# The Career prospects of overeducated Americans [*]

## *Preliminary*

Brian Clark[†]     Clément Joubert[‡]     Arnaud Maurel [§]

September 10, 2013

**Abstract**

Half of American and European workers have a level of education that does not match the level of education required for their job. Of these, a majority are overeducated, i.e. have more schooling than necessary to perform their job (see, e.g., Leuven & Oosterbeek, 2011). In this paper, we provide the first comprehensive analysis of overeducation over the life cyle in the US. We use data from the National Longitudinal Survey of Youth 1979 combined with the pooled 1989-1991 waves of the Current Population Survey to analyze overeducation status transitions and the corresponding effects on wages levels and growth. Overall, we find that overeducation is a persistent phenomenon at the aggregate and individual levels, with 71% of workers remaining overeducated after one year. Further, the hazard rate out of overeducation is strongly decreasing, with a drop of 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. Initial overeducation

[†]Duke University
[‡]University of North Carolina at Chapel Hill and IZA
[§]Duke University, NBER and IZA.

also has a negative impact on wages much later in the career, which points to the existence of scarring effects.

**JEL Classification: J24; I21**

# 1 Introduction

Half of American and European workers have a level of education that does not match the level of education required for their job. Of these, a majority have more schooling than necessary, and are deemed overeducated in the economics literature (see e.g., Leuven & Oosterbeek, 2011). As students, especially in the U.S., accumulate skyrocketing levels of college loan debt (845 billion dollars as of the second quarter of 2011, Source: NY Fed), there is a growing concern that expensive skills acquired in college may be underutilized in low-paying jobs. 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.

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. While the literature has paid little attention to this dimension, analyzing transitions in overeducation status and wages is key to disentangling the role played by labor market frictions versus other factors such as selection on unobservable characteristics, compensating differentials and career mobility prospects.[1] For example, if overeducation was due to search frictions only, one would expect this type of mismatch to be largely transitory. Conversely, selection on unobservables, compensating differentials or career mobility motives would generate persistence in the overeducation patterns.

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 US during the 1970's. Subsequent literature, recently surveyed by Leuven & Oosterbeek (2011), has

---

[1]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 find it optimal to be overeducated.

produced two key results using different datasets and measures of overeducation.[2] The first finding is that around 30% of workers have more schooling than is required by their job. Second, wage returns to surplus schooling average 4.3%, which corresponds to only about half of the returns to "required" schooling.[3]

Only few papers examine the evolution of overeducation over the life cycle, although, as argued above, dynamics are of clear interest in this context. US evidence is particularly scarce.[4] A notable exception is Rubb (2003) who provides evidence from the Current Population Survey (CPS) that overeducation displays a substantial degree of persistence, with less than 20% 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 time, the short length of the CPS panel dimension makes it difficult to address this question adequately. A recent study by Baert et al. (2013) points to such effects: they show that accepting an overeducated job has the effect of postponing access to a "matched" job for a sample of Flemish youth.

This paper fills the gap in the literature by providing the first comprehensive analysis of overeducation over the life cyle in the US. We use the National Longitudinal Survey of Youths 1979 (NLSY79), combined with the pooled 1989-1991 waves of the CPS, to analyze (1) the determinants of overeducation, (2) how overeducation incidence changes along the life cycle, and (3) how wage levels and returns to experience differ across overeducated, matched or undereducated individuals, and as a function of previous overeducation and undereducation spells. It is worth pointing out that, while the

---

[2]See, e.g., Hartog & Oosterbeek (1988); Verdugo & Verdugo (1989); Alba-Ramirez (1993); Kiker et al. (1997); Hartog (2000); Gottschalk & Hansen (2003) and Groeneveld & Hartog (2004).

[3]This finding is obtained by Leuven & Oosterbeek (2011) from an average of 151 studies.

[4]Using German, British, Canadian or Australian data a few studies have estimated panel wage regressions (Bauer, 2002; Frenette, 2004), dynamic random effect models of overeducation exit (Lindley & McIntosh, 2009; Mavromaras et al., 2013) or simply documented overeducation status transitions (Battu et al., 1999; Dolton & Vignoles, 2000). See Leuven & Oosterbeek (2011), section 4.3.

actual level of education is directly observed in the NLSY79, measuring the required level of education 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, for each given occupation in the 1980 3-digit Census Occupation classification, we compute the required level of education from the 1989-1991 CPS as the mode of the distribution of the levels of education among the individuals working in that given occupation. Relative to alternative approaches which have been used in the literature, and in particular those measuring the required levels of education with the General Educational Development scale provided by the Dictionary of Occupational Title (see, e.g., Hartog (1980) and Rumberger (1987)), this method is arguably more transparent and has the benefit of directly generating requirements in terms of years of education. Our findings regarding the incidence and determinants of overeducation are overall consistent with those obtained in the literature.

Depending on the aggregation level used for the educational attainment, we find that between 20% and 26% of workers are overeducated. To the extent that overeducated workers earn substantially less than those holding a job which match their educational level (around 4% less per year of overeducation), this suggests that overeducation accounts for a sizeable portion of wage dispersion among college graduates. In particular, this fuels concerns that, despite a large *average* college wage premium, high college loan balances will prove a formidable challenge for a large fraction of college graduates.

The NLSY79 allows us to document longitudinal patterns of overeducation for a representative sample of a cohort of American workers up to 18 years after labor market entry. Overeducation and total mismatch (defined as under- plus overeducation) decrease as workers progress through their careers, but remain sizeable 18 years after the first job. This suggests that frictions play a sizeable role but also that we need to bring other economic mechanisms into the fold to explain this long-term persistence. The

aggregate dynamics appear "favorable" to whites relative to blacks, and also, perhaps surprisingly, to women relative to men. That is, overeducation (resp. undereducation) decreases (resp. increases) faster among whites and among women. Overeducation is also a persistent phenomenon at the individual level, with 71.4% of workers remaining overeducated after one year. Blacks are again especially likely to remain overeducated (78.2% vs 70.7% for whites).

The hazard rate out of overeducation is also strongly decreasing, and drops by about 60% during the first 5 years spent overeducated. To investigate whether this reflects selection on unobservables or true duration dependence, we estimate a Mixed Proportional Hazard model of overeducation duration. While composition effects based on observables play next to no role, controlling for unobserved heterogeneity largely wipes out the duration dependence. This finding stresses the importance of controlling for dynamic selection on unobservables before drawing policy conclusions related to the impact of the duration of overeducation spells.

We then stratify the canonical augmented wage regression based on the unobserved types identified in the MPH model. This allows to show that individuals with lower propensities to exit overeducation also exhibit lower returns to schooling. We also find that past overeducation entails a wage penalty (reminiscent of scarring effects associated with unemployment spells), even after applying the above stratification. Indeed, when we look at median wages over the lifecycle, we do not find that initially overeducated individuals catch up, earningswise, with their matched peers. Rather, wage penalties associated with initial overeducation are persistent throughout the years covered by our sample.

The remainder of the paper is organized as follows. Section 2 describes the data used in the analysis and the construction of the overeducation measure. Section 3 describes the determinants of overeducation in our sample. Section 4 documents the longitudinal

patterns in the incidence of overeducation along the life cycle. Section 5 estimates a mixed proportional hazard model of overeducation duration allowing to separate true duration dependence from dynamic selection on observed and unobserved worker attributes. Section 6 presents some results pertaining to the effect of overeducation on wage levels and returns to experience, and Section 7 concludes.

## 2   Data and methodology

Our main data source is the National Longitudinal Survey of Youth 1979 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.[5] The main variables of interest are observed level of education, labor market earnings and occupation at the time of each interview. We also extract a set of covariates that describe the respondents and the job they occupy at the time of the interview such as ability measures, socio-demographic background, current household characteristics and non-wage job characteristics.

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. 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

---

[5]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.

cohort at that time. Required schooling in a given occupation code is defined as the mode of the distribution of the levels of education among the individuals working in that occupation.[6]

Relative to other approaches which have been used in the literature, and in particular those measuring the required levels of education with the General Educational Development scale (GED, see, e.g., **?**, 1980, Rumberger, 1987), this method is more transparent and has the benefit of directly generating requirements in terms of years of education. A limitation is that it is based on the distribution of schooling attainment among employed workers in a occupation, which is an equilibrium outcome of labor supply and demand decisions. Fluctuations in labor demand and supply will affect that distribution despite the fact that the actual skills required to perform the job have not changed. In that sense, our measure is strictly speaking one of required schooling rather than required skills; while correlated with, and informative of, the latter, it is best interpreted from the worker's perspective (i.e. in partial equilibrium) as the schooling required to be hired, rather than that required to perform the job. Alternatively, one can simply think about the workers we identify as overeducated as, literally, those for who have more schooling that the modal worker in their occupation (i.e. who are overeducated relative to their peers).

The 1989-1991 monthly CPS consists of roughly 50,000 households sampled for eight month periods, and potentially offers over 200,000 unique households [**CL: need to clarify this**]. One feature of the CPS is that it does not report unique identifiers for individuals within households. Hence, it is impossible to create a sample of unique individuals and their employment history over time. In order to approximate such a sample, we drop duplicate observations where a respondent from each household lives

---

[6]We match on the 1980 Census Occupation codes since they are both more recently revised than the 1970 codes, and are reported over a greater span of time than the 2000 codes. Since the NLSY79 only reports 1980 census codes for the 1982-2000 waves, we restrict our sample to these years.

in the same state, keeps the same education level, and remains in the same job class, industry, and occupation as identified by 1980 Census industry and occupation codes [**CL: I don't understand this**].[7] We also drop observations with missing occupation codes, those corresponding to individuals who did not complete their highest level of education, or who are enrolled in school.

After making these cuts, we are left with a pooled sample of 254,209 occupation and education level pairs. From this sample we compute the required level of education for each occupation as the mode of the distribution of the highest grade achieved within the occupation.[8]

We then match observed occupations in the NLSY79 to their corresponding required levels of education calculated from the CPS. Using the employment history of the 6,111 individuals that comprise the core civilian cross section of the NLSY79, we are left with a total 53,090 observations.[9]

Table 1 presents summary statistics for a selection of variables from our sample. As one would expect given the design of the NLSY79, our sample has roughly equal proportions of men and women, and representative proportions of Black, and Hispanic respondents. Of the 53,090 observations with employment status, 87.1% are employed, 7.9% are out of the labor force, and 5.0% are unemployed.

In most of our analysis, we use a collapsed measure of years of education into five categories: less than 12 years, 12-13, 14-15, 16-17, and over 18 years of completed

---

[7]We also constructed a dataset by sampling a random individual from each household and only taking one observation from their eight-month sample, but neither qualitatively changed our results.

[8]For all individuals in a Census occupation code we calculate the mode of the distribution of highest grade achieved. The mode of observed education level for 248 of the 486 occupations is based on fewer than 100 observations. In order to reduce the noise in calculating the mode for these occupations, we collapse the three-digit Census occupation code into its corresponding two-digit category, and recalculate the mode for all occupations by the two-digit classification. For occupations with fewer than 100 observations, we simply replace the calculated mode with that of its corresponding two-digit category.

[9]Individual item non-responses or skips result in a small fraction of missing observations for all but the core set of questions asked of the sample in the initial 1979 survey wave.

education. While we loose some information by using this aggregate schooling measure rather than the yearly one, the main advantage of this measure is that it should mitigate concerns about classification errors. Furthermore, this classification is natural since each category corresponds to high school dropouts, high school graduates, two-year and four-year college graduates, and graduate school. With this classification, in our sample, 11.27% have less than a high-school education, 56.19% are high school graduates, 12.36% have two years and 12.83% have four years of college education, and 7.35% have some graduate school experience.[10]

Figure 1 and Table 2 below reflect the distribution of overeducation, generated by subtracting an individual's observed highest level of completed education, from the level of education required by their occupation (both measured in years). This distribution is largely consistent with previous results obtained in the literature, as nearly half of all observations have a perfect match of observed and required education, while progressively smaller fractions exhibit education level mismatch. We examine the sensitivity of the incidence of overeducation to the use of our aggregate (versus yearly) measure of schooling. Noteworthy, the distribution follows a similar pattern, with approximately half of all observations exhibiting a perfect match, and symmetric proportions of over and undereducation that fall as the level of mismatch in years increases.

---

[10]Note that these proportions are taken from the sample of individuals who have permanently entered the labor market. We define the date of entry as the first survey year where (1) the individual is employed, (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 their highest level of education over the sample period 1982-2000. See figures 2 and 3 for a distribution of this variable in the sample.

# 3 The determinants of overeducation

We start by documenting the cross-sectional determinants of overeducation in our representative sample of American workers to determine whether they align with existing results in the literature. We estimate a probit regression on an indicator for overeducation status for the pooled cross sections of respondents.[11] We stratify the probit regression by schooling level (Tables 11, 12, 13).

Overall, our sample exhibits the same correlates of overeducation found in different contexts by previous studies. Workers with lower skills (conditional on their schooling attainment) and minorities (including women, but with the notable exception of hispanics) are more likely to be overeducated. Occupational hazards are also correlated with overeducation. This result does not support a compensating wage differential story, but it could still be consistent with it in the presence of strong ability bias.

First, AFQT scores exhibit a negative and significant relationship with overeducation at all schooling attainments. This result is consistent with prior findings in the literature, mostly in the European context, of a negative relationship between different measures of ability and overeducation probability (see, e.g., Allen & Van der Velden, 2001 and Chevalier & Lindley, 2009).

Also in accordance with existing studies, black and female college graduates (16 years of schooling) are more likely to be overeducated which might reflect discrimination or higher levels of frictions for these groups.[12] At other schooling levels, the female and black regressors are also either positive or insignificant. On the contrary, Hispanic college graduates are *less* likely to be overeducated while the reverse is true for Hispanics with 14 years of schooling. A stronger selection into college on unobservable ability

---

[11]This indicator is equal to one if an individual has a higher level of education than that required for their occupation.

[12]Those gender differences in overeducation rates, in connection with the existence of higher labor market frictions for females, are discussed in an early work by Frank (1978).

might explain this result given that college degrees are relatively less prevalent among this minority group.

Workers in hazardous jobs are also more likely to be overeducated at all three levels of schooling. This would appear to contradict a compensating wage differential story in which overeducated workers accepted low paying jobs because they present valuable non-pecuniary characteristics. This result complements previous studies that have found a negative correlation between overeducation and job satisfaction.[13] This can arguably be reconciled with compensating wage differentials if a form of ability bias is at work, whereby overeducated workers also tend to have low ability which gives them access to lower levels of both pecuniary and non-pecuniary job benefits.

# 4  The dynamics of overeducation

We now turn to a description of the careers of overeducated American workers. The NLSY79 allows us to document longitudinal patterns of overeducation for a representative sample of a cohort of American workers up to 18 years after labor market entry.

In order to evaluate the behavior of aggregate schooling mismatch over time, we will first consider the fraction of matched, over-, and under- educated individuals in the cohort, plotted against the number of years since labor market entry. Overeducation as well as total mismatch (over- plus undereducation) decrease along the NLSY79 workers' careers, but remain sizeable 18 years after the first job. This suggests that frictions play a role but the aggregate persistence in mismatch long after labor market entry calls for complementary mechanisms. The aggregate dynamics appear favorable to whites relative to blacks, and also, perhaps surprisingly, to women relative to men. That is, overeducation (resp. undereducation) decreases (resp. increases) faster among whites

---

[13]Hersch (1991),Korpi & Tåhlin (2009) Verhaest (2006b)

and among women.

We will then we focus of individual-level annual transitions between the overeducated, undereducated and matched states. Consistent with the cross-sectional patterns described earlier, we see that overeducation persistence is stronger among blacks and lower AFQT quartiles. We also take a first look at whether there might be an overeducation "trap": we show that the hazard rate out of overeducation decreases quickly with the time spent overeducated. We further investigate the issue of the duration-dependence of overeducation using hazard models in the next section.

## 4.1    Agregate incidence of schooling mismatch after labor market entry

In this section we describe how the total fraction of undereducated, matched and overeducated workers changes as years elapse after labor market entry. Note that we define labor market entry as the first year after the highest schooling level is completed by a given individual in the sample.[14]

Figure 5 reveals that the proportion of overeducated individuals decreases steadily from around 28% in the first year worked down to around 12% after eighteen years of experience. The graph suggests a break after seven years of labor market experience, at which point the incidence of overeducation appears to decrease faster than it did initially. Undereducation shows a steady increase from 6% to 13% which could suggest that a fraction of high performing individuals with low education levels get promoted to jobs with higher skill requirements. The total incidence of mismatch (under- plus overeducation) decreases from 34% to 25% over the first eighteen years of a career.

---

[14]In other words, we exclude years worked before observed school re-entry. Note that right-censoring should not be a major concern for school re-rentry to the extent that the youngest individuals in the NLSY79 are aged 35 in 2000 (last survey year in our sample).

We break down the previous graph by attained schooling in figure 6. This allows us to refine the above patterns, particularly for college graduates: the incidence of overeducation for that group decreases in the first 3 years after labor market entry before stabilizing until year ten and then decreasing again in the following years.[15]

Patterns across races are quite striking (figure 9). The increase in the fraction of undereducated white high school graduates is around 50% faster than their black counterparts', possibly reflecting weaker career advancement prospects for the latter. Black college graduates also do not exhibit the same gradual reduction in overeducation as white college graduates. Instead, the fraction of overeducated appears to increase for the first 9 years before stabilizing.

The evolution of mismatch over a worker's career appears rather favorable to women. Figure 10 shows that the fraction of female high school graduates who are undereducated (i.e who work in relatively high-skill jobs) is higher and increases faster than for male high school graduates. Similarly, the fraction of overeducated female college graduates is initially smaller than that of men and decreases more as well (from 50% to 35% against 60% to 55%). It is worth noting that these results may be partly driven by differential selection into employment across gender.

We proceed to compare the different AFQT quartiles in figures 11 and 12. The fraction of undereducated high school graduates is increasing in the AFQT score, as one might expect if schooling acts as a substitute for ability. The figures for college graduates show that the lowest quartiles see increasing levels of overeducation when it instead decreases for quartiles 2 and 3. Surprisingly, the fraction of overeducated

---

[15]Since these figures are based on years since entry into the labor market, we lose one observation at the end of the sample period for every year a respondent delays entering the labor market. This characteristic of the sample is especially important for respondents with high levels of education. In particular, 60.97% of those with 18 years of education enter the labor market after 1987, hence we are only able to see at most nine periods of labor market experience. Figures 7 and 8 describe the amount of right-censoring for the sample as a whole, and for each schooling category.

among college graduates in the highest quartile remains relatively stable.

## 4.2  Individual annual transitions

Table 3 presents the estimated probabilities of being unemployed, undereducated, overeducated or matched conditional on the previous year's status.[16] Overeducation is fairly persistent, with 71.4% of workers remaining overeducated after one year, while matched individuals have an 80.7% chance, and undereducated individuals a 38.6%, chance of remaining so. As another point of comparison, the unconditional incidence of overeducation is 20.8%.[17] It is worth noting that among college graduates, being matched is only slightly more persistent state than being overeducated (61.8% vs. 63.9%, see table 4).

Transition rates are broken down by gender and race in tables 5 and 6. Men are more likely than women to remain in a matched job (82.9% vs. 78.9%), undereducated (39.6% vs. 38.0%) but overeducation exhibits similar persistence across gender (72.1% vs 70.8%). Differences by race are sizeable: overeducation is significantly more persistent among blacks (78.2% vs. 70.7%).

As illustrated in table 7, the persistence of overeducation decreases monotonically with AFQT scores (from 83.2% for the lowest quartile, to 68.6% for the highest one ).

As a first step to investigate whether there might be an "overeducation trap", we also report the hazard rates out of each state over time in figure 4. The hazard rates out of overeducation, undereducation and matched employment decline with the length of the period considered. For example, after 2 years being overeducated, the probability of exiting that state drops from 29% to is 17% only. This number drops further to

---

[16]We use "matched" as a shorthand for "employed in a job that requires the level of schooling actually attained by the worker"

[17]It is also worth noting that, in our data, overeducation is found to be more persistent than unemployment, with 51.2% of workers remaining unemployed after one year.

14% and 9% after 5 and 10 years respectively. Overall, these patterns are suggestive of a negative duration dependence but could also result from compositional effects (permanent heterogeneity correlated with the hazard rate out of overeducation), as discussed in the next section.

# 5  Duration dependence vs. dynamic selection

The descriptive results discussed so far provide some suggestive evidence of duration dependence in the overeducation status, with a decrease from 14% to 7.3% in the exit rate from the first to the fifth year of overeducation. However, in order to investigate the role played by true rather than spurious duration dependence, we need to control for dynamic selection on worker attributes. We discuss below the econometric model that we use to estimate the extent of true duration dependence.

We assume that the duration of the first overeducation spell follows a mixed proportion hazard model (see Elbers & Ridder, 1982), where the baseline duration follows a Weibull distribution. Namely, the pdf. and cdf. of the duration of the overeducation spell, conditional on the unobserved heterogeneity $\nu_i$, are respectively given by:

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

Following Heckman & Singer (1984), we assume that the unobserved heterogeneity is distributed according to discrete distribution with $R$ points of support. The parameters $(\alpha, \beta)$ and the unobserved heterogeneity distribution are then estimated by maximizing the log-likelihood of the data, which is given by:

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

Where $t_i$ is the duration of the first overeducation spell, $T_i$ is the length of time to the end of the survey and $d_i = 1$ if individual $i$ leaves the overeducated state before the end of the survey. Finally, using the notations introduced above, the pdf. (resp. cdf) of the overeducation spell duration is given by $f\left(t | \mathbf{x}_i; \theta\right) = E_\nu(f\left(t | \mathbf{x}_i; \nu_i; \theta\right))$ (resp. $F\left(t | \mathbf{x}_i; \theta\right) = E_\nu(F\left(t | \mathbf{x}_i; \nu_i; \theta\right)))$.

It is interesting to compare the estimates for $R = 1$ (i.e. without unobserved heterogeneity) and $R = 2$ (i.e. with two unobserved types).[18] Table 8 report the estimation results corresponding to the homogeneous case with alternative sets of individual controls. The estimated $\alpha$ is well below one (at around 0.64), which means that the hazard out of overeducation is strongly decreasing in the duration of overeducation, even after controlling for *observed* heterogeneity. Age and AFQT scores positively affect the probability of exiting overeducation, while the existence of job hazards and a high Rotter scale score affect it negatively.

This is no longer the case once we allow for unobserved heterogeneity, as evidenced in table 9. Regardless of the set of controls used, we obtain values for $\alpha$ that are very close to one, signifying the absence of true duration dependence. The model identifies two groups of individuals, each accounting for close to half of the overeducated, where the first group has a very low hazard (type 1, 54.5% of the initially overeducated in the sample) while the second group is much likelier to exit overeducation quickly (type 2, 45.5% of the sample). As type 2s exit the sample of overeducated individuals, the probability that a random individual exits overeducation declines, as she is more and

---

[18]Note that we performed the estimation with higher number of types, but they resulted in higher BIC criteria than the model with two types while attributing similar values for the key parameters.

more likely to be a low hazard, type 1 individual.

# 6 Effects of overeducation on earnings

In this section we first show that our data exhibits returns to schooling, underschooling and overschooling that are consistent with the rest of the literature. We then stratify the canonical augmented wage regression based on the types identified in the MPH model (see previous section). This allows us to show that individuals with lower propensities to exit overeducation also exhibit lower returns to schooling. We also find that past overeducation entails a wage penalty (reminiscent of scarring effects associated with unemployment spells), even after applying the above stratification.

When we look at median wages conditional on initial over and undereducation, we show that wage penalties associated with initial overeducation are persistent throughout the years covered by our sample.

Finally, we document wage differentials associated with transitions into and out of overeducation and find them to be consistent with compensating differentials for women, but not for men.

## 6.1 Augmented log-wage regression

Estimates of the returns to overeducation are typically obtained by applying simple OLS to the following equation introduced by Duncan & Hoffman (1981):

$$\log w = \alpha^r S^r + \alpha^o S^o + \alpha^u S^u + X'\beta + \varepsilon$$

where $S^r$, $S^o$ and $S^u$ denote respectively the years of required schooling, years of overeducation[19] and years of undereducation,[20] $X$ a vector of controls (including experience and experience squared) and $\varepsilon$ a productivity shock. This model, which nests the standard Mincerian wage regression as a special case ($\alpha^r = \alpha^o = -\alpha^u$), 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 14 presents the estimated parameters for this model using our data.[21] While this abbreviated table only presents the results of a selection of important controls, the full regression also includes dummies for an employer's broad 1980 Census industry classification. The results from our sample are broadly consistent with results in prior literature. Namely, we see a return of 7.1% for each additional year of education, roughly half the rate of return, 3.2% for additional years of education over the required level for an occupation, and wage penalties for years of undereducation.

Table 14 also reveals a number of expected results. In particular, we see a gender wage gap of 15.8% between men and women, and a compensating wage differential for hazardous occupations. Likewise, increasing an individual's AFQT score by one decile is associated with an approximate 3% increase in wage; and we see wage premiums relative to the North Central region of 6.8% in the Northeast and 9.8% in the West, consistent with relatively higher costs of living in these regions. Lastly, measures for labor market experience all have the expected sign. Those with greater tenure, number of hours worked per year, and total labor market experience all have higher wages. Conversely, those with frequent and long unemployment spells,[22] and those frequently switching jobs during the year experience lower wage rates, all else equal.

---

[19]Difference, if positive, between completed and required years of schooling

[20]Difference, if positive, between required and completed years of schooling

[21]All wages inflation-adjusted to 2000 dollars, with the extreme 5% of values removed from the sample.

[22]As measured by total unemployment experience in weeks.

Finally, Table 15 presents the estimation results from a log-wage regression which is further augmented with first and second lags in overeducation status. A first interpretation of these estimates is that there are scarring effects associated with overeducation. Namely, being overeducated two years ago implies a reduction in wages of 3.6%. Being overeducated last year is also associated with a negative, though insignificant, coefficient. Interestingly, to the extent that overeducation is likely to be more frequent during recessions, these results provide a potential explanation for the scarring effects of graduating during a recession which have been recently discussed in the literature (see, e.g., Oreopoulos et al., 2012).[23] Alternatively, even though this channel is arguably mitigated through the use of cognitive and non-cognitive ability controls, these results could partly reflect a correlation between overeducation incidence and unobserved characteristics that affect productivity and wages. Under this hypothesis, even after exiting overeducation, a worker with past overeducation spells is more likely to have low unobserved ability which is picked up by the negative coefficients on lagged overeducation in the logwage regression.

## 6.2   Augmented wage regressions with heterogeneity types

We discuss below the estimation results corresponding to our augmented wage regression estimated separately for each of the unobserved types identified by the MPH procedure.[24] This approach complements the characterization of the two unobserved types discussed in section 5, by allowing in particular for heterogenous returns to required schooling, years of over- and under-education across the two types. Stratifying the re-

---

[23]See also Liu et al. (2012) who provide evidence from Norway that cyclical skill mismatch is an important mechanism behind the detrimental long-term impact of graduating during a recession.

[24]We predict the type of each individual in the sample using the posterior type probabilities implied by the estimated MPH 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.

gressions by heterogeneity types should also mitigate the endogeneity biases affecting the returns to schooling, to the extent that the types overlap with unobserved differences in ability. A limitation to be kept in mind before interpreting the estimates is that only the MPH sample (i.e. individuals with at least one overeducation spell) can be used in the procedure.

Estimation results, with and without lagged overeducation regressors, are reported in tables **??** and **??**. The first observation is that the types identified using only the duration data through the MPH model have markedly different returns to schooling. In both specification, individuals with low propensity to exit overeducation (type 1) have returns to required schooling that are twice as low as their high propensity counterparts. The wage regression constant, however, is higher for type 1s. This can be interpreted as type 1s having a comparative advantage in low schooling requirement jobs, which would explain their low propensity to exit these jobs.

Another important observation is that the returns to required schooling are overall lower than in the non-stratified regressions presented earlier. This is consistent with the idea that overeducation is correlated with low 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, type 2 individuals exhibit the usual pattern whereby returns to overeducation are about half as large as returns to required schooling. As for type 1 individuals, they appear to have almost no returns to years of overeducation in the first specification: they earn the same wages as if their schooling was that required by their jobs. However, the addition of overeducation lags reverts this result: years of overeducation now yield the same returns as years of required education. That is, overeducated individuals do not earn less than if they were matched. Also, the stratification by types does not remove the penalty for having been overeducated

21

(or undereducated) in previous periods. Interestingly, this pattern is consistent with education level mismatch leading to human capital depreciation, rather than with a screening model where previous mismatch would be used as a signal of low unobserved productivity by the employers.

## 6.3   Long-term earnings dynamics

Figure 13 describes the median hourly wage among workers in our four schooling attainment categories as they progress through their career. In particular, the graph exhibits a median college premium which grows from around 40% to around 60% after 10 years of market experiences and then back to around 40% at 18 year. Having "some college" warrants a smaller premium of about 15%, and the median workers in the "18 years" category earn about 15% more than the median college graduate. Once we condition this figure on initial overeducation status (in addition to actual schooling), we are able to provide a longitudinal picture to complement the usual results on returns to over- and undereducation. While the regressions found in the literature show the correlation between current overeducation and earnings, we show how initial overeducation affects earnings later in life. The striking result here is that the premia and penalties from initial undereducation and overeducation appear very persistent over time for the first three schooling categories (12, 14 and 16 years). Notably, this suggests the existence of long-lasting scarring effects of being initially overeducated, consistent with the accumulation of specific human capital.

The top two graphs in Figure 14 show that undereducation has different implications for the 12 years and 14 years schooling categories. For workers in the 12 years category, being initially undereducated does not improve earnings much early in the career. However, towards the end of our panel, initially undereducated individuals receive a premium of 20 to 25% compared to their "matched" peers. This corresponds to the

22

earnings earned by the median individual in the 14 years category. This is despite the fact that many of these initially undereducated high school graduates are working in jobs requiring up to 16 years of schooling. In other words they earn more than workers with their actual education, but less than other workers in the jobs they occupy. This is consistent with the returns to underschooling found in the literature. By contrast, individuals in the 14 years category who find themselves initially undereducated have a wage path that is similar to that of college educated individuals. That is, they do not appear to be penalized for their undereducation. In both cases, far from converging, the earnings of initially undereducated individuals remain consistently higher than their matched peers.

If we consider college graduates with different initial overeducation statuses, we have - as one might expect - that initially overeducated workers have an earnings path that lies significantly below that of initially matched individuals. Here again, the paths are significantly and persistently altered by the initial overeducation status.[25]

## 6.4   Wage differentials

We proceed to document wage differentials observed across transitions into and out of overeducation. To do so we compute wage rates associated with the job spells in which the respondent finds herself at each interview date. [**CL: Brian, could you provide relevant details regarding wage computation? Is it hourly wage rates or annual earnings?**] We then compute the percentage changes between the wage rate at interview t and t+1, and group them according to the pair of overeducation statuses at interviews t and t+1. We focus on four categories of transitions: from matched to matched, matched to overeducated, overeducated to matched and overeducated to

---

[25]Unfortunately the graph for the 18 year old category is made difficult to interpret by the large variability of the "matched" series.

overeducated.

The four corresponding distributions of wage differentials are reported in table 16. We further distinguish between three types of transitions. The non-CPS column contains cases in which one or more additional job spells occurred between the two interviews (occupations and wages are not available for spells that do not encompass an interview date). The unemployment column contains cases in which one or more unemployment spells of at least one month [**CL: Is it one month or more or less?**] occurred between the interviews. The rest of the transitions are termed "direct" transitions. We are primarily interested in the "direct" columns. The transitions in this column do not exhibit significant unemployment spells between jobs. As a result they are more likely to be voluntary - i.e. to correspond to "on-the-job" search - and thus reflect the pecuniary and non-pecuniary trade-off between the jobs at t and t+1.[**CL: cite Taber here**] Under this interpretation, they can be revealing of non-pecuniary preferences for overeducated vs. matched jobs.

The first observation is that the wage change distributions are very similar except for the "matched to overeducation" transitions. The latter sit to the left of the other three: wage increases are more likely and wage cuts are less likely. This is not consistent with the notion that overeducated workers are trading lower wages against higher non-pecuniary benefits. When we restrict to workers with a college degree, it is still the case that the "M to Oe" distribution is the furthest to the left among direct transitions. Separating males and females reveals interesting gender differences. The pattern described above is marked among men. Instead, women show an opposite pattern such that the "Oe to M" transitions exhibit relatively high proportions of wage increases. This suggests that contrary to men, women who voluntarily give up an overeducated job in favor of a matched job need to be compensated for it with higher wages. That is, women might find non-pecuniary benefits in overeducated jobs such as shorter hours.

24

# 7   Conclusion

Although economists and policy-makers have long been concerned about the determinants and wage effects of overeducation, little is still known about the dynamics of overeducation along the life cycle, particularly in the US. This paper uses data from the NLSY79 combined with the pooled 1989-1991 waves of the CPS to provide some of the first evidence on this question.

Overall, overeducation is a persistent phenomenon, particularly for black workers, with lasting negative effects on wages. These wage penalties are consistent with human capital depreciation and bear similarities with with "scarring" effects that have been found to accompany prolonged unemployment spells. This suggests that similar interventions as those that target the long-term unemployed could be directed at the long-term overeducated.

While the exit rate from overeducation decreases over the length of the overeducation spell, the estimation of a hazard model of overeducation duration that controls for both observed and unobserved heterogeneity suggests that true duration dependence is not strong. Rather, overeducation appears to be a different phenomenon for different groups of workers. On the one hand, it appears to be transitory for about half of our sample. Their situation would be best described with a search model that would naturally generate "frictional overeducation". On the other hand, for the rest of the sample overeducation is instead very persistent, whether it is because overeducated jobs present them with other benefits, or because their overeducation is compensated by lower ability so that they wouldn't fare better in matched jobs.

In sum, this study implies that overeducation is a complex phenomenon that involves a number of the classical ingredients of labor economics: human capital, search frictions, compensating wage differentials and ability differences. In order to quantify

the importance of these mechanisms and identify the policies which would be effective at reducing overeducation rates, one would ideally estimate a dynamic structural model nesting all of these different channels. This question is left for future research.

# 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.

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

Battu, H., Belfield, C. R. & Sloane, P. J. (1999), 'Overeducation among graduates: a cohort view', *Education economics* **7**(1), 21–38.

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

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

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.

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.

Groeneveld, S. & Hartog, J. (2004), 'Overeducation, wages and promotions within the firm', *Labour Economics* **11**(6), 701–714.

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.

Hartog, J. & Oosterbeek, H. (1988), 'Education, allocation and earnings in the Netherlands: Overschooling?', *Economics of Education Review* **7**(2), 185–194.

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', *The Review of Economics and Statistics* pp. 140–144.

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.

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.

Lindley, J. & McIntosh, S. (2009), 'A panel data analysis of the incidence and impact of over-education'.

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* (ahead-of-print), 1–23.

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.

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* pp. 629–643.

Verhaest, D. (2006b), 'The impact of overeducation and its measurement', *Social Indicators Research* **77**(3), 419–448.

# A  Tables and Figures

Table 1: Summary Statistics: Pooled Cross Section 1982-2000

| Variable | Mean | (Std. Dev.) | Min. | Max. | N |
|---|---|---|---|---|---|
| Black | 0.112 | (0.316) | 0 | 1 | 84,142 |
| Hispanic | 0.06 | (0.237) | 0 | 1 | 84,142 |
| Female | 0.511 | (0.5) | 0 | 1 | 84,142 |
| Age | 30.574 | (5.336) | 17 | 43 | 84,142 |
| AFQT | 0.51 | (0.276) | 0.01 | 0.99 | 79,493 |
| North Central | 0.298 | (0.458) | 0 | 1 | 74,581 |
| Northeast | 0.186 | (0.389) | 0 | 1 | 74,581 |
| South | 0.339 | (0.473) | 0 | 1 | 74,581 |
| West | 0.176 | (0.381) | 0 | 1 | 74,581 |
| Urban | 0.761 | (0.426) | 0 | 1 | 72,943 |
| Unemployment Rate | 6.137 | (1.548) | 3.65 | 11.083 | 74,581 |
| Born in The United States | 0.96 | (0.197) | 0 | 1 | 84,142 |
| Mother's Edu. (Yrs., 1979) | 12.061 | (3.43) | 0 | 20 | 77,176 |
| Father's Edu. (Yrs., 1979) | 11.766 | (2.613) | 0 | 20 | 80,541 |
| Lived With Both Parents | 0.63 | (0.483) | 0 | 1 | 84,142 |
| Number of Children Ever Born | 1.113 | (1.211) | 0 | 10 | 75,356 |
| Married, Spouse Present | 0.567 | (0.496) | 0 | 1 | 75,349 |
| Expected Education Level (Years) | 14.13 | (2.139) | 7 | 18 | 83,787 |
| Age at End of Education | 21.297 | (3.675) | 16 | 43 | 84,142 |
| Has GED (No HS Diploma) | 0.094 | (0.292) | 0 | 1 | 84,142 |
| Rotter Scale in 1979 | 8.412 | (2.385) | 4 | 16 | 83,500 |
| Sociability in 1985 | 2.899 | (0.671) | 1 | 4 | 79,356 |
| EEnrolled in High School | 0 | (0.006) | 0 | 1 | 75,318 |
| Not Enrolled, Less Than HS Education | 0 | (0.004) | 0 | 1 | 75,318 |
| Enrolled in College | 0.047 | (0.212) | 0 | 1 | 75,318 |
| Not Enrolled, HS Graduate | 0.953 | (0.212) | 0 | 1 | 75,318 |
| Highest Grade Completed (Collapsed) | 13.398 | (2.027) | 12 | 18 | 84,142 |
| Employed | 0.827 | (0.378) | 0 | 1 | 70,879 |
| Out of Labor Force | 0.118 | (0.322) | 0 | 1 | 70,879 |
| Unemployed | 0.042 | (0.201) | 0 | 1 | 70,879 |
| Weeks Unemployed | 1.75 | (6.067) | 0 | 52 | 74,037 |
| Weeks out of Labor Force | 7.282 | (15.62) | 0 | 52 | 74,037 |
| Individual Income (Year 2000 Dollars) | 27,603 | (23,005) | 308 | 216,200 | 62,806 |
| Hourly Wage (Year 2000 Dollars) | 12.802 | (6.802) | 1.92 | 55.76 | 63,215 |
| Family Poverty Status | 0.082 | (0.275) | 0 | 1 | 65,535 |
| Overeducation (Years) | 0.472 | (1.958) | −6 | 6 | 56,516 |
| Required Education (Yrs.) | 13.038 | (1.854) | 12 | 18 | 56,516 |
| Occupational Hazards | 0.156 | (0.935) | −0.700 | 2.861 | 68,448 |

Table 2: Overeducation Status Proportions

|              | Continuous | Categories |
|-------------:|:----------:|:----------:|
| Undereducated | 12.58% | 9.18% |
| Matched | 61.23% | 70.90% |
| Overeducated | 26.19% | 19.92% |

Note: Comparison based on a continuous measure of observed education and a measure collapsed into five educational levels.

Table 3: Transition Matrix

| $t-1$ | OE | UE | $t$ Matched | Unemp. |
|---|---|---|---|---|
| Overeducated | 71.4% | 8.5% | 16.0% | 4.1% |
| Undereducated | 14.2% | 38.6% | 43.2% | 3.9% |
| Matched | 6.1% | 7.1% | 80.7% | 6.0% |
| Unemployed | 8.0% | 4.3% | 36.5% | 51.2% |
| Total | 20.8% | 12.7% | 55.4% | 11.1% |

Note: Sample has completed education.

Table 4: Transition Matrix by Education

| (12 Years) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | | | | |
| Undereducated | | 36.9% | 59.3% | 3.7% |
| Matched | | 7.9% | 85.1% | 7.0% |
| Unemployed | | 3.9% | 45.7% | 50.4% |
| Total | | 13.0% | 73.8% | 13.2% |
| (16 Years) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| Overeducated | 61.8% | 2.0% | 32.2% | 4.0% |
| Undereducated | 24.5% | 23.6% | 46.2% | 5.7% |
| Matched | 30.4% | 3.4% | 63.9% | 2.2% |
| Unemployed | 23.3% | 1.6% | 17.8% | 57.4% |
| Total | 43.7% | 3.8% | 45.6% | 6.9% |

Note: Sample has completed education.

Table 5: Transition Matrix by Gender

| (Male) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | 72.1% | 7.6% | 18.1% | 2.1% |
| Undereducated | 15.6% | 39.6% | 41.8% | 2.9% |
| Matched | 7.7% | 5.4% | 82.9% | 3.9% |
| Unemployed | 8.7% | 4.2% | 47.9% | 39.2% |
| Total | 23.2% | 11.2% | 59.0% | 6.5% |
| (Female) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| Overeducated | 70.8% | 9.3% | 14.1% | 5.8% |
| Undereducated | 13.4% | 38.0% | 44.1% | 4.5% |
| Matched | 4.8% | 8.5% | 78.9% | 7.8% |
| Unemployed | 7.7% | 4.4% | 31.9% | 56.0% |
| Total | 19.0% | 13.7% | 52.9% | 14.5% |

Note: Sample has completed education.

Table 6: Transition Matrix by Race

| (White) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | 70.7% | 8.5% | 16.6% | 4.2% |
| Undereducated | 13.5% | 38.8% | 43.5% | 4.2% |
| Matched | 6.3% | 7.0% | 80.8% | 5.9% |
| Unemployed | 7.8% | 4.4% | 36.2% | 51.6% |
| Total | 20.6% | 12.6% | 55.6% | 11.1% |
| (Black) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| Overeducated | 78.2% | 8.7% | 10.2% | 2.9% |
| Undereducated | 20.4% | 36.6% | 41.4% | 1.6% |
| Matched | 4.5% | 8.6% | 79.7% | 7.2% |
| Unemployed | 9.0% | 3.5% | 39.6% | 47.9% |
| Total | 22.4% | 13.0% | 53.7% | 10.9% |

Note: Sample has completed education.

Table 7: Transition Matrix by AFQT Quartile

| (1st) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | 83.2% | 12.4% | 2.9% | 1.5% |
| Undereducated | 9.8% | 34.1% | 50.5% | 5.6% |
| Matched | 0.4% | 6.2% | 84.3% | 9.1% |
| Unemployed | 3.4% | 3.1% | 45.0% | 48.6% |
| Total | 8.4% | 10.6% | 66.3% | 14.7% |

| (2nd) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | 73.4% | 9.6% | 11.5% | 5.4% |
| Undereducated | 9.8% | 36.9% | 49.2% | 4.2% |
| Matched | 1.9% | 7.2% | 84.7% | 6.2% |
| Unemployed | 4.7% | 4.2% | 39.5% | 51.5% |
| Total | 12.0% | 12.5% | 62.6% | 12.9% |

| (3rd) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | 73.0% | 10.3% | 12.3% | 4.3% |
| Undereducated | 17.4% | 36.9% | 42.3% | 3.4% |
| Matched | 7.2% | 8.3% | 79.4% | 5.1% |
| Unemployed | 12.7% | 6.5% | 31.8% | 49.0% |
| Total | 24.8% | 14.6% | 51.2% | 9.4% |

| (4th) | | | $t$ | |
| $t-1$ | OE | UE | Matched | Unemp. |
| --- | --- | --- | --- | --- |
| Overeducated | 68.6% | 6.0% | 21.7% | 3.6% |
| Undereducated | 19.2% | 46.4% | 31.2% | 3.2% |
| Matched | 17.8% | 7.0% | 72.1% | 3.2% |
| Unemployed | 15.3% | 4.2% | 24.1% | 56.5% |
| Total | 36.1% | 12.6% | 43.4% | 7.9% |

Note: Sample has completed education.

Table 8: Proportional Hazard Model (Out of Overeducation)

| | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| $\alpha$ | 0.644*** | (0.0277) | 0.638*** | (0.0277) | 0.653*** | (0.0274) |
| Constant | −3.670*** | (0.2526) | −3.806*** | (0.1350) | −3.461*** | (0.0883) |
| Black | −0.209 | (0.0623) | −0.128 | (0.1230) | | |
| Hispanic | −0.168 | (0.1324) | 0.096 | (0.1473) | | |
| Female | 0.157 | (0.1543) | 0.134 ** | (0.0633) | | |
| Age† | 0.067*** | (0.0664) | −0.265*** | (0.0411) | | |
| AFQT‡ | 0.026* | (0.1920) | 0.446*** | (0.1348) | | |
| Northeast | 0.036 | (0.0950) | | | | |
| South | 0.121 | (0.0821) | | | | |
| West | −0.184 | (0.1013) | | | | |
| Urban | 0.004 | (0.0981) | | | | |
| Unemployment Rate† | 0.017 | (0.0965) | | | | |
| Born in The United States | −0.138 | (0.1109) | | | | |
| HH in SMSA, Central City | −0.072 | (0.0804) | | | | |
| HH in SMSA, Not Central City | 0.037 | (0.0389) | | | | |
| HH not in SMSA | −0.770 | (0.3721) | | | | |
| Tenure (1k Wks.) | −0.068* | (0.0352) | | | | |
| Hours Worked (1k Hrs./Yr.)† | 0.033 | (0.0331) | | | | |
| Number of Jobs (SLI)† | 0.052 | (0.0521) | | | | |
| Work Experience (1k Hrs.)† | −0.077 | (0.0466) | | | | |
| Total Unemp. (Wks.)† | 0.042 | (0.0375) | | | | |
| Mother's Edu. (Yrs., 1979)† | 0.018 | (0.0366) | | | | |
| Father's Edu. (Yrs., 1979)† | 0.032 | (0.0730) | | | | |
| Lived With Both Parents | 0.109 | (0.2228) | | | | |
| Has GED (No HS Diploma) | 0.332 | (0.1453) | | | | |
| Rotter Scale in 1979† | 0.030 | (0.0339) | | | | |
| RotterScore$^2$ | −0.054* | (0.0245) | | | | |
| Sociability in 1985† | 0.044 | (0.0322) | | | | |
| Occupational Hazards | −0.114* | (0.0494) | | | | |
| LL | 5,979.8 | | 6,002.3 | | 6,033.2 | |
| Observations | 1,973 | | 1,973 | | 1,973 | |

Standard errors in parentheses

† Values converted to Z-scores

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Mixed Proportional Hazard Model (Out of Overeducation)

| | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| $\alpha$ | 1.051 | (0.0584) | 1.079 | (0.0414) | 1.118* | (0.0367) |
| $P(V_1)$ | 0.518*** | (0.0541) | 0.568*** | (0.0255) | 0.577*** | (0.0207) |
| $P(V_2)$ | 0.482*** | (0.0541) | 0.432*** | (0.0255) | 0.423*** | (0.0207) |
| $V_1$ | 0.078*** | (0.0220) | 0.086*** | (0.0124) | 0.077*** | (0.0114) |
| $V_2$ | 1.993*** | (0.1824) | 2.203*** | (0.1122) | 2.261*** | (0.0998) |
| Constant | $-4.742$*** | (0.4355) | $-5.245$*** | (0.2291) | $-4.627$*** | (0.1230) |
| Black | $-0.422$*** | (0.1119) | $-0.039$ | (0.2457) | | |
| Hispanic | 0.096 | (0.2912) | $0.422**$ | (0.1980) | | |
| Female | 0.548* | (0.2730) | 0.212 | (0.1216) | | |
| Age† | 0.106 | (0.1381) | $-0.376**$ | (0.0663) | | |
| AFQT‡ | $-0.237$ | (0.2735) | $0.809**$ | (0.2470) | | |
| Northeast | $-0.262$ | (0.2190) | | | | |
| South | 0.139 | (0.1640) | | | | |
| West | $-0.141$ | (0.2052) | | | | |
| Urban | 0.059 | (0.1858) | | | | |
| Unemployment Rate† | 0.019 | (0.1975) | | | | |
| Born in The United States | $-0.151$ | (0.2479) | | | | |
| HH in SMSA, Central City | $-0.027$ | (0.1424) | | | | |
| HH in SMSA, Not Central City | $-0.109$ | (0.1432) | | | | |
| HH not in SMSA | $-0.823$ | (0.6114) | | | | |
| Tenure (1k Wks.) | $-0.102$ | (0.0580) | | | | |
| Hours Worked (1k Hrs./Yr.)† | $-0.005$ | (0.0622) | | | | |
| Number of Jobs (SLI)† | 0.085 | (0.0928) | | | | |
| Work Experience (1k Hrs.)† | $-0.151$ | (0.0863) | | | | |
| Total Unemp. (Wks.)† | 0.028 | (0.0630) | | | | |
| Mother's Edu. (Yrs., 1979)† | 0.082 | (0.0505) | | | | |
| Father's Edu. (Yrs., 1979)† | 0.009 | (0.1489) | | | | |
| Lived With Both Parents | $-0.005$ | (0.3564) | | | | |
| Has GED (No HS Diploma) | 0.688* | (0.2957) | | | | |
| Rotter Scale in 1979† | 0.015 | (0.0751) | | | | |
| RotterScore$^2$ | $-0.047$ | (0.0641) | | | | |
| Sociability in 1985† | $-0.017$ | (0.0693) | | | | |
| Occupational Hazards | $-0.144$ | (0.1225) | | | | |
| LL | | 5,893.3 | | 5,813.8 | | 5,840.5 |
| Observations | | 1,973 | | 1,979 | | 1,979 |

Standard errors in parentheses
† Values converted to Z-scores
‡ Unit interval
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Determinants of Overeducation

| | $\mathbb{1}\left\{Overeducated\right\}$ | |
|---|---|---|
| Black | 0.167*** | (0.0079) |
| Hispanic | 0.114*** | (0.0100) |
| Female | −0.018*** | (0.0048) |
| Age | 0.011*** | (0.0007) |
| AFQT‡ | 0.417*** | (0.0094) |
| Northeast | 0.006 | (0.0062) |
| South | 0.038*** | (0.0053) |
| West | 0.044*** | (0.0065) |
| Urban | 0.035*** | (0.0059) |
| Unemployment Rate | −0.013*** | (0.0018) |
| Born in The United States | 0.003 | (0.0115) |
| HH in SMSA, Central City | 0.021** | (0.0069) |
| HH in SMSA, Not Central City | −0.011* | (0.0052) |
| HH Not in SMSA | −0.016* | (0.0070) |
| Tenure (1k Wks.) | −0.049*** | (0.0111) |
| Hours Worked (1k Hrs./Yr.) | 0.039*** | (0.0032) |
| Number of Jobs (SLI) | 0.004 | (0.0028) |
| Work Experience (1k Hrs.) | −0.007*** | (0.0004) |
| Total Unemp. (Wks.) | −0.001*** | (0.0001) |
| Mother's Edu. (Yrs., 1979) | 0.011*** | (0.0008) |
| Father's Edu. (Yrs., 1979) | 0.012*** | (0.0010) |
| Lived With Both Parents | 0.050*** | (0.0047) |
| Has GED (No HS Diploma) | −0.111*** | (0.0082) |
| Rotter Scale in 1979 | 0.005 | (0.0052) |
| Rotter Scale$^2$ | −0.000 | (0.0003) |
| Sociability in 1985 | 0.022*** | (0.0032) |
| Occupational Hazards | 0.003 | (0.0027) |
| Constant | −0.571*** | (0.0417) |
| Adjusted $R^2$ | | 0.162 |
| F | | 205.872 |
| Observations | | 40,280 |

Standard errors in parentheses

Model includes controls for industry. Estimates not reported.

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Determinants of Overeducation (14 Years of Education)

| | $\mathbb{1}\{Overeducated\}$ | |
|---|---|---|
| Black | 0.053** | (0.0201) |
| Hispanic | 0.055* | (0.0226) |
| Female | −0.011 | (0.0130) |
| Age | −0.008*** | (0.0020) |
| AFQT‡ | −0.223*** | (0.0259) |
| Northeast | 0.008 | (0.0175) |
| South | 0.003 | (0.0141) |
| West | 0.044* | (0.0171) |
| Urban | 0.017 | (0.0166) |
| Unemployment Rate | 0.006 | (0.0051) |
| Born in The United States | 0.118** | (0.0365) |
| HH in SMSA, Central City | 0.052** | (0.0177) |
| HH in SMSA, Not Central City | 0.021 | (0.0137) |
| HH Not in SMSA | 0.070*** | (0.0193) |
| Tenure (1k Wks.) | 0.053 | (0.0292) |
| Hours Worked (1k Hrs./Yr.) | −0.009 | (0.0084) |
| Number of Jobs (SLI) | 0.013 | (0.0070) |
| Work Experience (1k Hrs.) | 0.002* | (0.0010) |
| Total Unemp. (Wks.) | 0.001*** | (0.0002) |
| Mother's Edu. (Yrs., 1979) | −0.001 | (0.0019) |
| Father's Edu. (Yrs., 1979) | 0.003 | (0.0026) |
| Lived With Both Parents | −0.013 | (0.0126) |
| Has GED (No HS Diploma) | −0.074** | (0.0284) |
| Rotter Scale in 1979 | 0.013 | (0.0144) |
| Rotter Scale$^2$ | −0.000 | (0.0009) |
| Sociability in 1985 | −0.025** | (0.0084) |
| Occupational Hazards | 0.068*** | (0.0071) |
| Constant | 1.018*** | (0.1127) |
| Adjusted R$^2$ | | 0.183 |
| F | | 34.742 |
| Observations | | 5,733 |

Standard errors in parentheses

Model includes controls for industry. Estimates not reported.

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Determinants of Overeducation (16 Years of Education)

| | $\mathbb{1}\{Overeducated\}$ | |
|---|---|---|
| Black | 0.071** | (0.0235) |
| Hispanic | −0.110*** | (0.0325) |
| Female | 0.042*** | (0.0124) |
| Age | 0.001 | (0.0024) |
| AFQT‡ | −0.119*** | (0.0307) |
| Northeast | 0.015 | (0.0160) |
| South | −0.025 | (0.0144) |
| West | 0.042* | (0.0184) |
| Urban | −0.041* | (0.0174) |
| Unemployment Rate | 0.013** | (0.0050) |
| Born in The United States | 0.045 | (0.0308) |
| HH in SMSA, Central City | 0.039* | (0.0171) |
| HH in SMSA, Not Central City | 0.007 | (0.0137) |
| HH Not in SMSA | −0.013 | (0.0206) |
| Tenure (1k Wks.) | 0.049 | (0.0310) |
| Hours Worked (1k Hrs./Yr.) | −0.012 | (0.0092) |
| Number of Jobs (SLI) | 0.015* | (0.0075) |
| Work Experience (1k Hrs.) | −0.001 | (0.0012) |
| Total Unemp. (Wks.) | 0.000 | (0.0003) |
| Mother's Edu. (Yrs., 1979) | −0.002 | (0.0021) |
| Father's Edu. (Yrs., 1979) | 0.001 | (0.0029) |
| Lived With Both Parents | −0.030* | (0.0147) |
| Has GED (No HS Diploma) | −0.245** | (0.0831) |
| Rotter Scale in 1979 | −0.023 | (0.0164) |
| Rotter Scale$^2$ | 0.001 | (0.0010) |
| Sociability in 1985 | 0.002 | (0.0090) |
| Occupational Hazards | 0.247*** | (0.0113) |
| Constant | 1.034*** | (0.1232) |
| Adjusted R$^2$ | | 0.212 |
| F | | 48.779 |
| Observations | | 6,735 |

Standard errors in parentheses

Model includes controls for industry. Estimates not reported.

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Determinants of Overeducation (18 Years of Education)

|  | $\mathbb{1}\{Overeducated\}$ | |
|---|---|---|
| Black | $-0.006$ | (0.0273) |
| Hispanic | $-0.009$ | (0.0316) |
| Female | $0.107^{***}$ | (0.0121) |
| Age | $-0.015^{***}$ | (0.0023) |
| AFQT‡ | $-0.344^{***}$ | (0.0350) |
| Northeast | $-0.025$ | (0.0170) |
| South | $0.004$ | (0.0150) |
| West | $-0.022$ | (0.0177) |
| Urban | $0.007$ | (0.0177) |
| Unemployment Rate | $-0.018^{**}$ | (0.0055) |
| Born in The United States | $0.076^{*}$ | (0.0305) |
| HH in SMSA, Central City | $0.044^{**}$ | (0.0165) |
| HH in SMSA, Not Central City | $0.016$ | (0.0137) |
| HH Not in SMSA | $-0.039$ | (0.0202) |
| Tenure (1k Wks.) | $0.157^{***}$ | (0.0319) |
| Hours Worked (1k Hrs./Yr.) | $-0.052^{***}$ | (0.0083) |
| Number of Jobs (SLI) | $0.002$ | (0.0076) |
| Work Experience (1k Hrs.) | $0.007^{***}$ | (0.0011) |
| Total Unemp. (Wks.) | $0.001^{***}$ | (0.0003) |
| Mother's Edu. (Yrs., 1979) | $0.001$ | (0.0020) |
| Father's Edu. (Yrs., 1979) | $-0.007^{**}$ | (0.0028) |
| Lived With Both Parents | $0.047^{***}$ | (0.0138) |
| Has GED (No HS Diploma) | $-0.215^{**}$ | (0.0735) |
| Rotter Scale in 1979 | $-0.038^{**}$ | (0.0141) |
| Rotter Scale$^2$ | $0.003^{**}$ | (0.0009) |
| Sociability in 1985 | $0.051^{***}$ | (0.0091) |
| Occupational Hazards | $0.052^{***}$ | (0.0137) |
| Constant | $1.613^{***}$ | (0.1185) |
| Adjusted $R^2$ | | 0.194 |
| F | | 26.759 |
| Observations | | 4,064 |

Standard errors in parentheses

Model includes controls for industry. Estimates not reported.

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Augmented log-wage regression

|  | ln (HourlyWage) | |
| --- | --- | --- |
| Years of Overeducation (Positive) | 0.032*** | (0.0016) |
| Years of Undereducation (Positive) | −0.040*** | (0.0025) |
| Required Education (Yrs.) | 0.071*** | (0.0015) |
| Black | 0.004 | (0.0081) |
| Hispanic | 0.045*** | (0.0104) |
| Female | −0.158*** | (0.0048) |
| Age | 0.003*** | (0.0008) |
| AFQT‡ | 0.301*** | (0.0103) |
| Northeast | 0.068*** | (0.0063) |
| South | −0.017** | (0.0054) |
| West | 0.098*** | (0.0067) |
| Urban | 0.016** | (0.0061) |
| Unemployment Rate | −0.020*** | (0.0019) |
| Born in The United States | −0.086*** | (0.0121) |
| HH in SMSA, Central City | 0.033*** | (0.0072) |
| HH in SMSA, Not Central City | 0.047*** | (0.0054) |
| HH Not in SMSA | −0.114*** | (0.0072) |
| Tenure (1k Wks.) | 0.242*** | (0.0112) |
| Hours Worked (1k Hrs./Yr.) | 0.005 | (0.0037) |
| Number of Jobs (SLI) | −0.057*** | (0.0032) |
| Work Experience (1k Hrs.) | 0.012*** | (0.0004) |
| Total Unemp. (Wks.) | −0.001*** | (0.0001) |
| Mother's Edu. (Yrs., 1979) | 0.004*** | (0.0008) |
| Father's Edu. (Yrs., 1979) | 0.002 | (0.0011) |
| Lived With Both Parents | −0.010* | (0.0049) |
| Has GED (No HS Diploma) | 0.025** | (0.0090) |
| Rotter Scale in 1979 | −0.005 | (0.0055) |
| Rotter Scale$^2$ | 0.000 | (0.0003) |
| Sociability in 1985 | 0.005 | (0.0033) |
| Occupational Hazards | 0.034*** | (0.0026) |
| Constant | 1.290*** | (0.0457) |
| Adjusted $R^2$ | | 0.429 |
| F | | 758.2 |
| Observations | | 30,243 |

Standard errors in parentheses

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Augmented log-wage regression with lagged overeducation status

|  | ln (HourlyWage) | |
| --- | --- | --- |
| Years of Overeducation (Positive) | 0.047*** | (0.0030) |
| Years of Undereducation (Positive) | −0.057*** | (0.0037) |
| Required Education (Yrs.) | 0.073*** | (0.0019) |
| Overeducated (t-1) | −0.018 | (0.0133) |
| Undereducated (t-1) | 0.045*** | (0.0135) |
| Overeducated (t-2) | −0.036** | (0.0119) |
| Undereducated (t-2) | 0.038** | (0.0119) |
| Black | 0.005 | (0.0099) |
| Hispanic | 0.042*** | (0.0125) |
| Female | −0.170*** | (0.0058) |
| Age | 0.001 | (0.0009) |
| AFQT‡ | 0.318*** | (0.0124) |
| Northeast | 0.075*** | (0.0076) |
| South | −0.015* | (0.0066) |
| West | 0.107*** | (0.0084) |
| Urban | 0.015* | (0.0072) |
| Unemployment Rate | −0.024*** | (0.0027) |
| Born in The United States | −0.099*** | (0.0149) |
| HH in SMSA, Central City | 0.037*** | (0.0088) |
| HH in SMSA, Not Central City | 0.055*** | (0.0064) |
| HH Not in SMSA | −0.119*** | (0.0086) |
| Tenure (1k Wks.) | 0.190*** | (0.0130) |
| Hours Worked (1k Hrs./Yr.) | −0.028*** | (0.0050) |
| Number of Jobs (SLI) | −0.066*** | (0.0047) |
| Work Experience (1k Hrs.) | 0.012*** | (0.0005) |
| Total Unemp. (Wks.) | −0.001*** | (0.0001) |
| Mother's Edu. (Yrs., 1979) | 0.004*** | (0.0010) |
| Father's Edu. (Yrs., 1979) | 0.002 | (0.0013) |
| Lived With Both Parents | −0.005 | (0.0060) |
| Has GED (No HS Diploma) | 0.043*** | (0.0121) |
| Rotter Scale in 1979 | −0.009 | (0.0067) |
| Rotter Scale$^2$ | 0.000 | (0.0004) |
| Sociability in 1985 | 0.008 | (0.0040) |
| Occupational Hazards | 0.031*** | (0.0031) |
| Constant | 1.472*** | (0.0559) |
| Adjusted $R^2$ | | 0.394 |
| F | | 389.6 |
| Observations | | 20,286 |

Standard errors in parentheses

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Wage Growth With Intermediate States

| | M to M | | | M to OE | | | OE to M | | | OE to OE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Direct | non-CPS | Unemp. | Direct | non-CPS | Unemp. | Direct | non-CPS | Unemp. | Direct | non-CPS | Unemp. |
| **Women** | | | | | | | | | | | | |
| $(\infty, -10\%)$ | 13.2% | 15.3% | 22.2% | 8.4% | 24.6% | 33.3% | 11.4% | 13.7% | 12.5% | 13.6% | 18.6% | 25.0% |
| $[-10\%, 5\%)$ | 22.7% | 11.4% | 12.5% | 19.1% | 6.6% | 6.7% | 19.6% | 9.8% | 6.3% | 19.4% | 11.0% | 12.9% |
| $[5\%, 20\%)$ | 23.2% | 15.6% | 15.2% | 28.5% | 14.8% | 20.0% | 26.7% | 8.8% | 43.8% | 23.6% | 12.2% | 15.2% |
| $[20\%, 35\%)$ | 15.2% | 13.7% | 15.4% | 16.2% | 14.8% | 0.0% | 12.1% | 9.8% | 18.8% | 16.8% | 16.6% | 12.1% |
| $[35\%, \infty)$ | 25.8% | 44.0% | 34.7% | 27.8% | 39.3% | 40.0% | 30.2% | 57.8% | 18.8% | 26.6% | 41.6% | 34.8% |
| **Men** | | | | | | | | | | | | |
| $(\infty, -10\%)$ | 13.0% | 12.6% | 18.7% | 13.2% | 30.8% | 42.9% | 8.3% | 11.1% | 18.2% | 14.9% | 21.8% | 17.6% |
| $[-10\%, 5\%)$ | 22.2% | 13.6% | 12.2% | 25.0% | 11.5% | 0.0% | 17.4% | 7.4% | 9.1% | 17.9% | 12.8% | 16.2% |
| $[5\%, 20\%)$ | 23.2% | 17.7% | 16.3% | 31.6% | 15.4% | 28.6% | 33.1% | 9.3% | 36.4% | 24.9% | 16.5% | 22.1% |
| $[20\%, 35\%)$ | 15.4% | 15.8% | 14.6% | 10.3% | 19.2% | 0.0% | 11.6% | 9.3% | 18.2% | 14.9% | 13.3% | 13.2% |
| $[35\%, \infty)$ | 26.2% | 40.3% | 38.1% | 19.9% | 23.1% | 28.6% | 29.8% | 63.0% | 18.2% | 27.3% | 35.6% | 30.9% |
| **16 Yrs. Edu.** | | | | | | | | | | | | |
| $(\infty, -10\%)$ | 9.7% | 13.0% | 17.6% | 7.3% | 27.1% | 36.4% | 12.2% | 15.4% | 13.3% | 16.6% | 26.2% | 36.7% |
| $[-10\%, 5\%)$ | 17.7% | 7.8% | 0.0% | 20.4% | 4.2% | 9.1% | 18.0% | 8.8% | 6.7% | 18.3% | 15.4% | 6.7% |
| $[5\%, 20\%)$ | 24.3% | 16.9% | 5.9% | 27.3% | 18.8% | 9.1% | 24.5% | 5.5% | 40.0% | 20.6% | 9.2% | 20.0% |
| $[20\%, 35\%)$ | 15.4% | 11.7% | 17.6% | 15.9% | 18.8% | 0.0% | 13.1% | 11.0% | 20.0% | 16.2% | 13.8% | 10.0% |
| $[35\%, \infty)$ | 32.9% | 50.6% | 58.8% | 29.0% | 31.3% | 45.5% | 32.2% | 59.3% | 20.0% | 28.2% | 35.4% | 26.7% |

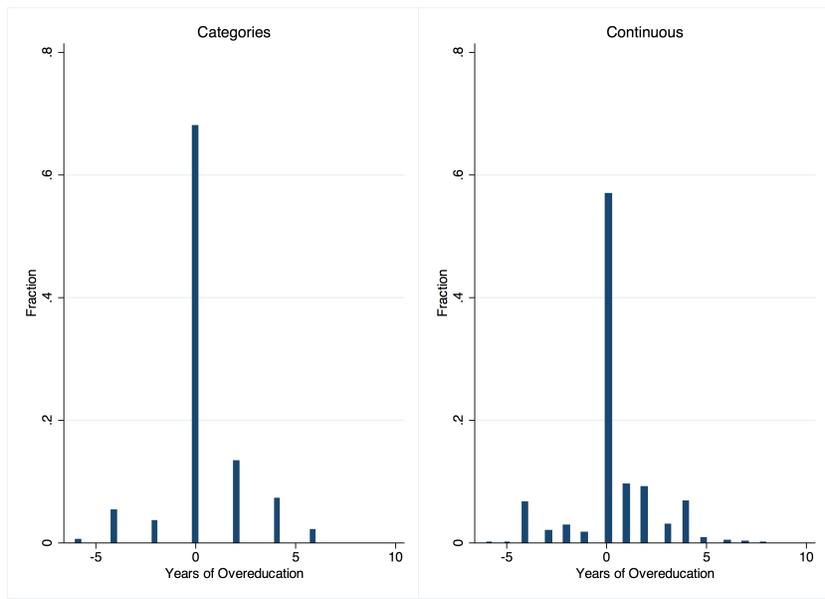Note: Sample has completed education.

45

Figure 1: Distribution of Overeducation

Table 17: Augmented log-wage regression with unobserved types

| | (Type 1) | | (Type 2) | |
| | $\ln\left(HourlyWage\right)$ | | $\ln\left(HourlyWage\right)$ | |
|---|---|---|---|---|
| Years of Overeducation (Positive) | $-0.005$ | $(0.0045)$ | $0.024^{***}$ | $(0.0038)$ |
| Years of Undereducation (Positive) | $-0.040^{***}$ | $(0.0077)$ | $-0.070^{***}$ | $(0.0108)$ |
| Required Education (Yrs.) | $0.036^{***}$ | $(0.0038)$ | $0.069^{***}$ | $(0.0034)$ |
| Black | $-0.019$ | $(0.0212)$ | $0.020$ | $(0.0196)$ |
| Hispanic | $-0.081^{**}$ | $(0.0247)$ | $0.119^{***}$ | $(0.0266)$ |
| Female | $-0.137^{***}$ | $(0.0116)$ | $-0.156^{***}$ | $(0.0111)$ |
| Age | $0.008^{***}$ | $(0.0020)$ | $0.010^{***}$ | $(0.0020)$ |
| AFQT‡ | $0.317^{***}$ | $(0.0281)$ | $0.276^{***}$ | $(0.0259)$ |
| Northeast | $0.088^{***}$ | $(0.0157)$ | $0.127^{***}$ | $(0.0152)$ |
| South | $-0.013$ | $(0.0138)$ | $0.035^{**}$ | $(0.0131)$ |
| West | $0.063^{***}$ | $(0.0169)$ | $0.107^{***}$ | $(0.0160)$ |
| Urban | $0.054^{**}$ | $(0.0168)$ | $0.005$ | $(0.0155)$ |
| Unemployment Rate | $-0.030^{***}$ | $(0.0048)$ | $-0.023^{***}$ | $(0.0049)$ |
| Born in The United States | $-0.154^{***}$ | $(0.0307)$ | $-0.001$ | $(0.0342)$ |
| HH in SMSA, Central City | $0.030$ | $(0.0166)$ | $0.048^{**}$ | $(0.0157)$ |
| HH in SMSA, Not Central City | $0.075^{***}$ | $(0.0126)$ | $0.075^{***}$ | $(0.0127)$ |
| HH Not in SMSA | $-0.081^{***}$ | $(0.0201)$ | $-0.101^{***}$ | $(0.0179)$ |
| Tenure (1k Wks.) | $0.127^{***}$ | $(0.0295)$ | $0.188^{***}$ | $(0.0270)$ |
| Hours Worked (1k Hrs./Yr.) | $0.001$ | $(0.0093)$ | $-0.028^{***}$ | $(0.0086)$ |
| Number of Jobs (SLI) | $-0.058^{***}$ | $(0.0077)$ | $-0.086^{***}$ | $(0.0072)$ |
| Work Experience (1k Hrs.) | $0.013^{***}$ | $(0.0011)$ | $0.010^{***}$ | $(0.0010)$ |
| Total Unemp. (Wks.) | $-0.003^{***}$ | $(0.0003)$ | $-0.001^{**}$ | $(0.0002)$ |
| Mother's Edu. (Yrs., 1979) | $-0.005^{*}$ | $(0.0020)$ | $0.002$ | $(0.0018)$ |
| Father's Edu. (Yrs., 1979) | $0.004$ | $(0.0027)$ | $0.010^{***}$ | $(0.0026)$ |
| Lived With Both Parents | $-0.024$ | $(0.0134)$ | $0.019$ | $(0.0124)$ |
| Has GED (No HS Diploma) | $-0.093^{*}$ | $(0.0452)$ | $0.027$ | $(0.0472)$ |
| Rotter Scale in 1979 | $0.005$ | $(0.0160)$ | $-0.037^{**}$ | $(0.0127)$ |
| Rotter Scale$^2$ | $-0.000$ | $(0.0010)$ | $0.002^{*}$ | $(0.0008)$ |
| Sociability in 1985 | $0.011$ | $(0.0088)$ | $0.023^{**}$ | $(0.0080)$ |
| Occupational Hazards | $-0.002$ | $(0.0101)$ | $0.042^{***}$ | $(0.0067)$ |
| Constant | $1.884^{***}$ | $(0.1224)$ | $1.175^{***}$ | $(0.1138)$ |
| Adjusted $R^2$ | | $0.354$ | | $0.365$ |
| F | | $96.1$ | | $115.3$ |
| Observations | | $5,200$ | | $5,975$ |

Standard errors in parentheses

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Augmented log-wage regression with unobserved types

| | (Type 1) $\ln(HourlyWage)$ | | (Type 2) $\ln(HourlyWage)$ | |
|---|---|---|---|---|
| Years of Overeducation (Positive) | 0.014* | (0.0058) | 0.030*** | (0.0048) |
| Years of Undereducation (Positive) | −0.029* | (0.0129) | −0.064** | (0.0210) |
| Required Education (Yrs.) | 0.028*** | (0.0045) | 0.058*** | (0.0042) |
| Overeducated (t-1) | −0.083*** | (0.0189) | −0.089** | (0.0289) |
| Undereducated (t-1) | −0.043 | (0.0311) | −0.056 | (0.0536) |
| Overeducated (t-2) | −0.114*** | (0.0167) | −0.097*** | (0.0277) |
| Undereducated (t-2) | −0.062* | (0.0270) | −0.108* | (0.0446) |
| Black | −0.026 | (0.0240) | −0.005 | (0.0239) |
| Hispanic | −0.105*** | (0.0281) | 0.096** | (0.0324) |
| Female | −0.129*** | (0.0137) | −0.161*** | (0.0137) |
| Age | 0.004 | (0.0025) | 0.006* | (0.0026) |
| AFQT‡ | 0.294*** | (0.0334) | 0.266*** | (0.0315) |
| Northeast | 0.096*** | (0.0183) | 0.149*** | (0.0185) |
| South | −0.029 | (0.0164) | 0.043** | (0.0159) |
| West | 0.075*** | (0.0206) | 0.120*** | (0.0201) |
| Urban | 0.056** | (0.0194) | 0.000 | (0.0181) |
| Unemployment Rate | −0.023*** | (0.0063) | −0.024*** | (0.0067) |
| Born in The United States | −0.140*** | (0.0356) | −0.045 | (0.0485) |
| HH in SMSA, Central City | 0.061** | (0.0198) | 0.061** | (0.0191) |
| HH in SMSA, Not Central City | 0.098*** | (0.0145) | 0.088*** | (0.0152) |
| HH Not in SMSA | −0.058* | (0.0236) | −0.085*** | (0.0216) |
| Tenure (1k Wks.) | 0.078* | (0.0335) | 0.153*** | (0.0311) |
| Hours Worked (1k Hrs./Yr.) | −0.025* | (0.0118) | −0.082*** | (0.0116) |
| Number of Jobs (SLI) | −0.054*** | (0.0112) | −0.086*** | (0.0110) |
| Work Experience (1k Hrs.) | 0.015*** | (0.0014) | 0.010*** | (0.0012) |
| Total Unemp. (Wks.) | −0.003*** | (0.0003) | 0.000 | (0.0003) |
| Mother's Edu. (Yrs., 1979) | −0.005* | (0.0023) | 0.001 | (0.0023) |
| Father's Edu. (Yrs., 1979) | 0.003 | (0.0031) | 0.013*** | (0.0031) |
| Lived With Both Parents | −0.024 | (0.0158) | 0.003 | (0.0151) |
| Has GED (No HS Diploma) | −0.060 | (0.0604) | −0.031 | (0.0664) |
| Rotter Scale in 1979 | 0.017 | (0.0188) | −0.047** | (0.0153) |
| Rotter Scale$^2$ | −0.001 | (0.0012) | 0.002* | (0.0009) |
| Sociability in 1985 | 0.009 | (0.0102) | 0.013 | (0.0097) |
| Occupational Hazards | −0.001 | (0.0124) | 0.048*** | (0.0081) |
| Constant | 2.156*** | (0.1444) | 1.850*** | (0.1463) |
| Adjusted $R^2$ | | 0.335 | | 0.327 |
| F | 48 | 53.6 | | 58.3 |
| Observations | | 3,553 | | 4,008 |

Standard errors in parentheses

‡ Unit interval

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
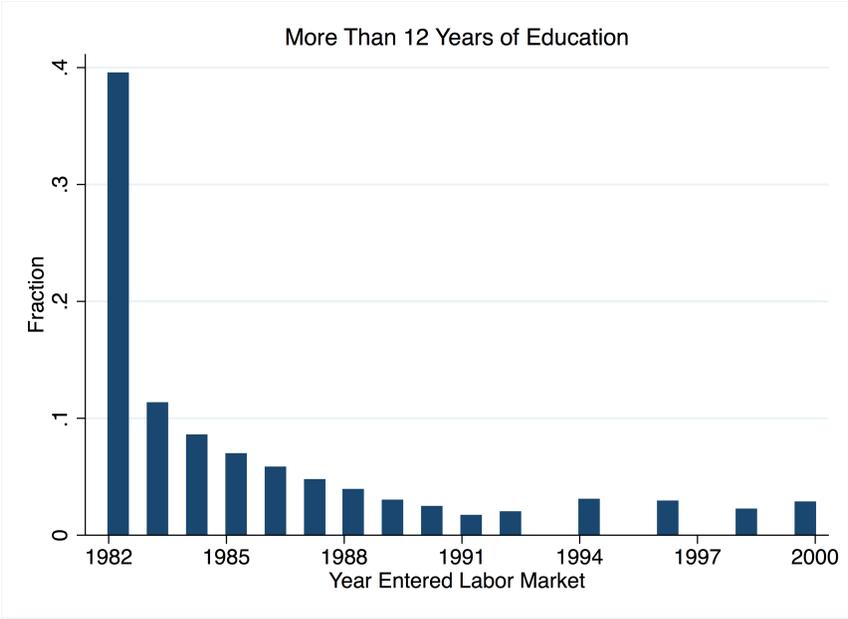
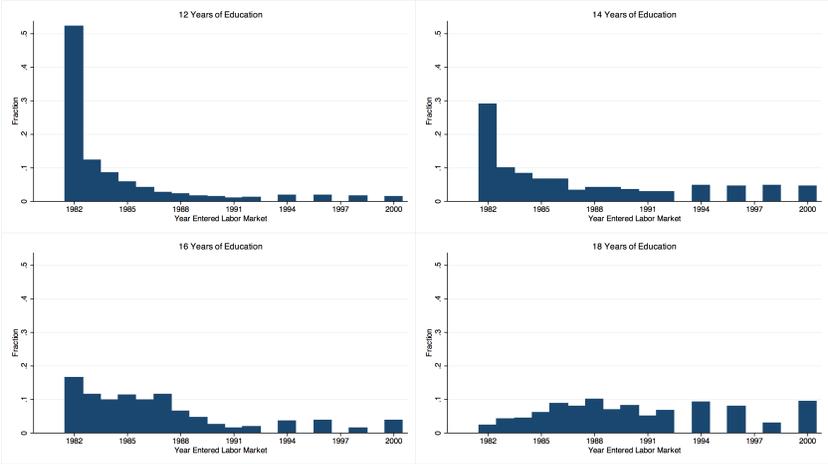Figure 2: Distribution of Labor Market Entry
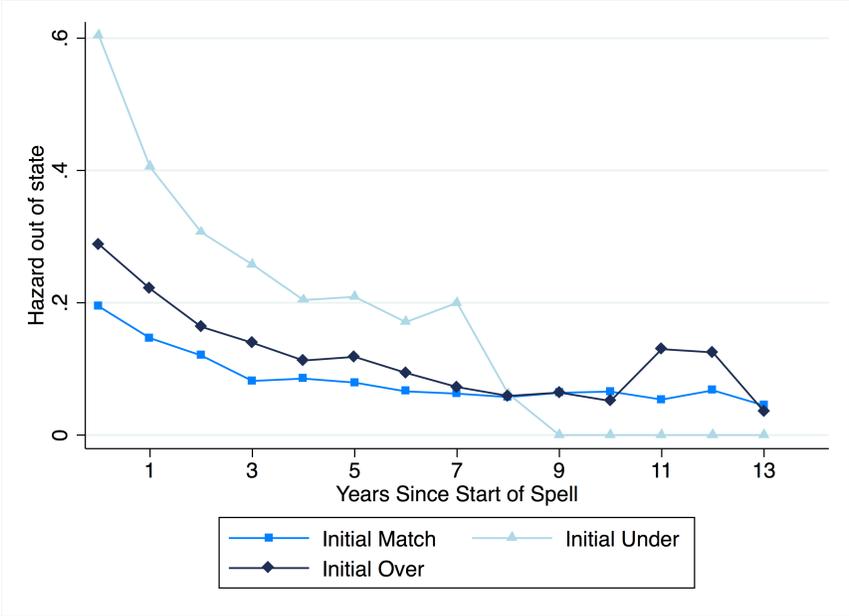


Figure 3: Distribution of Labor Market Entry

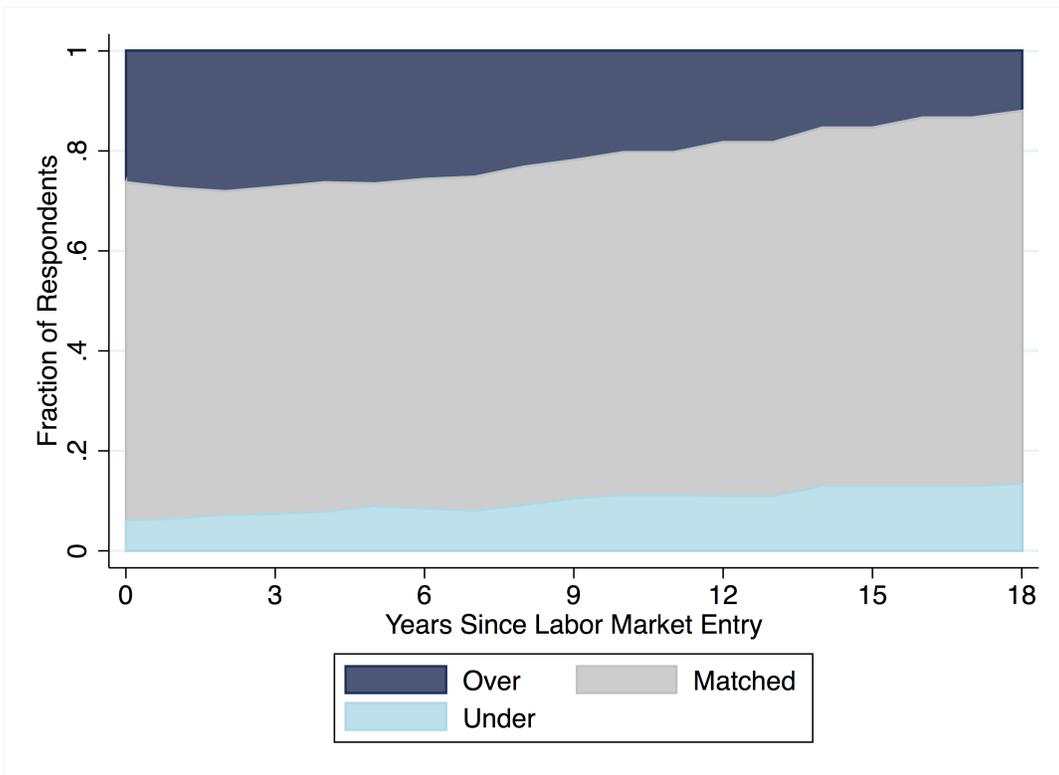Figure 4: Hazard out of Over- and Under-education

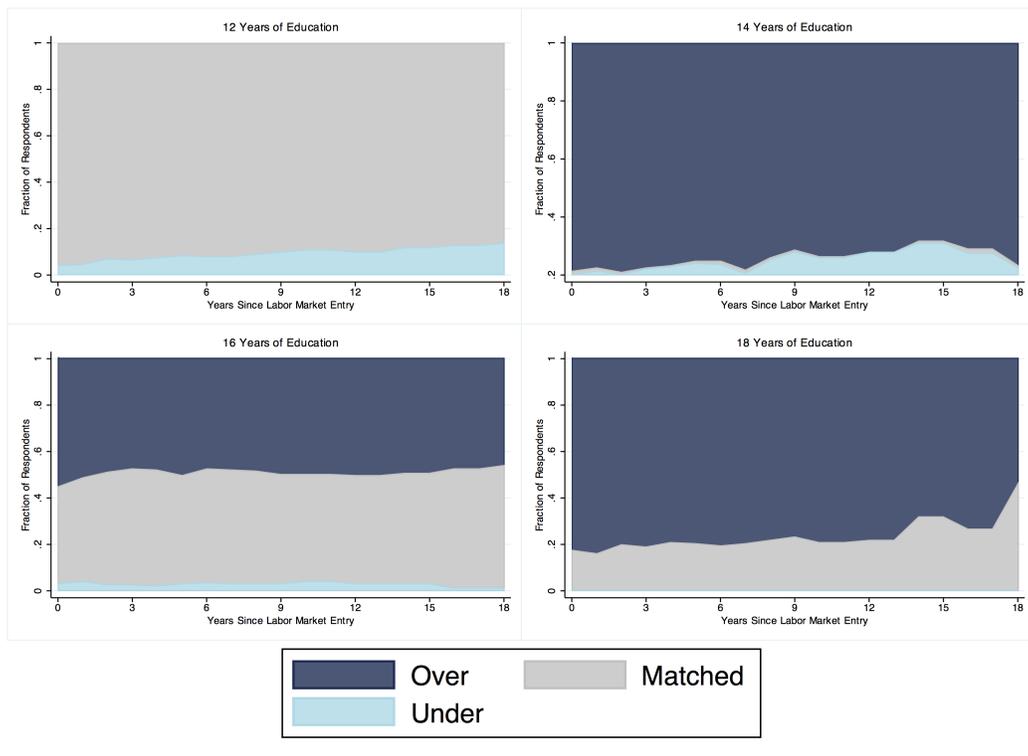Figure 5: Composition of Overeducation (Fraction of Respondents)

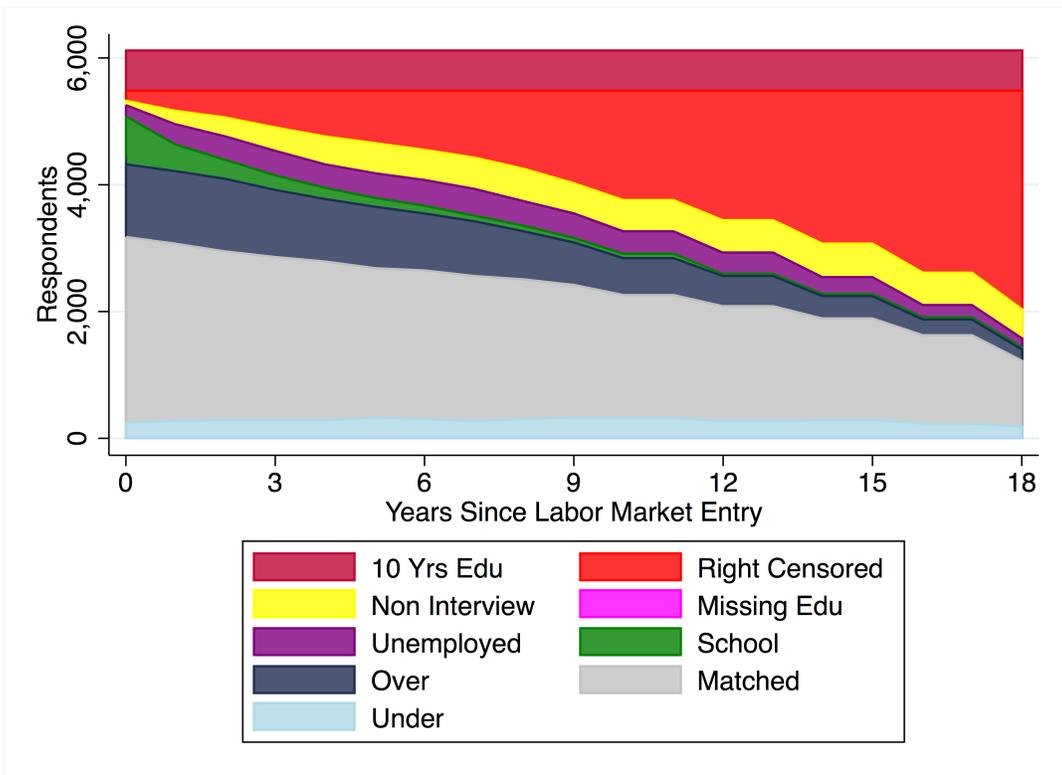Figure 6: Composition of Overeducation by Education Level (Fraction of Respondents)

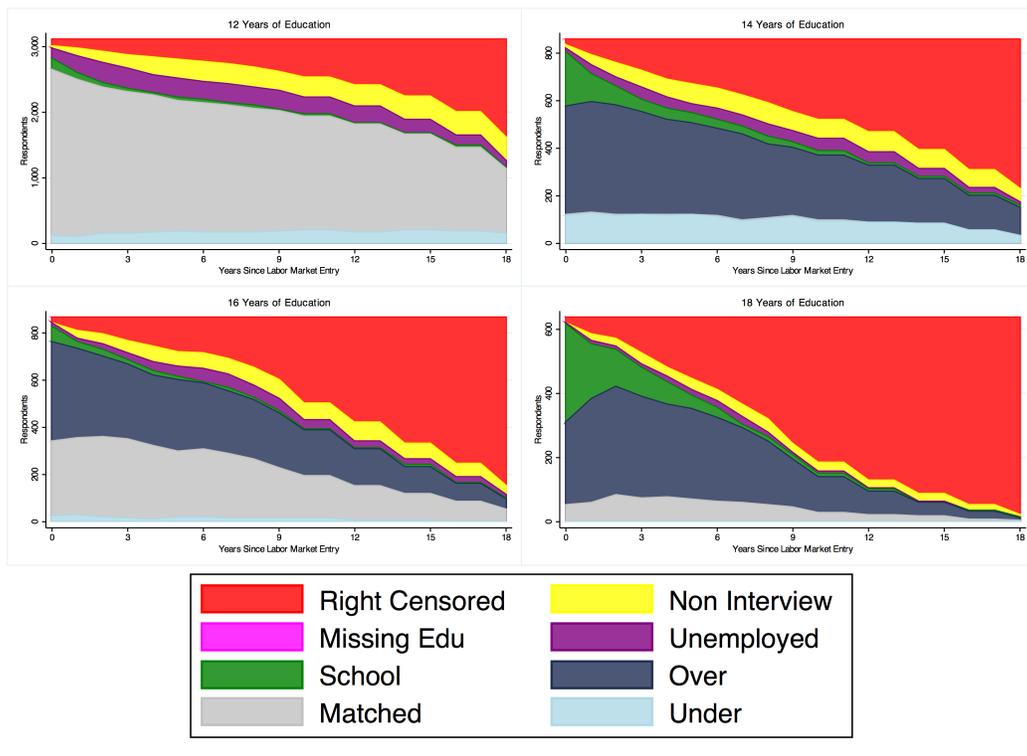Figure 7: Composition of Overeducation (Number of Respondents)

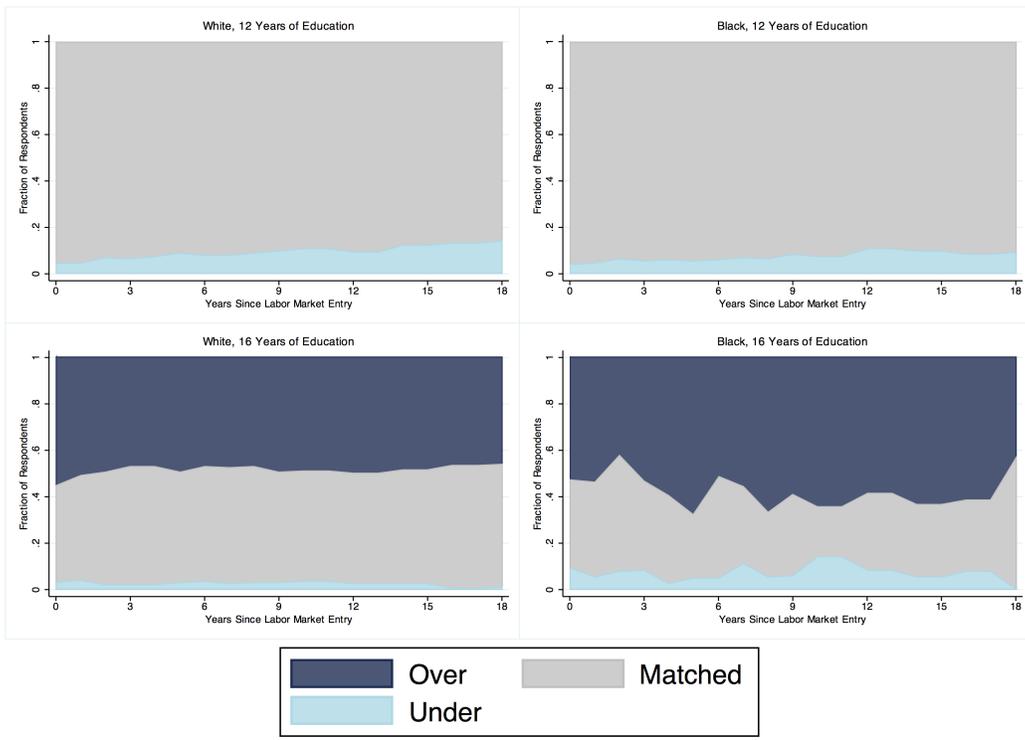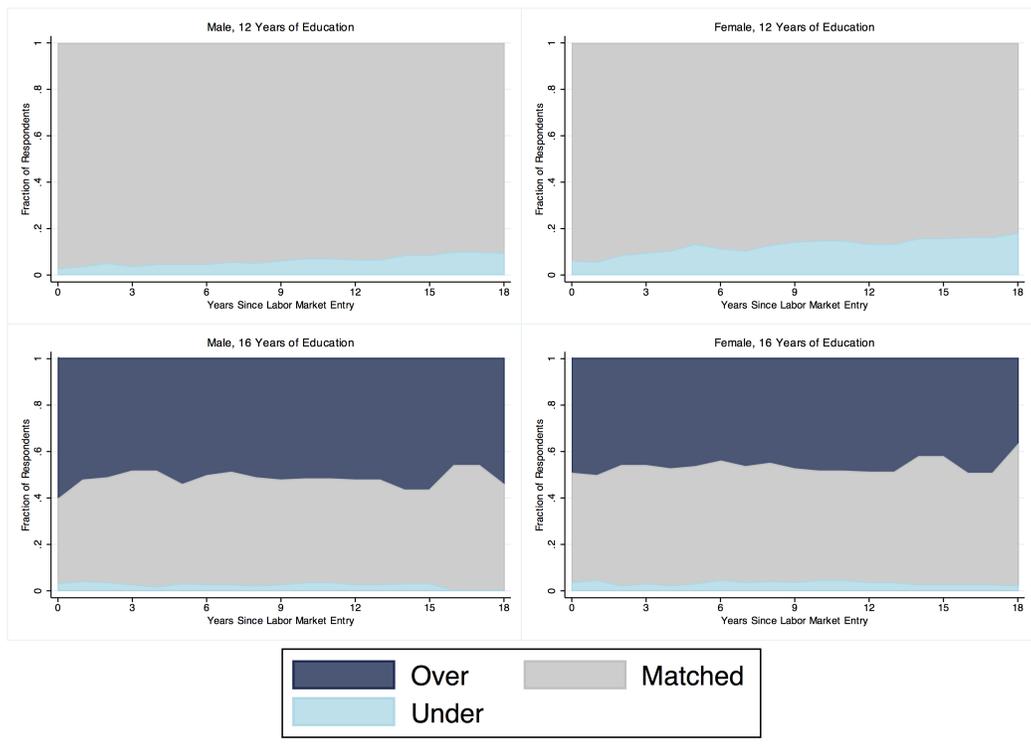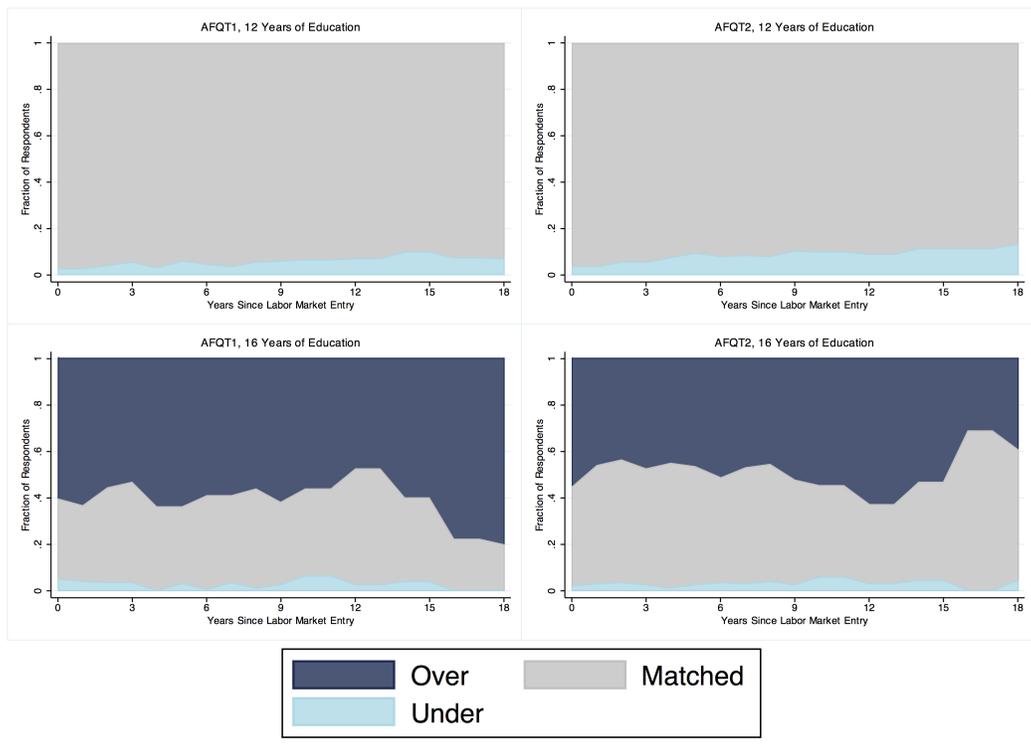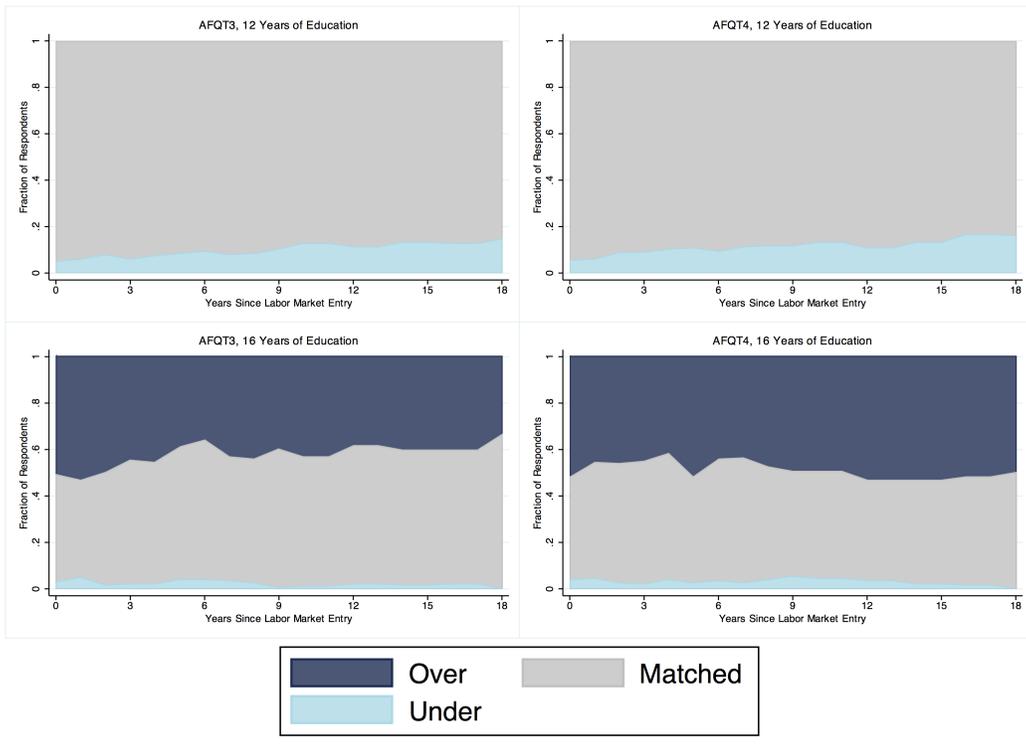Figure 8: Composition of Overeducation by Education Level (Number of Respondents)

Figure 9: Composition of Overeducation by Race (Fraction of Respondents)

Figure 10: Composition of Overeducation by Gender (Fraction of Respondents)

Figure 11: Composition of Overeducation by AFQT (Fraction of Respondents)

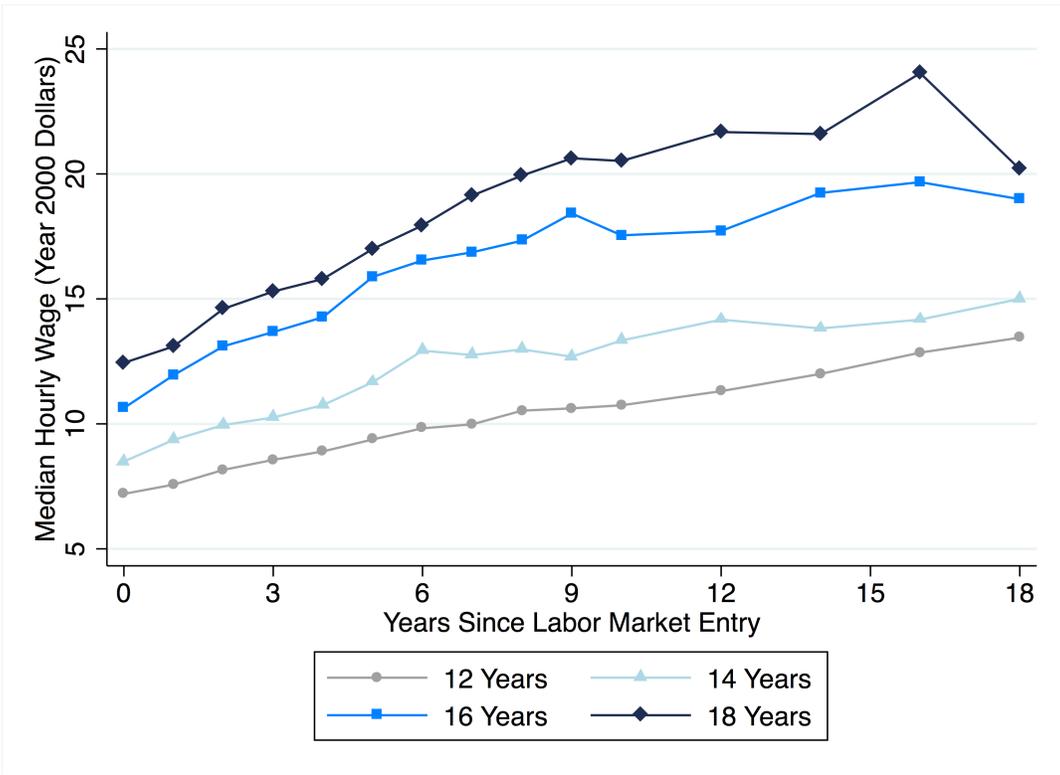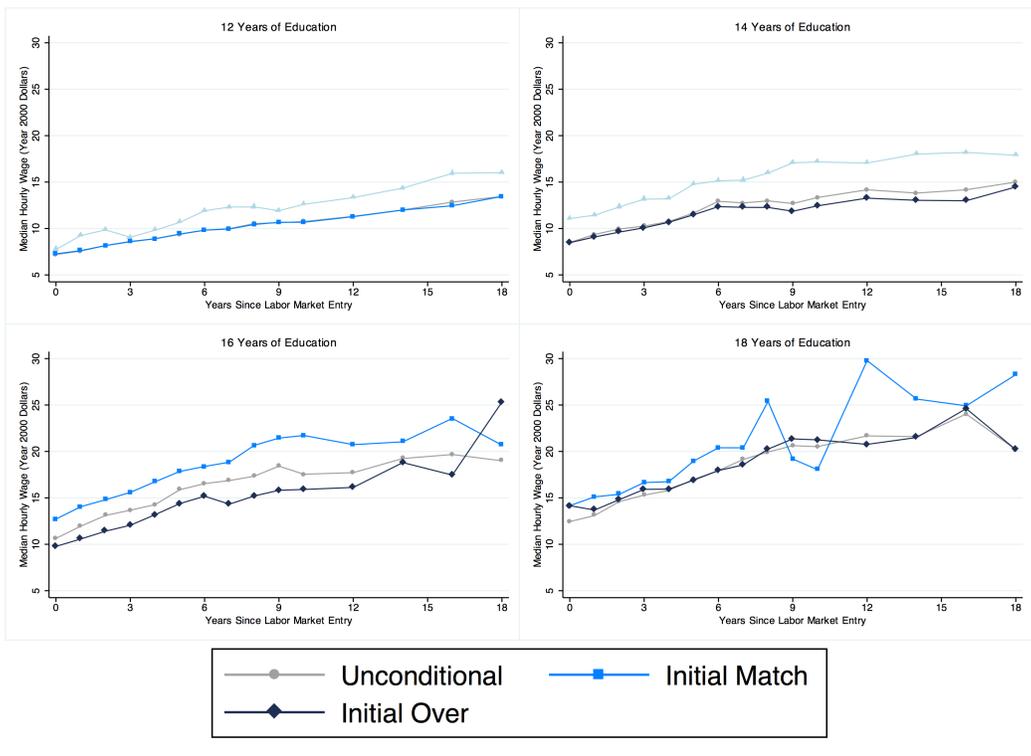Figure 12: Composition of Overeducation by AFQT (Fraction of Respondents)

Figure 13: Path of Median Hourly Wage

Figure 14: Path of Median Hourly Wage