant variable runs from zero to one, therefore, a standard panel regression will generate predicted values that lie outside the 0 to 1 bounded scale. The problem is overcome within a cross-sectional framework using a non-linear fractional logit model developed by Papke & Wooldridge (1996) with the same authors then extending and extending the approach to panel data in a later study (Papke & Wooldridge (2008)). Unfortunately, the panel fractional logit procedure is only currently available for the random effects estimator and in order to control for the impacts on time-invariant country specific unobserved heterogeneity we must augment the model with a Mundlak correction (Mundlak, 1978) which causes the model to approximate a fixed effects specification.[iv] The model to be estimated can be written as:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 \overline{X_{it}} + \alpha_i + \varepsilon_{it} \qquad (3)$$

Where `the Mundlak correction is operationalised through the inclusion of $\overline{X_{it}}$, which denotes the time averaged means of the independent variables in the models. However, within the fixed effect framework we cannot exclude the possibility that some of the right hand variables may be endogenous, particularly in case of controls related to labour market participation where the direction of causality could potentially work in either direction. In order to overcome this problem we adopt dynamic panel models using the GMM estimator developed by Arellano & Bond (1991) which accounts for both fixed effects and endogeneity. This is achieved by first differencing the model, which removes the fixed effect, and using lags to instrument the explanatory variables in the model[v]. Adopting this dynamic framework equation 2 becomes:

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \beta_j \Delta x_{ijt} + \Delta \varepsilon_{it} \qquad (4)$$

A drawback of the GMM model outlined above is that easily interpretable marginal effects are difficult to extract. Our general empirical approach is to use the Mundlak corrected random effects fractional logit model to measure the marginal effects of those variables indicated as being statistically significant within the more robust GMM estimation framework. In terms of our explanatory variables, these are chosen to reflect potential demand and supply-side factors that may potentially drive overeducation. Overeducation could arise due to the supply of educated labor outstripping demand, primarily as a result of the tendency of governments in developed economies to continually seek to raise the proportion of individuals with third-level qualifications. Alternatively, it may be that the quantity of educated labor does not exceed supply but that there exists imbalances in composition, i.e. individuals are being educated in areas where there is little demand, leading to people from certain fields of study being particularly prone to overeducation. We have sought to include in our models a number of controls that reflect the level and composition of both labour demand and supply as well as measures that are designed to reflect the degree of synchronisation between demand and supply-side factors and indicators of labour market flexibility. In keeping with the existing literature, business cycle effects are also controlled for by the inclusion of measures of per capital GDP and the unemployment rate. Nevertheless, we cannot easily reflect all potential explanations for overeducation within our models. For instance, labour demand and supply might be perfectly synchronized yet overeducation might still arise due to frictions arising from asymmetric information, institutional factors that prevent labour market clearance or variations in individual preferences related to either job mobility or work-life balance. Some recent literature has pointed towards the importance of factors related to job-search and preferences in determining overeducation (McGuinness, Whelan & Bergin 2016; McGuinness & Pouliakas,

forthcoming), however, no data was available to reflect such factors for the sample of countries included in our study.

Results

The sample for the study is restricted to employees in employment who work full-time. The average levels of overeducation for the EU 28, based on quarterly data, for the period 2001 to 2011 are reported in the first column of Table 1. The estimated rate of overeducation varies from 8 per cent in the Czech and Slovak Republics to 30 per cent and over in Ireland, Cyprus and Spain. Generally speaking, we observe the estimated incidence of overeducation to be lowest in the Eastern group such as the Czech Republic, Slovenia and the Slovak Republic, highest in the Periphery group, such as Spain and Ireland, with the Central group lying somewhere in the middle. There are, however, some exceptions to this general pattern, for instance, overeducation rates were relatively high in Lithuania and Estonia, while overeducation in Portugal was well below the level observed in other peripheral countries. In order to assess the consistency of our estimates, the second column of Table 1 provides a comparison with a number of point estimates for 2014 generated by Flisi et al. (2014), who applied a comparable approach to the OECD Programme for the International Assessment of Adult Competencies (PIAAC) data. Generally speaking our overeducation estimates match closely with those from the PIAAC based study, with the exception of the estimate for Denmark where a relatively large discrepancy exists. Figure 1 plots the rates over time for both the entire sample and then, separately, for the Central, Eastern and Peripheral European groups. The first thing that becomes obvious is that overeducation has remained remarkably stable, at just under 18 percent, across the EU 28. There is some evidence of some slight cyclicality with overeducation rates rising somewhat following the onset of the great recession in 2008 before falling off again in 2010. However, more variation is apparent when the data is analyzed separately for the Central, Eastern and Peripheral country groupings. The rate of overeducation is highest, at between 25 and 30 percent, in peripheral countries with the series appearing somewhat volatile with overeducation appearing to rise between 2003 and 2008 before falling thereafter. Overeducation in Central European countries ranged between 17 and 20 percent and, in contrast to the Peripheral group, overeducation appeared to rise somewhat in the aftermath of the great recession before falling back after 2010. Finally, the incidence of overeducation appears most stable in Eastern European countries, with the series appearing to fluctuate around a 15 percent average throughout the period. The patterns were similar when the relationships were analyzed separately by gender, with the exception of central European countries where the rate of overeducation appeared to rise more consistently for females since 2008 while remaining more stable for males.[vi]

[Table 1 about here]

[Figure1 about here]

Table 2 summarizes the relationship between male and female overeducation rates and the general direction of the trend observed in each country.[vii]. There is quite a diversity in both the general direction of the trends and the relationship between male and female overeducation. For just under half of the countries, overeducation appears to be trending upward over time. However, while the rate of increase for most appears quite slight, a much steeper slope is observed for countries in the Peripheral Group including Spain, Greece, Portugal and Italy and also in Poland. Furthermore, overeducation appears not to have risen in any observable way in 12 countries including Austria, Belgium, Germany, Denmark, Ireland, Iceland and Luxemburg while it has fallen over time in Cyprus, Croatia, Lithuania and Latvia. The incidence of female overeducation lies above that for males in the majority of countries, however, there are exceptions of this trend with a number of countries across each of Central, Eastern and Peripheral blocks exhibiting higher rates for males.

[Table 2 about here]

Given the diversity of country level trends in overeducation, it is of interest to assess the extent to which rates of overeducation have converged, or diverged, over the period. Divergence would be consistent with a situation whereby overeducation is becoming worse in areas where the level of mismatch was highest, while convergence suggests that there was more upward pressure on rates of overeducation in countries where it was initially low. From a technical perspective, convergence (divergence) is consistent with a negative (positive) and significant $\beta 1$ coefficient in the Barro regression from Equation 1. We test for the presence of ongoing convergence over the period 2003 Q1 to 2010 Q1. This time period was chosen so as to maximize the number of countries that could be included in the model, nevertheless, the results remain unchanged when the model was tested on a longer time-series including fewer countries[viii]. The coefficients from the Barro models are presented in Table 3 and indicate that ongoing convergence was a feature of the time-period. The results suggest that there is a tendency for countries to converge towards a common overeducation rate over time for all measures of overeducation. The results strongly support convergence, with the gap between the countries with the highest and lowest rates of overeducation closing by an average of 3 per cent per annum. Convergence was also detected when the analysis was carried out separately on male and female rates of overeducation, with the rate of catch up marginally higher for female educational mismatch.

[Table 3 about here]

It may be the case that the degree of on-going convergence varies among the three groups of countries with common structural, geographical and historical features. It is not possible to

estimate Barro regressions separately for our three groups as the sample size is too small. In order to overcome this difficulty, we assess the rate of on-going convergence by plotting the standard deviation of overeducation rates across countries, on the grounds that on-going convergence would be consistent with falling cross-county dispersion over time. Plotting the standard deviation across all countries confirms the results from Table 3 that on-going convergence did occur over the time period across the EU 28. However, once again, the aggregate picture appears to hide substantial variation. Convergence appears strongest within central and Peripheral EU countries and more modest among the Eastern Group (Figures 2). Nevertheless, Figure 2 also suggests that while convergence was a consistent feature across both the total sample and the individual country blocks between 2003 and 2010, there was evidence of an increase in dispersion post 2010 albeit to a level well below 2003 values.

[Figure 3 about here]

We now bring the analysis full circle by using the EU-LFS data to calculate a number of additional variables that can potentially explain variations in overeducation across countries over time. Specifically, for each country for each quarter, we calculate variables reflecting both the demand for and supply of labour such as the labor force shares of migrants, the employment shares of workers who are part-time, temporary, shares employed in various sectors (administration, sales and manufacturing), the unemployment rate and the participation rate. We also compute a variables to reflect the degree of symmetry between labour demand and supply, specifically, the ratio of workers employed in professional occupations to workers in low-skilled occupations. This is intended to pick up the effects of skill biased technological change which is generally associated with a shift in the relative demand away from high skilled to low skilled labor and in many countries a general hollowing out of mid-skilled occupations. In addition to the variables calculated from the individual labour force surveys we also derive some indicators from external data sources and, where necessary, annual is interpolated to quarterly data series. Information on GDP per capita and R&D spend[ix] was sourced from Eurostat and the Organisation for Economic Co-operation and Development (OECD), respectively. Information on the number of students enrolled in tertiary and vocational programmes was sourced from the OECD and this was standardized by age cohort using the LFS data.[x]

As we are dealing with panel data we adopt two specification approaches, a quasi-fixed effects fractional logit model that controls for time invariant country level heterogeneity and a GMM estimator that controls for the presence of both time invariant fixed effects and endogenous regressors (Arellano & Bond, 1991). The GMM estimator is considered the most robust, however, the fractional logit estimates provides us with a more accessible indication of the marginal effects associated with the impacts identified as significant under the GMM framework. It should be noted that the marginal effects generated by the fractional panel logit are very similar to those produced by the standard fixed effects equations, suggesting that the

use of a fractional outcome variable is not highly problematic in this instance[xi]. This is perhaps not surprising as the incidence of overeducation typically lies in the 10 to 30 per cent, implying that there is no clustering around the extreme values of 0 or 1. We estimate each model initially on the pooled country sample before measuring relationships separately by gender and then for the three country classifications. With respect to the GMM, as the Arellano-Bond GMM regression is run on first differences,we expect to find evidence against the null hypothesis of zero autocorrelation in the first-differenced errors at order 1. The AR(2) test on the residuals in first differences is used to detect AR(1) in the underlying levels variables. In the results presented in Tables 3 and 4, we reject no autocorrelation of order 1 and fail to reject no autocorrelation of order 2, therefore, showing that the Arellano-Bond model assumptions are satisfied. The models also pass the Sargan test for valid instruments.

Comparing the fixed effects and GMM models for the pooled sample, a number of factors appear to consistently influence cross-country variations in overeducation. Falls in unemployment tend to lead to increases in overeducation, suggesting that both forms of mismatch may be viewed as substitutes by workers. Given that both overeducation and unemployment can be considered forms of mismatch that are potentially subject to scarring effects that influence future labour market outcomes (Baert & Verhaest, 2014), workers may prefer to remain unemployed and continue job search, rather than enter employment and be overeducated if the scarring effect of the former outweighs that of the latter. Vobemer & Schuck (2015) argue that evidence of strong lock in effects 5 years after initially becoming overeducated supports the view that workers may prefer unemployment to overeducation. The fractional logit model suggests that a one percent increase in the unemployment rate will result in an average 0.15 percentage point fall in the incidence of overeducation within countries. This trade-off effect was observed in both the pooled and gender specific models. Overeducation was found to be negatively related to the share of female workers in the labour market, suggesting that factors that stimulate female participation may also simultaneously influence overeducation. The fractional logit model suggests that a one percentage point increase in the share of female employment will produce an average 0.65 percentage point fall in the incidence of overeducation within countries. This is a somewhat surprising results given that theories of differential overqualification would tend to suggest that higher shares of females in the labour market, particularly married females, the higher the rate of overeducation (Frank 1978; McGoldrick & Robst 1996). However, our data suggest that this is not always the case and lower female rates of overeducation persist in a number of countries (Table 1). One potential explanation is that the result may reflect variations in policy approaches that facilitate and encourage effective female participation. Specifically it is likely that we observe higher proportions of matched females, and lower rates of overeducation, in labour markets that pro-actively pursue policies, such as in the areas of childcare, that enable females to remain in the labour market without having to occupationally downgrade. This is supported in our data where we find that female overeducation rates are lower than those of males in a number of Nordic countries (Denmark, Iceland) and in Luxemburg and the Netherlands that are known to adopt more progressive policies towards childcare (Table 1). Conversely females will be more likely to remain

inactive in countries with little provision for childcare and a sustainable work-life balance. Interestingly the female employment share variable was significant in both the male and female models suggesting that the existence of progressive policies that facilitate female participation also have positive spill-over effects for males.

The composition of employment also appears to be an important feature with overeducation inversely related to the employment share of manufacturing, suggesting that either hiring procedures or requirements in sectors requiring vocational skills differ in important ways that alleviate mismatch. This hypothesis is further strengthened by the finding in the models that overeducation is higher in countries where there exists a heavier emphasis on academic enrolments, as opposed to vocational education, suggesting that the occupational specific nature of vocational qualifications and labour markets lead to a lower exposure to mismatch relative to academic qualifications and more general occupations. In terms of marginal effects, the fractional logit model suggests that a 1 percentage point increase in the share of manufacturing employment leads to an average reduction of 0.5 percentage points in the incidence of overeducation within countries. With respect to the expansion of academic enrolments, the models suggest that this is damaging only for females with a 1 point increase in the ratio generating a 0.08 percentage point increase in the incidence of overeducation. An increase in the share of high-skilled jobs in the economy (the ratio of workers in high to low skilled occupations) was found to reduce overeducation among males and increase it among females suggesting that the benefits of skill biased technological change (SBTC) are more skewed towards educated males who are potentially displacing high skilled females. These effects may be partially explained by gender differences in field of study among graduates, in particular, the lesser tendency among females to study STEM subjects. In line with this, the evidence suggests that SBTC is skewed towards new technologies (Acemoglu, 1998) which will tend to favor STEM graduates.

Variables such as the share of temporary and part-time workers were included in the model to capture the effects of labour market flexibility on overeducation on the basis that countries with higher employment shares of temporary and / or part-time workers are generally viewed as having an increased capacity to respond to labour market disequilibria. The GMM estimates from Table 5 indicate that these measures of flexibility are important determinants of overeducation for males only. Exactly why only male overeducation tends to be lower in more flexible labour markets is unclear, however, to the extent that females are more likely to trade overeducation for other job attributes such as job security and flexibility (McGuinness & Sloane, 2011), they may also be less likely to take advantage of the increased job opportunities that flexible labour markets offer in search of an improved match. Finally, in addition to using the fractional logit to estimate marginal effects, we also derive long run elasticity's from the GMM estimates, these are compared in Appendix 1 Table 1 and are broadly in line with each other, suggesting that the short run marginal effects and long run impacts of the key variables are broadly similar.

Table 6 estimates the models separately for each of our specified country groupings and a number of important differences become apparent. For central European countries the determinants of cross-national variations in overeducation reflect those of the pooled model with labour market flexibility, female participation all proving important. However, the unemployment rate is not a pertinent factor within the central European grouping with the results suggesting that the hypothesized trade-off driving the effect in the pooled sample is being primarily driven by Eastern European countries. The Eastern European country block is further distinguished by a positive relationship between overeducation and labour force share of migrants, suggesting that migrants in these countries are more likely to be sub-optimally positioned in the labour market. Within peripheral countries overeducation, while inversely related to the female share of employment, is positively related to the overall participation rate; furthermore, within the peripheral block, overeducation was lower in countries with a higher concentration of employment in the sales and hospitality sector which, again, is more reliant than average on vocational skills. Finally, in both peripheral and eastern European countries overeducation was lower in countries with a higher availability of vocational places; within peripheral countries overeducation is also higher in countries with a heavier emphasis on third-level education.

Summary and conclusions

Overeducation is known to be costly to workers and it also has negative implications for firms and the wider macro economy. To date, the vast body of research in the area has focused on examining the incidence and impacts of overeducation within countries. This paper attempts to examine patterns in overeducation between countries using a specifically designed panel dataset constructed from the quarterly Labour Force Surveys (LFS) of EU 28 countries over a twelve to fifteen year period. The descriptive evidence shows that there are wide variations in overeducation rates throughout Europe with rates generally highest in peripheral countries and lowest in eastern European states. It is not the case that overeducation is rising rapidly over time in all countries, where overeducation has been seen to be growing the trend has been very gradual; furthermore, overeducation rates were found to be static or falling in around in approximately fifty percent of EU 28 countries. Nevertheless, the evidence does point towards convergence in overeducation rates with countries exhibiting the lowest incidences of overeducation in 2002 experiencing the highest growth rates in overeducation over the 2003 to 2012 period. We estimate the overeducation rates in Europe converged at a rate of 3.3 percent per annum over the period with a similar result emerging when convergence in male and female rates was assessed. Further analysis revealed that convergence appears strongest within central and Peripheral EU countries and most modest among the Eastern Group. In terms of the factors that potentially drive cross-country variations in overeducation, a number of key variables emerged from our analysis. We found that overeducation was lower in central European countries with a higher female employment share which is suggestive of the important role of policies designed to facilitate females remaining in the labour market without having to occupationally downgrade. This

hypothesis was strengthened by the finding that females are less likely to be overeducated in Nordic countries and in Luxembourg and the Netherlands that have a strong tradition of equality legislation and childcare provision. Interestingly, we found that increased female participation lowers the male and female incidences, suggesting the existence of strong spill-over effects from equality legislation and / or childcare provision. Labour market flexibility was found to be an important mediating factor but only for male rates of overeducation in central European countries. There was also some evidence that overeducation and unemployment were treated as substitutes by female workers, however, the finding was restricted to Eastern Europe. The composition of labour demand also appears to be important with countries employing larger shares of labour in sectors reliant on vocational skills, manufacturing in central Europe and sales and hospitality in peripheral countries, experiencing lower rates of overeducation. Finally the nature of education provision appears to be important, particularly in peripheral and central European countries. There was strong evidence to suggest that overeducation tends to be lower the higher the availability of vocational educational options for young persons seeking post-secondary education and training in Eastern and Periphery country groupings. This reinforces the conclusions of McGuinness et Al. (2016) who found evidence to support the view that the acquisition of vocational and work related skills is an important determining factor in avoiding mismatch among university graduates.

The findings suggest that while overeducation may respond to policy variables, the impact of particular policies will tend to vary depending on specific labour market contexts. Nevertheless, the work does point to a number of areas where policy could play a role. The results regarding the balance between vocational and academic pathways suggests that workers with skills and competencies that are more directly identifiable as job related will be less likely to become overeducated. Therefore, a combined approach of improving the availability of vocational programmes to school leavers and increasing the practical aspects of more academic postsecondary courses is likely to yield positive results in most countries. The finding with respect to female participation is unique and suggests that the expansion of policies that allow females to remain active in the labour market without having to occupationally downgrade benefits both males and females in terms of lowering the rate of overeducation. Enhancing labour market flexibility and the capacity of the labour market to respond to shocks is another policy option for which there is some evidence here, however, there is also a risk that beyond a certain level, increased deregulation will reduce job quality in a way that actually stimulates overeducation. Finally, there are a number of policy levers that could not be considered here that are also open to government including removing information asymmetries between job seekers and employers (McGuinness et al. 2016; McGuinness & Pouliakas, *forthcoming*) and examining ways to enhance the flexibility of firms to more effectively accommodate shifts in the educational composition of labour supply.

References

Acemoglu, D. (1998). Why do new technologies complement skills? Directed technical change and wage inequality. *The Quarterly Journal of Economics*, 113(4), 1055-1089.

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

Baert, S., and Verhaest, D. (2014). Unemployment or overeducation: Which is a worse signal to employers? IZA Discussion Papers, 8312.

Barro, Robert J. (1997). *Determinants of Economic Growth: A Cross-Country Empirical Study*, Cambridge MA, MIT Press.

Croce, G. and Ghignoni, E. (2012) Demand and supply of skilled labour and overeducation in Europe: A country level analysis. *Comparative Economic Studies*, 54, 413–39.

Davia, M. A., McGuinness, S. & O'Connell, P. (2017). Determinants of regional differences in rates of overeducation in Europe, *Social Science Research,* 63, 67-80.

Dolton, P. and Vignoles, A. (2000). The incidence and effects of overeducation in the UK graduate labour market, *Economics of Education Review*, 19, 179–98.

Flisi, S., Goglio, V., Meroni, E., Rodrigues, M. and Vera-Toscano, E. 2014."Occupational mismatch in Europe: Understanding Overeducation and Overskilling for Policy Making" JRC Science and Policy reports, European Commission, Luxemburg, ISSN 1831-9424.

Frank, R. H. (1978). Why Women Earn Less: The Theory and Estimation of Differential Overqualification, *American Economic Review*, 68 (3), 360-373.

Frenette, M. (2004). The overqualified Canadian graduate: the role of the academic programme in the incidence, persistence and economic returns to overeducation. *Economics of Education Review* 23, 29–45.

McGoldrick, K.M. and Robst, J.(1996). Gender differences in overeducation. A test of the theory of differential overqualification. *American Economic Review*, 86, 280–84.

McGuinness, S. (2006) Overeducation in the labour market, *Journal of Economic Surveys*, 20, 387–417.

McGuinness, S., A. Bergin and A. Whelan. (2016) “Is there a Role for Higher Education Institutions in Improving the Quality of First Employment?”, *BE Journal of Economic Analysis and Policy.* Available online: https://doi.org/10.1515/bejeap-2016-0174, December 2016.

McGuinness, S. and K. Pouliakas (*forthcoming*). “Deconstructing Theories of Overeducation in Europe: A Wage Decomposition Approach”, *Research in Labor Economics*.

McGuinness, S. and Pouliakas K. (2016). Deconstructing theories of overeducation in Europe: a wage decomposition approach. IZA Discussion Paper 9698.

McGuinness, S. and Pouliakas, K. 2015. “Decomposing wage penalties of overeducated workers: Evidence from the Cedefop European skills and jobs survey”, research paper presented at the Cedefop/IZA workshop on skills and skill mismatch, 29-30 November 2015, Thessaloniki.

McGuinness, S and Sloane, P. J. (2011). Labour market mismatch among UK graduates: An analysis using REFLEX data. *Economics of Education Review*, 30(1), 130–145.

McGuinness, S. and Wooden, M. (2009). Overskilling, job insecurity and career mobility, *Industrial Relations*, 48, 265–86.

Mundlak, Y. (1978). On the pooling of time series and cross section data, *Econometrica*, 46 (1), 69-85.

Papke, L. E., and Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates, *Journal of Applied Econometrics,* 11, 619–632.

Papke, L. E. and Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates, *Journal of Econometrics*, 145, 121-133.

Quintini, G. (2011). Over-qualified or Under-skilled: A review of the existing literature. OECD Social, Employment and Migration working papers No.121.

Roodman, D. (2006). How to do xtabond2: an introduction to “Difference” and “System” GMM in Stata. Center for Global Development Working Paper Number 103.

Rubb, S. (2003) Overeducation: A short or long run phenomenon for individuals? *Economics of Education Review*, 22, 389–94.

Sloane, P.J. (2003). Much ado about nothing? What does the overeducation literature really tell us? in F. Buchel, A. de Grip, and A. Mertens (eds) Overeducation in Europe: Current Issues in Theory and Policy, Edward Elgar, Cheltenham.

Verhaest, D. & Van der Velden, R. (2012). Cross-country differences in graduate overeducation. *European Sociological Review*, 29(3), 642-653.

Voßemer, J., & Schuck, B. (2015). Better Overeducated than Unemployed? The Short- and Long-Term Effects of an Overeducated Labour Market Re-entry. *European Sociological Review*, 32, 251–265.

Tables

Table 1: Overeducation Rates: Comparison of ESRI Estimates from the EU Labour Force Survey (LFS) data averaged over 2001-2011 and estimates of Flisi et al. (2014) based on Programme for the International Assessment of Adult Competencies (PIAAC) data in 2014

	ESRI Estimate (2001-2011 average)	**Estimates based on PIAAC from EC (2014)**
Austria	0.19	0.23
Belgium	0.26	0.24
Bulgaria	0.11	
Cyprus	0.31	0.31
Czech Republic	0.08	0.12
Germany	0.18	0.22
Denmark	0.18	0.31
Estonia	0.24	0.26
Spain	0.30	0.34
Finland	0.14	0.17
France	0.17	0.17
Greece	0.28	
Hungary	0.13	
Ireland	0.33	0.33
Italy	0.24	0.24
Lithuania	0.25	
Luxembourg	0.17	
Latvia	0.19	
Netherlands	0.22	0.22
Poland	0.11	0.11
Portugal	0.18	
Romania	0.10	
Sweden	0.14	0.19
Slovenia	0.09	
Slovak Republic	0.08	0.10
UK	0.21	0.20

Table 2: Key Characteristics of Country Level Overeducation Series based on ESRI Estimates from EU-LFS data, 2001-2011 (27 Countries)

	Male >Female	Male < Female	Positive Trend	Negative Trend	No Trend
Austria		X			X
Belgium		X			X
Bulgaria		X			X
Czech			X		
Germany		X			X
Denmark	X				X
Estonia		X			X
Spain		X	X		
Finland		X	X		
France		X			X
Greece	X				X
Hungary		X	X		
Ireland					X
Iceland	X				X
Italy	X		X		
Lithuania	X			X	
Luxemburg	X				X
Latvia		X		X	
Netherlands	X				X
Norway			X		
Poland		X	X		
Portugal		X	X		
Romania			X		
Sweden		X	X		
Slovenia		X	X		
Slovak	X		X		
UK		X			X

Table 3: Barro Regression Results: Time Period Q1 2003 – Q1 2012 for 26 countries

Overeducation Shares	Coefficients
Total Overeducation	-0.033*** (0.009)
Female Overeducation	-0.036** (0.011)
Male Overeducation	-0.032*** (0.008)

Note: ***,**,* denotes significance at the 10%, 5% and 1% significance levels.

Table 4: Determinants of Overeducation – RE Panel Fractional Logit with Mundlak Correction

VARIABLES *(Lag Q1)*	(1) Overall	(2) Female	(3) Male
Overall Participation Rate	-0.13	-0.22*	-0.04
	(0.086)	(0.126)	(0.105)
% Migrants in Labour Force	0.06	0.04	0.06
	(0.036)	(0.030)	(0.046)
% Temporary Workers	-0.10**	0.01	-0.16***
	(0.050)	(0.083)	(0.055)
% Part-time Workers	-0.04	0.10*	-0.12
	(0.067)	(0.064)	(0.080)
% Female Workers	-0.65***	-0.74***	-0.59**
	(0.171)	(0.209)	(0.246)
Overall Unemployment Rate	-0.15**	-0.17***	-0.16*
	(0.065)	(0.060)	(0.089)
% Employed in Public Administration	0.54***	0.71**	0.44
	(0.222)	(0.348)	(0.277)
% Employed in Sales & Hotels	-0.08	0.17	-0.27
	(0.154)	(0.163)	(0.178)
Share of Manufacturing	-0.58***	-0.54***	-0.57***
	(0.117)	(0.156)	(0.125)
Ln GDP per capita	-0.02	-0.00	-0.03*
	(0.021)	(0.027)	(0.019)
R&D Expenditure (GERD)	0.00	0.00	0.00
	(0.006)	(0.005)	(0.007)
Ratio of Workers in High (2,3) to Low (7,8,9) SOC	-0.01	0.02	-0.02***
	(0.012)	(0.02)	(0.010)
Ratio of Vocational Students to Pop. Aged (15-19)	-0.02	-0.02	-0.02
	(0.022)	(0.034)	(0.019)
Ratio of Tertiary Students to Pop. Aged (20-24)	0.05*	0.08**	0.03
	(0.028)	(0.042)	(0.021)
Mundlak Correction	INC	INC	INC
Constant	INC	INC	INC
Observations	1,129	1,129	1,129
Number of countries	27	27	27
Prob > F	0.00	0.00	0.00

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Determinants of Overeducation - Arellano–Bond Linear Dynamic Panel-Data Approach (xtabond)*

VARIABLES	(1) Overall	(2) Female	(3) Male
Overall Participation Rate	-0.02	-0.07	0.02
	(0.057)	(0.083)	(0.059)
% Migrants in Labour Force	0.03	0.02	0.02
	(0.020)	(0.016)	(0.024)
% Temporary Workers	-0.04	0.02	-0.07**
	(0.026)	(0.054)	(0.034)
% Part-time Workers	-0.09**	-0.05	-0.10***
	(0.043)	(0.068)	(0.037)
% Female Workers	-0.33***	-0.32**	-0.33**
	(0.094)	(0.143)	(0.133)
Overall Unemployment Rate	-0.08**	-0.11**	-0.06
	(0.040)	(0.044)	(0.048)
% Employed in Public Administration	0.02	0.27*	-0.15
	(0.189)	(0.151)	(0.266)
% Employed in Sales & Hotels	0.00	0.16	-0.11
	(0.091)	(0.117)	(0.096)
Share of Manufacturing	-0.30***	-0.29**	-0.29***
	(0.075)	(0.120)	(0.065)
Ln GDP per capita	-0.02	-0.02	-0.02**
	(0.012)	(0.017)	(0.010)
R&D Expenditure (GERD)	0.00	-0.00	0.00
	(0.004)	(0.004)	(0.004)
Ratio of Workers in High (2,3) to Low (7,8,9) SOC	-0.00	0.01	-0.01*
	(0.009)	(0.014)	(0.007)
Ratio of Vocational Students to Pop. Aged (15-19)	-0.01	-0.01	-0.01
	(0.014)	(0.021)	(0.013)
Ratio of Tertiary Students to Pop. Aged (20-24)	0.02	0.04	0.01
	(0.018)	(0.027)	(0.014)
Lagged Total Overeducation	0.53***		
	(0.050)		
Lagged Overeducation Female		0.50***	
		(0.061)	
Lagged Overeducation Male			0.57***
			(0.043)
Constant	0.49***	0.47***	0.49***
	(0.109)	(0.147)	(0.099)
Observations	1,123	1,123	1,123
Number of countries	27	27	27
Prob > F	0.00	0.00	0.00
A-Bond Test AR (1), *p* (*H0: no autocorrelation*)	0.00	0.00	0.00
A-Bond Test AR (2), *p* (*H0: no autocorrelation*)	0.23	0.03	0.65
Sargan Test, *p* (*H0:overidentifying restrictions are valid*)	0.94	0.98	0.84

*Estimated with Robust Standard errorsin parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Determinants of Total Overeducation by Country Groupings - Arellano–Bond Linear Dynamic Panel-Data Approach (xtabond)*

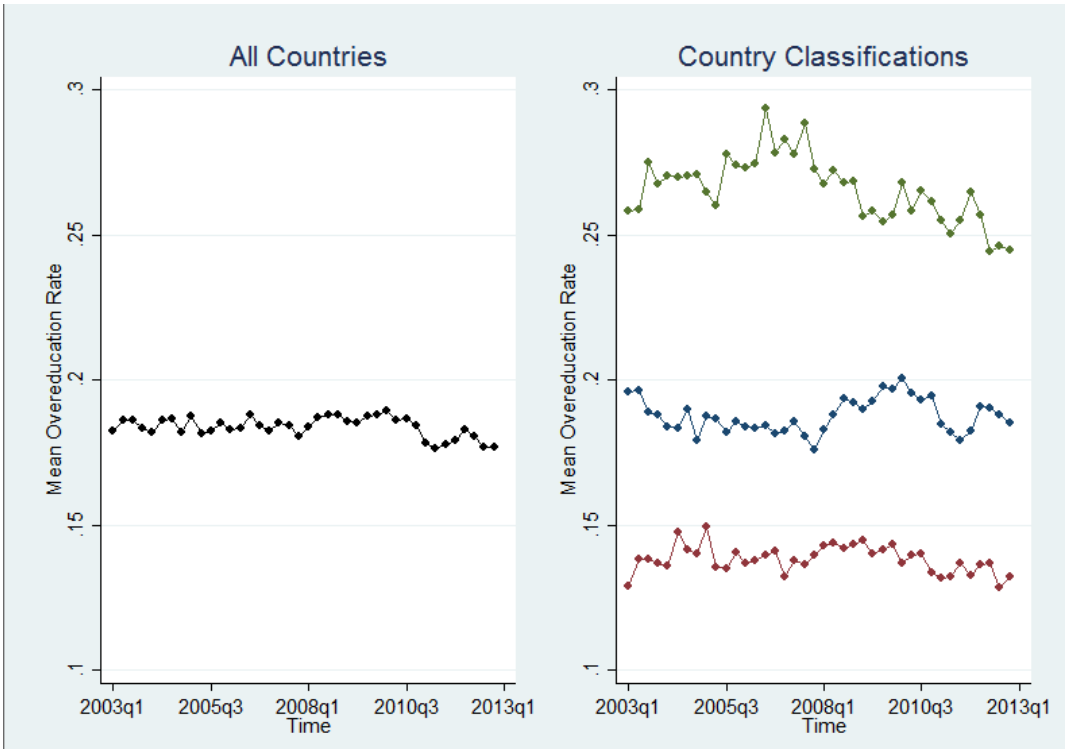
VARIABLES	(1) ALL	(2) Central	(3) Eastern	(4) Periphery
Overall Participation Rate	-0.02	-0.01	-0.11	0.51***
	(0.057)	(0.075)	(0.103)	(0.157)
% Migrants in Labour Force	0.03	0.01	0.06***	-0.08
	(0.020)	(0.020)	(0.018)	(0.054)
% Temporary Workers	-0.04	-0.10*	0.11	-0.04
	(0.026)	(0.055)	(0.076)	(0.039)
% Part-time Workers	-0.09**	-0.12**	0.09	-0.22
	(0.043)	(0.053)	(0.071)	(0.146)
% Female Workers	-0.33***	-0.59***	-0.22	-0.44***
	(0.094)	(0.193)	(0.133)	(0.106)
Overall Unemployment Rate	-0.08**	-0.10	-0.11**	-0.19
	(0.040)	(0.069)	(0.053)	(0.120)
% Employed in Public Administration	0.02	-0.05	0.18	-0.48
	(0.189)	(0.385)	(0.173)	(0.404)
% Employed in Sales & Hotels	0.00	-0.01	0.15	-0.71**
	(0.091)	(0.097)	(0.132)	(0.279)
Share of Manufacturing	-0.30***	-0.50***	-0.17	-0.44
	(0.075)	(0.075)	(0.104)	(0.389)
Ln GDP per capita	-0.02	-0.03	-0.01	0.02
	(0.012)	(0.024)	(0.012)	(0.036)
R&D Expenditure (GERD)	0.00	0.00	0.00	-0.01
	(0.004)	(0.003)	(0.004)	(0.012)
Ratio of Workers in High (2,3) to Low (7,8,9) SOC	-0.00	-0.01	0.02	0.02
	(0.009)	(0.011)	(0.018)	(0.036)
Ratio of Vocational Students to Pop. Aged (15-19)	-0.01	0.01	-0.06**	-0.69**
	(0.014)	(0.016)	(0.030)	(0.277)
Ratio of Tertiary Students to Pop. Aged (20-24)	0.02	-0.01	0.08	0.94**
	(0.018)	(0.021)	(0.051)	(0.395)
Lagged Total Overeducation	0.53***	0.51***	0.45***	0.22**
	(0.050)	(0.068)	(0.127)	(0.101)
Constant	0.49***	0.81***	0.32**	0.19
	(0.109)	(0.246)	(0.134)	(0.346)
Observations	1,123	510	406	207
Number of countries	27	12	10	5
Prob > F	0.00	0.00	0.00	0.00
A-Bond Test AR (1), *p* (*H0: no autocorrelation*)	0.00	0.02	0.02	0.05
A-Bond Test AR (2), *p* (*H0: no autocorrelation*)	0.23	0.29	0.11	0.17
Sargan Test, *p* (*H0:overidentifying restrictions are valid*)	0.94	0.82	0.79	0.44

*Estimated with Robust Standard errorsin parentheses

*** p<0.01, ** p<0.05, * p<0.1

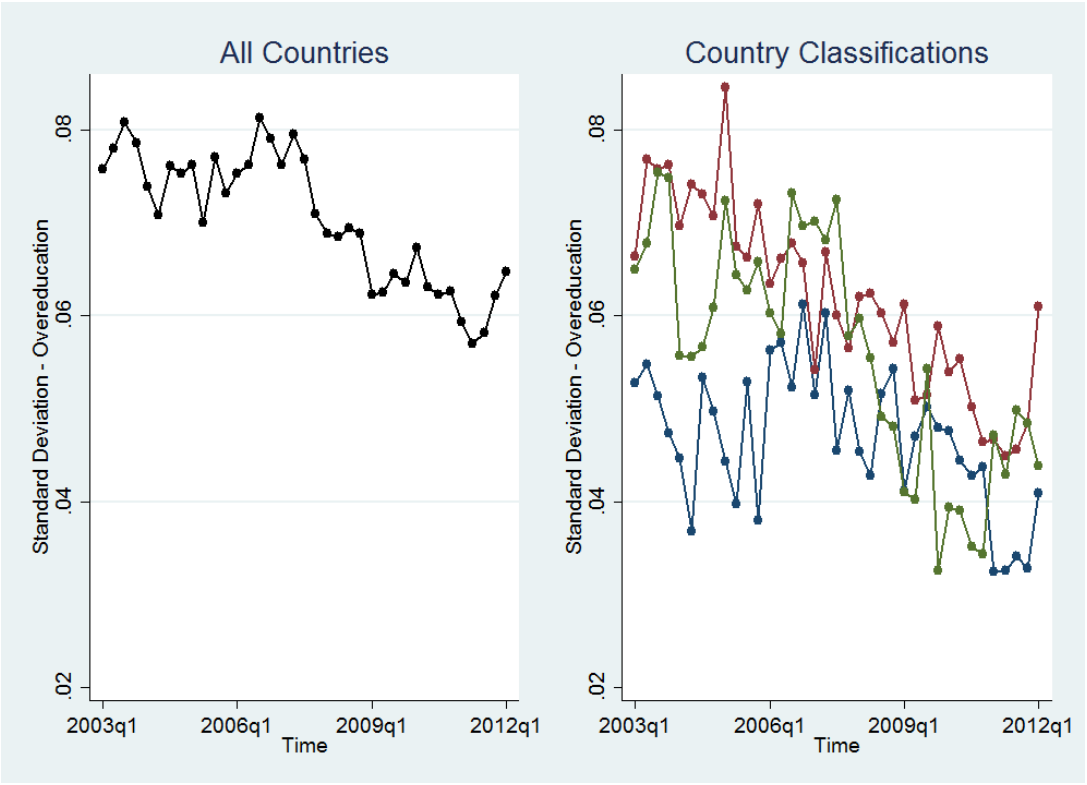
Figures

Figure 1: Mean Overeducation Rate (restricting to full-time employees), 2003-2013.



***Country Classification Legend:** 1=Central (Blue); 2=Eastern (Red); 3=Periphery (Green).*

Figure 2: Standard Deviation of Total Overeducation from Q1 2003 to Q1 2012



***Country Classification Legend:** 1=Central (Blue); 2=Eastern (Red); 3=Periphery (Green).*

Table A1: Estimates Long-run Effects derived from Arellano–Bond Linear Dynamic Panel Model (Xtabond) and estimated marginal effects from Panel Fractional Logit (quasi fixed effects model)

VARIABLES *Only those found to be significant in xtabond model*	**(1)** **A-BOND**[xii]	**(2)** **Frac logit**
% Female Workers	-0.69***	-0.65***
	(0.212)	(0.171)
Share of Manufacturing	-0.64***	-0.58***
	(0.144)	(0.117)
% Part-time Workers	-0.19**	-0.04
	(0.089)	(0.067)
Overall Unemployment Rate	-0.17**	-0.15**
	(0.079)	(0.065)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[i] Derived from factor analyses carried out on subjective variables.
[ii] Deviations of the observed rate from the natural rate .
[iii] The descriptive analysis and tests for long-run relationships include Cyprus, Croatia and Germany. These countries are excluded from later analysis because of missing or incomplete data.
[iv] The correction amounts to including the means values of each term as additional explanatory variables in the model.
[v] A problem with dynamic panel models is that, particularly for small T and large N, t first differencing creates a correlation between the differenced lagged dependant variable and the error term. A solution to this problem is to instrument the explanatory variables with their lagged values as the lags will be highly correlated with the explanatory variables but uncorrelated with the error term. Such instrumenting ensures that problem of endogenous regressor is eliminated. The Arellano Bond implements this process in a GMM contexts.See Roodman (2006) for a full discussion.
[vi] Results are available by request from the authors.
[vii] Please note that the direction of trend was assessed through a visual inspection of the data rather than through any formal statistical test.
[viii] Results are available by request from the authors.
[ix] Gross Domestic Expenditures on R&D (GERD) from the OECD was used.
[x] Some existing research has indicated that overeducation tends to be lower in countries with more developed vocational pathways (Mavromaras and McGuinness, 2012).
[xi] Results are available by request from the authors.
[xii] The long-run effect of the covariate is usually defined to be the sum of the current (and lagged coefficients, if used) divided by 1 minus the sum of the lagged coefficients on the dependent variable.