DISCUSSION PAPER SERIES

IZA DP No. 8313

The Career Prospects of Overeducated Americans

Brian Clark
Clément Joubert
Arnaud Maurel

July 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Career Prospects of Overeducated Americans

Brian Clark
Duke University

Clément Joubert
University of North Carolina at Chapel Hill and IZA

Arnaud Maurel
Duke University, NBER and IZA

Discussion Paper No. 8313
July 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 8313
July 2014

ABSTRACT

The Career Prospects of Overeducated Americans*

In this paper we analyze career dynamics for the large share of U.S. workers who have more schooling than their peers in the same occupation. We use data from the NLSY79 combined with the CPS to analyze transitions into and out of overeducated employment, together with the corresponding effects on wages. Overeducation is a fairly persistent phenomenon at the aggregate and individual levels, with 66% of workers remaining overeducated after one year. Overeducation is not only more common, but also more persistent among blacks and low-AFQT individuals. Further, the hazard rate out of overeducation drops by about 60% during the first 5 years spent overeducated. However, the estimation of a mixed proportional hazard model suggests that this is attributable to selection on unobservables rather than true duration dependence. Finally, overeducation is associated with lower current as well as future wages, which points to the existence of scarring effects.

JEL Classification: J24, I21

Keywords: overeducation, mismatch

Corresponding author:

Arnaud Maurel
Duke University
Department of Economics
213 Social Sciences Building, Box 90097
Durham, NC 27708-0097
USA
E-mail: apm16@duke.edu

* Clément Joubert and Arnaud Maurel gratefully acknowledge financial support from the W.E. Upjohn Institute for Employment Research through the Early Career Research Grant No. 13-141-02B. We thank Peter Arcidiacono, Pat Bayer, Donna Gilleskie, Joop Hartog, Lisa Kahn, Francis Kramarz, Sang Yoon (Tim) Lee, Brian McManus, Hessel Oosterbeek, Seth Sanders, Ken Wolpin, Shintaro Yamaguchi and the attendees of the 2013 North American Summer Meeting of the Econometric Society (Los Angeles), the workshop "Empirical Research in the Economics of Education" (Rovira i Virgili University, Oct. 2013), the workshop "Skill Mismatch: Microeconomic Evidence and Macroeconomic Relevance" (Mannheim, April 2014), and seminar participants at Duke, John Hopkins and UNC - Chapel Hill for useful comments and suggestions.

1 Introduction

As American students accumulate college loan debt (1.08 trillion dollars as of December 31, 2013, Source: NY Fed), there is a growing concern that expensive skills acquired in college may be underutilized in low-paying jobs. Existing cross-sectional studies estimate that around a third of American workers are "overeducated", i.e. have more schooling than necessary for their job (see Leuven & Oosterbeek, 2011). To the extent that these workers earn substantially less than their matched peers (returns to surplus schooling average 4.3%, or about half of the returns to required schooling), overeducation may account for a sizeable portion of within-schooling level wage dispersion.[1] At the aggregate level, overeducation could reflect skill mismatch, and be inefficient relative to an allocation of workers to jobs that require their actual level of education.[2]

In order to understand how much of a problem overeducation really is, it is crucial to go beyond the cross-sectional stylized facts and investigate longitudinal patterns. This paper provides the first analysis of the career dynamics of overeducated U.S workers. Specifically, we use the National Longitudinal Survey of Youth 1979 (NLSY79), combined with the pooled 1989-1991 waves of the Current Population Survey (CPS), to analyze (1) the determinants of overeducation, (2) how overeducation incidence changes along the career, and (3) how wages differ across overeducated, matched or undereducated employment, and as a function of previous overeducated employment spells.

[1]These returns to surplus schooling are obtained by Leuven & Oosterbeek (2011) by averaging estimates from 151 studies.

[2]Throughout the paper, we use "matched" as a shorthand for being neither in an overeducated or in an undereducated employment spell. Overeducated (undereducated) employment spells correspond to situations where individuals have more (less) schooling than "required" for their job (see page 5 for a discussion of the notion of required education). In this paper we use "overeducation" or "overeducated employment" in place of the more accurate "employment spell in which a worker is currently overeducated". While our analysis primarily focuses on overeducation, it is worth noting that undereducation is also relatively common, in particular among individuals who have completed 14 years of education, and, as such, has the potential to account for some of the wage dispersion within this schooling category.

While the literature has paid relatively little attention to the longitudinal dimension, analyzing transitions into and out of overeducated employment, together with their effects on wages, is key to disentangling the role played by labor market frictions versus other factors such as selection on ability, compensating differentials and career mobility prospects.[3] For example, if overeducation was due only to search frictions, one would expect this type of mismatch to be transitory and concentrated early in the career. Conversely, selection on ability, compensating wage differentials or career mobility motives would generate persistence in the overeducation patterns.[4] The individual persistence and duration dependence of overeducated employment, together with the wage penalties associated with it, are also important for the design of unemployment insurance and training programs. For instance, encouraging early exit from unemployment may push more workers into overeducated work with potentially negative long-term effects on earnings.[5]

The question of overeducation was first brought to the attention of economists and policy-makers by Freeman (1976), who argued that excess supply of college graduates was causing the decline in the college wage premium observed in the U.S. during the 1970's. While the cross-sectional properties of overeducation are well-studied (see, e.g., Verdugo & Verdugo, 1989; Alba-Ramirez, 1993; Kiker et al., 1997; Hartog, 2000), still little is known about the evolution of overeducation over the life cycle, although, as argued above, dynamics are of clear interest in this context. U.S. evidence is particularly

[3]The career mobility factor was initially investigated by Sicherman & Galor (1990). The general idea is that high-skilled workers may face higher promotion probabilities in low-skilled jobs. It follows that forward-looking individuals may choose to become overeducated.

[4]See Gottschalk & Hansen (2003) who show that, in a model with two sectors, two skills (college and non-college) and heterogeneous preferences for each sector, some college workers may choose to work in the non-college sector. Uncertainty in returns to schooling is another channel which could generate persistence in overeducation, since some individuals may find it *ex post* optimal to be overeducated (see Lee et al., 2014).

[5]In a recent paper, Baert et al. (2013) show that accepting an overeducated job, versus staying unemployed, has the effect of postponing access to a matched job for a sample of Flemish youth.

scarce.[6] A notable exception is Rubb (2003) who provides evidence from the CPS that overeducation displays a substantial degree of persistence, with around 30% of the individuals overeducated in year t switching to a job which matches their level of education in year $t+1$. While duration dependence and dynamic selection effects imply that these transition rates are likely to decrease over the length of the spell, the CPS panels are too short to address this question adequately.

It is worth pointing out that while we borrow the wording "required level of education" from the existing literature, defining and measuring that concept is not a trivial task. In this paper we use a statistical measure, in a similar spirit as, e.g., Verdugo & Verdugo (1989) and Kiker et al. (1997). Namely, we compute the mode of the distribution of schooling in the 1989-1991 CPS for each occupation in the 1980 3-digit Census Occupation classification. We also restrict the CPS sample used to compute the mode to individuals in the same birth cohorts as the NLSY79 respondents. The required levels of education are then defined as those within 15 percentage points of the schooling mode.[7] The typical overeducated worker in our sample has two or four years of college education, but is working as a secretary or a cashier, say, among a majority of high school graduates. Relative to alternative approaches in the literature, and in particular those measuring the required levels of education with the General Educational Development (GED) scale provided by the Dictionary of Occupational Titles (see, e.g., Hartog, 1980, and Rumberger, 1987), this approach is arguably more transparent and has the benefit of directly generating requirements in terms of years of education. As such,

[6] A few studies have used German, British, Canadian or Australian data to estimate panel wage regressions (Bauer, 2002; Frenette, 2004), dynamic random effect models of overeducation exit (Mavromaras et al., 2013) or simply document overeducation status transitions (Dolton & Vignoles, 2000).

[7] Using this type of conservative measure of mismatch allows us to mitigate the risk of misclassification, which could arise if two or more occupations with different required levels of education were aggregated up at the 3-digit level. Depending on the aggregation level used for the educational attainment, our measure of overeducation yields incidence levels of between 18% and 25% among all workers, and as much as about 40% among college graduates.

while our required schooling variable should be a pretty good indicator of the amount of schooling needed to perform well in an occupation, it is most cleanly interpreted as the schooling attainment of a typical worker in one's occupation.

We document longitudinal patterns of overeducation for the NLSY79 cohort up to 12 years after labor market entry. Overeducation incidence within the cohort decreases as workers progress through their careers, but remains sizeable 12 years after the first job. This suggests that, while frictions are likely to play a role, we need to bring other economic mechanisms into the fold to explain this long-term persistence. Overeducation is also a fairly persistent phenomenon at the individual level, with around 66% of overeducated workers remaining in overeducated employment after one year. We find that blacks and low cognitive ability workers (as measured by their AFQT scores) are not only more likely to be overeducated, but also less likely to switch into matched jobs. That is, the longitudinal dimension magnifies the cross-sectional black-white and cognitive ability gaps.

The hazard rate out of overeducated work is also strongly decreasing in overeducation duration, and drops by about 60% after five years. We estimate a mixed proportional hazard model (Elbers & Ridder, 1982) of overeducated employment duration to investigate whether this decreasing hazard rate reflects selection on unobservables or true duration dependence. While composition effects based on observable characteristics explain away some of the duration dependence, further controlling for unobserved heterogeneity largely wipes it out. In other words, the duration of the overeducated employment spell does not have a significant impact on the probability to exit overeducation. Instead, we identify large unobservable differences in the hazard rate: while overeducation is found to be very persistent for 30% of the sample, the rest is much more likely to exit quickly, consistent with a frictional view of overeducation.

We then revisit the classical augmented wage regression used in the overeducation literature.[8] In particular, we stratify the estimation based on the unobserved heterogeneity types identified in the duration model. We show that individuals with lower propensities to exit overeducated employment also exhibit lower returns to schooling. Importantly, we also find that past overeducation entails a sizeable wage penalty of between 2.6% and 4.2%, which persists over four years. Past overeducation is also associated with lower returns to tenure in future matched employment spells. This provides a likely candidate mechanism behind the negative wage effects of graduating during a recession recently discussed in the literature (Kahn, 2010, Liu et al., 2012, Oreopoulos et al., 2012, and Altonji et al., 2013), since overeducation is likely to be more frequent during recessions, consistent with the cyclical upgrading literature (Bils & McLaughlin, 2001).

The remainder of the paper is organized as follows. Section 2 describes the data used in the analysis and the construction of the required schooling measure. Section 3 discusses the determinants of overeducation in our sample. Section 4 documents the longitudinal patterns in the incidence of overeducated employment along the career. Section 5 estimates a mixed proportional hazard model of overeducated employment duration allowing us to separate true duration dependence from dynamic selection on observed and unobserved worker attributes. Section 6 presents results pertaining to the effect of overeducation on wages, and Section 7 concludes. All tables and figures are collected in the Appendix.

[8] Augmented wage regressions, pioneered by Duncan & Hoffman (1981) replace the usual years-of-education regressor with three terms: years of required education in the current occupation, years of education in excess of that required level and years of education below that required level. The corresponding coefficients are interpreted as returns to required education, returns to overeducation and returns to undereducation.

2 Data

Our main data source is the NLSY79 which is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979.[9] We pool the observations for the 6,111 individuals that comprise the core civilian cross section of the NLSY79, from the 1982 to the 1994 rounds, which results in 79,443 person-year observations.[10] Then we cut 5,947 person-year observations with a level of education that was unreported or less than 12 years; 22,272 where the individual had not entered the labor market permanently; and 6,228 that were non-interviews.[11] After making these cuts, we are left with a total of 44,996 observations corresponding to 4,895 distinct individuals.

The main variables of interest are the highest level of completed education, the occupation (measured using the 1980 3-digit Census code) and the hourly wage at the time of each interview.[12] Besides these, the variables used in our analysis include age, minority status, gender and place of birth, cognitive and non-cognitive skill measures,

[9]Of the 12,686 individuals interviewed in the initial 1979 wave, these data provide information for respondents on a yearly basis from 1979 to 1994 and biyearly afterwards.The initial wave is comprised of a core civilian cross-section of 6,111 and an oversample of 5,295 black, Hispanic, and economically disadvantaged individuals born between January 1, 1957 and Dec. 31 1964. This is further supplemented by a military sample of 1,280 individuals born in the same period. We only keep the core cross-sectional sample of the NLSY79 in order to maintain a consistent sample between the NLSY79 and the CPS.

[10]We restrict our sample to these years because it is the largest contiguous period where the NLSY79 reports the 1980 Census Occupation codes on an annual basis. We use the 1980 codes because they better reflect the set of occupations available over the period of interest than the 1970 codes.

[11]We define the date of entry as the first survey year where (1) the individual is employed in the civilian labor force, (2) works more than 26 weeks out of the year, (3) is not enrolled in school as of May 1st of the survey year, and (4) has reached her highest level of education over the sample period 1982-1994.

[12]In practice, we use the occupation and hourly wage corresponding to the current or most recent job at the time of the interview. When individuals hold multiple jobs at the same time, we use the occupation and wage corresponding to the job in which the respondent worked the most hours. We adjust for inflation by reporting all wages in constant dollars and then drop the top and bottom 2.5% of the reported person-year wages for every survey wave.

geographical location and the corresponding local unemployment rate, family characteristics, a measure of hazards associated with the current occupation and employment history (see Table 1).

2.1 Measuring required schooling

The NLSY79 does not have direct measures of required schooling in the job occupied by the respondent. In this paper, we use a statistical measure for the required level of education, in a similar spirit as, e.g., Verdugo & Verdugo (1989) and Kiker et al. (1997). Namely, for each given occupation in the 1980 3-digit Census Occupation classification, we compute the required level of education from the pooled monthly samples of the 1989-1991 waves of the CPS.[13] In order to obtain a sample that matches the NLSY79 sample, we restrict the age range within each year of the CPS to that of the NLSY79 cohorts at that time (see Appendix A for additional details on the CPS sample used in the analysis). Required schooling in a given occupation code is then defined as the mode of the distribution of the levels of education among the individuals working in that occupation.[14] For occupations such that the frequencies of two or more schooling levels are within 15 percentage points of each other, we choose to use a more conservative definition of over (and under)-education. Specifically, workers whose schooling attainments fall within the range defined by these schooling levels are classified as matched, while those with a higher (lower) level of education are defined as overeducated (undereducated). It is important to note that our results are robust

[13]See Yamaguchi (2012), who provides evidence based on task complexity that using occupations at the 3-digit level is key to properly capture the heterogeneity of jobs.

[14]Finite sample variability should not be a major concern here given the large number of observations (on average above 1,000) which are used to estimate the mode of each occupation. See, e.g., Dutta & Goswami (2010), who show that, for a Bernoulli distribution with sample size larger than 100, the mode of the empirical distribution matches the population mode with a probability close to 0.9.

to the choice of other cutoffs.[15] In order to mitigate concerns with classification error on attained schooling, we collapse our years of education variable into four categories: 12-13 years, 14-15, 16-17, and over 18 years of completed education. This classification is natural since each category, simply referred to as 12, 14, 16 and 18 years of schooling in the rest of the paper, corresponds to high school graduates, two-year and four-year college graduates, and graduate school.[16]

A clear advantage of this method is that it generates requirements directly in terms of years of education. This is in contrast to a common approach in the literature which maps occupations into skills first, and then skills into years of education (using, for example, the GED scale). In practice the latter approach is problematic, in particular since there is no clear consensus on the mapping between the skill content of occupations, as measured by the GED scale, and years of schooling (Leuven & Oosterbeek, 2011). On the other hand, one limitation of our measure of required schooling is that it is based on the distribution of schooling attainment among workers employed in a given occupation, which is an equilibrium outcome of labor supply and demand decisions.[17] Ultimately, one can simply think about the workers we identify as overeducated as those who have more schooling than their peers, in the sense of having more schooling than the modal worker in their current occupation.

[15] Specifically, we considered four alternative definitions of over (and under)-education, using (i) the mode only (i.e. no cutoff), (ii) a 5% cutoff, (iii) a 10% cutoff and (iv) a 20% cutoff. Our results were overall very similar to those obtained using our baseline definition.

[16] It is also possible that errors in the occupation codes recorded in each interview of the NLSY79 could generate artificial transitions between overeducation statuses. In that case, our estimates of overeducation persistence could still be interpreted as lower bounds.

[17] It follows that fluctuations in demand and supply will affect that distribution even if the actual skills required to perform the job have not changed.

2.2 Summary statistics

Table 1 presents summary statistics for the NLSY79 variables used in our analysis. Of the 44,996 observations in our final sample, 84.7% are employed, 10.6% are out of the labor force, and 4.8% are unemployed. 66.1% are high school graduates, 12.9% have two years and 14.8% have four years of college education, and 6.3% have some graduate school experience.

Figure 1 shows the distribution of years of overeducation, generated by subtracting an individual's observed highest level of completed education, from the level of education required by their occupation. More than half of all observations have a perfect match of observed and required education, while progressively smaller fractions exhibit 1 or more years of education level mismatch. Comparing the two panels shows that collapsing schooling attainment into four categories preserves the shape of the distribution. It also mitigates the concern that small errors in the measure of years of schooling attainment will generate overeducation status misclassification.

In Table 2, we further break down overeducation status by our categorical measure of attained schooling. It is apparent that overeducation is mechanically absent from the lowest schooling level (12 years of education) while undereducation is absent from the highest schooling level (18 years of education). Consistent with the existence of a relatively small number of jobs requiring 14 or 18 years of education, a very large fraction of the observations corresponding to those schooling levels exhibit education level mismatch. Although a larger share of jobs require 16 years of education, it is interesting to note that 37.4% of college graduates are overeducated, typically working in a job requiring 12 years of education. Table 3 lists the 10 occupations accounting for the most number of observations in overeducated employment. Secretaries and Sales workers account for the largest numbers of overeducated workers. For most of the

occupations in the table the modal worker has 12 years of schooling, and the typical overeducated worker has two or four years of college education. One notable exception is teachers, which typically have 16 years of schooling in the CPS.

3 The determinants of overeducation

Before moving on to the analysis of the career dynamics of overeducated workers, we start by documenting the cross-sectional determinants of overeducation in our sample. We report in Table 4 the estimation results from a Probit model, which allows the probability of overeducation to depend on a set of socio-demographic characteristics, ability measures, family characteristics, a measure of hazards associated with the current occupation and employment history.[18] We stratify the regression by schooling level, consistent with our focus throughout the paper on overeducation as a labor market, rather than educational, phenomenon.

AFQT scores exhibit a negative and significant relationship with overeducation at all schooling attainments. This negative relationship between cognitive ability and the likelihood of overeducation, which is in line with prior findings in the literature (see, e.g., Allen & Van der Velden, 2001 and Chevalier & Lindley, 2009), is consistent with ability and schooling attainment being labor market substitutes.

Also in accordance with existing studies, females are about 5 to 13% more likely to be overeducated than males. Women may place more value on non-pecuniary characteristics associated with low-requirement jobs, such as flexibility in hours worked or their proximity to the family house, making it easier to combine work and home production activities. Alternatively, this could reflect discrimination on the part of employers.

[18] In practice we use a partial maximum likelihood estimator, clustering standard errors at the individual level. The resulting inference is robust to serial correlation in the unobserved determinants of overeducation.

At any rate, given the existence of a substantial and persistent wage penalty of being overeducated (discussed in Section 6), this result implies that overeducation is an important aspect of the gender wage gap, which is absent from most of the literature on this question.[19]

Noteworthy, there is also a strong positive correlation between minority status and overeducation at the 14 years of schooling level. Blacks and Hispanics are 15.9% and 12.2%, respectively, likelier than whites to be overeducated among that group. Among college graduates and above, the relationship becomes insignificant, which is possibly a result of a strong selection into college based on unobservable skills that are negatively correlated with overeducation.

Lastly, the evidence regarding the theory that overeducated workers accept lower wages in exchange for better non-pecuniary job characteristics is mixed. Workers in hazardous jobs are actually more likely to be overeducated at all three levels of schooling. On the other hand, overeducated workers are more likely to hold several jobs (except for those with 18 or more years of schooling), which possibly reflects a higher flexibility for these overeducated jobs. These results complement previous studies that have interpreted a negative correlation between overeducation and job satisfaction as evidence against the compensating wage differential model of overeducation (see, e.g., Hersch, 1991, and Korpi & Tåhlin, 2009).

4 The dynamics of overeducation

Figure 2 displays the aggregate incidence of overeducated employment over the first 12 years of work for the NLSY79 cohort, for workers with at least some college. Overall, the incidence of overeducation decreases by about 12 percentage points, from 62.3% to

[19] One notable exception is the early analysis by Frank (1978).

50.4%, over the first 12 years of respondents' careers.[20] While the decline in aggregate overeducation rates as the career progresses does suggest that overeducation is in part frictional, the most striking feature of this graph is that the incidence of overeducation remains very high more than 10 years after labor market entry. Overall, this is a clear indication that overeducation is a persistent phenomenon.

In Figure 3, we further disaggregate this graph along several observable characteristics. Blacks do not exhibit the same reduction in overeducation as whites (Panels 1 and 2). Similarly, Panels 3 and 4 show that overeducation among females decreases much less than among males. Lastly, individuals at higher AFQT quartiles see a larger decline in overeducation incidence than those at lower quartiles (Panels 5 through 8). Overall, these results suggest that overeducated black, female and low-AFQT workers are less likely to receive and/or accept offers from matched jobs. These dynamic patterns therefore accentuate their already higher propensity to be overeducated.

Individuals patterns point to a similar story. Table 5 presents the fractions of individuals who are non-employed (defined as unemployed or out of the labor force), undereducated, overeducated or matched at interview time, conditional on the status reported during the interview one year before, for workers with at least some postsecondary education. Overeducation is also persistent at the individual level, with 65.9% of workers remaining overeducated after one year. This fraction is 20 points higher than the unconditional overeducation rate (46.3%). By comparison, non-employed individuals have a smaller 49.0% chance of remaining in that state.[21]

Transition rates are further broken down by gender and race in Tables 6 and 7. Men are

[20] Using a self-reported measure of overeducation, Dolton & Vignoles (2000) also find that its incidence among a 1980 cohort of U.K. university graduates decreased over time, from 38% to 30% 6 years after graduating.

[21] With a different measure of overeducation and using CPS data, Rubb (2003) obtains a level of persistence for overeducated individuals of 73%.

slightly more likely than women to remain in an overeducated job but also more likely to transition into matched jobs. Differences by race are sizeable: i) overeducation is more persistent among blacks relative to whites, ii) overeducated blacks are much less likely than whites to transition to a matched job, and iii) these matched spells are less persistent for blacks than for whites. Finally, as illustrated in Table 8, the persistence of overeducation decreases monotonically with AFQT scores. All of these patterns in individual transitions confirm that the aggregate persistence described earlier does not result from cancelling flows in and out of overeducation.

Table 5 also shows that transitions into overeducation are equally likely among workers who were undereducated, matched or non-employed in the previous year. Among males, however, non-employed workers are much likelier to transition into overeducation than matched or undereducated workers (Table 6). Across all categories of workers, but especially for black and low-AFQT workers, the non-employed are more likely to transition into overeducation than into matched jobs (Table 7 and 8). In addition, transitions into matched jobs are more common among the overeducated than the non-employed. Taken together, these patterns suggest that for some workers overeducation is a pathway from non-employment into matched employment.

Finally, the NLSY79 data allows us to go beyond annual transitions, and report the hazard rates out of overeducated employment as a function of the duration of overeducation (see Figure 4).[22] After 3 years being overeducated, the probability of exiting that state, defined as starting a new matched or undereducated employment spell, drops from 39% to only 20%. This number further drops to 15% and 10% after 5 and 10 years respectively. Overall, while this pattern is consistent with a negative duration dependence associated with overeducated employment, it could also result from compositional

[22]For any given year t, these hazard rates are computed as the number of individuals leaving overeducated employment in year t, divided by the number of individuals who are still overeducated at the beginning of year t.

effects (permanent heterogeneity correlated with the hazard rate out of overeducated employment). We attempt to tell these two effects apart in the next section.

5 Duration dependence versus dynamic selection

The results discussed so far provide some suggestive evidence of duration dependence in overeducation status, with a strongly decreasing hazard rate out of overeducation. However, in order to establish the role played by true, rather than spurious duration dependence, we need to control for dynamic selection on worker attributes.

Specifically, we assume that the duration of the first spell of overeducated employment is determined by a mixed proportional hazard model, where the baseline duration follows a Weibull distribution.[23] The probability distribution function (pdf.) and cumulative distribution function (cdf.) of the duration of the overeducation spell, conditional on the set of observed individual characteristics x_i and the unobserved heterogeneity ν_i, are respectively given by:[24]

$$f\left(t|\mathbf{x}_i,\nu_i,\alpha,\theta\right)=\exp\left(\mathbf{x}_i\theta\right)\alpha t^{\alpha-1}\nu_i\exp\left[-\exp\left(\mathbf{x}_i\theta\right)t^{\alpha}\nu_i\right], \tag{5.1}$$

$$F\left(t|\mathbf{x}_i,\nu_i,\alpha,\theta\right)=1-\exp\left[-\exp\left(\mathbf{x}_i\theta\right)t^{\alpha}\nu_i\right]. \tag{5.2}$$

Following Heckman & Singer (1984), we assume that the unobserved heterogeneity follows a discrete distribution with R points of support. The parameters (α,θ) and the unobserved heterogeneity distribution are then estimated by maximizing the log-

[23] By definition, this model is estimated on individuals who have at least one overeducated spell, which mechanically excludes individuals with only 12 years of schooling.

[24] While using a parametric specification allows us to get more precise estimates, it is important to note that the mixed proportional hazard model is identified nonparametrically from single-spell data (see Elbers & Ridder, 1982).

likelihood of the data, which is given by:

$$\ell = \sum_{i=1}^{N} \left[d_i \log f\left(t_i | \mathbf{x}_i, \alpha, \theta\right) + \left(1 - d_i\right) \log \left[1 - F\left(T_i | \mathbf{x}_i, \alpha, \theta\right)\right]\right]$$

where N is the number of individuals in the sample with at least one overeducated spell, $d_i = 1$ if individual i leaves the overeducated state before the end of the survey (0 otherwise), t_i is the duration of the first overeducation spell (observed if $d_i = 1$) and T_i is the length of time to the end of the survey. The pdf. of the overeducation spell duration is given by $f\left(t|\mathbf{x}_i, \alpha, \theta\right) = E_\nu\left(f\left(t|\mathbf{x}_i, \nu_i, \alpha, \theta\right)\right)$ and the cdf. is given by $F\left(t|\mathbf{x}_i, \alpha, \theta\right) = E_\nu\left(F\left(t|\mathbf{x}_i, \nu_i, \alpha, \theta\right)\right)$, where $E_\nu(.)$ denotes the expectation operator with respect to the distribution of ν. Throughout our analysis, we consider that an overeducation spell ends when the individual starts a new employment spell in an occupation which does not require less than his level of schooling.[25]

It is interesting to compare the estimates for $R = 1$ (i.e., without unobserved heterogeneity) and $R = 2$ (i.e., with two unobserved heterogeneity types).[26] Columns 2,4 and 6 of Table 9 report the estimation results corresponding to the case without unobserved heterogeneity with alternative sets of individual controls. The estimated α is well below one (between 0.77 and 0.84 depending on the specification), which means that the hazard out of overeducation is strongly decreasing in the duration of overeducation even after controlling for *observed* heterogeneity. In other words, while part of

[25]The NLSY79 allows us to compute the required level of education only for the current or most recent job at the time of each interview (the "CPS job"). To keep the exposition simple, the duration model we present in the text ignores the existence of job spells that might have occurred in between two interviews. We also estimated a model that explicitly takes into account these between-interview employment spells, treating the corresponding overeducation status as missing, which resulted in negligible differences in the estimation results (available upon request).

[26]We estimated the model with more than two types, but including these additional types did not significantly improve the fit of the model. It resulted in higher BIC criteria than the model with two types while attributing similar values for the key parameters.

this negative duration dependence is attributable to selection on observables, as shown by the increase in the estimated α parameter going from specification (3) to (1), the exit rate is still declining with the duration of the overeducation spell after controlling for an extensive set of observed characteristics.

By contrast, once we allow for unobserved heterogeneity, we obtain values for α that are very close to one, signifying the absence of true duration dependence (Table 9, columns 1, 3, and 5). In other words, the duration of the first overeducated employment spell does not seem to have a significant impact on the probability to exit overeducation. The estimation results point to the existence of two groups of individuals with markedly different dynamics: the first group has a low hazard (type 1, 28.9% of the ever-overeducated in the sample) while the second group is much more likely to exit overeducation quickly (type 2, 71.1%). As type 2s exit the pool of overeducated individuals, the probability that a random individual exits overeducation declines, as she is more and more likely to be a low-hazard, type 1 individual.

The ratio between the two unobserved heterogeneity parameters is $\frac{V_1}{V_2} = 0.11$, which implies that type 2s exit overeducation almost 10 times as fast as type 1s. Interestingly, the stark difference in hazard rates between the two unobserved groups suggests that overeducation follows different mechanisms in each case. The high exit rate in group 2 is consistent with a frictional view of overeducation. For the remaining, low-hazard third of the sample, it could be that their aptitude (after controlling for AFQT) is not sufficient for jobs that match their level of formal schooling. Alternatively, they may have preferences for non-pecuniary, unobserved job characteristics found in some of the jobs that require less schooling, thus translating into highly persistent overeducation. In the next section we explore these ideas further as we examine whether the two types differ in the wage patterns they exhibit.

The coefficients on observable characteristics allow us to complement and refine the analysis of the raw transition dynamics presented in Section 4. The negative coefficient on blacks in all specifications of the duration model implies that the lower exit rates out of overeducation for that group are robust to controlling for observable and unobservable heterogeneity. Similarly, the positive effect of AFQT scores on overeducation exit also carries over from the raw transitions into the mixed proportional hazard model. The coefficient for women is also significantly negative when the full set of observable characteristics is included in the proportional hazard model. However once unobserved heterogeneity is taken into account, being female no longer has a direct effect on the probability of exiting overeducation. This runs against the notion that discrimination would keep women in jobs where they are overeducated.

Lastly, an interesting finding is that time spent unemployed in the past has a negative effect on overeducation exit. This result complements a previous finding in the literature that longer time spent unemployed reduces the probability of finding a job (see e.g. Kroft et al., 2013). It also suggests that higher overeducation persistence is one of the mechanisms through which past unemployment affects future wages (see e.g. Schmieder et al., 2013).

6 Effects of overeducation on wages

6.1 Wage dynamics

While the regressions found in the literature focus on the cross-sectional correlation between current overeducation and wages, we examine whether initial overeducation is also associated with lower wages later in the career. Figure 5 describes the median hourly wage among workers with 14, 16 and 18 years of education as they progress

through their career, both unconditional and conditional on their overeducation status. The striking result here is that the penalties from overeducation at the start of the career appear very persistent over time for the first two schooling categories (14 and 16 years). Notably, this feature is consistent with the existence of long-lasting scarring effects of being overeducated at the start of the career. In the following sections, we provide additional evidence of these effects by showing that the negative association between past overeducated employment and wages still holds after controlling for observed heterogeneity (including AFQT scores and measures of non-cognitive skills), and for the unobserved heterogeneity types identified in the duration analysis.

6.2 Augmented wage regressions

The impact of overeducation on wages has been measured in the literature by applying OLS to the following log-wage equation introduced by Duncan & Hoffman (1981):

$$\log w_{it} = \alpha^r S_{it}^r + \alpha^o S_{it}^o + \alpha^u S_{it}^u + X_{it}'\beta + \varepsilon_{it}$$

where, for any given individual i in year t, S_{it}^r, S_{it}^o and S_{it}^u denote respectively the number of years of required schooling, years of overeducation (years of schooling above the required level) and years of undereducation (years of schooling below the required level), X_{it} a vector of controls (including ability measures, socio-demographic background characteristics, labor market experience and experience squared) and ε_{it} an idiosyncratic productivity shock. This model, which nests the standard Mincerian wage regression ($\alpha^r = \alpha^o = -\alpha^u$), allows for the estimation of separate wage returns to the (i) required years of schooling, (ii) years of overeducation, and (iii) years of undereducation.

Table 10, Panel 1, presents the pooled OLS parameter estimates for this model using

our data.[27] Unlike for the rest of our analysis, individuals with 12 years of schooling are part of the estimation sample. While none of these individuals are overeducated, including them helps identify the returns to the required years of schooling, years of overeducation and undereducation. The results from our sample are broadly consistent with prior existing studies. Namely, we see a return of 9.6% for each additional year of education, less than half the rate of return (3.8%) for additional years of education over the required level, and substantial wage penalties for years of undereducation (-6.9% for each additional year of undereducation).

These estimates also reveal a number of expected results. In particular, we see a gender wage gap of 15.5%, and a significant compensating wage differential for hazardous occupations. We also find significant wage premia relative to the North Central region of 9.7% and 6.9% for Northeast and West, respectively, consistent with higher average costs of living in these regions. Lastly, measures for labor market experience all have the expected sign. Those with greater occupation-specific tenure, and total labor market experience have higher wages. Conversely, those with frequent and/or long unemployment spells (as measured by total unemployment experience in weeks), and those frequently switching jobs during the year have lower wage rates, all else equal.

Following-up on the dynamic patters discussed in Section 6.1, we report in Panel 2 of Table 10 the estimation results from a log-wage regression which is further augmented with four lags in overeducation status. We also include interactions between lagged overeducation status and an indicator for the fact that the overeducation spell is ongoing. This disentangles the effect of past completed overeducation spells on wages when

[27]For the occupations such that the discrepancy between the frequency of the mode and the second most frequent schooling level is less than 15 percentage points, individuals whose schooling level falls with the range defined by the two most frequent schooling levels are assumed to be matched, and years of required schooling is set equal to their actual level of education. Individuals whose schooling level is higher (lower) than the upper (lower) bound of the aforementioned range are assumed to be overeducated (undereducated), and years of required schooling is set equal to that upper (lower) bound.

an individual is currently matched or undereducated, from the effect of the duration of an ongoing overeducated spell on current wages if an individual is still overeducated. Notably, we find wage penalties of 2.6-4.2% per year associated with up to four lags of overeducation. Those effects are both statistically and economically significant.[28] It is interesting to compare the magnitude of these scarring effects with those generated by past unemployment. While the penalty associated with the first lag of unemployment is found to be sizeably stronger than that associated with overeducation (7.6% vs. 3.9%), the penalty associated with past unemployment is also less persistent as further lags do not significantly affect wages.

These results show that the overeducation scarring effects suggested by the raw wage data remain after including a rich set of controls, including measures of cognitive and non-cognitive ability. To the extent that, consistent with the cyclical upgrading literature (see, e.g., Bils & McLaughlin, 2001), overeducated employment is likely to be more frequent during recessions, it follows that overeducation is a likely candidate mechanism behind the negative and persistent wage effects of graduating during a recession recently uncovered in the literature (Kahn, 2010, Liu et al., Oreopoulos et al., 2012, and Altonji et al., 2013).

Finally, Table 11 reports estimates from our log-wage regression model augmented with one lag in overeducation status, interacted with occupation-specific tenure.[29] The coefficient in front of the interaction term is estimated to be negative and significant, both statistically and economically. Interestingly, this implies that returns to tenure are lower for workers who have been previously overeducated. This result, along with the scarring effects discussed above, is suggestive of an adverse effect of overeducation

[28]Older lags, on the other hand, do not generate significant additional penalties (results available from the authors upon request).

[29]We only report the results from a regression with one lag in overeducation status since specifications with more lags yielded imprecise estimates.

not only on the level, but also on the on the growth of future wages.

6.3 Accounting for heterogeneity types

Next, we run our augmented wage regressions separately for each of the unobserved types identified from the mixed proportional hazard model.[30] This approach complements the characterization of the two unobserved types discussed in Section 5, by allowing the types to have different returns to required schooling, years of over and undereducation. Stratifying the regressions by heterogeneity types also mitigates the endogeneity biases that may affect the estimated returns to schooling, to the extent that the types overlap with unobserved differences in productivity.[31] A limitation to keep in mind before interpreting the estimates is that they are obtained with the same sample used for the mixed proportional hazard model, which is comprised of individuals with at least one overeducation spell.

Estimation results, with and without lagged overeducation regressors, are reported in Table 12. In both specifications, individuals with a low propensity to exit overeducation (Type 1) have returns to required schooling that are below their high propensity counterparts. This result is consistent with these individuals being of lower ability than type 2 individuals. It is also interesting to note that the returns to required schooling are overall lower than in the non-stratified regressions discussed earlier. This is consistent

[30] We predict the type of each individual in the sample using the posterior type probabilities implied by the estimated likelihood and that individual's characteristics. Specifically, we consider that the individual is of a given heterogeneity type if the posterior probability of being of this type is larger than 0.5.

[31] The instrumental variables approach is especially challenging in this context since a valid instrument must be provided for required schooling in addition to the usual instrument for schooling. The fixed effect approach can also be problematic. In this setting, the returns to schooling are identified off individuals changing jobs with different required schooling levels, and, in particular, switching across matched, undereducated and overeducated jobs. These transitions are likely to be correlated with changes in unobserved wage determinants. Besides, the fixed effect estimates may not be generalizable to those who never switch overeducation status or intensity.

with overeducation being negatively correlated with ability, since the sample used for the stratified regressions is composed of individuals with at least one reported overeducation spell. In both the specification with lags and without them, both types exhibit a similar pattern whereby returns to overeducation are much lower than returns to required schooling. The fact that all of these individuals decided to go to college in spite of the high probability of being overeducated and receiving relatively small monetary benefits to college education is consistent with the existence of a positive consumption value of college attendance (see, e.g., Carneiro et al., 2003 and D'Haultfoeuille & Maurel, 2013).

Lastly, the stratification by types does not remove the penalty for having been overeducated in the past. In particular, the coefficients associated with all four lags in overeducation status are negative and significant for type 1 individuals. This suggests that these penalties are not driven by unobserved ability differences between those who have and have not been previously overeducated. Overall, this strengthens our previous finding that overeducation produces scarring effects which are reminiscent of what has been identified in the case of long-term unemployment.

7 Conclusion

Although economists and policy-makers have long been concerned about the determinants and wage effects of overeducation, little is still known about its dynamics along the career. This paper combines data from the NLSY79 and CPS to provide the first analysis of the career dynamics of overeducated U.S. workers. Overall, we find overeducation to be a persistent phenomenon, particularly for blacks and low-ability workers, with sizeable and lasting negative effects on wages. These wage penalties bear similarities with the scarring effects that have been found to accompany prolonged unemployment

spells.

Controlling for dynamic selection on unobservable characteristics is key in this context. While the exit rate from overeducation decreases quickly over the duration of an overeducation spell, the estimation of a mixed proportional hazard model suggests that, after controlling for both observed and unobserved heterogeneity, true duration dependence is not strong. Rather, the propensity to exit overeducated employment appears to be very heterogeneous among workers.

We find that past unemployment increases the duration of future overeducation spells, thus indicating that overeducation is likely to be one the mechanisms through which the scarring effects on earnings associated with unemployment spells operate (see e.g. Schmieder et al., 2013). The scarring effects associated with overeducation could also account for some of the negative wage effects of graduating during a recession which have been recently uncovered in the literature (Kahn, 2010, Liu et al., 2012, Oreopoulos et al., 2012, and Altonji et al., 2013).

From a policy standpoint, since both overeducation and unemployment (see, e.g., Saporta-Eksten, 2014) are associated with negative and persistent wage shocks, a relevant question becomes whether a marginal increase in unemployment duration is more or less harmful, in terms of lifetime earnings, than entering an overeducation spell. The answer has bearing on the design of unemployment insurance - should early exit be encouraged at the cost of more mismatch? - and the appropriate evaluation of the performance of employment agencies.

In sum, our results suggest that overeducation is a complex phenomenon that involves a number of the classical ingredients of labor economics: human capital, search frictions, ability differences and, perhaps, compensating wage differentials. In order to quantify the importance of each mechanism and explore the effects of unemployment insurance

or schooling subsidy programs, a promising avenue would be to estimate a dynamic structural model of schooling and occupational choice that would nest these different channels, while allowing for correlated unobserved heterogeneity in job mobility and productivity.

A CPS data

The 1989-1991 monthly CPS survey has a sample target of 50,000 households split into eight representative subsamples, each of which is interviewed for the first and last four months of a 16-month period. In any given month, a new sample of 6,250 households is surveyed for the first time. As a result, the pooled monthly CPS data from January of 1989 through December of 1991 contain 268,750 unique households. From these pooled cross-sections we keep only observations in the age range spanned by the NLSY79 cohort, which leaves 795,631 observations. Then we drop observations where an individual is unemployed, does not report a Census occupation code, has a missing level of education, did not complete the reported level of education, or is enrolled in college. After making these cuts, we are left with a sample of 506,930 occupation and education level pairs, where the education level is defined as the highest grade achieved by the surveyed individual.

From this sample we estimate the required level of education for each occupation identified by its 3-digit Census occupation code. The required level of education is defined as the sample mode of the distribution of education levels among workers in the occupation. Then we match observed occupations in the NLSY79 to a required level of education based on their 3-digit occupation code. 125 out of 488 occupations are observed less than 100 times in our CPS pooled sample. In order to reduce the sampling variance of the corresponding required levels of education, we collapse these low-frequency occupations using two-digit codes rather than three-digit codes before applying the procedure described above. Less than 2% of our NLSY79 observations are in such occupations.

B Tables and Figures

Table 1: Summary statistics for pooled cross section (1982-1994)

Variable	Mean	(Std. Dev.)	N[1]
Required Education (Yrs.)	12.871	(1.751)	34,535
Overeducation (Yrs.)	0.365	(1.635)	37,087
Age	27.503	(3.904)	44,996
Black	0.112	(0.315)	44,996
Hispanic	0.058	(0.233)	44,996
Female	0.514	(0.5)	44,996
Born in The United States	0.963	(0.189)	44,996
AFQT	0.503	(0.274)	43,052
Rotter Scale in 1979	8.44	(2.371)	44,669
Sociability in 1985	2.894	(0.669)	43,897
12 Years of Education	0.661	(0.474)	44,996
14 Years of Education	0.129	(0.335)	44,996
16 Years of Education	0.148	(0.355)	44,996
18 Years of Education	0.063	(0.242)	44,996
GED (No HS Diploma)	0.09	(0.286)	44,996
Northeast	0.194	(0.395)	44,748
South	0.331	(0.471)	44,748
West	0.171	(0.377)	44,748
Urban	0.785	(0.411)	43,828
Unemployment Rate (in %)	6.814	(1.374)	44,748
HH in SMSA	0.767	(0.423)	42,708
Mother's Edu. (Yrs., 1979)	11.965	(3.387)	41,395
Father's Edu. (Yrs., 1979)	11.714	(2.557)	43,102
Lived With Both Parents	0.673	(0.469)	44,996
Employed	0.847	(0.36)	44,996
Out of Labor Force	0.106	(0.307)	44,996
Unemployed	0.048	(0.213)	44,996
Tenure (1k Wks.)	0.17	(0.173)	41,782
Number of Jobs	1.407	(0.867)	44,996
Work Experience (1k Hrs.)	17.222	(8.800)	29,743
Weeks Unemployed	2.291	(7.991)	44,275
Hourly Wage	12.717	(6.158)	38,717
Occupational Hazards	−0.138	(0.952)	37,889

Unemployment Rate (by Region) corresponds to the annual unemployment rate for each Census region (BLS Local Area Unemployment Statistics data). The measure of occupational hazards is drawn from a sample of the 1970 Census which included occupational characteristics from the Dictionary of Occupational Titles (DOT). An indicator for a DOT occupation having hazards was averaged within each 1980 Census occupation code; we then convert this percentage measure to a Z-score. Tenure is defined as the total number of weeks an individual has worked in a given occupation across all employers. The total number of jobs is measured since the last interview. Wages are measured in constant dollars.

[1] Person-year observations

Table 2: Overeducation status proportions by education level

	Yearly	Categories				
		All	12	14	16	18
Undereducated	15.9%	14.1%	4.7%	16.3%	2%	
Matched	59.5%	68%	95.3%	18.1%	60.5%	28.7%
Overeducated	24.6%	17.9%		65.7%	37.4%	71.3%

Sample excludes individuals that have completed less than 12 years of education.

Table 3: 10 most frequent occupations among overeducated workers

1980 Census Code	Occupation	Required Edu. Min.	Required Edu. Max.	Observations OE	Observations Total
313	Secretaries	12	12	420	1,525
274	Sales workers, other commodities	12	12	394	960
243	Supervisors and Proprietors, Sales Occupations	12	12	323	772
337	Bookkeepers, accounting, and auditing clerks	12	12	243	649
19	Managers and administrators, n.e.c	12	16	223	2,290
156	Teachers, elementary school	16	16	197	318
389	Administrative support occupations, n.e.c.	12	12	163	415
276	Cashiers	12	12	121	675
433	Supervisors, food preparation and service occupations	12	12	117	254
376	Investigators and adjusters, except insurance	12	12	113	243

The columns for required education correspond to the lowest (Min.) and highest (Max.) levels of required education that we calculate from the CPS. Note that these values will differ if the mode and second largest observed levels of education are separated by fewer than 15 percentage points, as in occupation 19, Managers and administrators.

Table 4: Probit model of overeducation status

	14 Years		16 Years		18 Years	
Black	0.159***	(0.0555)	0.035	(0.0673)	0.107	(0.1045)
Hispanic	0.122*	(0.0628)	−0.081	(0.0697)	−0.023	(0.0866)
Female	0.055	(0.0392)	0.125***	(0.0290)	0.105***	(0.0374)
Born in The United States	0.206**	(0.0838)	−0.016	(0.0624)	0.021	(0.0926)
AFQT†	−0.132***	(0.0511)	−0.148**	(0.0622)	−0.187*	(0.0995)
Rotter Scale in 1979	0.010	(0.0076)	0.006	(0.0065)	−0.003	(0.0077)
Sociability in 1985	−0.036	(0.0256)	−0.028	(0.0218)	0.018	(0.0285)
GED (No HS Diploma)	−0.051	(0.0829)	−0.282	(0.2360)	−0.095	(0.2206)
Northeast	−0.027	(0.0522)	0.002	(0.0355)	−0.005	(0.0529)
South	−0.011	(0.0416)	0.007	(0.0370)	−0.008	(0.0452)
West	−0.003	(0.0520)	0.006	(0.0426)	−0.021	(0.0492)
Unemployment Rate	0.016*	(0.0093)	0.003	(0.0078)	−0.019*	(0.0099)
HH in SMSA	−0.077	(0.0541)	−0.006	(0.0428)	0.032	(0.0504)
Tenure (1k Wks.)	0.078	(0.0976)	−0.056	(0.1017)	0.396***	(0.1176)
Work Experience (1k Hrs.)	−0.005	(0.0062)	−0.018***	(0.0056)	0.014*	(0.0073)
Work Experience^2 (1k Hrs.)	0.003	(0.0022)	0.003	(0.0019)	−0.006**	(0.0028)
Number of Jobs	0.031*	(0.0161)	0.053***	(0.0140)	−0.031*	(0.0175)
Total Unemp. (Wks.)	0.002***	(0.0006)	0.000	(0.0008)	0.002	(0.0014)
Occupational Hazards	0.130***	(0.0328)	0.279***	(0.0414)	0.223**	(0.0897)
χ^2		201.92		377.69		160.54
LL		-1,737.63		-2,235.03		-816.98
Observations		3,125		4,064		1,763

Model also includes controls for industry, age, mother and father's education level in 1979, a dummy for living with both parents in 1979, and a dummy for living in an urban area at the time of interview. Unemployment Rate (by Region) corresponds to the annual unemployment rate for each Census region (BLS Local Area Unemployment Statistics data). Entries correspond to marginal effects (standard errors in parentheses).

† Values converted to Z-scores

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Transition matrix

			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	65.9%	4.2%	25.1%	4.8%
Undereducated	29.0%	44.9%	20.8%	5.4%
Matched	28.5%	3.9%	64.1%	3.6%
Non-employed	30.8%	4.7%	15.5%	49.0%
Total	46.3%	7.5%	37.6%	8.6%

Sample has entered the labor market and has a schooling attainment of 14 years or more.

Table 6: Transition matrix by gender

(Male)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	66.3%	3.7%	26.5%	3.5%
Undereducated	28.9%	49.0%	18.8%	3.4%
Matched	29.5%	3.0%	65.2%	2.2%
Non-employed	49.2%	7.0%	18.8%	25.0%
Total	48.7%	6.7%	40.4%	4.2%
(Female)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	65.5%	4.7%	23.8%	6.0%
Undereducated	29.0%	42.3%	22.0%	6.6%
Matched	27.3%	4.7%	63.0%	5.0%
Non-employed	23.5%	3.7%	14.2%	58.5%
Total	44.1%	8.2%	35.0%	12.6%

Sample has entered the labor market and has a schooling attainment of 14 years or more.

Table 7: Transition matrix by race

(White)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	65.4%	3.9%	26.0%	4.6%
Undereducated	27.4%	44.4%	22.6%	5.6%
Matched	27.9%	3.7%	64.8%	3.6%
Non-employed	29.4%	4.6%	16.8%	49.2%
Total	45.5%	7.1%	39.0%	8.5%
(Black)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	70.1%	6.8%	17.2%	5.9%
Undereducated	40.0%	48.0%	8.0%	4.0%
Matched	35.8%	6.4%	54.1%	3.7%
Non-employed	40.4%	5.3%	7.0%	47.4%
Total	54.2%	11.2%	24.0%	10.5%

Sample has entered the labor market and has a schooling attainment of 14 years or more.

Table 8: Transition matrix by AFQT quartile

(Lowest)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	72.8%	6.2%	15.4%	5.6%
Undereducated	42.1%	42.1%	5.3%	10.5%
Matched	37.1%	9.7%	50.0%	3.2%
Non-employed	44.7%	2.6%	0.0%	52.6%
Total	58.0%	11.0%	19.3%	11.7%
(2nd)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	71.5%	6.3%	16.9%	5.3%
Undereducated	46.0%	34.9%	12.7%	6.3%
Matched	31.7%	1.8%	58.7%	7.8%
Non-employed	30.8%	0.0%	15.4%	53.8%
Total	52.9%	6.7%	28.5%	12.0%
(3rd)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	66.3%	4.1%	24.7%	5.0%
Undereducated	26.7%	45.9%	20.7%	6.7%
Matched	30.5%	4.4%	61.4%	3.6%
Non-employed	32.4%	9.4%	17.3%	41.0%
Total	46.9%	8.7%	36.3%	8.2%
(Highest)			t	
$t-1$	OE	UE	Matched	Non-emp.
Overeducated	62.4%	2.9%	30.5%	4.2%
Undereducated	21.8%	46.5%	28.9%	2.8%
Matched	25.6%	3.8%	68.0%	2.5%
Non-employed	23.8%	4.2%	19.0%	53.0%
Total	41.7%	6.4%	44.5%	7.4%

Sample has entered the labor market and has a schooling attainment of 14 years or more.

Table 9: Duration models (first overeducation spell)

	(1)				(2)				(3)			
	MPH		PH		MPH		PH		MPH		PH	
α	1.031	(0.0459)	0.836***	(0.0262)	0.989	(0.0347)	0.822***	(0.0261)	1.115***	(0.0404)	0.768***	(0.0261)
$P(V_1)$	0.289***	(0.0803)			0.218***	(0.0565)			0.472***	(0.0440)		
$P(V_2)$	0.711***	(0.0803)			0.782***	(0.0565)			0.528***	(0.0440)		
V_1	0.153***	(0.0615)			0.094***	(0.0557)			0.142***	(0.0243)		
V_2	1.344***	(0.1167)			1.253***	(0.0740)			1.768***	(0.1202)		
Constant	−3.432***	(0.3097)	−3.092***	(0.2324)	−3.024***	(0.3293)	−2.723***	(0.2041)	−4.086***	(0.1095)	−3.339***	(0.0880)
Black	−0.313*	(0.1860)	−0.265**	(0.1344)	−0.288*	(0.1653)	−0.287**	(0.1265)				
Hispanic	0.136	(0.2694)	0.015	(0.1665)	0.261	(0.3226)	0.015	(0.1610)				
Female	0.002	(0.0977)	−0.176***	(0.0664)	0.102	(0.0908)	−0.078	(0.0633)				
Born in The United States	−0.294	(0.2541)	−0.150	(0.1884)	−0.318	(0.3206)	−0.198	(0.1869)				
AFQT†	0.149***	(0.0547)	0.101***	(0.0390)	0.150***	(0.0522)	0.104***	(0.0370)				
Rotter Scale in 1979†	0.026	(0.0463)	0.027	(0.0324)	0.016	(0.0447)	0.022	(0.0319)				
Sociability in 1985†	−0.026	(0.0526)	0.010	(0.0319)	−0.012	(0.0513)	0.019	(0.0321)				
14 Years of Edu.	−0.733***	(0.1071)	−0.587***	(0.0769)	−0.829***	(0.1053)	−0.692***	(0.0739)				
18 Years of Edu.	−1.706***	(0.1297)	−1.388***	(0.0969)	−1.560***	(0.1269)	−1.290***	(0.0951)				
GED (No HS Diploma)	0.257	(0.3226)	0.153	(0.2489)	0.211	(0.3194)	0.102	(0.2463)				
Northeast	0.035	(0.1185)	−0.002	(0.0943)								
South	0.346**	(0.1453)	0.223***	(0.0822)								
West	0.035	(0.1407)	0.004	(0.1012)								
Unemployment Rate†	−0.073	(0.0510)	−0.066*	(0.0368)								
HH in SMSA	0.088	(0.1465)	0.101	(0.1087)								
Tenure (1k Wks.)	−0.230	(0.4995)	−0.010	(0.3710)								
Work Experience (1k Hrs.)†	−0.050	(0.0675)	−0.031	(0.0519)								
Work Experience SQ (1k Hrs.)†	−0.036	(0.0527)	−0.017	(0.0305)								
Number of Jobs†	−0.025	(0.0488)	0.016	(0.0333)								
Total Unemp. (Wks.)†	−0.166***	(0.0579)	−0.133***	(0.0467)								
Occupational Hazards	−0.207***	(0.0728)	−0.191***	(0.0465)								
LL		5,321.9		5,348.6		5,344.9		5,374.8		5,442.7		5,501.3
Observations		1,648		1,648		1,648		1,648		1,648		1,648

Model also includes controls for age, mother and father's education level in 1979, a dummy for living with both parents in 1979, and a dummy for living in an urban area at the time of interview. Models estimated on the first overeducation spells of individuals with at least one overeducation spell. Unemployment Rate (by Region) corresponds to the annual unemployment rate for each Census region (BLS Local Area Unemployment Statistics data). Standard errors in parentheses.

† Values converted to Z-scores

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Augmented log-wage regressions

	(1) $\ln(HourlyWage)$		(2) $\ln(HourlyWage)$	
Required Education (Yrs.)	0.096***	(0.0046)	0.108***	(0.0051)
Schooling Above Required (Yrs.)	0.038***	(0.0055)	0.066***	(0.0065)
Schooling Below Required (Yrs.)	−0.069***	(0.0065)	−0.080***	(0.0069)
Black	0.007	(0.0253)	0.003	(0.0256)
Hispanic	0.041	(0.0339)	0.046	(0.0338)
Female	−0.155***	(0.0148)	−0.148***	(0.0147)
Born in The United States	−0.049	(0.0390)	−0.049	(0.0389)
AFQT†	0.110***	(0.0126)	0.111***	(0.0128)
Rotter Scale in 1979	−0.004	(0.0029)	−0.004	(0.0029)
Sociability in 1985	0.006	(0.0104)	0.005	(0.0104)
GED (No HS Diploma)	0.001	(0.0256)	−0.010	(0.0254)
Northeast	0.097***	(0.0191)	0.101***	(0.0190)
South	−0.027	(0.0167)	−0.022	(0.0167)
West	0.069***	(0.0213)	0.076***	(0.0213)
Unemployment Rate	0.001	(0.0032)	0.000	(0.0032)
HH in SMSA	0.129***	(0.0185)	0.131***	(0.0186)
Tenure (1k Wks.)	0.250***	(0.0356)	0.264***	(0.0353)
Work Experience (1k Hrs.)	0.035***	(0.0030)	0.036***	(0.0030)
Work Experience^2 (1k Hrs.)	−0.008***	(0.0010)	−0.008***	(0.0010)
Number of Jobs	−0.060***	(0.0069)	−0.059***	(0.0069)
Total Unemp. (Wks.)	−0.001***	(0.0002)		
Occupational Hazards	0.045***	(0.0066)	0.044***	(0.0066)
Constant	1.052***	(0.1028)	0.962***	(0.1068)
Unemployed (t-1)			−0.076***	(0.0239)
Unemployed (t-2)			−0.028	(0.0182)
Unemployed (t-3)			−0.003	(0.0158)
Unemployed (t-4)			−0.038*	(0.0219)
Overeducated (t-1)			−0.039***	(0.0129)
Overeducated (t-2)			−0.042***	(0.0105)
Overeducated (t-3)			−0.039***	(0.0114)
Overeducated (t-4)			−0.026**	(0.0127)
Overeducated (t-1), Spell Ongoing			0.013	(0.0182)
Overeducated (t-2), Spell Ongoing			0.021	(0.0279)
Overeducated (t-3), Spell Ongoing			0.007	(0.0386)
Overeducated (t-4), Spell Ongoing			−0.007	(0.0589)
Adjusted R^2		0.372		0.375
F		112.6		79.9
Observations		12,775		12,775

Model also includes controls for age, mother and father's education level in 1979, a dummy for living with both parents in 1979, and a dummy for living in an urban area at the time of interview. Model estimated on individuals with 12 years of schooling or higher, at least four years after labor market entry. Unemployment Rate (by Region) corresponds to the annual unemployment rate for each Census region (BLS Local Area Unemployment Statistics data). Standard errors are clustered at the individual level.

† Values converted to Z-scores

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Augmented log-wage regression (returns to tenure as a function of past overeducation)

	$\ln(HourlyWage)$	
Required Education (Yrs.)	0.094***	(0.0041)
Schooling Above Required (Yrs.)	0.053***	(0.0047)
Schooling Below Required (Yrs.)	−0.067***	(0.0060)
Black	0.008	(0.0223)
Hispanic	0.037	(0.0286)
Female	−0.139***	(0.0126)
Born in The United States	−0.041	(0.0324)
AFQT†	0.110***	(0.0108)
Rotter Scale in 1979	−0.004	(0.0025)
Sociability in 1985	0.003	(0.0089)
GED (No HS Diploma)	0.004	(0.0209)
Northeast	0.088***	(0.0164)
South	−0.019	(0.0145)
West	0.088***	(0.0181)
Unemployment Rate	−0.003	(0.0025)
HH in SMSA	0.129***	(0.0161)
Tenure (1k Wks.)	0.321***	(0.0356)
Work Experience (1k Hrs.)	0.033***	(0.0020)
Work Experience^2 (1k Hrs.)	−0.008***	(0.0007)
Number of Jobs	−0.057***	(0.0053)
Occupational Hazards	0.041***	(0.0056)
Overeducated (t-1)	−0.047***	(0.0168)
Overeducated (t-1), Spell Ongoing	0.022	(0.0317)
Overeducated (t-1) x Tenure	−0.169***	(0.0638)
Overeducated (t-1), Spell Ongoing x Tenure	−0.030	(0.1086)
Unemployed (t-1)	−0.071***	(0.0174)
Constant	1.104***	(0.0911)
Adjusted R^2		0.379
F		142.5
Observations		18,950

Model also includes controls for age, mother and father's education level in 1979, a dummy for living with both parents in 1979, and a dummy for living in an urban area at the time of interview. Model estimated on individuals with 12 years of schooling or higher, at least 1 year after labor market entry. Unemployment Rate (by Region) corresponds to the annual unemployment rate for each Census region (BLS Local Area Unemployment Statistics data). Standard errors are clustered at the individual level.

† Values converted to Z-scores

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Augmented log-wage regressions stratified by types

	Type 1		Type 2		All		Type 1		Type 2		All	
	$\ln(HourlyWage)$		$\ln(HourlyWage)$		$\ln(HourlyWage)$		$\ln(HourlyWage)$		$\ln(HourlyWage)$		$\ln(HourlyWage)$	
Required Education (Yrs.)	0.071***	(0.0127)	0.084***	(0.0131)	0.077***	(0.0088)	0.071***	(0.0123)	0.083***	(0.0128)	0.077***	(0.0086)
Schooling Above Required (Yrs.)	0.020	(0.0146)	0.016	(0.0131)	0.019*	(0.0097)	0.033**	(0.0143)	0.025*	(0.0139)	0.033***	(0.0098)
Schooling Below Required (Yrs.)	−0.067***	(0.0189)	−0.052	(0.0315)	−0.063***	(0.0170)	−0.068***	(0.0188)	−0.055*	(0.0308)	−0.066***	(0.0168)
Black	0.019	(0.0558)	0.069	(0.0663)	0.046	(0.0418)	0.009	(0.0538)	0.063	(0.0649)	0.035	(0.0420)
Hispanic	0.042	(0.0664)	0.051	(0.1030)	0.026	(0.0560)	0.028	(0.0639)	0.067	(0.1050)	0.030	(0.0554)
Female	−0.081***	(0.0294)	−0.134***	(0.0371)	−0.104***	(0.0235)	−0.076***	(0.0286)	−0.135***	(0.0379)	−0.102***	(0.0233)
Born in The United States	−0.062	(0.0642)	−0.219**	(0.1035)	−0.114**	(0.0538)	−0.071	(0.0606)	−0.219**	(0.1010)	−0.114**	(0.0510)
AFQT†	0.188***	(0.0437)	0.200***	(0.0637)	0.190***	(0.0367)	0.183***	(0.0421)	0.197***	(0.0629)	0.180***	(0.0355)
Rotter Scale in 1979	−0.003	(0.0068)	−0.009	(0.0080)	−0.006	(0.0051)	−0.001	(0.0066)	−0.010	(0.0081)	−0.006	(0.0051)
Sociability in 1985	0.004	(0.0234)	0.014	(0.0287)	0.006	(0.0181)	−0.002	(0.0230)	0.013	(0.0286)	0.003	(0.0179)
GED (No HS Diploma)	−0.086	(0.0786)	0.359*	(0.1828)	0.026	(0.0845)	−0.092	(0.0778)	0.357**	(0.1681)	0.015	(0.0856)
Northeast	0.134***	(0.0393)	0.117**	(0.0529)	0.132***	(0.0316)	0.138***	(0.0384)	0.121**	(0.0532)	0.137***	(0.0313)
South	−0.006	(0.0362)	0.023	(0.0460)	0.006	(0.0287)	0.004	(0.0352)	0.022	(0.0459)	0.010	(0.0285)
West	0.001	(0.0449)	0.108**	(0.0514)	0.053	(0.0339)	0.009	(0.0435)	0.110**	(0.0516)	0.061*	(0.0336)
Unemployment Rate	0.004	(0.0078)	0.010	(0.0103)	0.007	(0.0062)	0.005	(0.0077)	0.010	(0.0103)	0.008	(0.0061)
HH in SMSA	0.114**	(0.0455)	0.098*	(0.0507)	0.107***	(0.0337)	0.125***	(0.0420)	0.098*	(0.0546)	0.111***	(0.0340)
Tenure (1k Wks.)	0.177**	(0.0785)	0.144	(0.1025)	0.148**	(0.0627)	0.216***	(0.0756)	0.141	(0.1007)	0.173***	(0.0616)
Work Experience (1k Hrs.)	0.035***	(0.0080)	0.044***	(0.0126)	0.041***	(0.0074)	0.039***	(0.0079)	0.047***	(0.0128)	0.044***	(0.0073)
Work Experience^2 (1k Hrs.)	−0.010***	(0.0024)	−0.014***	(0.0042)	−0.012***	(0.0023)	−0.011***	(0.0025)	−0.015***	(0.0043)	−0.013***	(0.0024)
Number of Jobs	−0.038**	(0.0167)	−0.096***	(0.0182)	−0.064***	(0.0123)	−0.038**	(0.0167)	−0.093***	(0.0186)	−0.062***	(0.0123)
Total Unemp. (Wks.)	−0.002**	(0.0007)	−0.001	(0.0007)	−0.001***	(0.0005)						
Occupational Hazards	0.023	(0.0180)	0.075***	(0.0184)	0.048***	(0.0141)	0.020	(0.0186)	0.072***	(0.0181)	0.047***	(0.0141)
Constant	1.177***	(0.2908)	0.942***	(0.3077)	1.106***	(0.2130)	1.450***	(0.2942)	1.042***	(0.3031)	1.281***	(0.2120)
Unemployed (t-1)							−0.130	(0.0789)	−0.071	(0.0730)	−0.111**	(0.0560)
Unemployed (t-2)							−0.038	(0.0579)	−0.050	(0.0544)	−0.051	(0.0403)
Unemployed (t-3)							0.013	(0.0519)	−0.019	(0.0517)	−0.017	(0.0366)
Unemployed (t-4)							−0.059	(0.0647)	−0.005	(0.0581)	−0.037	(0.0452)
Overeducated (t-1)							−0.054***	(0.0145)	−0.071**	(0.0301)	−0.058***	(0.0129)
Overeducated (t-2)							−0.064***	(0.0128)	−0.036	(0.0282)	−0.053***	(0.0113)
Overeducated (t-3)							−0.075***	(0.0133)	0.016	(0.0276)	−0.048***	(0.0119)
Overeducated (t-4)							−0.046***	(0.0156)	−0.025	(0.0271)	−0.036***	(0.0135)
Overeducated (t-1), Spell Ongoing							0.018	(0.0250)	0.026	(0.0279)	0.028	(0.0188)
Overeducated (t-2), Spell Ongoing							−0.020	(0.0390)	0.057	(0.0377)	0.013	(0.0277)
Overeducated (t-3), Spell Ongoing							−0.005	(0.0586)	0.000	(0.0454)	0.018	(0.0381)
Overeducated (t-4), Spell Ongoing							0.039	(0.0915)	−0.035	(0.0690)	0.006	(0.0567)
Adjusted R^2		0.240		0.354		0.281		0.267		0.355		0.295
F		13.1		17.2		25.2		10.6		13.0		19.1
Observations		2,527		1,379		3,906		2,527		1,379		3,906

Model also includes controls for age, mother and father's education level in 1979, a dummy for living with both parents in 1979, and a dummy for living in an urban area at the time of interview. Model estimated on individuals with at least one overeducated spell, at least four years after labor market entry. Types imputed based on the posterior type probabilities from the MPH model. Unemployment Rate (by Region) corresponds to the annual unemployment rate for each Census region (BLS Local Area Unemployment Statistics data). Standard errors are clustered at the individual level.

† Values converted to Z-scores

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

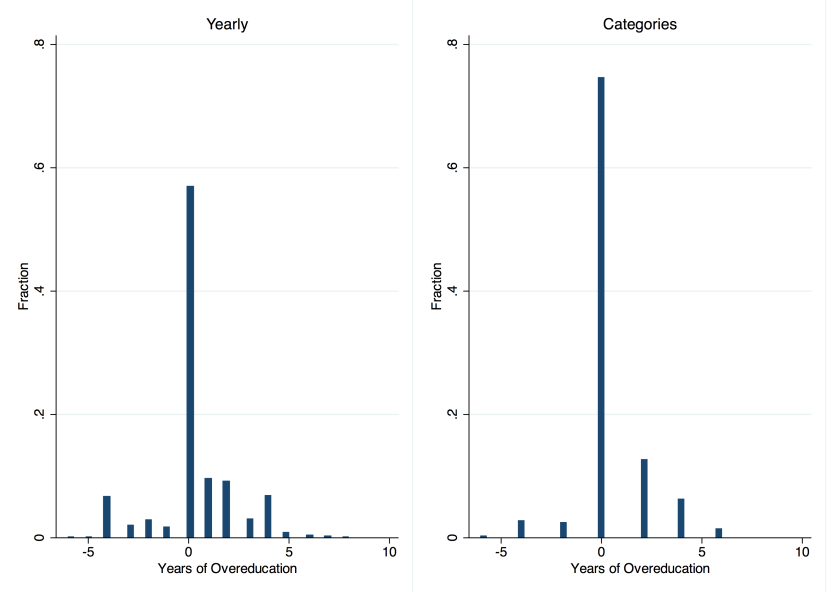

Figure 1: Distribution of overeducation

Figure 2: Composition of overeducation (fraction of respondents)

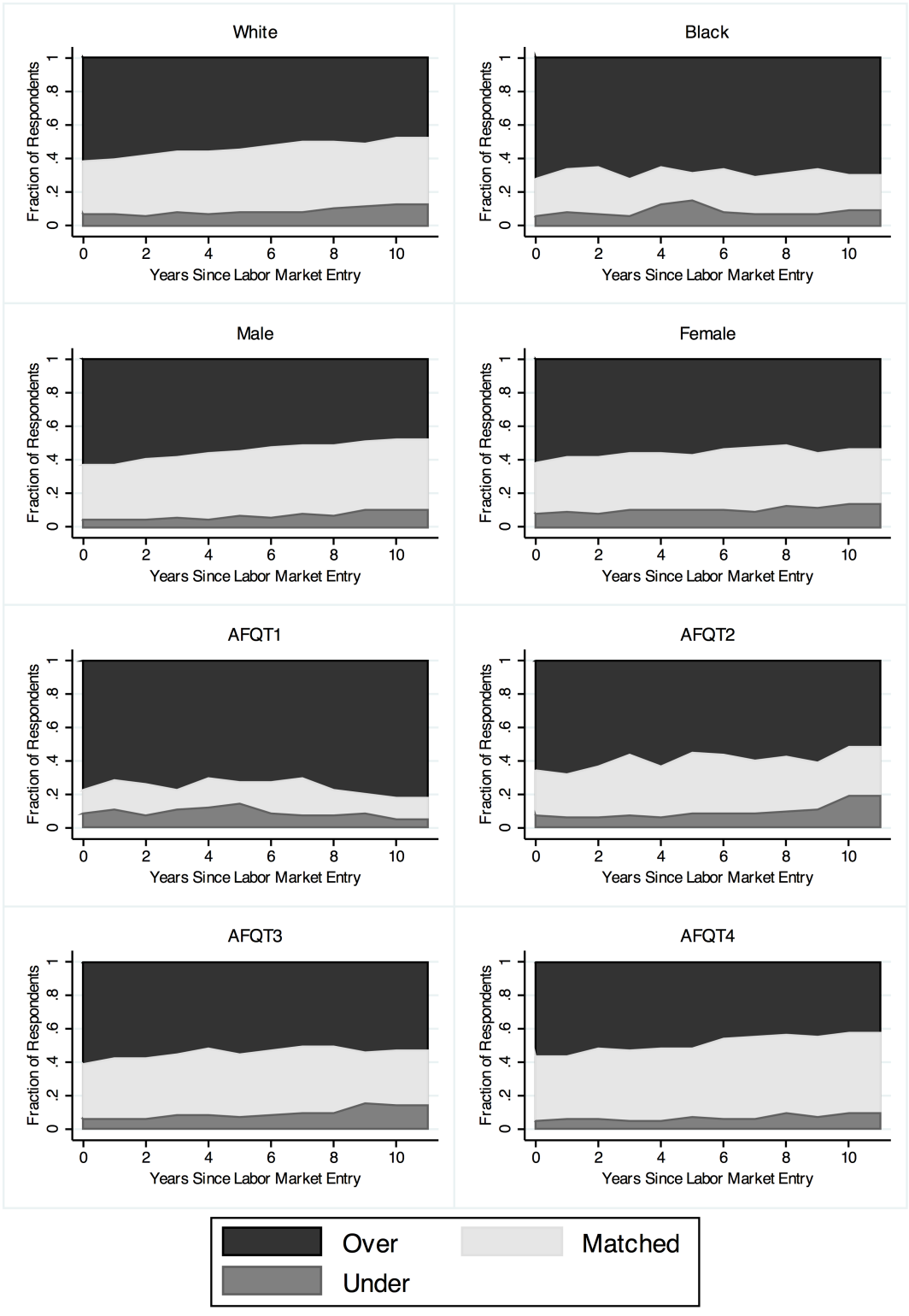

Figure 3: Composition of overeducation (over 14 years of education)

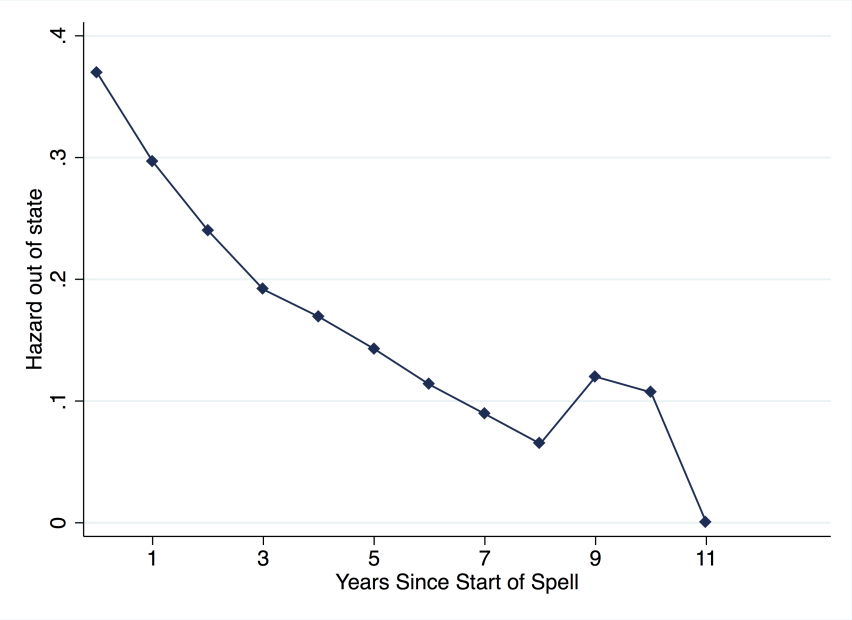

Figure 4: Hazard out of overeducation

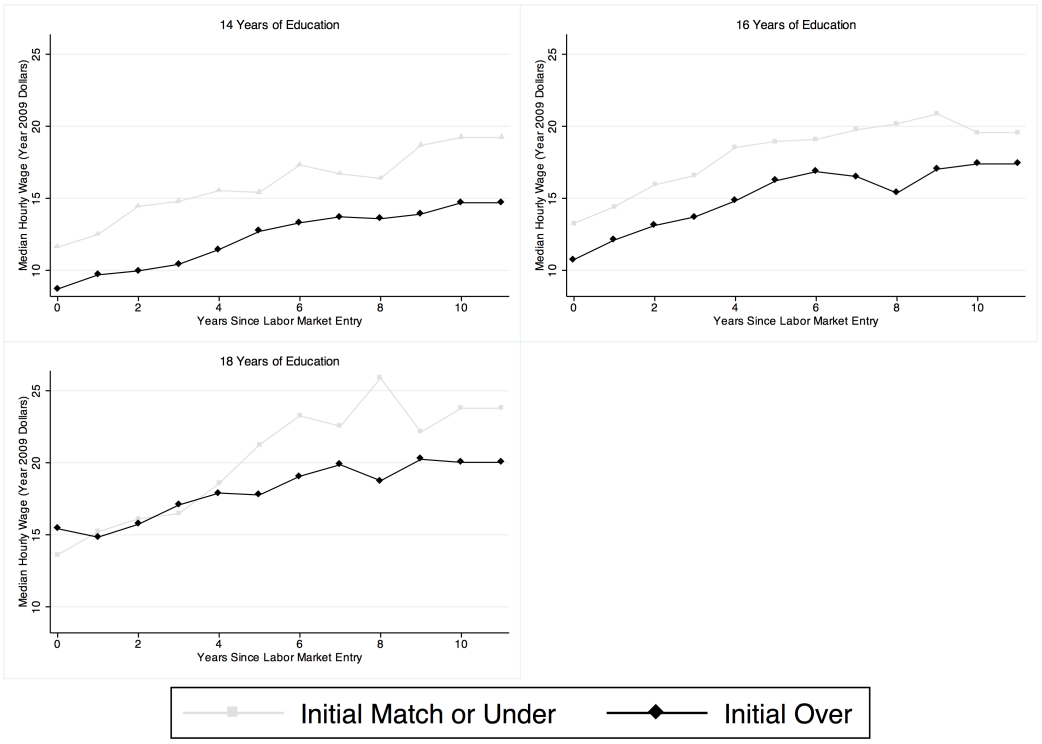

Figure 5: Path of median hourly wage

References

Alba-Ramirez, A. (1993), 'Mismatch in the Spanish labor market: Overeducation?', *Journal of Human Resources* **28**(2), 259–278.

Allen, J. & Van der Velden, R. (2001), 'Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search', *Oxford Economic Papers* **53**(3), 434.

Altonji, J., Kahn, L. & Speer, J. (2013), Cashier or consultant? Entry labor market conditions, field of study, and career success. Unpublished manuscript.

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

Bauer, T. K. (2002), 'Educational mismatch and wages: A panel analysis', *Economics of Education Review* **21**(3), 221–229.

Bils, M. & McLaughlin, K. (2001), 'Interindustry mobility and the cyclical upgrading of labor', *Journal of Labor Economics* **19**(1), 94–135.

Carneiro, P., Hansen, K. & Heckman, J. J. (2003), 'Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice', *International Economic Review* **44**, 361–422.

Chevalier, A. & Lindley, J. (2009), 'Overeducation and the skills of UK graduates', *Journal of the Royal Statistical Society: Series A* **172**, 307–337.

D'Haultfoeuille, X. & Maurel, A. (2013), 'Inference on an extended Roy model, with an application to schooling decisions in France', *Journal of Econometrics* **174**, 95–106.

Dolton, P. & Vignoles, A. (2000), 'The incidence and effects of overeducation in the UK graduate labour market', *Economics of Education Review* **19**(2), 179–198.

Duncan, J. & Hoffman, S. (1981), 'The incidence and effects of overeducation', *Economics of Education Review* **1**, 75–86.

Dutta, S. & Goswami, A. (2010), 'Mode estimation for discrete distributions', *Mathematical Methods of Statistics* **19**(4), 374–384.

Elbers, C. & Ridder, G. (1982), 'True and spurious duration dependence: The identifiability of the proportional hazard model', *Review of Economic Studies* **49**, 403–409.

Frank, R. (1978), 'Why women earn less: The theory and estimation of differential overqualification', *American Economic Review* **68**, 360–373.

Freeman, R. (1976), *The Overeducated American*, New York: Academic Press.

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), 29–45.

Gottschalk, P. & Hansen, M. (2003), 'Is the proportion of college workers in noncollege jobs increasing?', *Journal of Labor Economics* **21**(2), 449–471.

Hartog, J. (1980), 'Earnings and capability requirements', *Review of Economics and Statistics* **62**, 230–240.

Hartog, J. (2000), 'Over-education and earnings: Where are we, where should we go?', *Economics of Education Review* **19**(2), 131–147.

Heckman, J. & Singer, B. (1984), 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica* **52**, 271–320.

Hersch, J. (1991), 'Education match and job match', *Review of Economics and Statistics* pp. 140–144.

Kahn, L. B. (2010), 'The long-term labor market consequences of graduating from college in a bad economy', *Labour Economics* **17**(2), 303–316.

Kiker, B., Santos, M. & De Oliveira, M. (1997), 'Overeducation and undereducation: evidence for Portugal', *Economics of Education Review* **16**(2), 111–125.

Korpi, T. & Tåhlin, M. (2009), 'Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000', *Labour Economics* **16**(2), 183–193.

Kroft, K., Lange, F. & Notowidigdo, M. J. (2013), 'Duration dependence and labor market conditions: Evidence from a field experiment', *Quarterly Journal of Economics* **128**(3), 1123–1167.

Lee, D., Lee, T. & Shin, Y. (2014), The option value of human capital. Unpublished manuscript.

Leuven, E. & Oosterbeek, H. (2011), Overeducation and mismatch in the labor market, *in* E. Hanushek, S. Machin & L.Woessmann, eds, 'Handbook of the Economics of Education', Vol. 4, Elsevier.

Liu, K., Salvanes, K. & Sorensen, E. (2012), 'Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession', *IZA Discussion Paper No. 6820* .

Mavromaras, K., Mahuteau, S., Sloane, P. & Wei, Z. (2013), 'The effect of overskilling dynamics on wages', *Education Economics* **21**(3), 281–303.

Oreopoulos, P., von Wachter, T. & Heisz, A. (2012), 'The short- and long-term career effects of graduating in a recession', *American Economic Journal: Applied Economics* **4**, 1–29.

Rubb, S. (2003), 'Overeducation: A short or long run phenomenon for individuals?', *Economics of Education Review* **22**(4), 389–394.

Rumberger, R. (1987), 'The impact of surplus schooling on productivity and earnings', *Journal of Human Resources* **22**, 24–50.

Saporta-Eksten, I. (2014), Job loss, consumption and unemployment insurance. Unpublished manuscript.

Schmieder, J., von Wachter, T. & Bender, S. (2013), 'The causal effect of unemployment duration on wages: Evidence from unemployment insurance extensions', *NBER Working Paper No. 19772* .

Sicherman, N. & Galor, O. (1990), 'A theory of career mobility', *Journal of Political Economy* **98**(1), 169–192.

Verdugo, R. & Verdugo, N. (1989), 'The impact of surplus schooling on earnings: Some additional findings', *Journal of Human Resources* **24**(4), 629–643.

Yamaguchi, S. (2012), 'Tasks and heterogeneous human capital', *Journal of Labor Economics* **30**, 1–53.