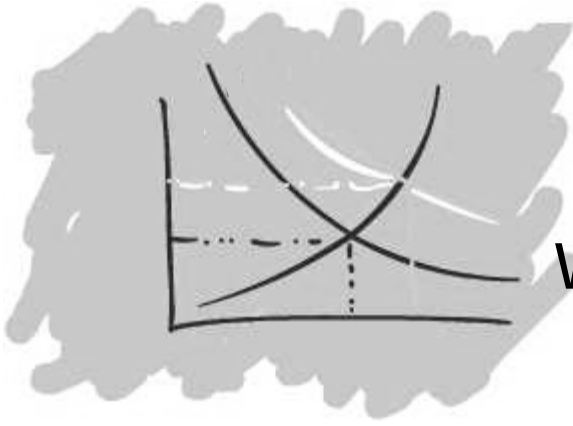# ECONOMIC ANALYSIS

WORKING PAPER SERIES

Overeducation Dynamics and

Personality

◆

Maite Blázquez Cuesta and Santiago Budría

Working Paper 5/2011

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

DEPARTAMENTO DE ANÁLISIS ECONÓMICO:
TEORÍA ECONÓMICA E HISTORIA ECONÓMICA

Overeducation Dynamics and Personality

Maite Blázquez Cuesta[*] – Universidad Autónoma de Madrid

Santiago Budría[**] – University of Madeira and CEEAplA

October 2011

Abstract

We use the 2000-2008 waves of the German Socioeconomic Panel to examine overeducation transitions. The results are based on a first-order Markov model that allows us to account for both the initial conditions problem and for potential endogeneity in attrition. We find that overeducation dynamics, especially the probability of entering overeducation, is significantly influenced by personality. Notwithstanding theses differences associated with individual heterogeneity, still there appears to be considerable overeducation persistence. Almost 18% of the overeducation risk is due to individual state dependence, i.e., the fact of having been overeducated in the previous year.

Keywords: Overeducation dynamics, trivariate probit model, personality traits.

JEL-Codes: C33, I21, J24

---

1. Introduction

A large body of the literature on overeducation has focused on its adverse consequences, mainly in terms of wage penalties, the general finding being that the rate of return to overschooling is positive but lower than the rate or return to adequate schooling. This wage-reducing effect appears to have been strikingly consistent over time and across countries[1]. In contrast, much less consensus has been reached about the main causes of overeducation, and especially about the issue of whether it is a temporary or permanent phenomenon. Several theories in the literature support the view of the phenomenon being a short-term problem. For instance, matching theory (Jovanovic, 1979) suggests that overeducation represents a poor match for workers because they are qualified to perform higher-level jobs. Over time, however, workers are expected to improve their job match. In the same line, the career mobility hypothesis supports that overeducation is part of a career path or insertion process in the labour market. According to this view, workers may enter jobs for which they are overeducated and later on move to jobs that match better their educational attainments (Groot, 1996; Groot and Maasen van den Brink, 1997; Sicherman, 1991; Dekker et al., 2002 and Büchel and Mertens, 2004).

Other theories consider overeducation as a more serious and long-lasting problem. This occurs for instance when the labour market is characterized by imperfect information – Spence's job-screening model – or when the presence of labour market rigidities induces workers to occupy jobs for which they are overqualified. In this line, the works of Dolton and Vignoles (2000) and Green et al. (2002) provide evidence that family responsibilities and/or regional immobility may be some of the reasons behind workers' decisions to voluntarily accept jobs for which they are overqualified. In this line, the work of Büchel and Battu (2003) shows that married women living in rural areas in Germany are more likely to be overeducated, though the gender differential is reduced when commuting times are accounted for. The view of overeducation as a form of allocative inefficiency is also supported by the assignment model (Sattinger, 1993). Under this approach, educational mismatches arise when

---

[1] Using meta-analysis weighting techniques, Leuven and Oosterbeek (2011) report, in the most recent survey on the subject, a return to a year of required schooling of about 9%, whereas a year of overschooling pays-off some 4.5%.

workers are not allocated to jobs in which they have a comparative advantage. Thus, overeducation would persist until a more efficient allocation of individuals to jobs could arise, through either improved matching processes or government policies intended to reduce such inefficiencies.

An alternative explanation of the overeducation phenomenon focuses on the presence of unobserved heterogeneity amongst individuals with the same level of qualification. Bauer (2002) and Chevalier (2003) argue that overeducation mainly reflects unobserved differences in personal characteristics like ability, motivation and other unmeasured skills. Studies that consider differences between the unobservables of workers with and without an adequate job match are, however, very scarce. Nonetheless, it may be the case that the overeducated have a weaker endowment of unobservable characteristics – such as motivation and ability – that influence, not only wage determination, but also the risk of remaining in a poor match permanently.

In this paper we attempt to shed new light on the main factors driving overeducation. This is done in a dynamic perspective based on overeducation exit and entry rates. Our particular contributions are two. First, using the 2000-2008 waves of the German Socioeconomic Panel (GSOEP), we investigate who is more likely to stay overeducated or enter overeducation, and derive measures of state dependence of the process. For this purpose we estimate a first-order Markov model that allows us to control for the endogeneity of initial overeducation status (non-random selection of people being initially overeducated) and for non-random attrition when modelling overeducation transitions. This estimating strategy, novel in the overeducation literature, closely resembles Stewart and Swaffield (1999) and Cappellari and Jenkins's (2004) modelling of the dynamics of low income transitions.

Second, the paper gives special emphasis to the role of personality traits when accounting for overeducation status. This is motivated by a corpus of recent papers documenting non-trivial relationships between non-cognitive skills and labour market outcomes, particularly earnings (Heckman et al., 2006, Semykina and Linz, 2007, Mueller and Plug, 2006, Heineck and Anger, 2010). Thus, in an attempt to capture individual heterogeneity more explicitly, this paper includes personality traits – the Big Five Inventory (BFI) – and locus of control (LOC)

measures as potential determinants of overeducation. In doing so, it provides pioneering estimates on the relationship between overeducation risk and non-cognitive skills.

The outline of the paper is as follows. Section 2 reviews some papers that analyze the extent of overeducation persistence. Section 3 focuses on the relationship between personality traits and labour market outcomes. Section 4 describes the data and the definition of overeducation used in this study. The econometric approach is described in Section 5. Empirical results are reported in Section 6. Section 7 presents the concluding remarks.

2. Overeducation persistence

The presence of state dependence in overeducation is important for policy reasons, as the cost of labour market mismatches for individuals depends on both the size of the wage penalty and on how long that penalty persists. If evidence emerges of serious mismatch in the labour market, with a substantial amount of people failing to obtain jobs that fully utilize their qualifications, then the policy of education expansion may need to be reviewed. This would be the case if overeducation is the consequence of rigidities and adjustment problems in the labour market. In such a case it would lead to serious market inefficiencies – with the excess of supply of highly educated workers creating a bumping-down process of highly qualified workers into lower qualified jobs, eventually crowding out lower educated workers into unemployment.

Although this issue has been recognized in many circumstances, careful and systematic study of the dynamics of overeducation has received relatively little attention in the literature, principally due to the lack of appropriate longitudinal data. Furthermore, the existing evidence is not very conclusive. Some papers provide indirect evidence of substantial state dependence in overeducation. For instance, Dolton and Vignoles (2000) use a survey of UK graduates (of programs normally requiring three or four years to complete) and find that a significant portion of graduates remain overqualified six years after graduation (30%). In the same line, Frenette (2004) uses a representative survey of Canadian graduates to explore whether overqualification among the highly educated is simply a short-term phenomenon or, in contrast, it is a long-term issue for certain graduates. The results show that overqualification is

highly state-dependent, with those who are overqualified two years out being far more likely to still be overqualified five years out. Finally, Mavromaras et al. (2009) estimate a dynamic random effects probit model using the Australia HILDA Survey 2001-2006. They find significant evidence of state dependence in overeducation, which is especially high for individuals with higher levels of education.

Other works have showed, in contrast, that overeducation mobility is relatively high. For instance, Groot and Maassen van den Brink (2004) use a longitudinal sample from the Netherlands to examine the determinants of overeducation entries and exits. They account for unobserved heterogeneity by estimating random effects probit models. At the same time, transitions in and out of overeducation are estimated with a Markov type model (fixed effects), putting special attention to the role of job mobility – distinguishing between job-to-job mobility and internal mobility. They find that job-to-job mobility increases the probability of exiting from overeducation, while internal mobility increases the risk of becoming overeducated. In the same line, Lindley and McIntosh (2010) use the British Household Panel Survey (1991-2005) to estimate the determinants of transitions out of overeducation in the UK. They find that the likelihood of overeducation is negatively related to job tenure, suggesting that overeducation is mostly a temporary phenomenon. However, for a minority of individuals it seems to be a reasonably permanent state.

In this paper we examine overeducation persistence and its main determinants using a trivariate probit model. This type of model allows us to take account of the fact that the set of individuals at risk of exiting overeducation, or the set at risk of entering overeducation, may not be a random sample of the population, an example of an 'initial conditions' problem (Heckman, 1981). At the same time, our estimates of overeducation transitions control for potentially non-random selection into the subsample of individuals who are observed in two consecutive years. This type of multivariate models has been used in studies of earnings mobility (Stewart and Swaffield, 1999; and Cappellari and Jenkins, 2004) and studies on employment transitions (Cappellari and Jenkins, 2008; Cappellari et al., 2010). However, and to the best of our knowledge, this type of technique has not been applied to analyze overeducation dynamics. This is precisely one of the main purposes of the paper.

3. The role of personality traits

In a comprehensive survey, Almlund et al. (2011) document important relationships between personality and a variety of life outcomes, including health, criminal activity, economic success and labour market outcomes. Organizational and industrial psychologists have shown that apart from job performance (see Barrick and Mount, 1991, for a meta-analysis and, more recently, Judge at al., 2007), personality is related in meaningful ways to job satisfaction (Judge et al., 2002) and career success (Kammeyer-Mueller et al., 2008).

The evidence is suggestive of many channels driving these relationships. Personality might be thought of as part of an individual's set of productive traits just as general or specific education or job-related training. In this context, some personality traits are expected to be rewarded in the labour market, while others are expected to be punished. Bowles et al. (2001) suggest that non-cognitive personality variables, such as attitudes towards risk, ability to adapt to new economic conditions, hard work, and the rate of time preference affect both earnings and the transmission of economic status across generations. Mueller and Plug (2006) use US data to estimate traditional Mincerian wage equations based on the human capital model extended to incorporate the Big Five Inventory of personality. Even though the overall contribution of these in explaining the variance in earnings is typically found to be modest, some relevant effects have been established. Substantial earnings disadvantages are associated with Agreeableness and Neuroticism, whereas Openness to experience is positively related to earnings. Overall, they find that non-cognitive abilities have effects comparable to those commonly found for cognitive ability. In Heckman et al. (2006) achievement test scores explain much more of the earnings variance but have similar effects on labour market outcomes as the measures of personality traits. Using pooled Dutch data, Nyhus and Pons (2005) find negative earnings effects for Agreeableness and Neuroticism. Their study documents diverging personality effects by education level, with the effects of Agreeableness and Neuroticism being larger and smaller, respectively, among university graduates. Heckman et al. (2006) and Semykina and Linz (2007) detect relevant effects of external LOC on earnings, although the later study primarily focuses on explaining male-female earnings differentials. In the same line, Heineck and Anger (2010) find that workers who score in the bottom 25% of the LOC scale are exposed to a wage penalty of up to 20% compared to

workers who score in the top 25% of the scale. In their setting, these effects are twice as large as the effects of workers' cognitive abilities.

Several explanations can be put forward to account for these relationships. Personality may affect labour market success indirectly through the type of occupation chosen. Judge et al. (1999) document non-trivial interactions between specific personality traits and occupational attainment across the life span. This evidence is well in line with sorting theories suggesting that some of the five personality dimensions may predict "extrinsic" career success if personality traits fit the psychological requirements of the job. Conscientiousness (Ham et al. 2009) and LOC (Heckman et al., 2006) predict sorting into occupations. Similarly, personality may be intertwined with the individual education path. Heckman et al. (2010) estimate a model of sequential educational choice and find that personality, as measured by participation in adolescent risky behaviours, primarily affects age 30 earnings through its effects on education. Still, they find that for a given educational attainment the effects of personality variables on outcomes are weak. In Borghans et al. (2011) substantial portions of the variance in achievement test scores and grades, which are often used as measures of cognition, are explained by personality variables.

Other mechanisms might be the mediating role that personality exerts on absenteeism (Störmer and Fahr, 2010), the likelihood of self-employment (Caliendo et al., 2008) and unemployment duration (Caliendo et al., 2010). Dohmen at al. (2009) highlight the role of reciprocity in generating better labour market prospects by showing that reciprocal workers are paid more, exert greater effort on the job, are more likely to be employed and report higher life satisfaction. Moreover, in a bargaining model context, personality differences may lead to differences in pay if certain traits tend to concede too quickly.

In sum, there is an emerging literature concerned with assessing and providing explanations for the conspicuous relationship between personality and labour market outcomes. One of the contributions of this paper is to extend the debate to the realm of educational mismatches and, more specifically, overeducation. To the extent that the incidence of overeducation has important implications for earnings and job satisfaction, this purpose is interesting in itself. Moreover, overeducation may exert a mediating role in the empirical relationships

documented so far. Although assessing this role is not the object of the present investigation, the results presented here represent a preliminary step.

4. Data and measurement of overeducation

Conducted in Germany since 1984, the GSOEP is a wide-ranging representative longitudinal study of households that contains a large set of personal, family and labour market characteristics of household members. The unit of analysis in the present paper are individuals. After dropping observations with missing values in the relevant variables we retain 71,321 observations.

Workers are classified as either adequately educated or overeducated according to a statistical approach based on the ISCO88 classification disaggregated on a 3-digit level. Required schooling is defined as a one standard deviation range around the mean level of schooling within an occupation. Workers are considered to be adequately educated if their actual education falls within this range and overeducated if their actual education is greater than one standard deviation above the mean for the specific occupation. Occupations with less than 10 observations in a year are excluded from the analysis. Alternative definitions with modal value and 2-digit occupations were also examined.

Table 1 contains the summary statistics of the estimating sample. The proportion of overeducated workers amounts to 14.4%[2]. The average educational attainment is 12.5 years of schooling. Tenure amounts to 11.2 years on average, with 26.7% of the sample having a temporary contract. 14.7% of the workers are self-employed and women account for 46.4% of the sample. Most individuals are married or live with a partner (67.5%), albeit a significant fraction is single (23.6%). Finally, 9.8% of the respondents declare that they have a "bad" or "very bad" health condition. In the regression stage, the continuous variables (age, years of completed education, tenure, number of children and adults at home) are entered in their

---

[2] This prevalence rate is towards the bottom range of the bunch of estimates reported in the literature. Using the modal value instead of the average years of schooling within occupations results into 26% of the sample being overeducated, a figure that comes closer to standards. We disregarded this alternative measurement method because empirically it results in a less demanding criterion for the definition of overeducation.

logarithm form.

4.1. Personality traits

All the items included in the Big Five Inventory, as well as the LOC and reciprocity items are to be answered on 7-point Likert type scales (1 — "disagree completely" (LOC)/"does not apply to me at all" (BFI) to 7 — "agree completely" (LOC)/"applies to me perfectly" (BFI)). The BFI represents a widely accepted approach to conceptualizing personality, as meta-analyses consistently support the construct validity of this approach (Costa and McCrae, 1992). After aggregating across items, the BFI provides a score for the five major traits that define human personality across cultures: Conscientiousness, a tendency to be organized, strong-willed, persistent, reliable, and a follower of rules and ethical principles; Neuroticism, the tendency to experience negative emotions such as anxiety and depression; Extraversion, the tendency to be sociable, warm, active, assertive, cheerful, and in search of stimulation; Agreeableness, the dimension of interpersonal relations, characterized by altruism, trust, modesty, and cooperativeness; and Openness to experience, the tendency to be imaginative, creative, unconventional, emotionally and artistically sensitive. Each of these factors is hierarchically defined by specific facets. The BFI questionnaire used in the 2005 wave of the GSOEP is based on 3 items per personality dimension. Despite psychologists typically work with longer questionnaires, the shortened version used in this paper, known as the BFI-S, has been validated against longer inventories[3]. A further issue in personality measures is the concern that variability in the resulting scores arise from measurement error. In our data, encompassing tests of internal consistency were satisfactory[4].

---

[3] The BFI-S items are: I see myself as someone who… i) worries a lot, ii) gets nervous easily, iii) is relaxed, handles stress well; iv) is communicative, talkative, v) is outgoing, sociable, vi) is reserved; vii) is original, comes up with new ideas, viii) values artistic experiences, ix) has an active imagination; x) is sometimes somewhat rude to others, xi) has a forgiving nature, xii) is considerate and kind to others; xiii) does a thorough job, xiv) does things effectively and efficiently, xv) tends to be lazy. Neuroticism: i)-iii), Extraversion: iv)-vi), Openness to experience: vii)-ix), Agreeableness: x)-xii), Conscientiousness: xiii)-xv). By construction, the score in a given dimension ranges from a minimum of 3 to a maximum of 21.

[4] A principal component analysis with varimax rotation was conducted. Factor analyses clearly replicated the Big Five factors by yielding a correlation matrix with five eigenvalues above unity. The five principal components accounted for 60.7% of the total variance. The Cronbach's alphas for the five dimensions were 0.607, 0.657, 0.625, 0.505 and 0.609, respectively. It must be noticed that for a given level of internal consistency, fewer items per dimension result into lower alphas (Mueller and

Locus of control (LOC) is measured externally and internally. People with an external LOC believe that his/her behaviour is guided by fate, luck, or other external circumstances, while those with an internal LOC believe that his/her behaviour is guided by his/her personal decisions and efforts. Within psychology, LOC is considered to be an important aspect of personality and, more importantly, the extent to which one finds social comparisons inspiring or threatening is known in the field of psychology to depend on whether one finds a sense of control over the dimension under evaluation (Lockwood, 2002). In the GSOEP, LOC is surveyed with 10 items of which four relate to internal and the other six measure external LOC[5]. Unfortunately, internal LOC was found to exhibit a very limited amount of construct validity in the data[6], meaning that the surveyed items are not at all appropriate for measuring the underlying scale. This forced us to excluded internal LOC from the analyses.

The average scores (in a 1-7 scale) of the different personality dimensions are reported in Table 2. In the regressions stage of the paper, these were normalized to a mean zero and unit variance. Figure 1 shows the density function of the normalized scores. The graphs show that only a few of the traits are not normally distributed, namely Conscientiousness and Agreeableness, which are left-skewed. As a final remark, the BFI and LOC information gathered in the 2005 wave of the GSOEP were not surveyed in any of the previous or subsequent years. To deal with this limitation, it is assumed thorough the paper that these constructs are constant over the 7-years period[7].

Plug, 2006). Hence, although these reliability coefficients are towards the lower range of admissible values, they point to a reasonable amount of internal consistency given the low (3) number of items per personality traits.

[5] These are: i) My life course depends on me, ii) Influence on social conditions through involvement, iii) Success takes hard work, iv) Doubt my abilities when problems arise; v) Haven't achieved what I deserve, vi) What you achieve depends on luck, vii) Others make the crucial decisions in my life; viii) Possibilities are defined by social conditions; ix) Abilities are more important than effort; x) Little control over my life. Internal LOC measured by i)-iv). External LOC measured by v)-x).

[6] The alpha reliability coefficient was as low as 0.201.

[7] This should not be seen as a stringent assumption, as it is generally accepted that adult's personality traits are fairly stable over time (Roberts and Del Vecchio, 2000, Costa and McCrae, 2002). In our sample, the respondents mean age is 41.5 years and they are interviewed during no more than 7 consecutive years, so that the potential interdependency between early life events and personality should not matter much.

4.2 Overeducation transitions

An informative way of looking at movements into and out of overeducation is in terms of conditional probabilities. Restricting the attention to employed individuals, Table 3 shows the probabilities of being overeducated in year t given an individual's state – either overeducated or not – at $t-1$. The transitions are pooled over the period 2000-2008.

As it can be observed, the likelihood of being overeducated in a given year differs substantially depending on overeducation status in the previous year. The persistence rate is remarkably large, with only 13.6% of workers exiting from mismatched jobs in a given year, whereas only 2.0% of those who were not overeducated one year are overeducated the following year. In other words, the next year's overeducation risk among the overeducated is as much as 84.4 percentage points (pp) higher than for those who are not overeducated. A second aspect that should be taken into consideration when analysing transitions is the high and non-ignorable attrition problem that most panel surveys frequently suffer. As other longitudinal data, the GSOEP surveys not only the original sample from the first wave, but also households and persons that entered the survey at later points in time. While the original sample members are augmented with the entrance of new members, there are other households that leave the survey for several reasons: survey-related reasons (unsuccessful follow-up and refusal) and reasons unrelated to the survey (moves abroad and deaths). The "Missing" column in Table 3 reveals that a non-trivial proportion of the sample used in this study were not retained from one year to the next. Furthermore, the retention rates slightly vary between those who were initially overeducated (15.8%) and those who were not (17.1%).

The descriptive analysis provides preliminary evidence of a strong state dependence effect in the dynamics of overeducation. Broadly speaking, state dependence is defined as the degree to which the effect of any initial endowment (often referred to as individual characteristics) on an outcome may be attenuated or accentuated by the continued presence of that outcome. However, it does not necessarily imply that this state dependence, observed in aggregate terms, is true at the individual level. It may be the result of *heterogeneity*, where certain individual characteristics increase the probability of being overeducated. This would create

the appearance of state dependence in the aggregate transition probabilities if some of the relevant characteristics exhibit persistence over time (e.g. education). Alternatively, or in addition, there may be 'true', or structural, state dependence for individuals. That would be the case if being overeducated in one period increases *in itself* the probability of being over-educated in the next period, even relative to another individual with identical characteristics who was not overeducated at the previous period. Distinguishing between omitted heterogeneity and structural state dependence is of paramount importance: while state dependence due to heterogeneity can be influenced by changing individual characteristics, structural state dependence may be harder to tackle. Although addressing this issue is difficult – it requires additional information, such as the availability of suitable instruments – this distinction must be explicitly modelled in the estimating framework for the correct understanding of outcomes that persist over time. This is precisely described in the next section.

5. The econometric approach

This section considers the modelling of transitions into and out of overeducation. Hazard regressions models have been commonly used in the literature to study dynamics in different fields, such as employment and poverty. Generally, in this type of models unobserved factors are assumed to be independent of entry to the state, and the problem of attrition is commonly ignored.

However, when modelling transition probabilities, it is important that the 'initial conditions' problem is addressed (Heckman, 1981). It can be viewed as an endogenous selection problem: unobserved factors affecting transitions between states may be correlated with factors determining the original state. At the same time, one of the potentially most damaging effects on the value of panel data is the presence of biasing attrition, i.e. attrition selectively related to outcome variables of interest. As Atkinson and Micklewright (1991) point out, even a loss of 5% of the sample each year will eventually have a substantial cumulative impact on the proportion of the original sample remaining in the panel. If attrition bias is selective and concentrated among certain types of individuals, it may lead to biased estimations.

5.1 The model

We consider a sample of individuals in employment situation at period $t-1$. The risk of overeducation can be written in terms of an unobserved latent overeducation propensity o^*_{it-1} which is a linear function of observable characteristics plus an error term normally distributed:

$$o^*_{it-1} = \beta' z_{it-1} + \varepsilon_{it-1}, \quad \varepsilon_{it-1} \sim N(0,1) \tag{1}$$

where $i=1....N$ indexes individuals, z_{it-1} is the vector of explanatory variables that includes personal, household and job characteristics, and ε_{it-1} is an error term that can be written as the sum of an individual-specific component plus an orthogonal white noise error: $\varepsilon_{it-1} = \xi_i + \vartheta_{it-1}$. Let $O_{it-1} = I(o^*_{it-1} > 0)$ be a dummy variable indicating whether the individual i is over-educated in year $t-1$, where the indicator function $I(\ )$ equals one if its argument holds, and zero otherwise.

Transitions into and out of overeducation are only possible if the individuals are also observed in period t. We denote r^*_{it} as the sample retention propensity between years $t-1$ and t:

$$r^*_{it} = \psi' x_{it-1} + \tau_{it}, \quad \tau_{it} \sim N(0,1) \tag{2}$$

where x_{it-1} is a vector of explanatory factors, and the error term $\tau_{it} = \mu_i + \lambda_{it}$ is the sum of a normal individual component, μ_i, plus a normal orthogonal white noise, λ_{it}. Let $R_{it} = I(r^*_{it} > 0)$ be a binary indicator of permanence in the sample for each individual.

The third component to be specified is that characterising overeducation status in period t, which is observed only if $R_{it} = 1$. Let the latent propensity of overeducation be characterized by:

$$o^*_{it} = [O_{it-1}\delta'_1 + (1 - O_{it-1})\delta'_2]m_{it-1} + v_{it}, \quad v_{it} \sim N(0,1) \tag{3}$$

where the vectors δ_1, δ_2 and m_{it-1} are column vectors, and the error term v_{it} is the sum of an individual specific effect plus an orthogonal white noise: $v_{it} = \varphi_i + \zeta_{it}$. We define an indicator variable $O_{it} = I(o_{it}^* > 0)$ for overeducation in year t. This specification allows the impact of explanatory variables on current overeducation to differ according to the individual status in the previous period. Thus, Eq. (3) allows us to examine the conditional probability of the event and, therefore, overeducation transitions[8].

Finally, we allow the unobservable factors in Eqs. (1)-(3) to be jointly distributed as a trivariate normal with zero means, unit variances, and the following three unrestricted correlations:

$$\rho_1 = corr(\varepsilon_{it-1}, \tau_{it}) = cov(\xi_i, \mu_i)$$
$$\rho_2 = corr(\varepsilon_{it-1}, v_{it}) = cov(\xi_i, \varphi_i) \qquad (4)$$
$$\rho_3 = corr(\tau_{it}, v_{it}) = cov(\mu_i, \varphi_i)$$

Thus, the distribution of unobserved heterogeneity is parameterized via the cross-equation correlations. Correlation ρ_1 summarizes the association between unobserved individual specific factors determining base year overeducation status and sample retention. A positive (negative) value of this correlation term indicates that individuals who were more likely to be initially overeducated are more (less) likely to be in the sample next year. The correlation between unobserved factors affecting both current and previous overeducation status is captured by ρ_2. A positive (negative) sign indicates that individual who were more likely to be initially overeducated are more (less) likely to remain overeducated one year after, compared to those who were initially non-overeducated. Finally, ρ_3 is the correlation between unobserved individual specific factors determining retention propensities and current status.

5.2 Overeducation transition probabilities

Our main interest is on overeducation transition probabilities. The estimation of a trivariate probit allows us to predict the probability of being overeducated at t conditional on being

---

[8] A more complex analysis would estimate a simultaneous model including another equation for educational achievement, insofar as it may be related to personality traits. Unfortunately, the resulting model would be cumbersome and require additional instruments that render the analysis unfeasible with the available data.

overeducated at $t-1$ (the overeducation persistence rate), and the probability of being overeducated at t conditional on being non-overeducated at $t-1$ (the overeducation entry rate). These probabilities are characterized by the following expressions:

$$p_{it} \equiv Pr(O_{it} = 1|O_{it-1} = 1) = \frac{\Phi_2\left(\delta_1'm_{it-1}, \beta'z_{it-1}; \rho_3\right)}{\Phi(\beta'z_{it-1})} \tag{5}$$

$$e_{it} \equiv Pr(O_{it} = 1|O_{it-1} = 0) = \frac{\Phi_2(\delta_2'm_{it-1}, -\beta'z_{it-1}; -\rho_3)}{\Phi(-\beta'z_{it-1})} \tag{6}$$

where Φ and Φ_2 are, respectively, the univariate and bivariate standard normal cumulative distribution functions. A feature of first-order Markov models is that they provide closed-form solutions for the spell duration of the phenomenon under consideration. It can be shown that, in our case, the mean duration of an overeducation spell is given by $1/(1-p_{it})$, where $(1-p_{it})$ defines the overeducation exit rate, whereas $1/(e_{it})$ represents the mean duration of a non-overeducation spell. Finally, it is worth noting that while the aggregate transitions represented in Table 3 only refer to the subsample with $R_{it} = 1$, the transitions probabilities of Eqs. (5) and (6) are estimated also for the subsample of individuals who are not observed in year t (those with $R_{it} = 0$).

5.3 The likelihood function

The trivariate probit regression model includes an endogenous dummy variable (lagged overeducation), endogenous switching (in year t equations), and endogenous selection (of year t equations with respect to attrition). Thus, the likelihood function involves normal integrals of various dimensions, the largest being three. The contribution to the log-likelihood for each individual i with overeducation status observed in period $t-1$ is:

$$\begin{aligned} logL_i = {} & O_{it-1}R_{it} \log[\Phi_3(k_i\delta_1'm_{it-1}, n_i\psi'x_{it-1}, q_i\beta'z_{it-1}; k_in_i\rho_3, k_iq_i\rho_2, n_iq_i\rho_3)] \\ & +(1 - O_{it-1})R_{it} \log[\Phi_3(k_i\delta_2'm_{it-1}, n_i\psi'x_{it-1}, q_i\beta'z_{it-1}; k_in_i\rho_3, k_iq_i\rho_2, n_iq_i\rho_1)] \\ & +(1 - R_{it}) \log[\Phi_2(n_i\psi'x_{it-1}, q_i\beta'z_{it-1}; n_iq_i\rho_1)] \end{aligned}$$

where $k_i \equiv 2O_{it} - 1$, $n_i \equiv 2R_{it} - 1$, $q_i \equiv 2O_{it-1} - 1$

The estimation of the model requires evaluation of multivariate normal probability distribution functions. Our sample data consists of repeated observations on the same individual across successive periods of time. These repeated observations mean that the *i.i.d.* assumption is violated. To account for this we use a pseudo maximum likelihood estimator (PSML). There are methods in the statistics literature that allow for arbitrary correlations between observations within the same sample cluster (Huber, 1967; and Binder, 1983). In such methods, the sample log-likelihood is a "pseudo-likelihood" (Gourieroux and Monfort, 1996), from which a robust variance estimator of the parameter estimates can be derived using Taylor-series linearization. Thus, we evaluate the trivariate standard normal distribution function using simulation methods based on the Geweke-Hajivassiliou-Keane (GHK) simulator with 250 random draws.[9]

5.4 The covariates

We consider personal, demographic and job characteristics as the main determinants of overeducation. This includes gender, age, marital status, education, health status, variables referring to household composition (number of adults and children at home) and job characteristics (type of contract and job tenure). The main novelty in this respect is the inclusion of explicit controls for the respondent's personality traits as measured by the BFI and LOC constructs described above. This refinement is intended to capture individual heterogeneity more explicitly and to provide pioneering estimates of the impact that personality exerts on the likelihood of (entering and exiting) overeducation. To account for education-related specific effects of personality, these indicators are interacted with the individual level of schooling. All equations include year fixed effects.

All covariates in the Eq. (3) were measured using their values at year $t - 1$. Identification of the model requires some exclusion restrictions. First, we require variables that affect initial overeducation status and retention while having no effect on transitions, i.e., variables entering the z_{it-1} or x_{it-1} vectors but not m_{it-1}. As instrument for sample retention we

---

[9] The GHK simulator works by taking draws from upper truncated univariate standard normal distributions, and then recursively computing a multivariate probability value using Cholesky's factorization. For maximization, we use the modified Gauss-Newton routine implemented by Stata's *ml* command jointly with a cluster option to provide robust estimators.

include a binary variable indicating whether the individual was a GSOEP original sample member. With this identifying restriction we assume that these individuals are more stable survey members compared to joiners, and that sample membership status is orthogonal to the overeducation transition propensity. Thus, vector x_{it-1} includes all the variables in m_{it-1} plus the original panel member indicator. Two set of instruments were explored for base year overeducation. The first set referred to the respondent's parental socioeconomic background as measured by the parent's occupation when the respondent was aged between 14 and 15. These variables have been previously found to have a significant effect on the probability of having a poor labour market status as measured by low pay (Stewart and Swaffield, 1999, Cappellari and Jenkins, 2004). Unfortunately, the validity of these instruments was not supported by our data. As an alternative, we use the respondent's quality of relationship with parents. In particular, vector z_{it-1} contains all variables included in m_{it-1} plus a set of 12 binary variables, including variables to indicate missing information, describing the frequency (ranging from "very often" to "never") of unpleasant arguments and fights with father and mother when the individual was between 14 and 15 years old. The assumption being made here is that a poor relationship positively affects the likelihood of initial overeducation, whereas it does not affect the transition probabilities conditional on the previous period state.

6. Results

For expositional purposes, the discussion of the results is done in four stages. Firstly, we report on the estimated correlation between the unobservable component of the three equations and the encompassing exogeneity tests. Secondly, we concentrate on the effects of the explanatory variables on the overeducation transitions. Thirdly, we assess the extent of state dependence, differentiate between individual or 'true' state dependence and aggregate state dependence. Finally, we move on to explore the role of individual heterogeneity in explaining overeducation persistence, giving special attention to respondents' non-cognitive skills.

6.1 Testing the model specification
In order to assess the exogeneity of both initial conditions and sample retention, we test for the separate and joint significance of the correlation coefficients associated with the two

selection equations. Table 4 reports the estimated correlation structure of the unobserved factors jointly with the tests of ignorability of attrition and initial conditions.

The correlation between unobservables affecting retention and initial overeducation (ρ_1) is negative, indicating a lower retention propensity among the initially overeducated relative to the non-overeducated. Still, this statistic fails to be significant at conventional levels. More interestingly, the correlation between unobservables affecting initial overeducation and conditional current overeducation status (ρ_2) is positive and statistically significant. This supports the existence of unobserved factors that are positively related to the likelihood of being initially overeducated and, at the same time, increase the risk of overeducation persistence. As for the correlation between retention and transitions (ρ_3), the coefficient is positive but imprecisely estimated.

Exogeneity of initial conditions would imply that ρ_1 and ρ_2 are jointly zero, but such a hypothesis is strongly rejected (Wald test $p < 0.000$). On the other hand, exogeneity of retention would imply $\rho_1 = \rho_3 = 0$. In this case, the null hypothesis is not rejected, a result that is unsurprising given that ρ_1 and ρ_3 are imprecisely estimated. Finally, the test of joint significance of the three correlation coefficients indicates that they are jointly significant with a p-value of 0.000 or, to put it differently, that the two endogenous selection processes cannot be ignored when describing the dynamics into and out of overeducation. Still, inspection of the pair-wise tests warns that endogeneity operates mainly via a correlation with initial conditions.

Regarding the validity of the instruments, the results of Table 4 show that the relationship quality with parents and sample membership variables can be excluded from the transition equation separately (p-values of 0.166 and 0.218, respectively) and jointly (p-value = 0.178). On the other hand, the relationship variables are jointly highly significant (p-value = 0.034) in z_{it-1}, whereas original panel membership is a relevant covariate for retention (p-value = 0.000). Thus, according to the tests for inclusion, the validity of the proposed instruments is supported by the data.

6.2. The effect of explanatory variables on transition probabilities

Table 5 reports the maximum likelihood estimates for the variables in m_{it-1}, discriminating between the determinants of overeducation persistence (first column) and the determinants of overeducation entry (second column).[10] According to the test reported in the last row of the table, the equality of the two set of estimates ($\delta_1 = \delta_2 = 0$) is strongly rejected. Thus, there is a statistically relevant state dependence effect, according to which initial overeducation condition importantly determines the respondent's next period status.

As expected, the dimension most closely related to overeducation is attained schooling. More educated individuals are significantly more likely to be overeducated, and this occurs regardless of their status at $t-1$. However, this is not the general pattern. Most variables exert a differential effect depending on whether the individual was initially overeducated or not. This is the case of singlehood, which significantly reduces the probability of overeducation persistence (-0.161) while rendering the probability of overeducation entry unaltered (t-ratio = 0.75). The role of gender is exactly the opposite: women are less likely to enter overeducation (-0.095) but as likely as men to remain overeducated (t-ratio = -0.03).

Interestingly, we detect some differences in the dynamics of overeducation by personality groups. Specifically, we find that scoring high in Conscientiousness, Extraversion and external LOC, and scoring low in Openness significantly decreases the probability of remaining overeducated. Interestingly, the impact of these personality traits is strongly moderated by the individual level of schooling. The interaction terms between schooling, Conscientiousness, Extraversion and external LOC are significant at conventional levels and indicate that personality differences are relatively less important when accounting for overeducation persistence among the high educated. Among the non-overeducated at $t-1$ (second column), Conscientiousness an Extraversion are less relevant dimensions, whereas the beneficial effects of external LOC on the probability of overeducation are higher by a factor of almost 2 (-0.0761 against -1.265). Among this group, the role of personality differences is also significantly moderated by schooling levels.

---

[10] The estimates for initial overeducation and retention are available from the authors upon request.

As for the remaining variables, a higher number of adults in the household significantly raises the probability of remaining overeducated, while it leaves the probability of entering overeducation practically unaltered. Similarly, the self-employed are more likely to remain overeducated, while they are as likely to enter overeducation as the remaining workers. Having a temporary contract, by contrast, has a positive effect on the entry probability and a non-significant effect on the likelihood of overeducation persistence. Similarly, the beneficial effects of tenure on the overeducation probability are more conspicuous among the non-overeducated at $t-1$.

6.3 State dependence

This section is aimed at disentangling the puzzle of the main forces causing overeducation persistence. First, we address the question of how much state dependence there is in the conditional probability of remaining overeducated. The probabilities being considered above are aggregate probabilities. They do not necessarily imply that this state dependence observed in aggregate holds at the individual level. If certain individual characteristics increase the probability of an individual remaining overeducated, then observed state dependence may be the result of heterogeneity: if some of these characteristics exhibit persistence over time (e.g. education), this will create the appearance of state dependence in the aggregate transition probabilities even if such an effect is absent in individual transition probabilities

In addition, there may be individual state dependence: being overeducated in one period may in itself increase the probability of being overeducated in the next period, even relative to another individual with identical characteristics who was not overeducated in the first period. That would be the case, for example, if a spell of overeducation negatively affects an individual's perception of his market value and discourages him from applying for jobs commensurate with his level of education. Similarly, holding a mismatched job may alter workers' preferences or motivations in such a way as to make them more likely to remain in that segment of the labour market. This would be particularly the case if the mismatched status comes in exchange of other forms of compensation (commuting time, human capital accumulation, flexibility). A third explanation has to do with human capital losses: having excess education may cause the depreciation of human capital not currently being used,

thereby keeping an individual's productivity low and reducing the probability of achieving a good match in the future.

The calculations needed to assess the extent of aggregate against individual state dependence are reported in Table 6. Following Stewart and Swaffield (1999), for each individual we calculate the predicted conditional probability of being overeducated at t given being overeducated at $t-1$ (Eq. (5)). These are then averaged over first those overeducated at $t-1$ and then those non overeducated at $t-1$. The resulting figures, 0.862 and 0.169, respectively, are reported in rows 4 and 5 of the table. The difference between the two figures, immediately below (0.694), is the contribution that is *not* due to state dependence, insofar as these probabilities are conditional on an identical overeducation condition for both groups of individuals.

The state dependence effect is then calculated as the difference between the average probability of being overeducated at t conditional on being overeducated at $t-1$ over the sample who were non-overeducated at $t-1$ (0.169), and the raw aggregate probability of being over-educated at t over the same sample (0.020). This gap is due to differences in the initial overeducation status among otherwise similar individuals and, as such, it captures the extent of individual state dependence. According to the estimates, being initially overeducated increases *in itself* the risk of remaining over-educated in the next period. Still, the contribution of individual state dependence in the estimated models is moderate, with 17.6% of the difference in aggregate probabilities being due to the fact of having been overeducated at $t-1$, holding characteristics fixed.

6.4 Individual heterogeneity: the role of personality

This section complements the analysis by exploring the role of personality traits in explaining overeducation persistence and entry rates, as well as average and median lengths of time spent in and out of overeducation. The magnitude of the differential between mean and medians captures the degree of dispersion in spell lengths among individuals with identical observable characteristics.

We proceed as follows. First, we replace z_{it-1} in (1) by its mean over the subsample of overeducated individuals, and we compute the predicted probability of either remaining overeducated at period t – the persistence rate – and the predicted probability of entering overeducation – the entry rate. Taking this as the reference outcome, we then compute the same probabilities augmenting the corresponding value of each dimension of personality by 1 standard deviation, while holding the rest of covariates fixed at their mean values. The difference between these two probabilities captures the impact of the BFI and external LOC variables on overeducation transitions. Furthermore, and in order to account for the relative importance of personality, we compute the same difference in probabilities when years of schooling are increased by 1 standard deviation. Finally, we describe the joint effect of changes in schooling and personality.

The results are reported in Table 7. For the reference overeducated individual, the probability of remaining overeducated is 89.2%, whereas the probability of entering overeducation is 53.8%. These figures imply a state probability of overeducation of 83.3%, a mean (median) overeducation spell of 9.3 (6.1) years and mean (median) time between overeducation spells of 1.9 (0.9) years. As it can be observed, the probability of remaining overeducated increases by 5.5 pp when we simulate an increase of 1 standard deviation in average years of schooling. The impact on the entry rate is even larger (26.0 pp). These impacts are remarkable and highlight the close relationship between attained schooling and overeducation status. The close relationship between personality traits and overeducation is perhaps more surprising. A 1 standard deviation increase in Conscientiousness decreases the persistence and entry rates by 10.5 and 31.0 pp, respectively, relative to the reference individual, resulting in 46.5 pp shorter overeducation spells (4.95 against 9.26 years). An identical change in Extraversion results into even lower persistence and entry rates (-13.9 and -38.3 pp, respectively, relative to the reference individual), shorter average overeducation spells (-53.5 pp) and longer non-overeducation spells (62.1 pp). Despite to a lower extent, scoring high on External LOC has also a beneficial effect on the overeducation risk. By contrast, Neuroticism, Agreeableness and, especially, Openness are associated with higher overeducation risks. This is mainly due to sensitive increases in the corresponding entry rates, which are between 5.0 and 51.5 pp above that of the reference individual, while differences in persistence rates are much smaller.

All in all, the results show that personality differences play an important role in explaining overeducation transitions, being especially responsible for large differences in the overeducation entry rate. In the last rows of Table 7 we consider simultaneous changes in education and personality traits. This is done by reporting variations in probabilities following a 1 standard deviation increase in the schooling level *and* a similar increase in a given personality trait. There is a non-trivial interaction between schooling and personality, as suggested in the previous sections. For example, being above average in terms of schooling and Conscientiousness does not result in a higher overeducation risk relative to the reference individual. The corresponding statistics are indeed very close, with persistence and entry rates 2.0 and 3.2 pp lower, respectively, than that of the reference individual. This result suggests that a particular personality trait – Conscientiousness – may more than offset the effects of attained schooling on the risk of overeducation. This is also the case of Extraversion and, to a lower extent, external LOC. By contrast, the 10.5 pp higher state probability of overeducation among the more educated is barely affected by a higher score in Neuroticism (11.6 p) and Agreeableness (12.4 pp). Finally, the simultaneous effect of schooling and Openness on the overeducation risk is indeed lower than the sum of the independent effect of these variables.

A final lesson from Table 7 is that entry rates exhibit a greater amount of individual heterogeneity than persistence rates. This suggest that a substantial part of the variation in the state probability of overeducation is due to differences in the overeducation spell duration and not to differences in the lengths of time spent out of overeducation. The last column in the table shows that absolute differences in the latter case are relatively modest.

7. Concluding remarks

Using the 2000-2008 waves of the German Socioeconomic Panel (GSOEP), this paper shows that structural state dependence in overeducation is moderate. This was done estimating a first-order Markov models that allows us to control for the endogeneity of initial overeducation status (non-random selection of people being initially overeducated) and for non-random attrition.

The estimation of conditional probabilities shows that there appears to be substantial heterogeneity in the rates of movement into and out of overeducation. Apart from the effect of personal, demographic and job characteristics, our results reveal that overeducation dynamics are significantly influenced by personality traits. In this respect, the paper contributes to the emerging literature that relates non-cognitive skills with labour market outcomes. In particular, we find that some dimensions of personality like Conscientiousness, Extraversion and external LOC exert a higher impact on overeducation transitions – especially regarding entry rates – than that caused by attained schooling. This result is in line with other works in the literature suggesting that overeducation mainly reflects unobserved differences in personal characteristics like ability or motivation.

Notwithstanding the substantial differences in overeducation transitions associated with individual heterogeneity, the results uncover a non-trivial state dependence in overeducation. In particular, we find that 17.6% of the observed persistence in overeducation is due to the fact of having been overeducated in the previous year, holding characteristics fixed. This implies that a non negligible amount of overeducated people fails to move to more adequate jobs regardless of their personal characteristics, including non-cognitive skills. This observation is supportive of the view that rigidities and/or adjustment problems in the labour market make overeducation a serious and long-lasting problem. Tackling against educational mismatches would, therefore, require the necessity of reviewing the policy of education expansion.

References

Almlund, M., A.L. Duckworth, J.J. Heckman and Kautz, T.D. 2011. *Personality Psychology and Economics*, NBER Working Papers 16822, National Bureau of Economic Research, Inc.

Atkinson, A.B. and J. Micklewright. 1991. Unemployment Compensation and Labour Market Transitions: A Critical Review. *Journal of Economic Literature* 29: 1679-1727.

Barrick, M.R and M.K. Mount. 1991. The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology* 44(1): 1-26.

Bauer, T. 2002. Educational mismatch and wages: a panel analysis. *Economics of Education Review*, 21(3): 221–229.

Binder, D.A. 1983. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* 51: 279-292.

Borghans, L., B. Golsteyn, J.J Heckman and J.E Humphries. 2011. *Identification Problems in Personality Psychology*. IZA Discussion Paper 5605.

Bowles, S., H. Gintis, Herbert and M. Osborne. 2001. The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature* 39(4): 1137-1176.

Büchel, F., and H. Battu. 2003. The theory of differential overqualification: does it work? *Scottish Journal of Political Economy* 50: 1-16.

Büchel, F. and A. Mertens. 2004. Overeducation, Undereducation, and the Theory of Career Mobility. *Applied Economics* 36(8): 803 – 816.

Caliendo, M., F. Fossen and A. Kritikos. 2008. *The Impact of Risk Attitudes on Entrepreneurial Survival*. IZA Discussion Paper No. 3525.

Caliendo, M., D. Cobb-Clark and A. Uhlendorff. 2010. *Locus of Control and Job Search Strategies*. IZA Discussion Paper No. 4750.

Cappellari, L. and S. Jenkins. 2004. Modelling low income transitions. *Journal of Applied Econometrics* 19: 593-610

Cappellari, L. and S. Jenkins. 2008. Transitions between unemployment and low pay. *Research in Labour Economics* 28: 57-79.

Cappellari, L., R. Dorsett and G. Haile. 2010. State dependence and unobserved heterogeneity in the employment transitions of the over-50s. *Empirical Economics* 38(3): 523-554.

Chevalier, A. 2003. Measuring mismatch. *Economica* 70: 509-531.

Costa, P. T. and R.R McCrae. 2002. Looking backward: Changes in the mean levels of personality traits from 80 to 12. In D. Cervone and W. Mischel (Eds.), *Advances in personality science*, New York: Guilford Press: 219–237.

Dekker, R., A. de Grip and H. Heijke. 2002. The effects of training and overeducation on career mobility in a segmented labour market. *International Journal of Manpower* 23(2): 106-125.

Dohmen, T., A. Falk, D. Huffman and U. Sunde. 2009. Homo Reciprocans: Survey Evidence on Behavioural Outcomes. *The Economic Journal* 119(536): 592-612.

Dolton, P. and A. Vignoles. 2000. The incidence and effects of overeducation in the U.K. graduate labor market. *Economics of Education Review* 19: 179–198.

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.

Gourieroux, C. and A. Monfort, A. 1996. *Simulation based econometric methods*, Oxford: Oxford University Press.

Green, F., S. McIntosh and A. Vignoles. 2002. The utilization of education and skills: evidence from Great Britain. *The Manchester School* 70(6): 792–811.

Groot, W. 1996. The incidence of, and returns to overqualification in the UK. *Applied Economics* 28: 1345-1350.

Groot, W. and H. Maassen van den Brink. 1997. Allocation and the Returns to Over-education in the UK. *Education Economics* 5(2): 169–183.

Groot, W. and H. Maassen van den Brink. 2004. The dynamics of skill mismatches in the Dutch labour market. In Büchel, F., de Grip, A. and Mertens, A. (eds). *Overeducation in Europe: Current Issues of Theory and Policy*, Cheltenham, UK. Edward Elgar: 49-65.

Ham, R., P.N. Junankar and R. Wells. 2009. *Occupational Choice: Personality Matters*. IZA Discussion Paper No. 4105.

Heckman, J.J. 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press.

Heckman, J.J., J. Stixrud and S. Urzua. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24(3): 411-482.

Heckman, J.J., J.E., Humphries, S. Urzua and G. Veramendi. 2010. *The Effects of Schooling on Labor Market and Health Outcomes*. Unpublished Manuscript. University of Chicago, Department of Economics.

Heineck, G. and S. Anger. 2010. The returns to cognitive abilities and personality traits in Germany. *Labour Economics* 17(3): 535-546.

Huber, P.J. 1967. The behavior of maximum likelihood estimators under non-standard conditions. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, University of California Press, Berkeley CA.

Jovanovic, B. 1979. Job Matching and the Theory of Turnover, *Journal of Political Economy* 87: 972-990.

Judge, T.A., C. Higgins, C.J. Thoresen and M.R. Barrick. 1999. The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology* 52: 621-652.

Judge, T.A., D. Heller and M.K. Mount. 2002. Five-Factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology* 87: 530-541.

Judge, T.A., and A. Erez. 2007. Interaction and intersection: The constellation of emotional stability and extraversion in predicting performance. *Personnel Psychology* 60: 571-594.

Kammeyer-Mueller, J.D., T.A. Judge and R.F. Piccolo. 2008. Self-esteem and extrinsic career success: Test of a dynamic model. *Applied Psychology: An International Review* 57: 204-224.

Leuven, E. and H. Oosterbeek. 2011. Overeducation and mismatch in the labor market. In E. Hanushek, S. Machin and L. Woessmann (eds.). *Handbook of the Economics of Education*. Elsevier.

Lindley, J. and S. McIntosh. 2010. *Is the Over-Education Wage Penalty Permanent?* Technical report, University of Surrey DP 01/10.

Lockwood, P. 2002, Could It Happen to You? Predicting the Impact of Downward Comparisons on the Self. *Journal of Personality and Social Psychology* 82(3): 343-358.

Mavromaras, K., S. McGuinness and Y.K. Fok. 2009. *Overskilling dynamics and education pathways*. IZA Discussion Paper No. 4321.

Mueller, G. and E. Plug. 2006. Estimating the effects of personality on male and female earnings. *Industrial and Labor Relations Review* 60(1): 3–22.

Nyhus, E.K. and E. Pons. 2005. The effects of personality on earnings. *Journal of Economic Psychology* 26: 363–384.

Roberts, B.W. and W.F. DelVecchio. 2000. The rank-order consistency of personality from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin* 126: 3-25.

Sattinger, M. 1993. Assignment Models of the Distribution of Earnings. *Journal of Economic Literature* 31: 831-880.

Semykina, A and S. Linz. 2007. Gender differences in personality and earnings: Evidence from Russia. *Journal of Economic Psychology* 28(3): 387-410.

Sicherman, N. 1991. Over-education in the Labor Market. *Journal of Labor Economics* 9(2): 101-122.

Stewart, M. and J. Swaffield. 1999. Low pay dynamics and transition probabilities. *Economica* 66: 23-42.

Störmer, S. and R. Fahr. 2010. Individual Determinants of Work Attendance: Evidence on the Role of Personality. IZA Discussion Paper No. 4927.

Tables and Figures

Table 1. Summary statistics

	Mean	SD
Overeducated	0.144	0.351
Age	41.450	10.340
Tenure	11.200	9.838
Temporary contract	0.267	0.443
Self-employed	0.147	0.354
Woman	0.464	0.499
Years schooling	12.540	2.716
No. of adults	2.205	0.833
No. of children	0.709	0.971
Married	0.675	0.469
Single	0.236	0.425
Divorced	0.077	0.267
Widow	0.012	0.108
Badhealth	0.098	0.298

Table 2. Summary statistics – Personality traits

	Mean	SD
Conscientiousness	5.936	0.910
Neuroticism	3.967	1.218
Extraversion	4.829	1.130
Agreeableness	5.459	0.973
Openness	4.496	1.200
External LOC	3.545	0.878

Table 3. Annual overeducation persistence, entry and exit rates and attrition (%)

Overeducation status in period *t-1*	Overeducation status in period *t*		
	Non-overeducated	Overeducated	Missing
a) Stayers in the panel			
Non-overereducated	98.0	2.0	
Overeducated	13.6	86.4	
All	85.9	14.1	
b) All workers			
Non-overereducated	81.2	1.7	17.1
Overeducated	11.5	72.7	15.8
All	71.4	11.7	16.9

Table 4. Estimates of model correlations, and model test statistics

Correlation between unobservables affecting:	Estimate	|t-ratio|
Base year overeducation status and attrition (ρ_1)	-0.0191	-1.39
Base year overeducation and current overeducation status (ρ_2)	0.6057	6.62
Attrition and conditional overeducation(ρ_3)	0.0369	0.41

Null hypothesis for tests	Test statistic	p-value
Exogeneity of initial conditions: $\rho_1 = \rho_2 = 0$	44.01	0.000
Exogeneity of sample attrition: $\rho_1 = \rho_3 = 0$	2.31	0.316
Joint exogeneity: $\rho_1 = \rho_2 = \rho_3 = 0$	67.42	0.000
Exclusion of relationship quality from overeducation transition (d.f.=24)	30.60	0.166
Exclusion of sample membership from overeducation transition (d.f.=2)	3.04	0.218
Exclusion of relationship quality and sample membership from overeducation transition (d.f.=26)	32.46	0.178
Inclusion of relationship quality in base year overeducation (d.f.=12)	22.37	0.034
Inclusion of sample membership in attrition equation (d.f.=1)	75.24	0.000

Table 5. Trivariate probit model estimation: probability of overeducation at *t*

Covariates (measured at *t-1*)	Over-educated a *t-1* δ_1	|t-ratio|	Non over-educated at *t-1* δ_2	|t-ratio|
Individual characteristics				
Woman	-0.002	-0.03	-0.095	-2.41
Ln (age)	-0.060	-0.04	0.286	0.19
Ln (age^2)	0.010	0.05	-0.102	-0.47
Ln (years education)	4.050	12.53	3.748	17.47
Single	-0.161	-2.26	0.041	0.75
Divorced	0.086	0.95	-0.044	-0.61
Bad health	-0.022	-0.30	-0.030	-0.49
Conscientiousness	-0.780	-1.77	-0.327	-1.44
Neuroticism	0.186	0.43	0.267	1.17
Extraversion	-0.916	-2.18	-0.391	-1.67
Agreeableness	0.210	0.51	0.110	0.50
Openness	1.539	3.47	0.980	3.98
External LOC	-0.761	-1.66	-1.265	-5.34
Education * Conscientiousness	0.306	1.90	0.140	1.58
Education * Neuroticism	-0.071	-0.45	-0.103	-1.16
Education * Extraversion	0.329	2.13	0.150	1.65
Education * Agreeableness	-0.065	-0.43	-0.035	-0.40
Education * Openness	-0.587	-3.58	-0.369	-3.85
Education * External LOC	0.332	1.96	0.511	5.45
Household characteristics				
Ln (adults)	-0.155	-2.53	0.052	1.17
Ln (children)	-0.017	-0.33	-0.010	-0.25
Job characteristics				
Temporary	-0.082	-1.29	0.123	2.70
Self-employed	-0.224	-3.58	-0.013	-0.25
Ln (tenure)	0.221	3.79	0.028	0.59
Ln (tenure^2)	-0.084	-4.35	-0.051	-3.21

No. of observations	71321
Log-likelihood	-46161.28
Model chi-square (d.f=153)	18963.55 ($p < 0.000$)
No state dependence $\delta_1 = \delta_2$ (d.f.=38)	631.27 ($p < 0.000$)

Table 6. Aggregate and individual state dependence

	$Prob(Over\ at\ t|Over\ at\ t-1)$
Raw aggregate probabilities of overeducation at t conditional on	
Overeducation at *t-1*	0.862
Non overeducation at *t-1*	0.020
Difference	0.842
Model predicted probabilities: $Prob(Over\ at\ t|Over\ at\ t-1)$	
Average for the sample of overeducated at t-1	0.862
Average for the sample of non over-educated at t-1	0.169
Difference	0.694
State Dependence	0.148
	(17.6%)

Table 7. Overeducation risk variation by selected personal characteristics

	Persistence Rate	Entry Rate	Pr(over)	Overeducation spell duration		Non-overeducation spell duration	
				Mean	Median	Mean	Median
Overeducated reference individual	0.892	0.538	0.833	9.26	6.06	1.86	0.90
Percentage points variation following:							
+ 1 std in years of schooling	5.49	26.02	10.46	83.05	87.94	-20.65	-31.86
+ 1 std deviation in Conscientiousness	-10.54	-31.04	-22.26	-46.53	-49.35	45.01	66.56
+ 1 std deviation in Neuroticism	0.90	5.02	2.02	8.00	8.47	-4.78	-7.23
+ 1 std deviation in Extraversion	-13.90	-38.29	-29.32	-53.45	-56.70	62.05	91.39
+ 1 std deviation in Agreeableness	1.79	7.99	3.66	17.39	18.42	-7.40	-11.23
+ 1 std deviation in Openness	9.53	51.49	16.78	369.57	391.17	-33.99	-54.24
+ 1 std deviation in External LOC	-3.48	-20.45	-9.36	-22.30	-23.63	25.70	38.23
+ 1 std deviation in schooling and Conscientiousness	-2.02	-3.16	-3.31	-14.29	-15.14	3.26	4.91
+ 1 std deviation in schooling and Neuroticism	6.17	29.74	11.60	103.77	109.87	-22.92	-35.51
+ 1 std deviation in schooling and Extraversion	-4.48	-10.97	-8.27	-27.03	-28.64	12.32	18.43
+ 1 std deviation in schooling and Agreeableness	6.61	32.53	12.35	120.41	127.48	-24.54	-38.14
+ 1 std deviation in schooling and Openness	11.10	65.61	18.87	1100.00	1164.16	-39.62	-65.16
+ 1 std deviation in schooling and External LOC	2.69	10.41	5.20	28.57	30.26	-9.43	-14.33

Figure 1.Distribution of personality traits