Over-education in Cyprus: Micro and Macro Determinants, Persistence and State Dependence. A Dynamic Panel Analysis

Christiana Charalambidou* and Steven McIntosh**

Cyprus Centre for Business Research

**Corresponding author: Department of Economics, University of Sheffield, 9 Mappin

Street, Sheffield, S1 4DT, UK. Email: s.mcintosh@sheffield.ac.uk

Abstract: We use longitudinal panel data from the EU-SILC for the period 2005-2011 to examine the micro and macro determinants, persistence and dynamics of over-education. The key contribution in terms of the determinants of over-education is the inclusion of macro-level explanatory variables to control for both aggregate supply and aggregate demand labour market conditions. Their effects are found to be strongly significant and to have the expected sign. The article also disentangles the effect of past over-education experience on the likelihood of current over-education. Results demonstrate that over-education is not only a long-run phenomenon for many workers but also that current over-education is largely due to past circumstances of the individual with this state dependence present at all career stages.

Keywords: over-education; state dependence;

Acknowledgments

Introduction

Over-education is commonly defined as the situation whereby an individual has a higher level of attained education than that required by his/her job. Most existing research focusses on the effect that over-education has on individual-level outcomes, or on the characteristics of individuals associated with the likelihood of being over-educated, in an attempt to explain why the phenomenon occurs. It is to this second literature that this paper contributes. In particular we go beyond the standard analysis looking at individual correlates of over-education status, to consider two additional issues: to what extent is the likelihood of over-education affected by previous experience of over-education, and what role do macro-level factors play in explaining over-education? These questions are answered in the context of a study of Cyprus, which has rarely been the focus of research in this area.

We therefore consider the dynamic properties of over-education, in particular the possibility of state-dependence, i.e. examine whether lagged over-education is found to be significantly affecting the probability of the following year's over-education even after controlling for background factors that initially caused the over-education (Mavromaras et al., 2012). At the same time, by engaging in a dynamic analysis of the determinants of over-education, the article will isolate the main determinants of over-education within a methodological setting that copes with unobserved heterogeneity and state dependence (Boll et al., 2016). As an extension, and in order to examine whether state-dependence in over-education differs based on the career stage one is in, separate regressions will also be run on sub-samples of respondents at different stages of their careers.

In addition, we go beyond the examination of micro level determinants and, by incorporating a number of both demand and supply side macro-level variables, attempt to link

the micro and macro literature[1] on over-education and examine how these labour market conditions affect the likelihood of over-education. The effect of overall macroeconomic conditions on the probability of over-education has attracted markedly less attention in the literature. This paper undertakes to fill this gap and to widen the understanding of the effect of changing macro conditions on the likelihood of over-education by taking into account labour market conditions both at labour market entry and at the time of the survey.

The occurrence of over-education is an important issue to understand, as it can represent an inefficient allocation of resources due to an underutilisation of (educational) skills (Linsley, 2005) and can therefore be costly not only for the individual but also for the firm and the economy as a whole. More specifically, at an individual level, apart from the wage losses that over-educated workers are likely to suffer in terms of diminished returns to their educational investments compared to matched individuals with comparable education (McGuiness, 2006), over-educated workers may also endure lower levels of job (e.g. Tsang et al. 1991; Battu et al., 2000) and life (e.g. Piper, 2015) satisfaction, may experience a cognitive decline (De Grip et al., 2008) and are also found to have poorer mental and physical health (e.g. Tsang and Levin, 1985).

For firms, the presence of overeducated workers may imply productivity losses, for example via lower employee satisfaction (Allen and van der Velden, 2001) and higher intention to quit (e.g. Tsang et al., 1991) or via higher rates of turnover that translate into lost investments in training, screening and recruitment (e.g. Sloane et al. 1999). At the macro level, over-education may translate into lower national welfare and wasted tax revenues if individuals are equipped with non-productive education (McGuiness, 2006).

---

[1] See Morano (2014) for a similar approach.

On the other hand, another strand of the literature views the phenomenon as a statistical artefact resulting from either inadequate measurement techniques or from the absence of adequate controls within the ordinary wage equation context (McGuiness, 2006), or as temporary and not associated with high costs. If over-education is indeed found to be temporary and a path or a stepping stone towards a more productive job matching one's education, then the costs of over-education for all parties involved are expected to be minimal. Determining whether over-education leads to further over-education or not is therefore a very important issue.

These questions are studied in the context of the Cyprus. Cyprus is a country characterised by a remarkably strong demand for higher education and like other small economies relies on its human capital as a key factor in production. This means that the quality of its human resources is of crucial importance for economic growth (Bacchus, 2008). The percentage of secondary school graduates who chose to pursue further studies during the years of analysis fluctuated between 76-82%.[2] According to the Cyprus Department of Higher and Tertiary education[3], one of their main goals over the past years has been to increase the number of people attending Higher Education in Cyprus. To encourage such participation, the University of Cyprus was established in 1992, as well as another two state universities plus private universities. As a result of these reforms and increased expenditure on education, the number of students in Cyprus has increased rapidly over the last two decades.

From the student perspective, Menon et al. (2012) explain that the high demand for Higher Education in Cyprus has been associated with the desire of students to improve their

---

[2] Statistics of Education (2010/11) Report, Statistical Service of Cyprus (2011) : Summary table IX: Percentage of graduates of upper secondary level who pursue further studies in Cyprus and abroad, 1991/92-2010/11 %

[3] http://www.highereducation.ac.cy/en/

employment prospects in the island's small labour market. At the same time, the country's small-sized economy, combined with the great number of university graduates limit employment and career opportunities for young graduates (Bacchus, 2008).

Combined, the above factors make Cyprus an interesting case study for analysing the phenomenon of over-education, as a failure to correctly match workers to jobs could potentially be very costly for the over-educated individuals, firms and the economy as a whole. Moreover, in the context of the recent recession that has considerably impaired the employment prospects of younger graduates, the risk of a strong persistence of over-education and hence of enduring economic costs, increases as it becomes increasingly difficult to find a good match.

A potential issue that may affect an analysis of over-education in Cyprus is that of emigration. A small island economy such as that of Cyprus is likely to lose a higher proportion to emigration than larger economies that can offer a broader range of employment. Of particular relevance to our study, it may be those who cannot find appropriate employment for their qualifications, who are more likely to emigrate. If the emigration occurs before the individuals are first observed in our sample, then this could lead to an under-estimate of the extent of over-education. More important for our analysis of persistence of over-education would be the case where such individuals chose to emigrate during the years that they are due to be included in the survey. It is the case, however, that the descriptive statistics considered later show that the characteristics of the sample remain consistent across waves (Table 1), suggesting that the sample remains representative, and attrition is not a major issue.

Flisi et al. (2017) report the rate of over-education in a number of countries, using information from the OECD Programme for the International Assessment of Adult Competencies (PIAAC). They show that Cyprus has a 20% rate of over-education, which is above the EU average of 13% and ranks third amongst the countries studied, behind Spain and Ireland.

Previous work investigating the dynamics of over-education includes Dolton and Silles (2001) and Scherer (2004), who both demonstrate a positive relationship between initial and current over-education status. Baert et al. (2013) using Flemish data also find that it takes longer for young graduates who accept a job below their level of qualifications to get a matched job than what it would have taken them had they continued to be unemployed.

Evidence for state-dependence in over-education has also been provided, for example by Mavromaras et al. (2012) and Mavromaras and McGuiness (2012) using Australian panel data and dynamic random effects probit models. Similar methods and results have been found by Frenette (2004) for Canada, Kiersztyn (2013) for Poland and Boll et al. (2016) for Germany. In terms of the empirical literature examining the effects of macro labour market conditions on over-education, this has mostly focused on how labour market conditions such as unemployment at the time of graduation (labour market entry) and at present affect the probability of over-education and future career paths. Evidence that labour market conditions at time of entry can counter-cyclically influence later mismatch has been provided by, for example, Kahn (2010) for the US, and Liu et al. (2016) for Norway. Other papers in this area present analysis based on pooled samples of a number of countries. For example, Verhaest and Van der Velden (2013) consider 13 European countries and Japan, to show that entering the labour market during a recession continues to affect the quality of the first job match and up to five years later. Similarly, McGuinness *et al.* (2018) use a panel of the EU28 countries to show, amongst other things, that female employment share is inversely related to the incidence of over-education. Similarly, labour market flexibility, higher unemployment and a greater focus on vocational education were also all shown to be negatively related to over-education.

Evidence for an effect of current labour market conditions is harder to find. Dolton and Silles (2001) examine unemployment as a possible cause of over-education among college graduates in the UK, but do not find a statistically significant result. Morano (2014) finds a

similar result, though when interacting unemployment and age group indicators, does find that the coefficients of the interaction terms between unemployment and the 20-24 and 25-29 age groups are positive and statistically significant.

Our primary contributions to this literature are fourfold. First is the application of the analysis to a country of interest that has rarely been examined in the literature. Second, the existing research on state-dependence is extended to consider this phenomenon at different career stages. Third, the range of macro variables studied is extended beyond the ones on unemployment rates that are typically used, to here additionally include variables indicating the aggregate supply side of the market. Lastly, these issue have not been researched using the data set used here, to a description of which we now turn.

The remainder of the present paper is structured as follows: Section 2 describes the data and the derivation of the over-education variable and offers some descriptive statistics, Section 3 discusses the methodology, Section 4 outlines and discusses the results and Section 5 offers some concluding remarks.

2. Data and Descriptive Statistics

The data used in this paper come from the European Statistics on Income and Living Conditions (EU-SILC) that is coordinated and released by the statistical office of the European Union (Eurostat). The EU-SILC is a multi-dimensional instrument used to undertake analyses on poverty, inequality and deprivation. It focuses on income but also covers housing, material deprivation, labour, health, demography and education so as to allow studying the multidimensional approach of social exclusion. The EU-SILC provides two types of annual data: cross-sectional annual data with variables on income, poverty, social exclusion and other living conditions, and longitudinal data pertaining to individual-level changes over time,

observed over a four year period.[4]

In this paper, the longitudinal data files for Cyprus were used. Cyprus joined the EU-SILC in 2005, and it is the only individual level longitudinal survey currently available for that country. Respondents who successfully completed a personal interview in all four years of their respective panel are kept in the data set so as to create balanced sub-panels with the maximum number of observations per respondent. The four four-year panels: 2005-2008, 2006-2009, 2007-2010 and 2008-2011 available for Cyprus are used. The resulting dataset is therefore a seven-year unbalanced panel with four sub-panels.

Though these data are relatively old, one advantage of studying this time period is that the financial crisis hit Cyprus later than in most other countries, largely having its impact in 2013. We can therefore undertake the analysis free of a large negative shock to the economy in the middle of the sample period, which would otherwise clearly have affected the interpretation of the results for macro-level variables, as well as influencing the likelihood of finding state dependence.

Appendix Table A1 investigates the attrition between the first and the fourth year of data for each of the four original sub-panels by showing the mean value of the background characteristics of the individuals in the sample as well as the initial and final sample sizes of each panel. As can be seen, the proportions of the various background characteristics are relatively stable throughout the years of each panel, confirming the representative nature of the final sample.

Given that the focus of the present paper is people over-educated for their jobs,

---

[4]More information on the EU-SILC can be accessed at the following link:

http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/introduction#

observations of respondents who are: (i) students, (ii) soldiers[5] (ii) retired, unemployed or inactive/disabled, and/or (iv) self-employed or family workers[6] for the full four years in which they are interviewed, are excluded from the sample.[7] Individuals over the age of 40 in their first round of the survey are also dropped. By restricting the focus onto a specific age group of people that is nevertheless not too narrow, the risk of capturing a cohort effect of over-education, given the rise in the overall qualification level of the population, is reduced. Following the calculation of the over-education variable, those with attained education below secondary level (i.e. pre-primary and primary education) were also dropped from the sample.[8]

Following the above sample selection rules, the final sample size retained is 1062-1490 observations per cohort with the total sample size being 5333 observations across 1617 respondents. For the panel regressions, anyone with fewer than two observations is dropped from the sample. This is 186 observations, bringing the sample size down to 5147 observations and 1431 individuals.

Measurement of Over-education

Due to the fact that empirical research depends on the availability of appropriate data, no uniform measurement of over-education exists (Verhaest and Omey, 2006). On the contrary, a number of both subjective and objective measures have been used in the over-education

---

[5] Serving their 18 month compulsory military service

[6] The self-employed and family workers are excluded as their level of mismatch cannot be easily and reliably identified

[7] Observations for individuals who worked in at least some of the panel years are included.

[8] Respondents with pre-primary and primary education were only used when estimating the modal education level by occupation and the over-education variable but were then dropped from the sample as over-education exclusively affects people with a higher educational attainment. However, results are robust to first dropping those with pre-primary and primary education and then calculating over-education, and to keeping them in the sample throughout.

literature over the years (see Groot et al., 2000). With subjective measures, individuals self-assess the skills/education required to do their job whereas with objective measures job requirements are either inspected, or an individual's education is set against that of a reference group (frequently constructed based on broad occupational categories) (Piper, 2015).[9] None of these methods of measuring over-education are without limitation.[10] For example, the self-assessment methods are likely to be problematic for quantitative analysis given the subjective nature of the workers' responses as to what education and/or skills are required for his/her job. In terms of the objective methods, one of their main downsides is that the possible heterogeneity of jobs within occupations cannot be taken into account.

Over-education in this paper is measured via the Realised Matches (RM) approach whereby the individual's education level is compared to the mean or modal education level of workers in each occupation. More specifically, the modal educational level for each occupation group is determined within the pooled set of observations across years[11] so as to serve as a proxy or an indication of the required level of education within each occupation. Even though the choice of the mode rather than the mean here is due to the structure of the education level information in the data (categorical rather than continuous), the mode is usually preferred anyway. This is because it is less sensitive to outliers and technological change and does not entail the arbitrary use of one standard deviation above the mean when identifying who is over-educated (Flisi et al., 2017). The small size of the Cypriot economy with its narrow range of sectors means that there were only 27 2-digit occupations observed in the data set. In a handful

---

[9] See Verhaest and Omey (2006)

[10] See Groot et al. (2000); Verhaest and Omey (2006) for more details.

[11] Here, the assumption that occupations have a fixed level of required education over time is made. Given that the total time period in this paper is only seven years, it is safe to assume that jobs have fixed requirements during this period.

of these, the number of observations was too small to produce a reliable estimate of the modal education level. In such cases, the small occupations were merged together to produce larger, more aggregated occupation categories. The outcome of this was classification with 20 occupational groups.[12] Thus most of the occupations used remained at the two-digit level, which is the level typically used in over-education research.[13]

The over-education binary variable is then equal to 1 if the highest level of education attained by the respondent at the time of the survey is above the modal, or norm, level of education within their occupation (i.e. a respondent is over-educated) and equal to 0 if the highest level of education attained by the respondent is less than or equal to the modal level of education within their occupational group (i.e. a respondent is not over-educated).

Due to the fact that the RM approach tends to be rather sensitive to cohort effects, the calculation of the required education by occupation and hence the over-education variable was based on the respondents who were aged up to 40 years old in their first survey year rather than on the whole sample regardless of age, so as to avoid capturing the cohort effect.[14]

One shortcoming of measuring over-education using the RM approach is that it does not take heterogeneity of jobs within occupation codes into account (Sloane et al., 1999). Given the focus here on the dynamics of over-education, this means that observed over-education transitions are restricted to between-occupation changes. In other words, the present paper only

---

[12] The use of more aggregate job titles for the derivation of the realised matches norm is not unusual in the literature and guarantees a satisfactory number of observations within each occupation. According to Verhaest and Omey (2010), although at the cost of more heterogeneity within job titles, using more aggregate job titles mitigates biases that are related to the more detailed occupations but not to the aggregate.

[13] Restricting the analysis to two-digit occupations has little effect on the results since most of the sample satisfies this criteria anyway,

[14] Results are robust to small changes in this cut-off age point.

takes into consideration changes in over-education status as a result of occupation changes whereas another route out of over-education could also take place via a changed job role within one's occupation. We therefore may not observe all movements out of over-education, though most job changes that require a significant actual increase in education/skill level might be expected to involve a change in occupation code, and so should still be picked up in our analysis.

Another limitation of the RM method that has been mentioned in the literature is that the quality of education is difficult to take into account (Sloane et al., 1999). The fact that the education variable used in this paper refers to the highest level of education attained rather than simply years of education as is often found in the over-education literature (e.g. Piper, 2015), mitigates this concern to some extent.

Amongst the final employee sample, as defined earlier, 16% of observations are revealed to be in an over-educated state. This proportion is highest amongst the 25-29 age group (30%) followed by the 35-39 year olds (25%) and the 30-34 age group (20%), while the two smallest groups, the youngest (20-24) and oldest (40-44) age groups, have the smallest percentages of over-educated workers, 15% and 9% respectively. By year of observation, the proportion who are over-educated remains similar (15-17%) in all years from 2005-2010, before rising to 22% in 2011.

Comparing over-education states in adjacent years to investigate transitions, the vast majority of individuals do not transition in any one year. The data show that 96.3% of those over-educated in one year are still over-educated the next year whereas only 3.7% of those who are over-educated in one year find a matched job in the next year. This is strong evidence suggesting that over-education is a permanent state/long-run phenomenon for the great majority of respondents who have an over-education experience and could also be a signal of a state dependency in over-education.

Nevertheless, there are transitions observed in the data to investigate, as shown in Table 1, which shows patterns of over-education status (1=overeducated, 0=not overeducated) observed amongst individuals over time. The most frequently observed pattern is not to be overeducated in all periods. On the other hand, 15% of the sample stay over-educated during all survey rounds for which they are observed, while 4% of the sample are overeducated in at least one period but *not* in all periods.

Turning to the other variables used in the analysis[15], micro determinants of over-education are grouped into personal and job characteristics and work history variables. Over-education might be higher for women if they act as 'tied stayers' or 'tied movers' (Mincer, 1978) and decrease with age as workers acquire more work experience. Activity limitation because of health problems may increase the likelihood of over-education. Being married, even though it has been found to increase the chances of over-education due to considerations such as relocation because of a partner's job (Dolton and Silles, 2001), is expected to have a less clear effect in the case of Cyprus, due to the small country size and minimal travelling distances.

The job characteristics controlled for in the over-education equation are: part-time and temporary work arrangements which are both expected to increase the likelihood of over-education. According to Dolton and Silles (2011), people working in part-time jobs may not be able to fully use all forms of human capital including qualifications attained while those working under a temporary contract may see such jobs as a 'quick fix' rather than a permanent life-long career, increasing the chances of accepting jobs for which they are over-educated.

Given the strong correlation between such job characteristics and other characteristics included in the equation, in particular gender and also potentially age, then one question could

---

[15] Descriptive statistics available from authors on request

be whether the inclusion of such job characteristics captures some of the explanatory power of these key individual characteristics. We will therefore investigate the effect of the inclusion and exclusion of job characteristics on other coefficients.

In terms of work history, the proportion of the past year spent in unemployment, recent entry into the labour market and voluntary versus involuntary job changes are investigated. Time spent in unemployment may serve as a negative signal for prospective employers as human capital may depreciate during this time, or alternatively may indicate individuals waiting for an appropriate job rather than taking the first on offer, so that its expected sign is ambiguous. Workers who voluntarily changed job in the past year are expected to have a lower likelihood of over-education than involuntary movers, as they are more likely to have found themselves a good match before quitting their old one.

An innovation of the analysis is that a number of macro variables are also included. It is expected that a higher unemployment rate at the start of paid employment increases the likelihood of current over-education if workers initially accept a job for which they are over-educated due to limited opportunities for matched work and then fail to find a good match after that (Dolton and Silles, 2001). Current labour market conditions could also affect the likelihood and persistence of over-education, for example a high level of unemployment forcing people to stay in jobs for which they are over-educated, as opportunities for matched work are limited.

The number of people with an equivalent level of qualifications is included to control for the competition faced by individuals in the labour market. The annualised change in the supply of individuals by educational category and sex during the years of the study is on average 2%. Males (females) with lower and upper secondary education and post-secondary (non-tertiary) education have an annualised increase of 0.8% (-0.8%), while males (females) with first and second stage tertiary education have an annualised increase of 3.1% (5.1%), demonstrating the large rise in the number of tertiary level graduates.

Finally, the employment share by occupation is included as a proxy for labour demand within one's occupation, and is expected to be positively related to the opportunities of finding a good match and escaping over-education both within and outside the firm.

3. Methodology

Dynamic probit models have been motivated in different ways in the literature. In this paper, the primary reason for using dynamic specifications is to examine whether over-education in the previous period increases the likelihood of over-education in this period (and hence the coefficient of the lagged dependent variable is of primary interest). Nevertheless, the use of dynamic estimation is also necessitated by the high over-education persistence observed in the aggregate over-education, as well as observed serial correlation present in the idiosyncratic error term of the static models[16] which suggests that static panel analysis estimates are not efficient and possibly misspecified (Piper, 2013).

The Dynamic Random Effects Probit model is demonstrated by the latent equation:

$$OE^*_{it} = \gamma_i \, OE_{i\,t-1} + x'_{it} \, \beta + c_i + u_{it} \qquad (1)$$

Where i=1,....,N denotes individuals observed over t=1,...T periods. OE^*_{it} is the latent dependent variable for being over-educated. The observed outcome is a dichotomous indicator of the latent variable defined as:

$$OE_{it} = 1 \text{ if } OE^*_{it} > 0 \qquad \textbf{(if over-educated)}$$

---

[16]Tested using Wooldridge's (2002) test for serial correlation, implemented using the Stata command attributed to Drukker (2003), which rejects the null of no first order autocorrelation.

$OE_{it} = 0$ if $OE^*{}_{it} \leq 0$ (if not over-educated)

$OE_{i\,t\text{-}1}$ represents the lag of the dependent variable, with γ_i being the coefficient of interest to be estimated. x'_{it} is a vector of (time-varying and non-time-varying) explanatory variables and time/year dummies associated with observation i at time t, c_i is the individual-specific random component capturing the effect of time-invariant individual unobserved heterogeneity and u_{it} is an idiosyncratic error term associated with each observation i at time t and follows a normal distribution N $(0, \sigma^2{}_u)$

Adding dynamics to a model alters the interpretation of the equation (Greene, 2008). More specifically, in the absence of the lagged variable, the independent variables reflect the full set of information behind the observed outcome while in its presence, the entire history of the right-hand-side variables is incorporated in the equation. This means that any measured effect is conditional on this history; in this case, any impact (and therefore the interpretation) of the independent variables is representative of the effect of new information, whereas the lagged dependent variable reveals the effect of the past (Greene, 2008).

A potential problem stemming from the above dynamic model specification is that, when modelling a dummy variable, OE_{it}, that is a function of the lagged dependent variable, $OE_{it\text{-}1}$, the lagged dependent variable may be correlated with the error term. More specifically, given that a person's employment history in the data is not observed from the very beginning, there is a risk that the initial value arising from a person's first observation in the sample is conditional on observed or unobserved variables in the unknown past of that person. This means that the initial value of a respondent's over-education might be affected by his or her (unobserved) previously held over-education status (Boll et al., 2016). In other words, in a dynamic equation, any unobserved heterogeneity could be picked up by the lagged dependent variable, so that whatever made people over-educated in the first place could still be making them over-educated at present. This could lead to a high persistence and spurious state

dependence in over-education. This is known as the initial conditions problem (Heckman 1981; Blundell and Bond 1998; Arellano and Carrasco 2003).

Three methods have been suggested in order to correct for the fact that in a dynamic setup y_{i0} is likely to be correlated with unobserved heterogeneity c_i affecting y_{it}. The initial conditions problem was first examined by Heckman (1981), followed by less computation-intense estimators by Orme (1997), Arulampalam and Stewart (2009) and Wooldridge (2005). Given that the three methods' performance in the context of dynamic probit models is equivalent (Arulampalam and Stewart, 2009) and the fact that Woodridge's approach is simpler to implement in practice, similar to what is often done in the literature, this is the preferred method applied here. Wooldridge (2005) suggests including the individual's over-education outcome in year t=1 as an additional covariate that captures part of the unobserved heterogeneity between persons.

Another problem arising from equation (1) above, is the unrealistic assumption of the random effects specification that requires independence between the covariates and the unobserved heterogeneity term. Rather than switch to a fixed effects specification that is known to produce inconsistent estimates in a dynamic setting when the number of time periods is small (Nickell, 1981), we instead use the Mundlak (1978) correction, which provides a solution to this problem from within the random effects framework, correcting for the effects of any unobserved characteristics that do not vary over time, with ability likely to be one of the most important. In particular, the Mundlak correction assumes that the relationship between c_i and the means of the time-varying x-variables can be written as $c_i = \overline{x}'_i \delta + \varepsilon_i$, where ε_i~iid follows a normal distribution and is independent of x_{it} and u_{it} for all i and t, and δ is the vector of coefficients on the individual-specific variable means $\overline{x}'_i$. In practice, the Mundlak correction is applied by including the individual means of each of the time-varying variables that are assumed to be correlated with the unobserved heterogeneity on the right hand side of the

regression equation hence permitting an interpretation of the point estimates as being pure within-person effects (Boll et al., 2016). In the case of the determinants of over-education, the individual means over age, years of work experience and marital status are included on the right hand side of the random effects probit regression. Nevertheless, it has to be noted that even though individual means of all time variant covariates that are suggested to be potentially correlated with individual unobservable heterogeneity are included in the regression, the possibility that time invariant variables are also correlated with individual unobservable traits cannot be ruled out (Boll et al., 2016).

The Mundlak correction is in this case combined with the Wooldridge initial conditions correction and expressed by:

$$OE^*_{it} = \gamma_i OE_{i\,t-1} + x'_{it} \beta + \bar{x}'_i \delta + \theta OE_{i\,1} + \varepsilon_i + u_{it} \qquad (2)$$

Equation (2) above, is expected to reveal the true extent of over-education state dependence.

4. Results

Table 2 reports the marginal effects of the Wooldridge dynamic probit model with Mundlak corrections. The left column presents the results when only the micro-level variables are included in the regression while the right column reports the results when both micro and macro-level regressors are incorporated. The dependent variable is over-education (equal to 1 if over-educated).

As can be seen from the regression results in Table 2, consistent with existing literature, being female increases one's chances of being over-educated while being limited in activity because of health reasons is not found to significantly affect the probability of over-education. Age and work experience are also not found to significantly affect the probability of over-education, which means that when controlling for previous over-education status, age and experience are largely irrelevant. Squared terms for age and years of work experience were

also included but were not found to attract significant coefficients.

Being married on the other hand is found to be associated with a lower probability of over-education. This can be explained by the possibility that married individuals can afford to wait longer until they find a matched job due to the fact that they can rely on their partner's income until they do so.

Turning to job characteristics, having a temporary contract is found to be positively related to the chances of over-education while the part time effect is statistically insignificant. The proportion of the past year spent in unemployment and recent entry from unemployment are not found to have a significant effect on one's chances of over-education while recent entry into the labour market from other inactivity is found to lower the probability of being over-educated. This may be because such individuals have waited until they found a matched job to enter the labour market, rather than accept a job for which they are over-educated. Voluntary job change is not found to significantly affect the probability of over-education, while involuntary job change is found, as expected, to be associated with a significantly higher probability of over-education.

One question asked in Section 2 was whether the job characteristic variables were capturing some of the explanatory power of individual characteristics such as gender and age, given the strong correlation between them. We experimented with dropping the temporary and part-time variables, either individually or jointly, but in no cases did the gender and age coefficients or their statistical significance status change.

As for the main focus of the paper, what is clear from Table 2 is the highly significant, positive effect of past years' over-education on this year's over-education, which demonstrates that over-education is a self-perpetuating state. According to Mavromaras and McGuinness (2012), the intuition behind this finding is that the negative impact of those characteristics that were responsible for becoming over-educated in the first place will be heightened via the

continued presence of over-education, thus reinforcing the labour market disadvantages associated with over-education.

Specifically, the lag variable of over-education is found to be increasing the probability of over-education this year by 91 percentage points, pointing to a strong state dependence of over-education. This result suggests that there is little movement into or out of over-education, which is consistent with the descriptive statistics on transitions discussed earlier. This is a large effect, however previous studies in the literature also find significant over-education state dependence. For example, Kiersztyn (2013) reports odds ratios for state dependence in the range of 4.3 to 7.45. Mavromaras and McGuiness (2012) also find considerable state dependence of over-skilling in Australia.

One of the main reasons for the persistence of over-education found in the results is that people do not move jobs much. Nevertheless, the fact that the regression results presented in Table 2 control for job change means that, even holding job change constant (e.g. amongst those who change jobs), those who were over-educated last year are still more likely to be over-educated this year.

Lastly, the coefficient of the initial conditions variable is positive and significant meaning that being over-educated in the first instance a respondent is observed increases his/her chances of being over-educated at present by 2 percentage points.[17] Hence, not only over-education in the last period causes over-education this period but also over-education in the first instance a person is observed in the survey also carries over or causes over-education at present.

---

[17] As a robustness test regressions were also run after dropping those with only two observations in the sample (so just leaving people with three and four observations) so as to eliminate the possibility that initial over-education and the lag of over-education are measuring over-education in the same period. Results are almost identical when this is done.

An interesting extension to the results[18] is to examine whether the extent of an individual's over-education matters. Analysis of the data revealed that all over-educated individuals' qualifications were either one or two levels above the required education level for their occupation. When we placed two separate dummy variables for being over-educated in the previous period by one and two education levels respectively (the reference category remaining as those not over-educated), then both were strongly associated with the likelihood of being over-educated in the current period. It is true that being more seriously over-educated (by two levels) was associated with a higher marginal effect on remaining over-educated than being less seriously over-educated (by just one level), with marginal effects of 88 and 83 percentage points respectively, but it is clear that being over-educated at all, to any extent, is the most relevant factor, the difference being the two marginal effects not being statistically significant at even the 10% significance level.

Turning to the macro determinants of over-education, initial unemployment at the start of paid employment as well as the unemployment rate at the time of the survey do not have a significant effect on the probability of over-education.[19] On the other hand, the annualised change in the labour supply, by educational category and sex, used to serve as an indication of the level of worker competition in the labour market, and the annualised change in the employment share by occupation and sex, serving as a proxy for labour market demand, are both strongly and significantly related to the probability of over-education, and have the expected sign. Specifically, the probability of over-education increases by 0.35 percentage points as the relevant labour supply increases by 1 percentage point, while a 1 percentage point

[18] We are grateful to an anonymous referee for suggesting this additional analysis.

[19] Given that the 16-19 age group was small and unemployment for that age group was significantly higher than for other groups a robustness check when employees below 20 years are not included in the sample was also run. Results from this regression are robust.

increase in the employment share by occupation is associated with a 0.10 percentage points lower likelihood of over-education.[20]

Over-Education State Dependence by Career Stage

Having concluded that state dependence exists in the data and in order to examine whether this self-perpetuating nature of over-education differs depending on the career stage one is in, separate regressions were run restricting the sample to each of three career stages, again using the random effects probit model with Mundlak and initial conditions correction. The aim here was to use a direct measure of the career stage one is in, rather than use age as a proxy, given the diverse educational and work profiles of people of the same age in the sample. For this reason, and since a direct measure of work experience is available in the data, the three career stages were defined based on this variable as follows: 0-3 years of work experience (Early career); 4-9 years of work experience (Early to Mid-Career); 10-20 years of work experience (Mid career).[21] Table 3 shows the marginal effects for the lag of the over-education variable and the initial over-education coefficient for each career stage[22], and reveals that current over-education is only statistically significantly related to initial over-education for those in early-mid careers. Ignoring the early careers group, where the initial effect may be difficult to distinguish from the one-year lagged effect, these results suggest that the effect of being initially over-educated does eventually wear off as individuals progress in their careers.

In terms of the lag of over-education, being over-educated in the past year affects this year's over-education at an increasing rate as work experience rises. Thus, the self-perpetuating

---

[20] The results are almost identical when one macro variable at a time is entered in the regression.

[21] Few in the sample have more than 20 years of work experience, given the sample is restricted to those aged below 40.

[22] Full regression results are available from the authors on request.

nature of over-education is present at all career stages, so that over-education is a permanent phenomenon and that once in it, it is not easy to escape. This becomes increasingly so, the longer an individual has been in the labour market.[23] Being over-educated at a later stage in one's career might send a more negative signal to prospective employers than at an earlier career stage, impeding the chances of over-educated employees escaping over-education and finding a good match. Another channel via which this finding may be explained is that people with more work experience might have been over-educated for longer than people who are just entering the labour market or employees with only 2-3 years of work experience and who might have knowingly accepted a mismatched job on a short-term basis while they gain on-the-job experience. Unlike these early career-stage workers, employees in their early-mid or mid careers might become habituated to their jobs, lowering their on-the job search for a good match and hence having a higher probability of being over-educated in the next year.

5. Concluding Remarks

The present paper employs panel data from the EU-SILC for the period 2005-2011 and examines the factors that affect the probability of being over-educated for one's job. This is done in a dynamic setting that controls for initial conditions and unobserved heterogeneity. In this way, the paper contributes to the literature on the determinants of over-education in general and to the very limited over-education literature specific to Cyprus in particular. Furthermore, the dynamic modelling used in examining the determinants of over-education and the panel nature of the data set used make this paper, to the best of our knowledge, the first of its kind for Cyprus.

---

[23] The career stage 3 marginal effect is significantly larger than the career stage 1 marginal effect.

The first main finding of the paper is that there is strong evidence that over-education is highly self-persistent. Second, when considering how over-education is related to macroeconomic conditions, the results show that the likelihood of over-education varies positively with the supply of similarly-qualified individuals, and negatively with the employment share of the individual's occupation. Such demand and supply factors are more closely related to over-education likelihood than general indicators of labour market slack such as unemployment rates, which are not found to be significantly related to over-education.

Taken together, the implications of these two findings are that education policy that creates a mismatch between the supply of labour by education level and the existing demand at the same levels can create the existence of over-education, which can then become a persistent state for the individuals affected. Thus, over-education is likely to act as a trap rather than as a stepping stone to matched employment, which, the results also demonstrate, is true for workers at all career stages. This is likely due to a scarring effect of over-education, leading to long-term labour market disadvantage.

Policy should therefore be directed towards preventing entry into over-education and discouraging people from accepting mismatched jobs as a career strategy, rather than finding measures to correct it at a later stage or hoping it will correct itself. This could be done, for example, if policies that facilitate entry of young educated people into jobs commensurate with their education are enhanced, for example by subsidising part of their salaries. Such policies provide a motive for firms to employ young graduates directly into matched jobs, hence preventing them from accepting jobs for which they are over-educated and being trapped as a consequence. Similarly, there is a need for programmes specifically designed to offer employment and hence work experience and industry-specific knowledge alongside early career counselling and correct matching by government job centres to help individuals stay out of over-education.

References

Allen, J., & Van der Velden, R. (2001). Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, 53, 434-452.

Arellano, M., & Carrasco, R. (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics*, *115*, 125-157.

Arulampalam, W., & Stewart, M. (2009). Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators. *Oxford Bulletin of Economics and Statistics*, 71, 659-681.

Bacchus, M. (2008). The education challenges facing small nation states in the increasingly competitive global economy of the twenty-first century. *Comparative Education*, 44, 127-145.

Baert, S., Cockx, B., & Verhaest, D. (2013). Overeducation at the start of the career: stepping stone or trap? *Labour Economics,* 25, 123-140.

Battu, H., Belfield, C., & Sloane, P. (2000). How well can we measure graduate over-education and its effects? *National Institute Economic Review*, 171, 82-93.

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

Boll, C., Leppin, J., & Schömann, K. (2016). Who is overeducated and why? Probit and dynamic mixed multinomial logit analyses of vertical mismatch in East and West Germany. *Education Economics*, 24, 639-662.

De Grip, A., Bosma, H., Willems, D., & Van Boxtel, M. (2008). Job-worker mismatch and cognitive decline. *Oxford Economic Papers*, 60, 237-253.

Dolton, P., & Silles, M. (2001). Over education in the graduate labour market: some evidence from alumni data. Centre for the Economics of Education, London School of Economics and Political Science.

Flisi, S., Goglio, V., Meroni, E., Rodrigues, M., & Vera-Toscano, E. (2017). Measuring occupational mismatch: overeducation and overskilling in Europe-evidence from PIAAC. *Social Indicators Research*, 131, 1211-1249.

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

Greene, W. (2008). *Econometric Analysis*. 6th Edition. New Jersey: Prentice Hall

Groot, W., & Van den Brink, H. (2000). Overeducation in the labor market: a meta-analysis. *Economics of Education Review*, 19, 149-158.

Heckman, J. (1981). Heterogeneity and state dependence. In *Studies in Labor Markets* (pp. 91-140). University of Chicago Press.

Kahn, L. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17, 303-316.

Kiersztyn, A. (2013). Stuck in a mismatch? The persistence of overeducation during twenty years of the post-communist transition in Poland. *Economics of Education Review*, 32, 78-91.

Liu, K., Salvanes, K. & Sørensen, E. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84, 3-17.

Linsley, I. (2005). Causes of overeducation in the Australian labour market. *Australian Journal of Labour Economics*, 8, 121-143.

Mavromaras, K., & McGuinness, S. (2012). Overskilling dynamics and education pathways. *Economics of Education Review*, 31, 619-628.

Mavromaras, K., Mahuteau, S., Sloane, P., & Wei, Z. (2012). The persistence of overskilling and its effects on wages. National Centre for Vocational Education Research

McGuinness S. (2006). Overeducation in the Labour Market. *Journal of Economic Surveys*, 20, 387-418.

McGuinness, S., Bergin, A., & Whelan, A. (2018). Overeducation in Europe: trends, convergence, and drivers. *Oxford Economic Papers*, 70, 994-1015.

Menon, M., Pashourtidou, N., Polycarpou, A., & Pashardes, P. (2012). Students' expectations about earnings and employment and the experience of recent university graduates: Evidence from Cyprus. *International Journal of Educational Development*, 32, 805-813.

Mincer, J. (1978). Family migration decisions. *Journal of Political Economy*, 86, 749-773.

Morano, C. (2014). The determinants of overeducation: Evidence from the Italian labour market. *Investigaciones de Economía de la Educación*, 9, 681-698.

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

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49, 1417-1426.

Orme, C. (1997). *The Initial Conditions Problem and Two-Step Estimation in Discrete Panel Data Models*. Mimeo, University of Manchester.

Piper, A. (2013). *A Note on Modelling Dynamics in Happiness Estimations* (No. 49364). University Library of Munich, Germany.

Piper, A. (2015). Heaven knows I'm miserable now: overeducation and reduced life satisfaction. *Education Economics*, 23, 677-692.

Scherer, S. (2004). Stepping-stones or traps? The consequences of labour market entry positions on future careers in West Germany, Great Britain and Italy. *Work, Employment and Society*, 18, 369-394.

Sloane, P., Battu, H., & Seaman, P. (1999). Overeducation, undereducation and the British labour market. *Applied Economics*, 31, 1437-1453.

Tsang, M., & Levin, H. (1985). The economics of overeducation. *Economics of Education Review*, 4, 93-104.

Tsang, M., Rumberger, R. and Levin, H. (1991), The Impact of Surplus Schooling on Worker Productivity. *Industrial Relations*, 30, 209–228.

Verhaest, D. & Omey, E. (2006). The impact of overeducation and its measurement. *Social Indicators Research*, 77, 419-448.

Verhaest, D. & Omey, E. (2010). The determinants of overeducation: different measures, different outcomes? *International Journal of Manpower*, 31, 608-625.

Verhaest, D., & Van der Velden, R. (2013). Cross-country differences in graduate overeducation. *European Sociological Review*, 29, 642-653.

Wooldridge, J. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20, 39-54.

Table 1: Patterns of Over-education History

Over-education History Patterns	Frequency	%
00	302	5.87
000	465	9.03
0000	3436	66.76
0001	20	0.39
001	12	0.23
0011	40	0.78
01	4	0.08
0100	12	0.23
011	15	0.29
0111	32	0.62
10	2	0.04
100	18	0.35
1000	20	0.39
101	3	0.06
11	54	1.05
110	3	0.06
1100	12	0.23
111	129	2.51
1110	4	0.08
1111	564	10.96

Notes: Table shows patterns of overeducation status (1=overeducated, 0=not overeducated) for the periods in which individuals are observed. The frequency (%) column shows the number (percentage) of individuals in the sample who are observed with that pattern.

Table 2: Micro and Macro-level Determinants of Over-Education Regression Results

	Marginal Effects	Marginal Effects
Age	**0.005**	**0.004**
	(0.005)	(0.005)
Work Experience	**-0.003**	**-0.002**
	(0.005)	(0.005)
Married	**-0.06****	**-0.058****
	(0.03)	(0.029)
Female	**0.10****	**0.006**
	(0.005)	(0.007)
Health limitation	**-0.003**	**-0.002**
	(0.008)	(0.008)
Temporary Contract	**0.01***	**0.013**
	(0.007)	(0.008)
Part time	**0.012**	**0.015**
	(0.009)	(0.010)
Past Year unemployment	**-0.04**	**-0.027**
	(0.039)	(0.036)
Over-education at t-1	**0.91*****	**0.89*****
	(0.02)	(0.030)
Recently employed		
a) from unemployment	**-0.001**	**-0.002**
	(0.015)	(0.013)
b)from other inactivity	**-0.017****	**-0.015 ***
	(0.009)	(0.010)
Job Change		
a) Self-induced	**-0.002**	**-0.003**
	(0.008)	(0.008)
b)Employer and other reasons	**0.03***	**0.023***
	(0.015)	(0.014)
Initial Over- education	**0.02****	**0.015****
	(0.008)	(0.007)
Initial unemployment by year of labour market entry and sex	**-**	**0.001**
		(0.002)
Unemployment rate by age group and sex	**-**	**0.001**
		(0.001)
%change in the employment share by occupation	**-**	**-0.10 *****
		(0.04)
%change in the supply of graduates by educational category and sex	**-**	**0.35*****
		(0.106)
N	**3646**	**3646**

Notes: Standard errors clustered around personal ID in parentheses; Significance denoted by: *** p<0.01, ** p<0.05, * p<0.1; Year dummies and individual means over age, years of work experience and marital status as per the Mundlak correction are included but coefficients omitted from the table of results here.

Table 3: State Dependence of Over-Education by Career Stage Regression Results

	Career Stage 1 (0-3 years of work experience)	Career Stage 2 (4-9 years of work experience)	Career Stage 3 (10-20 years of work experience)
Over-education at t-1	0.88*** (0.04)	0.90*** (0.05)	0.97** (0.04)
Initial over-education	0.01 (0.02)	0.023* (0.02)	0.004 (0.01)

Notes: Standard errors in parentheses; Significance denoted by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Table A1: Attrition Check-Descriptive statistics in the First and Fourth Round of Each Sub-Panel in the Original Dataset

Panel:	2005-2008		2006-2009		2007-2010		2008-2011	
	Round 1=2005 n=2259	Round 4=2008 n=1780	Round 1=2006 n=2258	Round 4= 2009 n=1753	Round 1= 2007 n=2195	Round 4=2010 n=1672	Round 1=2008 n=1974	Round 4= 2011 n=1488
Age	44.58 (17.9)	47.48 (17.6)	45.1 (17.9)	49.1 (17.9)	45.09 (18.23)	48.28 (18.07)	46.40 (18.08)	49.75 (17.84)
Education:								
Pre-Primary	0.08 (0.27)	0.09 (0.28)	0.08 (0.27)	0.09 (0.28)	0.07 (0.26)	0.08 (0.27)	0.07 (0.25)	0.07 (0.26)
Primary	0.20 (0.40)	0.22 (0.41)	0.19 (0.39)	0.20 (0.40)	0.20 (0.40)	0.21 (0.41)	0.22 (0.41)	0.23 (0.42)
Lower secondary	0.12 (0.33)	0.10 (0.27)	0.11 (0.31)	0.08 (0.27)	0.12 (0.33)	0.09 (0.28)	0.12 (0.32)	0.08 (0.26)
Upper Secondary	0.38 (0.49)	0.37 (0.48)	0.38 (0.48)	0.36 (0.48)	0.38 (0.48)	0.37 (0.48)	0.35 (0.48)	0.36 (0.48)
Post-sec. non-tertiary	0.02 (0.13)	0.02 (0.15)	0.02 (0.15)	0.02 (0.14)	0.02 (0.12)	0.01 (0.12)	0.02 (0.13)	0.01 (0.12)
1st/2nd stage tertiary	0.20 (0.40)	0.22 (0.41)	0.22 (0.42)	0.25 (0.44)	0.21 (0.41)	0.24 (0.43)	0.23 (0.42)	0.26 (0.44)
Work Experience	22.62 (15.16)	23.54 (15.49)	22.62 (15.25)	24.42 (15.43)	23.24 (15.58)	24.53 (15.54)	24.14 (15.31)	25.4 (15.2)
Married	0.66 (0.47)	0.69 (0.46)	0.65 (0.48)	0.69 (0.46)	0.65 (0.48)	0.66 (0.47)	0.65 (0.48)	0.67 (0.47)
Female	0.51 (0.5)	0.52 (0.5)	0.53 (0.5)	0.52 (0.5)	0.53 (0.5)	0.53 (0.5)	0.51 (0.5)	0.52 (0.5)
Health limitation	0.25 (0.43)	0.21 (0.41)	0.20 (0.40)	0.23 (0.42)	0.24 (0.42)	0.23 (0.42)	0.23 (0.42)	0.28 (0.45)

Note: Standard deviation in parenthesis