

A Work Project, presented as part of the requirements for the Award of a Master Degree in
Economics from the NOVA School of Business and Economics.

Overeducation: Dynamics, Wounds and Scars

Vítor Gabriel Pereira Santos 20621

A Project carried out on the Master in Economics Program, under the supervision of:
Professor Pedro Portugal

January 2018

Overeducation: Dynamics, Wounds and Scars

Abstract

The present study investigates the extent to which the effects of entering the labor market in overeducated positions pertrain throughout workers careers. Adopting a time-to-event modeling approach, we analyze overeducation persistency and evaluate the existence of state duration dependence. We further investigate future earnings effects associated with past overeducation employment. Our results report strong entrapment effects rising from overeducation and for overeducated workers to yield lower returns for their schooling investment than their equally matched peers. We also conclude the effects of overeducation to perpetuate to future job allocations, with workers having an overeducation employment background reported as suffering wage losses ranging from 3.5% to 14%.

Keywords: Overeducation, Duration, State Dependence, Scarring wage effects

I Introduction

For Portugal, the last three decades presented a profound transformation of the educational system, largely expressed by considerable increases in both participation and completion rates at secondary and tertiary educational levels (OECD, 2014). Alongside with soaring educational levels are the concerns over the possible incapacity of the labor market to adjust to the growing pool of highly educated workers.

The latter fear is supported by a large body of empirical literature for Europe, claiming the existence of a large share of the labor force allocated into jobs requiring less education than workers have (overeducated) and for the incidence of the mismatch to be particularly high for Southern European countries (Boll *et al.*, 2016). For Portugal, explicitly, evidence reveals nearly 30% of Portuguese workers in the private sector to be overeducated, this percentage being slightly lower (21%) for newly hired workers (Araújo and Carneiro, 2017).

The educational mismatch has been evidenced to be a costly phenomenon, with overeducated workers reported having lower levels of job satisfaction (Tsang, 1987; Allen and van der Velden, 2001) and lower returns to their schooling investment (Hartog, 2000). At an aggregate level, overeducation may also describe an inefficient usage of available resources, with

evidence reporting negative associated effects in terms of short-run developments of the per capita output (Morgado *et al.*, 2014).

Although the previous effects have been presented as an accurate measure of disadvantage for overeducated workers, little has been the consideration by the literature on its longitudinal patterns. Nevertheless, for the existent literature covering its dynamics, overeducation is suggested to scar workers careers, with a large chunk of workers not only to prevail as overeducated for long periods but also to significantly delay entry into matching employment by accepting such job positions.

The present study intends to unfold the extent to which the effects of entering the labor market as overeducated pertrain throughout workers careers. We analyze overeducation persistency, some of its continuance determinants and state duration dependence to characterize its longitudinal dimension. We further investigate associated earnings effects at entry (wound) as well as the way these may perpetuate to subsequent job allocations (scar). By doing so, this study contributes to the Portuguese literature on Overeducation which, to the best of our knowledge, has concentrated solely on cross-section earnings' outcomes rising from the mismatch.

The reminder of this study is organized as follows: section II presents a brief overview of the literature on overeducation; section III summarizes the developed approaches to measure overeducation; section IV and V describe the adopted empirical framework together with data description and definition used to identify overeducated workers; section VI documents the obtained results and section VII concludes.

II Literature Review

Concerns over a growing pool of overeducated workers (i.e., having more schooling than necessary for their job) dates to the late 1970s with Freeman's (1976) pioneering work on US graduates. The author argued the excess of US graduates supply to be eventually preceded

by a slump in college wage premium, shedding out the phenomenon likelihood to persist.

By conceiving an augmented version of the canonical Mincerian earnings equation, the Over-Required-Under (ORU) model, Duncan and Hoffman (1981) grounded the possibility for estimation of earnings effects rising from educational mismatches. Their findings from the 1976 Panel Study of Income Dynamics (PSID) reported a 6.3% return to years of education required for the job, 2.9% for surplus and -4.2% return for deficit years of education. Hence, albeit having a wage premium over peers at their workplace, overeducated workers were presented as suffering a wage loss when compared to their peers finding adequate matches.

A myriad of studies have replicated Duncan and Hoffman's specification using cross-sectional data, reporting the pay loss to be robust across time, place and employed measure of required education¹. In the overeducation literature for Portugal, Kiker *et al.* (1997) present similar results, with a 7.6% return for years of schooling required, 4.8% for surplus and -5.6% for deficit years of education.

The stylized facts of these cross-sectional data studies were nonetheless criticized by its implicit exogeneity assumption of the mismatch. In other words, if individuals with the same educational credentials are not perfect substitutes, the assignment into overeducation may be non-random, compromising the unbiasedness of the OLS earnings estimates. To account for the latter, in recent years, several authors (e.g. Frenette, 2004; Bauer, 2002) applied panel data techniques to control for unobservable individual productivity differences, confirming the OLS return estimates on both over and undereducation years to be downward biased. Nevertheless, and while controversy exists regarding its magnitude, the still observed estimated earnings loss appears to represent a real measure of disadvantage experienced by overeducated workers (McGuinness and Pouliakas, 2017).

Yet, what the wage return findings fail to account is the possibility that overeducated workers may be only temporarily overemployed before switching to a job requiring their schooling level. In addition, lower unobserved ability, compensating non-pecuniary job amenities and

¹See Rumberger, 1987; Sicherman and Galor, 1990; Cohn and Khan, 1995; Daly; Bauer, 2002; Rubb, 2003; Linsley, 2005

career mobility considerations could rationalize apparent overeducation without necessarily implying a suboptimal schooling choice (Clark *et al.* 2014). To fully evaluate how much of a problem overqualification is, it becomes crucial to analyze its longitudinal dimension.

Sicherman and Galor (1990) and Robst (1995) findings of US overeducated workers experiencing higher career mobility in low-skilled jobs than their equally matched peers supported a frictional view of overeducation, with the authors concluding overeducation to be a stepping stone towards career upgrading. Their stepping stone hypothesis was nevertheless challenged by Rubb (2003), which presents substantial persistency in US overschooling by reporting a meager 20% annual outflow from overeducation into matching positions. By analyzing two UK graduate cohorts, Battu and Sloane (2000) estimate the educational mismatch to remain stable at 60% after 11 years since graduation. Buchel (2002) findings for Germany follow similar patterns, with overeducated German workers experiencing not only less upward mobility but also lower relative wage growth rates.

Although mixed, the body of evidence supporting persistency of overeducation seems to largely outweigh the temporary character hypothesis. On the other hand, and despite its acknowledged importance, evidence on the scarring effects of overeducation over workers careers is rather scarce.

Mavromaras and McGuinness (2012) estimates from a dynamic random effects probit model add to the longitudinal mismatch literature by presenting substantial state dependence, the latter being particularly high for individuals with higher education degrees. Schult and Buchel (2004) adopt parametric hazard models applied to data from two West German cohorts (1965 and 1971) to test negative occurrence dependence. The authors find individuals accepting jobs for which they are overeducated not only to significantly reduce their exit likelihood to matching employment but also be more prone to downward job mobilities.

The extent to which entering an overeducated job accelerates or delays transition into suitable positions has also been tested by Baert *et al.* (2013). Employing data from young Flemish unemployed graduates and adopting a timing of events approach to account for possible non-

random selection into overeducation, the authors find early overeducation to strongly retard transitions into adequate positions and for the negative effect to be permanent for both the short and long-term unemployed graduates.

To disentangle true duration dependence from selection on workers unobservables, Clark *et al.* (2014) apply a Mixed Proportional Hazard model on combined data from the 1979 US National Longitudinal Survey of Youth and the 1989-1991 waves of the Current Population Survey. Contrarily to previous findings on scarring overeducation effects, the authors' findings suggest the existence of negative duration dependence to be largely explained by individual unobservable heterogeneity, thereby concluding for the duration of overeducation employment not to significantly impact workers chances to leave the mismatch state.

III Measuring Overeducation

Educational mismatches arise when completed years of education exceed or fall behind job required schooling (S_i^r).² The starting point for identifying the mismatch requires estimation of schooling years required to perform the job. Literature of over and undereducation has measured it using either an objective or a subjective approach.

Subjective approaches rely on information provided by workers, either by inquiring the individual on his/her job match or to report the educational requirements of the job and indirectly define it. These have the advantage of being up to date and specific to the job of the respondent. Adversely, a worker's assessment is inevitably subjective and sensitive to measurement errors as individuals may overstate educational requirements to inflate their position or to simply reproduce current hiring standards.

Objective approaches use objective measures such as the educational level of peers working in the same position or professional evaluations to determine S_i^r . Literature relies on two of

² $S_i = S_i^r + \max(0, S_i^o - S_i^r) - \max(0, S_i^r - S_i^u)$, S_i as the observed completed years of education, S_i^r schooling years required for the job, $\max(0, S_i^o - S_i^r)$ an indicator variable representing surplus/overeducation years and $\max(0, S_i^r - S_i^u)$ representing deficit/undereducation years.

these instructions: the job analysis (JA) and the realized matches (RM) method. The JA relies on evaluation by professional job analysts who specify the required level of education for occupational classifications. In the RM method, the observed distribution of workers education at a given occupation is used to assess job qualification requirements, with both the mean and modal value of each occupational title being used to measure required schooling years for the job. When the mean is used, the average level of education of a given occupation is defined as the required schooling level for that occupation, with the individual classified as overeducated if his/her level of education exceeds the average by at least one standard deviation . In alternative, the required schooling level in a given occupation has been also determined as the mode of the completed schooling levels of people working in that job, with an overeducated worker being the one with a higher education than the mode.

Although the job analysis is conceptually superior due to its explicit objectivity, it constructs on the assumption that jobs within the same occupational classification have equivalent job requirements and, its costly and timely implementation, makes it likely to quickly become obsolete. The realized matches method has the advantage of being always available as it is directly computed from data. Yet, it contains observations on the equilibrium realized by demand and supply forces, thus already reflecting the allocation outcome of an assignment model. Furthermore, it may be sensitive to cohort effects, its design arbitrary and will generally find symmetry on the incidences over the types of mismatch (Hartog, 2000).

Altogether, the aforementioned methods convey advantages and disadvantages, and, to the best of our knowledge, no optimal methodology has been developed so far. The role of availability of data in determining the preferability of one measure over the other is discussed by Hartog (2000): “You use what is available”.

IV Empirical Framework

On this section we heavily rely on Jenkins' (1999) contributions to survival analysis to investigate overeducation persistency and to model the effect of duration in an overeducated position on the length of time until departure from overeducation and to two possible destination routes. We start by introducing the taxonomy used in the survival literature and follow with the adopted approach.

Let T be a non-negative random variable representing the individual elapsed time until departure from an overeducated employment spell, with a cumulative distribution function $F(t)$ and probability density function $f(t)$.

The cumulative distribution function is known in the survival literature as the failure function and returns the probability that an overeducation employment spell has occurred by duration t : $F(t) = Pr(T \leq t)$, with t the elapsed time since entry to an overeducated employment spell at time 0. The complement of the failure function is the survival function, summarizing the probability of remaining in an overeducated employment spell just before duration t or, in an analogous manner, the probability that an overeducated employment spell has not ended by duration t : $S(t) = Pr(T \geq t) = 1 - F(t)$.

The slope of the failure function $F(t)$ is the probability density function $f(t)$, which returns the (unconditional) probability of an individual having an overeducation employment spell of length exactly t or the 'probability' that an individual will leave an overeducated employment spell in an infinitesimal interval of time Δt :

$$\lim_{\Delta t \rightarrow 0} \frac{Pr(t < T \leq t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t}$$

An useful characterization of the distribution of T is given by the hazard function $h(t)$. Applied to the present context, the difference between the probability density and the hazard function results from the former presenting the concentration of departures from an overeducated employment spell at each instant of time whereas the latter summarizes the same concentration

but conditioning this on survival in overeducation up to that moment in time. The probability density function may then be thought as returning an instantaneous rate of departure whereas the hazard function describes a conditional one, thereby accounting the possibility for the rate of departure from an overeducated employment spell to keep changing with permanency in that state.

Formally, in continuous time, the hazard function is described as: ³

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

The numerator from above presents the probability for an overeducated employment spell to end in the interval $[t, t + \Delta t]$, given that it has not ended before, and the denominator the width of that interval. Dividing one by the other and taking the limit as the width goes down to zero yields the rate of overeducated employment departures at each instant in time. The hazard out of overeducation may be then interpreted as the instantaneous risk of exit for each individual at each t .

Although the underlying survival process is continuous in time, our data time intervals are partitioned into unit lengths, years. The interval boundaries are the positive integers $j = 1, 2, 3, 4, \dots$, with each given interval j being $(T_0, T_1], (T_1, T_2], \dots, (T_{j-1}, \infty[$. The data is observed on a discrete basis, with $T \in (1, 2, 3, \dots, j)$, and $T = t$ representing an exit from overeducation during $(T_{t-1}, T_t]$. The latter grounds the choice of discrete time hazard models followed by the current analysis.

To model the instantaneous risk of exit from overeducation, the hazard out of overeducation, we resort to the discrete proportional hazard model, accounting for the presence of a vector of

³The relationship between the hazard $h(t)$ and both the density $f(t)$ and the survival $S(t)$ function may be derived using Baye's rule. The conditional numerator in the limit expression may be defined as the ratio of the joint probability that T lies in $[t, t + \Delta t]$ and $T \geq t$ to the probability that $T \geq t$. The former may be written as $f(t)\Delta t$ for small Δt whereas the latter is the survival function by definition. Using Bayes rule: $h(t) = Pr(t < T \leq t + \Delta t | T \geq t) = \frac{Pr(t < T \leq t + \Delta t \cap T \geq t)}{Pr(T \geq t)} = \frac{Pr(t < T \leq t + \Delta t)}{Pr(T \geq t)}$. Applying the limit:

$$\lim_{\Delta t \rightarrow 0} \frac{Pr(t < T \leq t + \Delta t)}{Pr(T \geq t)} = \frac{f(t)}{S(t)}$$

covariates that may affect the time spent in that state. The hazard at time t for an individual with covariates X_i equals the product of a baseline hazard function $h_0(t)$, describing the risk for individuals with $X_i = 0$, and the relative risk $exp(X_i\beta)$, presenting the proportionate increase/decrease in the probability of leaving an overeducation position associated with the set of attributes X_i .

The estimation of the log likelihood function requires the choice of the functional form of the discrete hazard. The aforementioned nature of data implies the appropriate form of the hazard out of overeducation employment to be the complementary log-log functional form (Prentice and Gloeckler,1978):

$$h_i(t|X_i) = 1 - [1 - h_0(t)]exp(X_i\beta)$$

With the aim of analyzing the existence of duration dependence, we further extend the latter by including as additional covariate $\lambda(t)$, the natural logarithm of overeducation duration, capturing the elasticity of departure from an overeducated position with respect to elapsed duration in that state:

$$h_i(t|X_i) = 1 - [1 - h_0(t)]exp(\lambda(t) + X_i\beta) \tag{1}$$

Following the model in equation one, the hazard out of overeducation employment depends on the time spent (in years) as overeducated, captured by $\lambda(t)$, as well as on a set of individual covariates . We estimate the latter with no assumption on the distribution of the baseline hazard as to avoiding misspecification problems of choosing its wrong shape.

Hazard estimates from model of equation (1) can be significantly biased if unobserved individual heterogeneity (UIH) is not considered (Lancaster, 1979). Controlling only for observable differences, individuals with unobserved characteristics v_i (i.e. motivation, ability, family background) associated with higher exit rates leave overeducation faster. Consequently, for higher durations, the pool of survivors becomes increasingly composed of overeducated workers whose unobserved attributes are associated with lower exit rates (lower hazards). Thence,

any (positive) negative overeducation duration dependence will be (under)overestimated.

To handle UIH we follow Heckman and Singer's (1984) mass-point approach, accounting for an arbitrary distribution for individual heterogeneity by assuming each worker to belong to one of a number of different types $z = 1 \dots, z$, identified from their variation in duration conditional on observed characteristics. We accommodate the presence of two different types of overeducated workers, differing between themselves in unobserved characteristics affecting transitions out of overeducated positions. Although membership of each type is unobserved, the resulting hazard function incorporates an additional term (m_{type}), allowing for its intercept to vary with the different estimated heterogeneity type:

$$h(t|X_i) = 1 - [1 - h_0(t)]exp(m_{type} + \lambda(t) + X_i\beta) \quad (2)$$

with $m_{type=1}$ normalized to zero and mass point 2 equal to $m_{type=2} + \beta_0$.

We further replicate the previous specifications to assess the effect of overeducation employment on the hazard out to two specific destination routes: matching or underschooling positions.

With the aim of increasing our alertness over the scarring effects from overeducation employment throughout workers careers, we further extend our analysis to identify possible scarring wage effects from previous overeducation positions and how these may change with the time spent by the individual as overeducated.

By identifying individuals making the transition out of overeducation positions, we estimate an augmented version of Duncan and Hoffman's (1981) log-wage specification by introducing as explanatory variables lags of previous overeducation duration:

$$\log(w_{itj}) = x_i\beta + \delta_r S_i^r + \delta_o S_i^o + \delta_u S_i^u + \delta_\alpha Dur + \epsilon_i$$

with $\log(w_{itj})$ the natural logarithm of hourly earnings, x_i containing individual covariates, S_i^r years of required education for the job, S_i^o years of overeducation, S_i^u years of undereducation and Dur identifying duration from previous overeducation employment.

V Data Description and identification of overeducation

V.I Quadros de Pessoal (QP)

Our analysis is based on data collected from 1986 to 2013 in *Quadros de Pessoal* (QP), a longitudinal matched employer-employee-job title data set. Data are collected in October of every year by the Portuguese Ministry of Labor and Social Solidarity, covering all establishments with at least one wage earner. The mandatory status of the survey enhances data quality by mitigating problems commonly associated with panel data (i.e. attrition or measurement error). Detailed information at worker (i.e. earnings, occupation, tenure, work schedule, gender, education), firm (i.e. number of workers, main economic activity (CAE), location, legal status, turnover) and affiliated establishments (location, economic activity, employment, sales) is available.

At QP, each worker and firm in the data base is assigned with a unique identifying number. A unique feature of QP derives from the availability of data on collective wage agreements, such that each worker is assigned in each year to a specific job title and the corresponding collective agreement. The unusual richness of data allows us to track individuals over the years and to match them with firms, collective agreement and job title held in each year.

V.II Identifying Overeducated workers

The design of QP allows for identification of overeducation only through objective approaches. We resort to the realized matches (RM) approach and follow Kiker *et al.* (1997) on using the observed modal value of education to determine the required level of schooling for a given job title. Although the underlying choice between the mean and the modal value on the RM approach relies on some degree of arbitrariness, the mode is generally preferred by being less vulnerable to outliers and technological changes (Meroni *et al.* 2016). In addition, QP collects information regarding the collective agreement ruling the wage dimension of the match

between employer and employee.

The availability of the specific job title held by the worker, within each collective agreement, allows for refinement from the broad classification of occupations traditionally used by studies applying RM. The former translates into an elevated level of disaggregation, such that similar occupations are considered as distinct job titles if their wages are settled within different collective agreements (Portugal *et al.* 2016).⁴ The employed required schooling level is determined as the mode of the distribution of education amongst individuals working in a given job title, in each year. An individual is thus identified as overeducated if his/her educational diploma exceeds the observed distributional education mode of his/her job title on a specific year.

To estimate the impact of elapsed overeducation duration at job entry level on the hazard out of overeducation and on the transition into matching and underschooling employment, a set of restraints to the initial data were made. First, the sample was restricted to those individuals entering QP for the first time. An additional control, defined by individuals having a reported lower than 12 months of tenure, was set to excise individuals entering QP for the first time without being first job entries. Workers with higher than 30 years of tenure and those working in Agriculture, Fishery and Mining sectors were also excluded. At least, only individuals entering the labor market for the first time as overeducated were analyzed and followed over time.

The final sample for our first model (no destination route specified) is composed by 1,647,281 entry-level overeducated workers and 347,178 firms. From these, 862,248 (52%) remained in overeducation positions throughout our 27-year period of analysis. The estimation of the hazard out to the two aforementioned destination routes was performed by restricting the initial sample to those leaving an overeducation employment spell to each of the destination types.

⁴For instance, within QP a secretary in the financial services industry has a distinct job title from a secretary in the textile industry. Under the traditional broad classification (i.e. ISCED 1997 3-digit occupational level), both job titles would be considered as belonging to the same occupational group, resulting most likely in erroneous assessments about the mismatch. One expects the two different job titles to require distinct levels of education due to the distinct complexity of tasks performed. The latter effect is captured by the present study using the modal level for each job title, in each year, under a specific collective agreement.

Here, the final sample is composed by 785,033 individuals, where 73% left overeducation into a matching position and 27% entered underschooling employment.

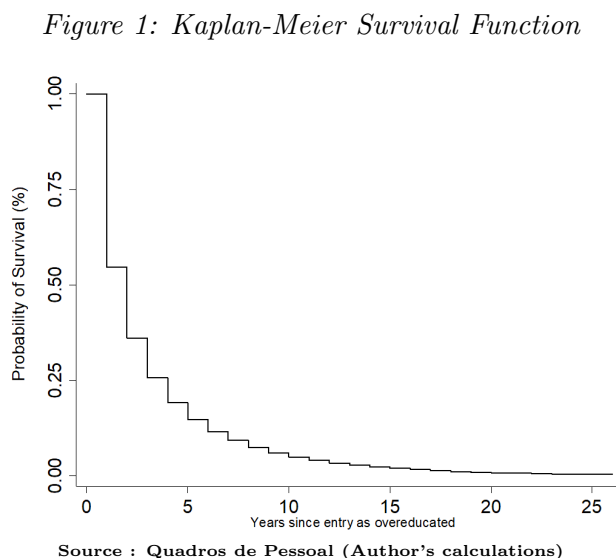
The average age for the overeducated worker is 31 years and nearly half has either upper secondary or tertiary educational levels. Heterogeneity is found regarding the amount of years spent in overeducated positions (between 1 up to 26 years), with higher permanence observed for individuals with higher diplomas.

Even if no sizeable differences are found with respect to gender overeducation incidence (57% males and 43% females), the gender distribution across sectors of activity is rather despair, with overeducated male workers being majorly allocated across construction and freight-transport activities and women into manufacturing, hospitality or retail super and hypermarkets jobs.

VI Empirical Results

VI.I Longer stay, harsh leave?

Figure 1. reports the non-parametric Kaplan-Meier Survival function of yearly transitions out of overeducation.



One year after entry, the probability for an individual to remain in an overeducated position falls sharply to nearly 50%. The median duration until exit from overeducation lies beneath 2 years and although the majority of workers leave 5 years since entry, each additional year spent as overeducated yield minor increases on the probability of exiting the mismatch status, suggesting the likelihood to leave overeducation to be decreasing with overeducation continuance.

Despite consistent with negative duration dependence, this pattern could also result from the presence of individual permanent heterogeneity correlated with the hazard out of that state, hence limiting any causality assessment to be based on the former simple descriptive depiction. To assess the role of individual and firm features on driving overeducation persistence, as well as to test for the possibility of negative duration dependence, the estimation of equations (1) and (2) was performed. The Model of equation (2) accounts for the possibility of two UIH types to be present, allowing to disentangle between “true” and spurious duration dependence. The selected covariates age, gender, firm size and a nationality dummy variable, with value 1 if the individual is born in Portugal and 0 otherwise, allow us to replicate already established longitudinal career patterns of overeducation employment for the Portuguese context.

The corresponding model estimates are presented in Table 1. The negative coefficient on age suggests older workers to experience higher persistence of overeducation employment. The negative relationship between exit likelihood and age may follow from the expected positive correlation between age and job tenure, with older workers higher accumulation of firm-specific human capital investments producing strong “*lock-in*” effects into bad positions (Pissarides, 1994).

The hazard out overeducation is 5% higher for males when compared to females. The manful-advantage in exiting overeducation employment may relate to the literature findings reporting higher overschooling prevalence for females. The latter could be the result of women placing more value on non-pecuniary job amenities associated with low-requirement jobs (i.e. flexibility of hours worked) or reflect labor market gender discrimination by employers. In addition,

Table I. Hazard out of Overeducation

Variables	(Eq.1)		(Eq.2)	
	Coefficient	std	Coefficient	std
<i>Age</i>	-0.0185***	(0.00094)	-0.0203***	(0.0011)
<i>Age^2</i>	0.0004***	(0.000013)	0.0004***	(0.00001)
<i>Male</i>	0.0461***	(0.002562)	0.0500***	(0.00295)
<i>Portuguese</i>	0.0113***	(0.002623)	0.0178***	(0.00299)
<i>Firm Size</i>	-0.0797***	(0.002542)	-0.0741***	(0.00289)
<i>log(t)</i>	-0.5608***	(0.001903)	-0.39282***	(0.00467)
<i>Prob (Type I)</i>	-		0.30516***	(0.00902)
<i>Prob (Type II)</i>	-		0.69483***	(0.00902)
<i>Intercept Type II</i>			0.39***	(0.123)
<i>Intercept Type I</i>			-0.7911***	(0.021)
<i>Constant</i>	-0.4032***	(0.01551)	-	
<i>Control. Year Effects</i>	YES		YES	
<i>Control. Regional Effects</i>	YES		YES	
<i>Control. Sector Effects</i>	YES		YES	
<i>Observations</i>	4,391,959		4,391,959	
<i>Number of individuals</i>	1.647.281		1.647.281	

Equation (1) corresponds to the model not controlling for UIH and model of equation (2) to the mass-point approach specification. The former accounts for the effect of individual unobserved permanent heterogeneity on duration by allowing the presence of distinct worker types, identified from variation in duration conditional on observed characteristics.

as proposed by Frank (1978), women’s higher overeducation persistence could also mirror the fact that, in dual earner households where men are the prime income source, women’s labor choice location is subdued to the husband’s labor market prospects. Given the weight placed by the literature on the link between overeducation persistence and spatial factors, the former could substantiate our findings of women’s higher prospects to remain in a mismatching position ⁵.

The transition process out of overeducation employment is slightly higher for the native population. Being born in Portugal increases one individual’s exit likelihood by roughly 2%,

⁵See McGoldrick and Robst (1996) and Buchel and van Ham (2000)

ceteris paribus. Although we cannot ascertain from our data on whether foreign individuals have acquired formal schooling outside Portugal, several premises can be placed to explain these results. The higher mismatch persistence for foreign workers could be explained by their shortage of specific labor market rewarded human capital characteristics (i.e. proficiency in the native language). This is also in line with literature findings of higher overeducation prevalence for immigrants, whom accordingly substitute their lack of native labor market valued attributes with higher educational credentials. The former findings may also be grounded on labor market discrimination, even though one must be careful on claiming discrimination *per se*. Employers may prefer to recruit national workers based on their superior knowledge about the average quality of nationally educated workers, henceforth increasing the tendency of confounding recruitment-based discrimination with imperfect information (Nielsen, 2007). The size of the firm appears to strongly impact the exit likelihood of overeducated workers, with those entering overeducated positions in larger firms being less likely to leave. Larger firms are acknowledged to possess well-defined job ladders, internal job-upgrades and promotions based on seniority (Fairris, 2004). The existence of such internal labor market mechanisms, allowing long-term advancement opportunities within the firm, may provide an incentive for workers to remain in mismatched positions. In addition, the former may also find explanation on larger firms having higher fringe benefits (Edmiston, 2007). The lower exit likelihood from overeducated positions in larger firms may then reflect workers preferences for these non-pecuniary amenities, in resemblance to findings reporting workers accepting low-paid jobs in exchange for larger benefits (Olson, 2002).

Referring to the estimate of interest for measuring state dependence, $\log(t)$, the results from both model equations (1) and (2) report the elasticity of exiting overeducation employment with respect to elapsed duration to be negative and statistically significant, in support of negative duration dependence. Controlling for UIH in specification (2) significantly decreases the elasticity estimate, indicating part of the negative duration dependence in the model of equation (1) to be attributable to selection on unobservables. Nevertheless, and even control-

ling for permanent heterogeneity, the reported estimate underlines the prevalence of negative state dependence, with a 10% increase in time spent as overeducated decreasing one worker's exit likelihood in 3.9%.

The mass-point approach model of equation (2) accomodates the presence of two distinct types of workers, with markedly different overeducation dynamics. The majority of overeducated workers are identified as belonging to Type II (70%), whereas 30% is identified as being a Type I. The higher intercept for Type II indicates the hazard for this type to be higher than that of Type I, other things being equal. In other words, the latter suggests the majority of the overeducated work force (70%) to leave overeducation employment in a faster way relatively to Type I people.

The significant difference between the hazards of the two UIH mass-points suggests overeducation to follow distinct mechanisms for each type.

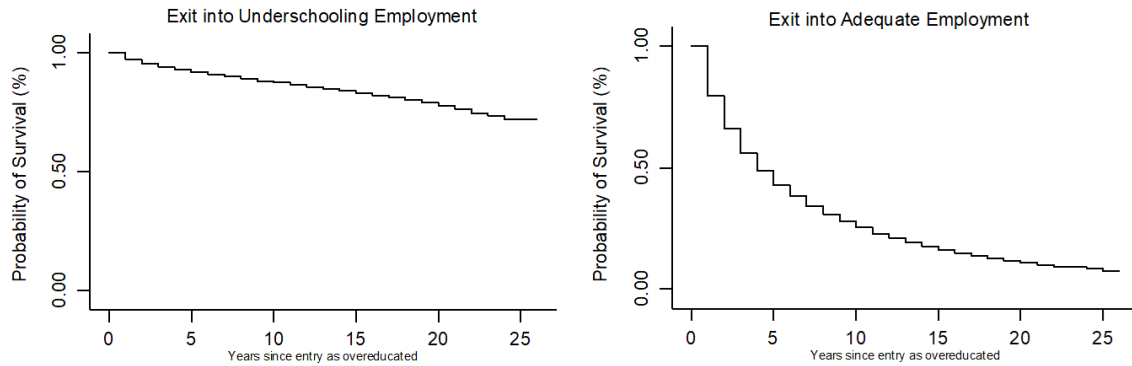
The high exit rates for Type II overeducated workers is consistent with a temporary view of overeducation. Accordingly, overeducation employment may result from a sluggish adjustment of firm processes to the available workforce or to exist only as long as it takes for workers to find an appropriate match (Alba-Ramirez, 1993). The decision to become temporarily overeducated may also mimic one individual's human capital investment strategy, whereby he/she would accept an overeducated job as to acquire the necessary on-the-job experience before upgrading into a matching position (Sicherman and Galor, 1990).

For the remaining, low-hazard, 30% overeducated workers, it could be that their ability is not sufficient for jobs that match their level of formal schooling, averting the upgrading transition to take place (Clark *et al.* 2014).

In what follows we analyze the results if considering the possibility out to two distinct exits, with Figure 2. reporting the corresponding Kaplan-Meier survival estimates.

The median duration until exit out of overeducation into a matching position is 5 years (right figure). On the left side of Figure 2., we consider exits from overeducation into under-schooling job positions. This may characterize, as an example, a post-secondary non-tertiary

Figure 2: Kaplan-Meier Survival Function by destination type



Source : Quadros de Pessoa (Author's calculations)

female working as a cashier at a retail store upgrading her position into a managerial one ,where the modal value of education is a college degree. At the defined duration interval, the corresponding survival function does not converge to zero but assumes and maintains some positive value. This might possibly represent a situation in which the risk of exiting to that destination is defective. Expressed in another way, when considering departure to that specific route, we may find overeducated workers with a non-zero probability of ending up in an overeducated employment spell which lasts forever (Addison and Portugal, 2003).

Failure to implement the mass point approach model controlling for UIH for the two exit destinations indicates inability to differentiate between distinct types of individuals making the transition to each of the states. In other words, the sort of individuals exiting overeducation into adequate employment/underschooling employment is homogeneous with respect to unobservable behavior affecting the transitions into that state.

The hazard estimates for the two exit routes are presented in Table II. Adversely to the former hazard results, and those considering exit into matching positions, the positive coefficient on age in the specification considering exit into underschooling suggest older workers to enjoy higher changeovers when this destination is considered. The transition from overeducation into underschooling positions might possibly characterize the labor market, either internal or external, compensating older workers experience for the lack of required formal schooling,

Table II. Hazard out to Matching or Underschooling employment

Variables	Exit into matching position	<i>std</i>	Exit into underschooling position.	<i>std</i>
<i>Age</i>	- 0.0461***	(0.0008)	0.038***	(0.0013)
<i>Age</i> ²	-0.0007***	(0.0001)	-0.0004***	(0.0002)
<i>Male</i>	0.0689***	(0.0024)	- 0.014***	(0.0040)
<i>Portuguese</i>	0.0943***	(0.0024)	-0.0693***	(0.0040)
<i>Firm Size</i>	0.04673***	(0.0014)	-0.009***	(0.0025)
log(<i>t</i>)	-0.9403***	(0.0015)	-0.9822***	(0.00263)
<i>Constant</i>	-1.36***	(0.0145)	-2.32***	(0.0232)
<i>Control. Year Effects</i>	YES		YES	
<i>Control. Regional Effects</i>	YES		YES	
<i>Control. Sector Effects</i>	YES		YES	
<i>Observations</i>	715,914		254,123	

Model (1) and Model (2) correspond to the hazard estimate out of overeducation into matching and underschooling positions, respectively, estimated using the complementary log-log functional form. The used sample comprehends only individuals who have left overeducation into each of these positions, ruling out re-entries into overeducation.

endorsing this boundary upgrading of job positions to occur.

Whichever mechanism favors the Portuguese nationality on the hazard out of overeducation, it appears to reverse when considering exit into underschooling positions, with foreign-born individuals presenting a 6% higher exit likelihood than those born in Portugal, *ceteris paribus*. For the reminder covariates, their effect is in line with those obtained in previous specifications, irrespectively on the destination venue.

The reported elasticity estimate hints a more severe negative duration dependence effect from overeducation employment for the two havens, with a 10% increase in duration reducing the exit likelihood, either to an under or matching position, by more than 9%. This stronger reported negative duration dependence effect may nevertheless translate the inability to control for different UIH types, with individuals remaining as overeducated for longer durations inducing an overestimation of the duration dependence effect.

VI.II Wage effects from overeducation employment : Wounds and Scars

In what follows, we address our pooled OLS estimates as in Duncan and Hoffman's (1981) log-wage equation (ORU Model), allowing for the estimation of separate wage returns for years of school required for the job, surplus and deficit years:

$$\log(w_{itj}) = x_i\beta + \delta_r S_i^r + \delta_o S_i^o + \delta_u S_i^u + \epsilon_i$$

where w_{itj} denotes hourly earnings for individual i in year t in job title j , S_i^r years of required education for the job, S_i^o years of overeducation, S_i^u years of undereducation, x_i a vector of individual control variables and ϵ_i the error term.

An augmented wage regression identifying past overeducation spells is also presented as to test for the existence of scarring wage effects of previous overeducation employment.⁶

Unlike our previous analysis, the employed sample for the log-wage regressions included not only those starting the career as overeducated but also those entering matching and under-schooling positions. These are included as to properly identify the returns for required, over and undereducation years. To disentangle between the wage effects of being on a still ongoing overeducation spell from those of past overeducation employment, individuals on a continuing overeducation spell were excised from the sample.

In Table III, ORU estimates for the separate earnings returns are consistent with those reported on Portugal by Kiker *et al.* (1997). In other words, when compared to co-workers who have just the required education to perform the job, the average overeducated Portuguese worker yields a 4.7 % return for the extra year of schooling above the required. Adversely, and relative to co-workers who have the required education, for undereducated workers each year of undereducation translates into an earnings loss of nearly 7%. Hitherto, and even if delivering economic value, each year of overeducation yields only half of the return of those just

⁶For the estimation of the two log-wage regressions we assume the error term ϵ_i to be cross-sectionally and serially independent, with zero mean and constant variance.

Table III. Scarring wage effects of overeducation employment

Variables	Over-Required-Under (ORU)		Augmented Wage Regression	
	ln (Hourly Wage)	std	ln (Hourly Wage)	std
<i>Age</i>	0.0452***	(7.67e-05)	0.0464***	(7.89e-05)
<i>Age</i> ^2	-0.000426***	(1.01e-06)	-0.000443***	(1.03e-06)
<i>Male</i>	0.290***	(0.000360)	0.289***	(0.000358)
<i>Tenure</i>	0.00114***	(2.08e-06)	0.00110***	(2.07e-06)
<i>Required Schooling</i> (δ_r)	0.0988***	(6.51e-05)	0.0991***	(6.50e-05)
<i>Surplus Schooling</i> (δ_o)	0.0473***	(8.04e-05)	0.0476***	(8.03e-05)
<i>Under Schooling</i> (δ_u)	- 0.0712***	(8.64e-05)	- 0.0716***	(8.61e-05)
<i>Duration</i>				
<i>2 Years</i>	-		-0.0352***	(0.0117)
<i>3 Years</i>	-		-0.0613***	(0.0124)
<i>5 Years</i>	-		-0.0738***	(0.0128)
<i>10 Years</i>	-		-0.139***	(0.0129)
<i>Constant</i>	-1.815***	(0.00154)	-1.593***	(0.0130)
<i>R-squared</i>	0.501		0.502	
<i>Control. Year Effects</i>	YES		YES	
<i>Observations</i>	41,590,015		34,290,843	

The sample for both specifications consists of individual yearly observations from 1986 to 2013. Both specifications correspond to the pooled OLS estimates of Duncan and Hoffmans log-wage regression, with the augmented wage specification including lags of past overeducation duration as covariates. For the latter, only individuals who experienced before but terminated an overeducation employment spell were considered.

required for the job (9.8%), indicating overeducated workers to earn less than their equally educated peers who find adequate matches. The remaining covariates present the expected effects, with a gender wage gap of nearly 30% and a positive return on experience (as proxied by age) and on job seniority.

Noticeably, the results from the augmented wage regression reveal persistent and statistically significant current wage penalties associated with past overeducation employment. The height of the effect increases substantially with duration in overeducation, with a current estimated pay loss of 3.5% for individuals with 2-year and of 14% for those with a 10-year overeducation employment background.

Distinct hypotheses can be offered to unveil the rationale ruling out a negative association between overeducation and future wages. First, overeducation employment may represent a period of skills underutilization. It is possible for individuals to suffer deterioration of their human capital whilst working as overeducated (de Grip *et al.*, 2008), and for the magnitude of depreciation to increase with duration in that state. Thence, the individual's marginal product, determined by his/her level of accumulated human capital, will continuously decline with overeducation duration. If workers are paid according to their marginal product (Becker, 1964), the reported pay losses are entirely consistent with the Human Capital view of the labor market.

The same reasoning could apply if overeducated workers are less productive *per se* and thus, even by changing into matching positions, employers will tend to pay them less (Groot, 1996). Therefore, to the extent that individuals who remain overeducated for longer are also the less able ones, the reported scarring wage effects may reflect the inability to control for residual unobserved ability components. If this is the case, then caution shall be exercised when interpreting the former wage effects in the standard manner, as they may suffer from the classic omitted variable bias.

Lower future wages for previously overeducated workers could nevertheless derive from stigmatization from employers. In similarity to the hiring model explored in Blanchard and Diamond (1994) for jobless duration, employers may rank employees on the recruitment process considering their overeducation duration. Higher durations may thus signal lower productivity or other unobservable undesirable features for prospective employers, shedding light on both our estimated scarring wage effects and also on our previous findings of higher durations delaying transitions into proper matches.

VII Conclusions

The present study combines data from a longitudinal matched employer-employee-job title data set to investigate the existence of scarring effects of overeducation employment along workers careers. Targeting first-job entries into overeducated positions, we analyze its persistence, examine the existence of negative duration dependence as well as associated entry and future wage effects.

Although nearly half of workers exit overeducation into a matching position 5 years after entry, the estimation of a mass-point approach hazard model identified two distinct types of overeducated workers. The mismatch appears to follow different mechanisms for both types, with overeducation employment expected to be long-lasting for nearly 30% of the overeducated population.

Men are found to leave overeducation positions quickly than women and the exit advantage is extended to workers born in Portugal. Older workers are unfavored in the hazard out of overeducation, presenting higher odds of having enlarged periods in that state.

The exit elasticity estimate with respect to elapsed duration is found negative and statistically significant, grounding the existence of negative state duration dependence. Overeducation appears as a trap and this effect is estimated as even harsh when considering matching positions as the destination venue, with a 10% increase in overeducation duration associated with a higher than 9% lower exit likelihood.

Estimation of separate wage returns for required, surplus and deficit years of schooling, reveals workers entering overeducation to yield nearly half of the return (4.7%) for their schooling investment than their equally educated peers located in matching positions. Extending the log-wage specification to include duration of previous overeducation employment, we observe not only overeducated workers to suffer a pay loss at entry but also to have significant future wage penalties associated with past employment on that state. The height of the scarring wage effect increases with elapsed duration, with an estimated pay penalty ranging between

3.5% up to 14 % for a 2-year and 10-year overeducation background, respectively.

The latter findings bear similarity to those reported to accompany prolonged unemployment spells and may relate to the findings of McGormick (1990) and Verhaest & Baert (2014) on overeducated workers to be stigmatized and for overeducation to provide an even stronger negative signal to prospective employers than unemployment.

In sum, entering overeducated positions may translate into a trap for a considerable portion of workers. We find overeducation as a disadvantageous form of employment, with overeducated workers not only suffering a pay loss by entering into such positions but for its effects to perpetuate to subsequent job allocations, with past overeducated employment associated with significant future earning losses.

VIII References

- (1) Addison, J. T., & Portugal, P. (2003). Unemployment Duration: Competing and Defective Risks. *The Journal of Human Resources*, 38(1), 156.
- (2) 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*.
- (3) Araújo , I & Carneiro, A. (2017). Educational mismatches and wages: evidence from a matched employer/employee dataset. In Portuguese Stata UGM (Sept 15).
- (4) Baert, S., Cockx, B., & Verhaest, D. (2013). Overeducation at the start of the career: Stepping stone or trap? *Labour Economics*, 25, 123140.
- (5) Battu, H., & Sloane, P. J. (2000). Overeducation and crowding out in Britain. In *The Overeducated Worker*. Cheltenham: Edward Elgar., 157175.
- (6) Bauer, T. K. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review*, 21(3), 221229.
- (7) Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: Columbia University Press.
- (8) Blanchard, O., & Diamond, P. (1994). Ranking, Unemployment Duration, and Wages. *Review of Economic Studies*, 61, 417434.
- (9) Boll, C., Sebastian, J., Rossen, A., & Wold, A. (2016). Overeducation New Evidence for 25 European Countries.
- (10) Buchel, F. (2002). The Theory of Differential Overqualification: Does it Work? *Scottish Journal of Political Economy*, 50, 1-16.
- (11) Buchel, F. & van Ham, M. (2003). Overeducation, regional labor markets, and spatial flexibility. *Journal of Urban Economics*, 53(3), 482-493.

- (12) Clark, B., Joubert, C., & Maurel, A. (2014). The career prospects of overeducated Americans. *IZA Journal of Labor Economics*, 6(1).
- (13) Cohn, E., & Khan, S. P. (1995). The wage effects of overschooling revisited. *Labour Economics*, 2, 6776.
- (14) de Grip, A., Bosma, H., Willems, D., & Boxtel, M. van. (2008). Jobworker mismatch and cognitive decline. *Oxford Economic Papers*, 60, 237253.
- (15) Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*.
- (16) Edmiston, K. (2007). The Role of Small and Large Businesses in Economic Development. *Economic Review*, 92(2), 7397.
- (17) Farris, D. (2004). Internal Labor Markets and Worker Quits. *Journal of Economy and Society*, 43(3), 573-594.
- (18) Frank, R. H. (1978). Why women earn less: The theory and estimation of differential overqualification. *The American Economic Review*, 68(3), 360373.
- (19) Freeman, R. (1976). *The Overeducated American*. New York: Academic Press.
- (20) 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(1), 2945.
- (21) Groot, W. (1996). The incidence of, and returns to overeducation in the UK. *Journal of Applied Economics*, 28(10), 13451350.
- (22) Hartog, J. (2000). Over-education and earnings: Where are we, where should we go? *Economics of Education Review*, 19(2), 131147.
- (23) Heckman, J., & Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, 52 (No.2), 271320.
- (24) Jenkins, S., Steinberg, D., Preston, D., Clarkson, D., & Colla, P. (1999). Survival analysis, (July), 375424.
- (25) Kiker, B. F., Santos, M. C., & de Oliveira, M. M. (1997). Overeducation and undereducation: Evidence for Portugal. *Economics of Education Review*, 16(2), 111125.
- (26) Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. *Econometrica*, 47, 939956.
- (27) Linsley, Ingrid, (2005), Causes of Overeducation in the Australian Labour Market, *Australian Journal of Labour Economics (AJLE)*, 8 (2), 121-143.
- (28) Buchel, F & M, P, (2004). Career prospects of overeducated workers in west Germany. *European Sociology Review*, 20(4), 321331.
- (29) Mavromaras, K., & McGuinness, S. (2012). Overskilling dynamics and education pathways. *Economics of Education Review*, 31(5), 619628.
- (30) McCormick, B. (1990). A Theory of Signalling During Job Search, Employment Efficiency, and Stigmatised Jobs. *Review of Economic Studies*, 57, 299313.

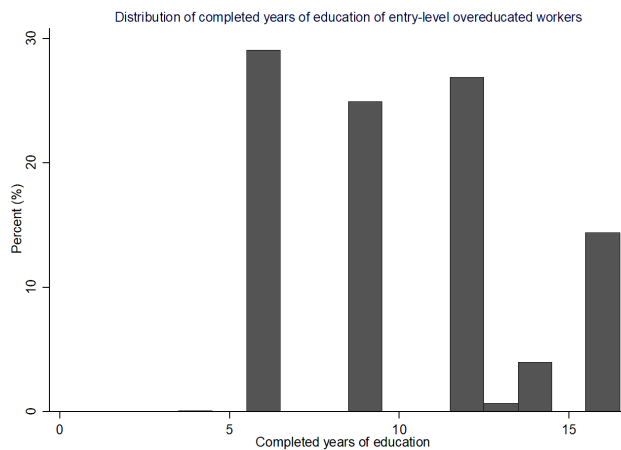
- (31) McGoldrick, K, & Robst, J (1996) .Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification. *American Economic Review*, 86, 280-284
- (32) McGuinness, S. (2006). Overeducation in the Labour. In *University of Kent* 20(3), 387419.
- (33) McGuinness, S., & Pouliakas, K. (2017). Deconstructing Theories of Overeducation in Europe: A Wage Decomposition Approach. In *Research in Labor Economics* (9698), 81127.
- (34) Meroni, E. C., Flisi, S., Goglio, V., Rodrigues, M., & Vera-Toscano, E. (2016). Measuring Occupational Mismatch: Overeducation and Overskill in Europe Evidence from PIAAC. *Social Indicators Research*, 131(3), 12111249.
- (35) Morgado, A., Sequeira, T. N., Santos, M., Ferreira-Lopes, A., & Reis, A. B. (2014). Measuring Labour Mismatch in Europe. *Social Indicators Research*, 129(1), 161179.
- (36) Nielsen, C. P. (2007). Immigrant Overeducation: Evidence from Denmark. In *Danish Institute of Governmental Research Immigrant*.
- (37) Olson, C. A. (2002). Do Workers Accept Lower Wages in Exchange for Health Benefits? *Journal of Labor Economics*, 20(S2), 91114.
- (38) OECD Outlook, E. P. (2014). *Education Policy Outlook*, (April), 126.
- (39) Pissarides, C. (1994). Search Unemployment with On-the-job Search. *The Review of Economic Studies*, 61(3), 457475.
- (40) Portugal, P., Guimares, P., & Cardoso, A. R. (2016). What drives the gender wage gap? A look at the role of firm and job-title heterogeneity. *Oxford Economic Papers*, 68(2), 506524.
- (41) Prentice, R. and L. Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57-67.
- (42) Ramirez, A. A. (1993). Mismatch in the Spanish labor market. *The Journal of Human Resources*, 28(2), 259278.
- (43) Robst, J. (1995). College quality and overeducation. *Economics of Education Review*, 14(3), 221228.
- (44) Rubb, S. (2003). Overeducation in the labor market: A comment and re-analysis of a meta-analysis. *Economics of Education Review*, 22(6), 621629.
- (45) Rumberger, R. W. (1987). The impact of surplus schooling on productivity and earnings. *Journal of Human Resources*, 22, 2450.
- (46) Sicherman, N., & Galor, O. (1990). A Theory of Career Mobility. *Journal of Political Economy*, 98(1), 169192.
- (47) Schult, M. & Buchel, F. (2004). Overeducation and human capital endowments. *International Journal of Manpower*, 25 (2), 150-166.
- (48) Tsang, M. (1987). The impact of underutilization of education on productivity: A case study of the U.S. Bell companies. *Economics of Education Review*.
- (49) Verhaest, D., & Baert, S. (2014). Unemployment or overeducation: Which is a worse signal to employers?. *IZA* (8312).

Appendix

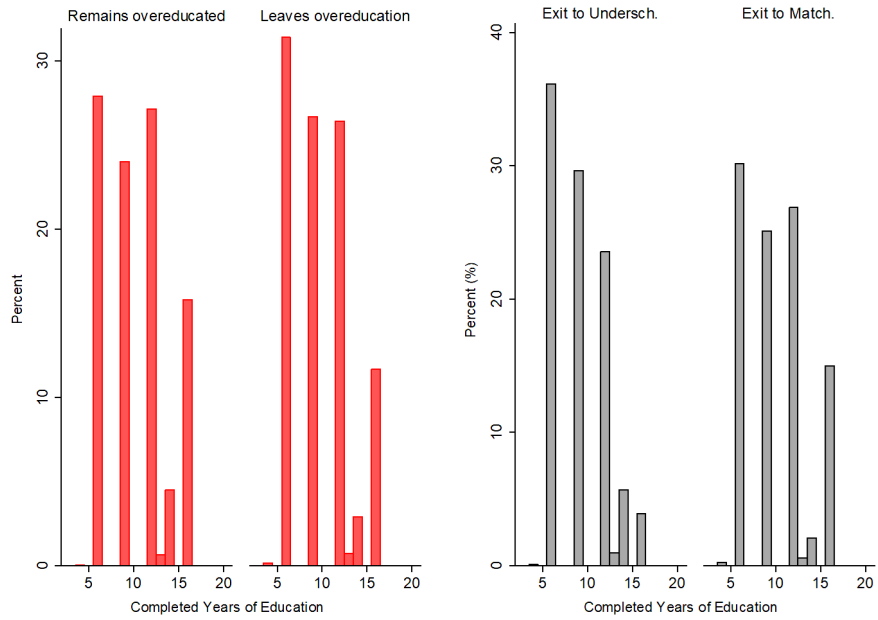
Appendix I. Overeducation incidence by Sectors of activity (left) and District location (right)

Top 4 Sectors of Activity for overeducated workers, by gender		Top 5 districts with higher % of overeducated population
<i>Males</i>	<i>Females</i>	Lisboa
Construction of buildings	Manufacture of ready-to-wear outwear	Porto
Construction of residential and non-residential buildings	Hotels with restaurant	Braga
Freight Transport by road	Social assistance to the elderly	Aveiro
Construction and Civil Engineering	Retail Sales in Super and Hypermarkets	Setúbal

Appendix II. Distribution of completed years of education of entry-level overeducated workers



Appendix III. Distribution of completed years of schooling of individuals who leave and those remaining as overeducated (left) and for those who leave by destination route (right)



Appendix IV. Duration in overeducated positions, by school diploma

