What Explains the Racial Gaps in Task Assignment and Pay Over the Life-Cycle? *

Limor Golan

Washington University in St. Louis,

Jonathan James

California Polytechnic State University,

Carl Sanders

Washington University in St. Louis

October 30, 2018

preliminary and incomplete

Abstract

Over the life-cycle the black-white wage increases substantially. We document that white workers are assigned initially to occupations with higher complex task requirements. The accumulated earnings gap after 15 years is about 117K. Empirically, about 53% of this gap can be accounted for after controlling for observable pre-market characteristics (AFQT, and schooling). To address the question of why the gaps increase over time and in order to account for dynamic selection, we develop a dynamic Roy model of employment and occupation choice and learning which nests discrimination. We then develop a two-step estimation method to recover the model's structural parameters, and use them to simulate counterfactual exercises that allows us to decompose the different factors affecting these gaps.

About 27% of the total pay gap after 15 years is explained by higher entry costs of black workers into complex-task occupation. In addition, equating the initial pre-market skills and the initial mean differences in beliefs (which reflect differences in initial unobserved skills) explain about half of the total earnings after 15 years. This reflects the importance of sorting of workers into jobs over time in

*

explaining the long-term earnings gaps, and the fact that it is not predicted by initial skill differences. Our estimates indicate that the variance of beliefs is increasing substantially more for white workers, reflecting possibly unobserved promotion gaps, training or an increase in the initially unknown skills over time. While the pre-market skill gaps (observed and unobserved) between black and white workers are clearly an important factor in the labor market outcome gaps, frictions and discrimination are important determinants of the life-cycle labor market outcome gaps.

1 Introduction

The wage gap between black and white workers increases substantially over the life-cycle and is correlated with an increase in the gap in the nature of tasks that black and white workers preform. At the same time, the gap in the complexity of skill in the occupations that black and white workers are assigned to increases.[1] Occupations with higher complex-skill requirement have substantially higher employment rate and are associated with higher pay. The net result of these gaps is that the total labor market earnings of black workers are dramatically lower than those of similar white workers. In our data, we estimate that white workers earn around $382,000 in the first 15 years of their careers compared to $265,000 for equivalent black workers.

In this paper we develop and estimate a dynamic model of employment, occupation choice and pay. In the model, workers wages increase on average due to sorting into increasingly better matches and accumulation of experience in the different occupations. The dynamics of the pay gap is determined by initial skill gaps (observed and unobserved) as well as labor market discrimination. We use the estimated model to quantify the role of the different factors in the observed racial pay gap dynamics. Previous research has focused on the racial wage gap and unemployment, but there is little research examining the interplay between occupation assignments, labor market experience and wages and its role in the increasing racial pay gap.[2] To the best of our knowledge this is the first paper to analyze and estimate the mechanism underlying the dynamics of the racial pay, occupation and experience gap accounting for dynamic selection.[3]

In this paper, we quantify the racial earnings gap and analyze its mechanisms using a life-cycle model of labor force participation and occupational choice. We first document racial differences in wages, labor force participation, earnings, and occupational choice over the career. While much of the wage gap is explained by

[1] See Figure 1.

[2] Speer (2015) documents the increase in earnings gap and demand for the different types of skills measured by tests scores in different occupations. He does not estimate the mechanism underlying the dynamics of the pay, occupation and experience gaps.

[3] See O'Neill (1990), Neal and Johnson (1996), Fryer (2011)

differences in pre-market skills, we find that differences in labor force participation and occupational choices remain after controlling for those skills, and as a result, about half of the total earnings after 15 years is not explained by initial observed differences in education and test scores. To account for these facts, we develop a model in which workers are endowed with both observed and unobserved skills, which can have different distributions across races. These include differences in education, other background demographics observed to the econometrician, and a skill unobserved by the econometrician. Additionally, there is a skill to the worker and the employer which they learn about over time. Our model extends the standard learning model, by allowing for a race-specific deterministic increase in unobserved skill. This results in changes over time in the accuracy of the signal received. Each period, a worker chooses to work and, if he works, between ten occupations, with different returns to different types skills. There are costs of entry into employment which vary by occupation as well as the cost of switching occupations depends on race and education. To the extent that there are differences in hiring and offers made to black and white workers in different occupations, it is captured by these differences in non-pecuniary part of the utility function. The model is a dynamic Roy model in which workers learn about their unobserved skills, accumulate experience in different occupations and sort into occupations based on their skills. Thus wages increase over time due to accumulation of experience in different occupations as well as due to learning about the unknown skill and sorting into a better match. Due to costs if entry into occupations, occupation switching costs or random taste shocks, workers may not be employed in the occupation that maximizes their expected life-time earnings.

The learning component of the model reflects the potential role that imperfect information plays in the racial wage gap, a force emphasized by Oettinger (1990). We allow for both the mean beliefs as well as the speed of learning to vary by race; the latter is due to differences in the accuracy of signals obtained while the former can reflect initial skill differences. The speed of learning depends on both the variance of the skill distribution and the noise; since the beliefs vary by race. Thus, this model incorporate a statistical discrimination component as in Phelps (1972). Since our model is dynamic, if employers obtain less accurate signals on their minority employees, the information friction in the model can potentially generate differences in employment and occupational choices and sorting and therefore part of the racial earnings gaps can be due to larger "mismatch" of workers to jobs. Moreover, if black workers have higher entry costs into "complex-skill" occupations, they might be less likely to sort into occupations that maximize their life-time earnings. At the same time, it is possible that it is the initial gaps in observed and unobserved skills that account for both the gap and the increase in the gap.[4] Our counterfactual analysis address the question of, the

[4]Keane and Wolpin (1997) as well as other papers find that initial observed and unobserved skill can account for the majority of the increase in the wage variance when workers are older.

relative importance of role of discrimination, initial (pre-market) skill gaps and preference differences and entry costs.

We estimate the model using matched National Longitudinal Survey of Youth 1979 (NLSY79) and Dictionary of Occupational Titles (DOT) data. We construct occupation classifications based on the DOT data, following Autor et al. (2003) and Antonovics and Golan (2012), so that our occupational groupings differ in the return to skills (however, we only use the index in order to classify the 3-digit occupations into 10 groups). We then develop and use a multistage estimation technique to estimate the model's structural parameters. Because of the computational burden of learning models, the estimation is conducted in two stages estimation in the spirit of Hotz and Miller (1983) and Arcidiacono and Miller(2011). Since the unobserved heterogeneity in our model is continuous (we do not observe the initial beliefs and signals that workers and employers observe), we maximize the likelihood of observing wages, employment and occupation choices simultaneously in the first stage, using approximation of the CCP's to correct for dynamic selection. These choice probabilities in the likelihood function are derived from the approximated valuation function and serve primarily as a selection correction device to produce unbiased estimates of the wage and skill distribution parameters. However, we do not estimate the individual CCP's which depend on each individual beliefs.[5] In the second stage we estimate the preference parameters. We exploit the property of finite state dependence of tour model to use the representation first developed in Altug and Miller (1998) (also used in Arcidiacono and Miller (2011) Gayle and Golan (2012)) to reduce the dimensionality of the model and use simulation estimation to solve for the fixed point of the model[6]

Our estimates confirm that the returns to the unknown skill is higher in occupations with higher complex-skill requirement, but the return to "known" skills are not monotonic in our occupation task-complexity ranking. We find that the occupation specific skills play an important role in wages, which suggest that if there is state dependence in occupational assignment over time, then the first occupation can predict wage increase over and above the"pre-market" characteristics. Our model allows for a race-specific deterministic increase in unobserve skill. This can reflect training or assignment to jobs that depend on this skill more heavily that we do not observe. We find that there is a small initial differences in mean beliefs and that it increases substantially over time. We are unable to separate between discrimination in training or task assignment we do not observed (promotions for example) and simply a faster increase in skills of more able workers. Lastly, the initial variance of beliefs is slightly larger but the gap increases substantially after

[5]Our estimator is different from the learning models estimators developed in James (2011) and Hincapie (2017) because we assume that wages are observed with a measurement error. As a result the beliefs cannot be recovered directly from the wage residuals

[6]See the algorithms in Arcidiacono et al. (2013), which is similar to the algorithm in Aguirregabiria and Mira (2002).

10 years. The larger variance of beliefs of white workers is consistent with Phelps (1972) model and may indicate faster learning about white workers, or assignment of these workers into tasks that depends more on the unobserved skills.

Our counterfactual exercises decompose the earnings gaps into wage offers and differences in non-wage factors. In the wage equation, are interested in determining the importance of both observed skills (e.g. AFQT as in Neal and Johnson (1996) or AFQT plus education as in Lang and Manove (2011)). We first simulate the model assigning black workers the same AFQT and education distributions that white workers have. We find that while these skills explain 36-37% of the experience and does not change much between the first 5 and 15 years in the labor market, it explains less than a quarter of the occupational task-complexity gap after 15 years. Overall it accounts for about half of the total accumulated earnings in the first five years and 37% of the gap after 15 years. Since the, unobserved skills of workers black and white workers might be different when they enter the labor market even if we condition on AFQT and education, we conduct a counterfactual equalizing the initial beliefs gaps of black and white workers and the observed skills. Closing the initial beliefs and pre-market observed skills closes 70% of the total earnings gap and about a half of the gap in total earnings after 15 years. Sorting into jobs in which workers have comparative advantage over time reduces the impact of the pre-market observed and the initial beliefs gap in our model. The The reason is that accumulated experience and differences in occupational sorting have larger impact on earnings later in the life-cycle.

About a half of the accumulated pay gap, 60% of the experience gap and $70-75\%$ of the task assignment gaps are unexplained by initial skills and man beliefs differences. We therefore conduct a counterfactual that equates the occupation entry costs of black and white workers. We find that about 75% of the experience gap is explained by the preference costs differences and 27% of the total earnings gap after 15 years is explained by these difference. Lastly, we conduct a counterfactual that equalizes all beliefs and a counterfactual that equalizes only the initial beliefs gap. We find the the latter closes 14% of the total earnings gap after 15 years and the former closes 27% of the total earnings gap. Thus, the evolution of beliefs plays an important role in the increase of the total earnings gap. It also explains about 10% of the employment gap but not the occupational gaps.

Related Literature

The economics literature emphasizes the importance of pre-labor-market factors in black/white wage gaps and labor market outcomes. See Altonji and Blank (1999), Cameron and Heckman (2001), and Carneiro et al. (2005), and Fryer (2011) for surveys of economic analyses of racial labor market gaps, including the empirical relationships between pre-market characteristics such as education, test scores, and family background with a variety of labor market outcomes.

Analysis of the racial wage gap between workers with similar human capital stocks became a focus of work after O'Neill (1990) and Neal and Johnson (1996) documented the relationship between pre-market skills and labor market gaps. In particular, Neal and Johnson found that controlling for test scored on the US Armed Forces Qualification Test (AFQT) accounted for 70 % of the racial wage gap for males. Neal and Johnson did not include education in their analysis due to concerns about endogeneity, but Carneiro et al. (2005) argue that AFQT scores themselves are affected by pre-existing socio-economic gaps, and Lang and Manove (2011) show that the gap conditional on AFQT increases after controlling additionally for education.

There has been less recent focus on the role of post-entry differences in the experiences of white and black workers. The fact that similar white and black workers are employed at different rates was originally discussed in the context of long term trends in racial wage gaps: Brown (1984), Chandra (2000), Juhn (2003), and Western and Pettit (2005) all found evidence that more black workers than white workers dropped out of the labor force between 1940 and 1990, and higher-wage black workers were more likely to drop out than higher-wage white workers, increasing the measured wage gap. In the context of a single cohort, Eckstein and Wolpin (1999) emphasize the difference between actual and potential wage offer distributions, making the point that observed wages can either under- or over-estimate discrimination. Antecol and Bedard (2004) find that including measures of actual (rather than potential) labor market experience close even more of the gap in the Neal and Johnson (1996)-type wage specification.

There are other economic theories of racial wage gaps that do not require systematic differences in average skills. For example, Oettinger (1996) develops and tests a model of statistical discrimination; assuming blacks have less precise signals of ability predicts an increasing wage gap with experience, even if both races have equal mean ability. He finds that blacks have lower gains to mobility which causes wage gap to rise with experience. Altonji and Pierret (2001) develop a test for statistical discrimination and do not find evidence supporting the hypothesis that employers discriminate based on easily observed characteristics such as education.

Given the importance pre-market skills have in the economic literature, there are also a number of papers considering the early childhood formation of skills. Without specific attention to racial differences, Cunha et al. (2010) consider the problem of estimating the skill formation of children. In the context of racial gaps specifically, Gayle et al. (2014) find that returns to early investment in children's education is larger for blacks; however, labor market wage gaps affect educational achievement by increasing the labor supply of black mothers and decreasing time investment in children and their educational attainments.

2 Model

We consider an economy with infinitely-lived, risk-neutral workers and firms with a common discount factor β. Workers can be either black or white, where $r_i \in \{\text{White, Black}\}$ denotes the race of individual i. Workers differ in the set of skills they possess. Each worker has a vector of a skill $x_{i,0}$, fixed over time and known at labor market entry. Workers also have an unknown time invariant skill θ_i, drawn from a population distribution that is normal with race-specific means and variances, $\theta_i \sim N\left(\mu_r, \sigma^2_{r,\theta}\right)$ Workers enter the labor market with beliefs about θ_i: from their perspective, at time 0 their θ_i is a normally distributed random variable with expected value $E_0\left[\theta_i\right] \equiv b_{i,0}$ and variance $\sigma^2_{i,0}$. The initial level of expected unknown skills $b_{i,0}$ is heterogenous across workers even in the first period, and $b_{i,0}$ is distributed $N\left(\mu_r, \sigma^2_{r,b}\right)$. We do not take a stand on why initial expected skills are heterogenous across workers in the first period, but the model is consistent with workers receiving different amounts of information during schooling or at the job interview stage and the amount of that information differing across races.

There are J occupations in the economy indexed by j, and in each occupation there are many identical firms. Occupations vary by the degree to which workers' output depends on both the known and the unknown skills. At each age a, workers choose an occupation or non-employment, $j \in \{0, 1, ..., J\}$, with $j = 0$ denoting non-employment. Let choices 1 through J be occupations ranked from lowest skill requirements (1) to highest (J) with an indicator $I_{ija} = 1$ if occupation j is chosen by worker i at age a and 0 otherwise. Choices are mutually exclusive: $\sum_{j'=0}^{J} I_{ij'a} = 1$. We denote accumulated experience (number of years working) at age a by $t_{i,a} \equiv \sum_{a'=1}^{a} (1 - I_{i0a'})$.

Unobserved ability evolves over time according through an on-the-job training mechanism, with growth rates given by a known set of parameters $\{\beta_t\}_{t=1}^{T}$. A worker at age a with $t_{i,a}$ years of actual experience then has $\theta_{i,a} = \beta_{t_{i,a}}\theta_i$. To interpret this, if a worker has t years of experience at age a and works during a period, in the next period his unobserved ability will have increased by the scaling factor β_{t+1}, while if he does not work in the next period he has $\theta_{i,a+1} = \theta_{i,a}$. If β_t is increasing in t, the effective levels of unobserved ability will fan out over time across workers, with higher unobserved ability workers seeing the fastest skill growth. Another interpretation is that over time task that workers perform change over time, even within occupations. If β_t increases on average, it means that the output or signal in the tasks depend more on the unobserved ability, if it decreases, it means that the signals from the tasks performs either depends less on the ability. If over time for example individuals work in larger teams the signals may be less accurate etc.

After each period, a worker who was employed sees a signal with information on their unobserved skills:

$$\tilde{z}_{ija} = \beta_{t_{i,a}} \theta_i + \varepsilon_{ia} \tag{1}$$

where ε_{ia} is drawn $N\left(0, \delta^2\right)$ and the worker only sees the realization of $\tilde{z}_{ija}$ but not its individual components. Non-employed workers see no signal and do not update their beliefs. Since β_t is a known set of parameters, we can rewrite the signal as

$$z_{ija} \equiv \frac{\tilde{z}_{ija}}{\beta_{t_{i,a}}} = \theta_i + \frac{\varepsilon_{ia}}{\beta_{t_{i,a}}} \tag{2}$$

The variance of the observed signal is $\delta^2_{i,t(a)} = \dfrac{\delta^2}{\beta^2_{t_{i,a}}}$. If β_t is increasing in t, then the signals become more informative with more experience, and within a period, a larger β_t implies faster learning. In estimation allow β_t to depend on race, but for the simplicity here we omit the race and the individual subscripts (i,r) and the age subscript in t.

We can write the expectation and variance of θ_i for individual with actual experience t as

$$b_{it} = E(\theta_i | z_{ij0}, .., z_{ijt}) = b_{i,t-1} + \frac{\delta_t^{-2}}{\sigma_{t-1}^{-2} + \delta_t^{-2}} (z_{ijt} - b_{i,t-1}). \tag{3}$$

Where the precision is given by:

$$\sigma_t^{-2} = \sigma_0^{-2} + \sum_{t'=1}^{t} \delta_{t'}^{-2} = \sigma_0^{-2} + \delta^{-2} \sum_{t'=1}^{t} \beta_{t'}^2 \tag{4}$$

In general if δ_t is constant over time, the deviation of the signal from the beliefs receive lower weight as experience increases, thus updating of beliefs decreases over time. However, in our model the variance of the noise changes, thus the weights can also increase if β_t increases sufficiently. Note that the signal to noise ratio is the same across occupations, but we allow it to vary by race to capture potential sources of discrimination. Learning is perfectly correlated across occupations as the unknown skill is general, and the speed of learning is the same across occupations (except for non-employment). In addition, if employers discriminate against minority workers by providing less training or assigning them to tasks in which they accumulate less human capital, it will imply slower learning. That is, lower δ and larger β_t imply higher precision. Note that that the precision is larger at age a the more experience a worker has.

The beliefs have a normal distribution. Assuming rational expectations, $b_{it} \sim N(b_{i0}, \dfrac{\sigma_t^4}{\sigma_t^2 + \delta_t^2})$. Notice that the dispersion of beliefs is smaller when the variance of noise of the signal δ_t^2 is smaller, and is larger when the precision is smaller. Therefore, if there is smaller. Intuitively, as the noise variance becomes arbitrarily large, there is no learning, and the beliefs don't change at all. Note that in contrast to standard learning models in which the noise variance is constant over time (the precision increases with experience), the variance might increase if δ_t^2 declines sufficiently (that is β_t increases sufficiently).

Wages are the worker's expected productivity above given the information at the beginning of the period. Occupations differ in the returns to the types of known and unknown skills. Worker i's wage at age a with known pre-market skills x_{i0}, years of actual (working) experience $t_{i,a}$, expected value of the unknown skill of b_{ia}, and in occupation j is given by

$$W(x_{i0}, exp_{ia}, b_{ia}, j) = a_{j0} + a_{j1}(x'_{i0}\gamma_1) + a_{j2}(exp_{ia}\gamma_2) + a_{j3}(\beta_{ta}b_{ia}) \tag{5}$$

To reduce the number of parameters we aggregate different components of the initial skill x_{i0} using a single-index form with weight vector γ_1. The variation in returns to skills is captured by the parameters a_{j0} through a_{j10}. Let exp_{ia} denote a vector of experience in each occupation. Similarly we create a composite experience index that aggregates the different occupation-specific experience into one index, allowing the weights on experience in each occupation to differ. Denote by γ_2 the vector of weights on each of occupation specific experience.

Within the framework of our model, the only direct race-specific coefficient is in the β_t which affect the expected output and its growth over time. Differences in initial means β_0 can reflect discrimination but can also reflect differences in average unobserved skill distributions and we cannot distinguish these two interpretations. For periods $t > 1$ differences in deterministic growth of the skill can reflect unobserved differences in tasks black and white workers perform within occupation (as we only have 10 categories) or different training provides for black and white workers. In addition, the variance of the wage component that depends on the beliefs may increase or decrease over time depending on the evolution of β as well as the variance of the beliefs and can create race differences in the dispersion of the beliefs.

Worker make choices to maximize expected lifetime utility. In a given period, the workers state variables are their stocks of known skills, level of actual experience, expected value and variance of the unknown skill, and their previous occupational choice. We denote the entire state vector of worker i at age a as $S_{ia} = \{x_{i0}, x_{ia}, b_{ia}, \sigma^2_{i,a}, I_{i,a-1}\}$.

The utility function is separable in consumption and a non-pecuniary components, but not across time as the non-pecuniary utilities depend on current and past labor force participation and occupation choices. As in standard discrete choice models, each period there is an extreme value type-1 preference shock, independent across individuals, occupations, and time, that is associated with each occupation η_{ija}. There is no borrowing or savings, so consumption equals income each period, and per-period utility of choice j given state vector S_{ia}, wages W, non-pecuniary flow utility function $u_{n,j}$, and shocks η_{ija} is given by

$$u_j(S_{ia}, \eta_{ija}) = \alpha W_{ija} + u_{n,j}(S_{ia}) + \eta_{ija} \tag{6}$$

We allow the non-pecuniary flow utility function to depend on race and education. This term will captures cost of entry into employment and the different occupations, and these costs vary by race and education. Any other frictions not included in the model that make it more costly for certain groups to get a job in different occupations will be captured in these non-pecuniary costs/benefits.

Define the optimal valuation function by

$$V(S_{ia}) \equiv E_{\eta_{ia}} \mathrm{MAX}_j \{u_j(S_{ia}, \eta_{iaj}) + \beta E[V(S_{ia+1})|S_{ia}, j]\} \tag{7}$$

The valuation function demonstrates the dynamic tradeoffs in the worker's problem. First, the current utility depends on the the current pay in the occupation chosen. Even without non-pecuniary benefit considerations, the occupational choice that maximizes expected lifetime pay may not be the one that maximizes current pay. This is because of the option value: if the variance of the beliefs $\sigma^2_{i,a}$ is large relative to the noise δ^2_a, faster learning allows for more efficient job assignment and therefore increases expected lifetime earnings. This feature is standard in learning models (see Miller (1984)). If there are racial differences in the speed of learning, it can create lifetime income gaps due to less efficient matching or "mismatch" of one race.

Second, if there are other frictions in addition to the informational frictions, such as discrimination in hiring or lower offer rates for black workers in certain occupations, it will be captured as differences in non-pecuniary benefits in the utility from choosing certain occupations or transitioning from non-employment to employment. This will lead to some workers choosing lower-wage occupations or unemployment in some periods even if that hurts their long-run earnings. In the model, this behavior will be treated as individually rational even if it leads to lower wages, but we cannot distinguish between non-wage-maximizing behavior that is driven by differences in preferences versus differences in opportunities.

3 Data and Descriptive Statistics

Sample Selection and Variables Used

Our sample contains yearly panel data on U.S. birth cohorts from 1957-1964 from the National Longitudinal Survey of Youth 1979 (NLSY79). In particular, our base sample uses the 3,177 white males from the cross-sectional sample, while we draw 1,451 black males from both the cross-sectional sample and an oversample of minorities. From the NLSY79, we use the labor market histories on wages, hours worked, "three-digit" occupation (the most disaggregated category from the 1970 Census coding system), completed years of education, as well as the youths' ASVAB (Armed Services Vocational Aptitude Battery) and Rotter Locus of Control (hereafter Rotter) scores from 1980.

The ASVAB and Rotter pre-market test scores have been widely used in the racial wage gap literature at least since O'Neill (1990) and have been found to be pivotal to the size of the measured racial wage gap. About 90% of the NLSY79 sample took the ASVAB in 1980. The ASVAB is a battery of 10 tests on subjects such as mathematics, reading, and vocational knowledge such as electronics. We use the constructed aggregate AFQT (Armed Forces Qualification Test) score, a widely-used composite of reading and math scores, as our single dimensional measure of performance on the ASVAB battery. The Rotter test is meant to measure "individual differences in a generalized belief in internal-external control", where internal-external control is the concept of "whether [a] person perceives [a] reward as contingent on his own behavior or independent of it" (Rotter 1966). Our use of the Rotter score attempts to control for differences in attitudes and expectations that may differ across individuals but not be reflected in more standard academic test scores.

We restrict our sample to those born after 1960 so they do not take the tests during college, and we drop wage observations in the bottom 1% and top 99% as outliers, as well as individuals with missing test scores. We also cannot use data on any individual who does not have any valid wage observations. Our final sample consists of 891 black workers with an average of 8.2 valid wage/occupation observations, and 1,564 white workers with an average of 9.9 valid wage/occupation observations.

For labor force participation, we consider a worker as full-time if they usually work 30 or more hours per week at their main job. For the summary statistics and model, we must identify the first year workers are in the labor market (that is, either "employed" or "unemployed and looking for a job"). To do this, we consider a worker to enter the labor market whichever is earlier, either: 1. the first time we observe him as a full-time worker after the year corresponding to the last year of completed education, or 2. the third year after completion of schooling even if he has never had a full-time job. The concerns with using the first post-schooling period as the first working period are that, first of all, it can be difficult to conclusively identify when schooling is completed, and second, that many workers seem to take time off from the labor force on completions of schooling. Choosing a first period closer to the estimated end of schooling would miss fewer workers who are unemployed after graduation, but also consider more workers not in the labor force as unemployed. Ultimately, different rules for constructing the first labor market period do not change our analysis significantly, since we are primarily concerned with gaps at 10 to 20 years in the labor force.

Occupational Complex Task Construction

There are approximately 400 three-digit occupational codes in the 1970 Census coding system used in the NLSY79. To reduce the dimensionality of the occupation space, we create a unidimensional measure of the cognitive difficulty associated with the average job within each occupation. To do this, we follow Antonovics and Golan (2012) which uses data from the Dictionary of Occupational Titles, 4th Edition; it contains a large number of measurements of occupational characteristics. We run Principal Component Analysis on the high-dimensional set of occupational measurements and take the first principal component, which seems to correspond to tasks related to non-routine or complex tasks which are classified as occupations with higher requirements for skills that are hard to observe in Antonovics and Golan (2012). We then construct our measure, which we call α, of complex tasks required in the occupation by dividing it into 10 categories. By construction, then, $\alpha \in [0, 1]$, and higher α correspond to occupations estimated to perform more complex tasks tasks. We first category is all occupations with value of α between 0 and 0.1, the second is between 0.1-0.2,.. and the last category is between 0.9-1.

Labor Market Racial Gaps

Summary statistics for education, premarket test scores, wages, occupations, and work experience are shown in Table 1. The first two columns list the pooled cross-sectional averages, and the second two show the differences in standard deviations across white and black workers. As has been well documented, white workers enter the labor market with significantly higher measured skills: almost 1 additional year of education and one standard deviation higher AFQT scores. These pre-market differences are reflected in labor market outcomes: white workers earn an average of $4 per hour more than black workers, are 12 percentage points more likely to be full time workers, work in occupations almost one standard deviation higher in cognitive tasks, and correspondingly have more previous labor market experience in high cognitive task occupations.

Racial wage gaps in wages and labor force participation have been documented extensively before, see

Fryer (2011). One less-examined labor market outcome that differs between black and white workers as a consequence of the combination of wages and participation are total lifetime earnings. Wage differences must be considered conditional on working, so if there are labor force participation gaps, only considering wages understates the wealth differences between black and white workers. First we will consider the already-known gaps between white and black workers in wages, labor force participation, and occupational choice, and then we document that the differences in earnings – even controlling for observable pre-market skills – are large.

The simultaneous existence of both wage and labor force participation gaps can be quantified by considering gaps in career earnings by race. Estimated earnings gaps by race are shown in Table 3. To generate predicted earnings, we perform three steps: first, interpolate wage data for workers missing wages in a year using a fixed effects regression; second, aggregate annual earnings (wages times 40 hours per week, 50 weeks per year) over time within worker careers; and finally, regress total earnings in a year on a quadratic of years since labor market entry, interacted with race. The predicted career earnings by race at years 5 and 15 in the labor market are shown in the panel labeled "Unconditional" in Table 3 Consistent with the wage and labor force participation gaps, the predicted career earnings gap is \$36,000 dollars after 5 years in the labor market, as black workers make around \$8,000 less per year (\$4 per hour unconditional wage gap times 2000 annual hours) than white workers and will work about 0.4 fewer years over the first 5 years in the carer. By 15 years in the labor market, the estimated earnings gap is \$150,000 and black workers have 2.23 fewer accumulated work years than whites.

The second panel of Table 3 repeats the same analysis as the top panel, but predicts earnings as a function of AFQT, Rotter, and Education, all interacted with race, as well as the quadratic in time since labor market entry. We consider the predicted earnings if all workers had the white mean AFQT and Rotter scores and exactly 12 years of education. Closing the observed gap in pre-market skills certainly significantly reduces career earnings gaps. At 5 years in the labor market, the gap in predicted career earnings falls from \$26,000 dollars to \$5,000, a fall of 81%. On the other hand, at 15 years in the labor market including these observed pre-market skills closes a smaller percentage of the gap, 53%, and the remaining gap of \$55,000 is still over two years worth of earnings for the average full time worker. While equating observed pre-market skills closed over half the wage gap, it only closed a quarter of the gap in total years as a full-time worker, which mechanically means the amount pre-market skills closes the earnings gap must be less than the amount it closes the wage gap. For overall welfare, an earnings deficit of \$74,000 for black workers compared to whites is quite large. In our model and counterfactuals, changes in the career earnings gaps will be how we measure the joint effect of reducing the black/white wage gap as well as the black/white labor force participation gap.

Occupations and the Racial Gap

As shown in Table 1 The wage age profile as well as the occupation gap increases substantially. As discussed above, part of the reason that the accumulated earnings gaps increases is because of differences in employment and accumulation of labor market experience for black and white workers. To further analyze the role of the racial gaps in occupation in the accumulated earnings gaps Figure A describes the probability of being employed in the next period, conditional on currently being employed by occupation and race. As clear, from the figure, the probability of employment rises substantially with the occupational task complexity. The probability of employment conditional on occupation is lower for black workers, but it is smaller than the employment probability gaps between low and high task complexity occupations.

Furthermore, job transitions increases earnings on average, and the increase is larger when workers move to higher rank occupations (see Antonovics and Golan 2012). Since job transitions are associated with wage growth over the life-cycle we document the average occupation complexity index change for workers who change occupations. With the exception of occupations with an index above 0.8, black workers transition into lower complexity occupations.

Lastly, we denote the importance of the occupation in the first years of employment. As noted by Antonovics and Golan (2012) the initial occupation assignment has larger predictive power of wage growth over time than education and initial wages. Table 4 presents regression results by race of wages on experience, AFQT, Education, initial occupation index (α_0) and the interaction of experience with AFQT, Education and α_0. Figure 5 depicts the predicted wage growth by initial occupation (in the regression we use the actual level of the occupation index however, in the figure we aggregate it to 5 occupation groups). Workers who started at the top quintile receive a premium of close to $1 over and above workers who start in the 3rd quintile (after controlling for education and AFQT scores difference). By age 40 the gap doubles. The effect is smaller for black individuals. The differences can be because occupational assignment masks unobserved skills. In addition, if the occupational transitions later on depend on the initial assignment the differences in wage growth can reflect occupation differences in returns occupation specific skills. Both explanations can affect the observed racial gap in these patterns and we further explore it in the estimation.

4 Empirical Implementation, Estimation, and Identification

4.1 Empirical Implementation

We use the following functional forms in the estimation. For expositional clarity of the estimation and identification section we denote in the state variables S_{ia} separately from the beliefs $b_{i,a}$ since the latter is unobserved by the econometrician.

Preferences First, since we selected workers who work full time, we do not focus on intensive margin of labor supply and we estimate an hourly wage equation. In the utility from consumption we therefore use $u_1(c_{ia}, j) = \alpha w(S_{ia}, b_{ia}, j)$, with α being the weight of wages relative to non-pecuniary forces in utility. The non-pecuniary component of the utility is given by

$$
\begin{aligned}
u_2(S_{ia}, j) &= \kappa_0 * age * (1 - I_{0a}) + \kappa_1 * (1 - I_{0a}) * I_{0,a-1} + \kappa_2 * (1 - I_{0a}) * (I_{i0jt}) * \sum_{j' \neq j} (1 - I_{ij',t-1}) \\
&\quad + \theta_{j0} + \theta_{j1}\mathrm{HGC}_i + \theta_{j2}\mathrm{Black}_i \qquad \text{for } j \in \{1, .., 10\} \\
u_2(S_{ia}, 0) &= 0 \qquad \text{for } j = 0
\end{aligned}
$$

The utility from non-employment is normalized to zero. The κ's in the model represent general entry costs, which are not occupation specific. The costs of working κ_0 depend on age and may include a trend in labor market experience, for example κ_0 is the relative utility from participation as a function of years of in the labor market.[7] The term κ_1 is the cost of moving from non-employment to employment and κ_2 is the cost of switching occupations. It capture the additional costs to an individual that work in occupation j (denote current experience by t) and worked in a different occupation j' in the last year he was employed (marked by experience t-1) relative to an individual who worked in the same occupation.

The parameters θ_{j0} are the non-pecuniary benefit/cost from working in occupation j, an occupation specific components of utility; We allow for education and race to affect the non-pecuniary utility from the choice of the occupation. The parameters θ_{j1} and θ_{j2} capture these differences.

Skills and Wages Second, the productivity of that skill varies by occupation as well. It is therefore, part of the composite skill $x'_{it}\gamma$ defined in equation 5. We allow the distribution of the unobserved skill to vary by race, both the variance and the means. That is, $E(b_{i0})$ and β_t as well as σ_0^2 vary by race.

Third, we assume the wage are observed to the econometrician with an iid measurement error. This

[7] We do not include time effect in addition to age effect due to the difficulty in identification of year and cohort effects.

assumption has implication on identification as well as estimation.[8] Denote the wage realization for individual i by w_{ija}. The measurement error is given by: $\zeta_{ijt} \equiv w_{ija} - W_{ja}$. It is assume to have a normal distribution $\zeta_{ijt} \sim N(0, \sigma^2_{\zeta,j})$. We allow the variance of the measurement error to vary by occupations.[9]

4.2 Estimation

The estimation is conducted in two stages. Because observed wages are a function of unobserved beliefs, measured with error, the wage parameters and the skill distribution parameters are identified outside of the restrictions of the dynamic programming problem. This facilitates a two-stage estimation in the spirit of Hotz and Miller (1993) , in the first stage, the choice probabilities can be approximated by a flexible logit. However, these CCP's are a function of the individuals beliefs which we do not observe. (Arcidiacono and Miller, 2011, Sec. 6) develop a two stage CCP estimator to address the problem of unobserved heterogeneity. However, the unobserved heterogeneity is a discrete variable, while the beliefs in our model are continuous. We therefore use the flexible logit (written as a function of the unobserved beliefs) and write the likelihood of observing wages, employment and occupation choices simulaneously. These choice probabilities in the likelihood function are derived from the approximated valuation function and serve primarily as a selection correction device to produce unbiased estimates of the wage and skill distribution parameters. Integrating over the unobserved beliefs allows us to estimate the beliefs distribution and wage function parameters, but not the individual CCP's which depend on each individual beliefs.

Given the parameters from the first stage, in the second stage we estimate the structural utility parameters by solving the dynamic programming problem. Given the large dimensionality of the state space, we use the Hotz and Miller (1993) , representation of the value function using conditional choice probabilities (CCPs), which allows us to write the value in terms of a conditional choice probability and the next period conditional value function. Because the state variables in our model only depend on a finite number of lagged choices, our model posses finite state dependence. Thus we can apply the insights of Altug and Miller (1998) and Arcdiacono and Miller (2011) such that we only need to forward simulate the CCPs two periods ahead, after which the remaining component of the expected future value term differences out. In contrast to these papers, however, we did not estimate the individual CCP's in the first stage, only the distribution of beliefs by race. Therefore, we use all the estimated parameter from the first stage to estimate the models using simulated maximum likelihood. We can simulate data from the model, drawing beliefs from the estimated distribution

[8] For example, we cannot use the estimators of learning models in James (2011) and Hincapie (2017) which assume that wages are observed without measurement erros, since the beleifs can be estimated directly from the wage equation.

[9] This can capture differences in reporting error if contracts vary by occupations.

of beliefs, using this representation of the value function and the property of finite state dependence; this reduces the computational burden as we do not need to simulate all paths. This two-stage procedure is outline below.

4.2.1 First Stage

In the model, the state-space for individual i in time period t is defined by the vector S_{it}, b_{it}. Here we write explicitly the beliefs separately from the observed systematic state variables because they are unobserved and must be integrated out in estimation. Define the choice j conditional value function absent the i.i.d non-pecuniary shock η_{ijt} as

$$v_j(S_{ia}, b_{ia}) + \eta_{ija} = u_j(S_{ia}, b_{ia}, \eta_{ija}) + \beta \int_{b'} V(S_{a+1}, b_{a+1}|b_{ia}, S_a, j) f_j(b'|b, S_a) db' \tag{8}$$

Note that the only stochastic persistent state variable is the beliefs. The transition function of beliefs depend on j because no working implies that beliefs do not change. Thus $f_j(b'|b, S_a)$ is $\mathrm{b}_{it} \sim N(b_{i0}, \frac{\sigma_t^4}{\sigma_t^2 + \delta_t^2})$, and it can be written recursively in terms of initial variance and actual experience and the sequence of β_t's.

The conditional value function is a non-linear function of the observed state variables S_{it} and the beliefs which are not observed by the econometrician. In the first stage we estimate the parameters of the wage equation and the distribution of beliefs which is normal with mean $E(b_a)$ and variance given by equation described above. If we were to observe all the individuals working in each period then under rational expectation $E(b_a) = E(b_0) = E(\theta)$ which is the true mean of skill in the population. To correct for selection of workers participating as well as workers selecting into each occupation. Under the assumption of type 1 extreme-value distribution of the taste shocks, the CCP's are given by

$$P_j(S_{ia}, b_{ia}, \omega) = \left(\prod_{j=0}^{10} \exp([v_j(S_{ia}, b_{ia}))^{I_{iaj}}\right) / \left(1 + \sum_{j=0}^{10} \exp(v_j(S_{ia}, b_{ia}))\right) \tag{9}$$

where ω is the vector of the models parameters: wage equation and beliefs distributions and utility function parameters. We can write the choice probabilities as a function of the observed and unobserved state variables. If the beliefs were observed, we could have estimated the CCP's from the data, however, the beliefs are not observed to the econometrician. We therefore write the likelihood of observing a combination of wages and occupation choices. In addition, in this stage, we estimate the distribution of beliefs by race. We approximate the value function with a flexible polynomial function of the state variables $v_j(S_{ia}, b_a) \approx$

$[\Upsilon(S_{ia}, b_a))]' \omega_j$, where the function $\Upsilon(\cdot)$ produces a vector of the state variables, higher order terms, and their interactions and ω_j is the choice specific parameters governing the approximation. That is we can write the probability of the observed choices as

$$\tilde{P}_j(S_{ia}, b_{ia}, \omega) = \left(\prod_{j=0}^{10} \exp([\Upsilon(S_{i0})]'\omega_j)^{I_{iaj}} \right) / \left(1 + \sum_{j=0}^{10} \exp([\Upsilon(S_{ia}, b_a)]'\omega_j) \right) \tag{10}$$

Given A_i observations of choices for individual i, a history of choices is summarized by the sequence of choices and wages. The likelihood of the sequence of observed choices for individual can be written as the product of the choice probabilities. In addition, the wages are observed with a measurement error which we assume to be iid across individuals, choices and time and is normal with mean 0.

The likelihood of all observed wages unconditional on unobserved beliefs, $L_i(I_i, W_i|S_i)$ is given by

$$L_i = \int_{b_0 \ldots b_A} \prod_{a=1}^{A_i} \tilde{P}_{a,j}(S_{ia}, b_{ia}) f_j(S_{a+1}, b_{a+1}|S_a, b_a) \frac{1}{\sigma_{j\zeta}} \phi(\frac{w_{ija} - W_{ja}}{\sigma_{j\zeta}}) db_0 ... db_{A_i} d\zeta_1 .. \zeta_{A_i} \tag{11}$$

where $\phi(\cdot)$ is the probability distribution function of a standard normal, standardized measurement error. The first stage log-likelihood function is given by

$$ll_{\text{first-stage}} = \sum_{i=1}^{N} \log L_i(I_n, W_n|\omega, S_n)$$

The estimates from this stage are the wage equation parameters and variance of the measurement error: $\mathbf{a}, \gamma, \sigma^2_\zeta$. In addition we estimate the expected values of the beliefs and covariances of beliefs across periods. Note that the variance evolves with actual experience only. The variance-covariance matrix of beliefs Σ_r has diagonal of $Var(b_{it}(r))$ and the off diagonal elements $Cov(b_{it}(r), b_{i,t-k})$. In addition, we estimate the parameters ω of the value function approximation from the CCP's. However, these parameters are not used anywhere in the estimation as the only role of them is to correct for selection.

4.2.2 Identification of Beliefs (incomplete)

The estimation above does not require us to impose the updating formulas that govern the evolution of the beliefs. The Bayesian learning process implies that the beliefs have a normal distribution and restrictions on the Variance-Covariance matrix. In the CCP's the individual state variables are the beliefs each period, and the entire job choice and experience history. Given initial distributions the variances are a deterministic function of the experience. Moreover, there is a unique mapping between the learning parameters and the Var-Cov matrix Σ_r; thus these parameters are sufficient for approximation of the individual CCP's as they allow individuals to form beliefs about future wages that are consistent with the learning model. To this extend, the CCP's and the wage and beliefs distributions are identified requiring the assumption of the normal distributions of wage residuals and measurement error.

We used in the first stage the following restrictions from the learning model:

1. The beliefs are only updated (both mean and variance) when individuals work.

2. The mean of the beliefs $E(b_{ia})$ of the *wage offer* distribution remains constant over time. That is, given rational expectations, and assuming that the beliefs are consistent with the underlying ability distribution, the mean of the beliefs over the entire population remain constant over time. However, due to selection into the market the mean of the beliefs of people who work are different.

3. The wage residuals contain the expression $\lambda_j * E(b_{ia}) * \beta_{t,r} + \zeta_{ijt}$. For white workers, we normalized $\beta_{t,w} = 0, \forall t$. This is because the returns to experience are occupation specific in the wage equation, and in general, the average increase in productivity is not identified separately. We can identify the differences in the increase in average productivity with experience for black workers because the race coefficient is excluded from the wage equation and only appears in the beliefs. Therefore, we identify the differences in change over time in the unobserved expected productivity, $\Delta\beta_{t,r}$.

4. The CCP's are approximated around the individual beliefs b_{ia}. While the variance evolves deterministically and depends only actual experience (and race), for each individual the beliefs depend on the unobserved sequence of signal realizations: $z_{i0}, .., z_{iA}$. We can write the beliefs in Equation 3 as a weighted sum if the initial beliefs, the average value of signals and experience

5. Given the estimates of the wage equation, mean beliefs, and the variance covariance matrix of beliefs we can estimate the variance parameters $\delta_0, \delta, \sigma, .$ We have over identifying restrictions. We can form moments using our estimates of Σ to recover $\Delta\beta_{t,r}, \delta, \sigma_0^2$.

6. The wage equation returns to skills, observed and unobserved in the different occupations are identified up to a normalization. We set the returns to all skill in occupations 6 to be 1.

4.3 Second Stage

In the second stage we need to directly construct the choice probabilities based on the dynamic optimization problem to facilitate estimation of the parameters of the utility function Θ. Under the assumption that the contemporaneous shock to utility, η is distributed iid type-I extreme value, the choice probability implied by the structural model is

$$\Pr(I_{ijt}|S_{it},\alpha,\theta,\kappa) = \frac{\exp(v_j(S_{it},\Theta))}{\sum_{j'=0}^{3}\exp(v_j(S_{it},\Theta))} \tag{12}$$

Where the value function v_j is defined in Equation 8 The expectations of the conditional value function are taken with respect to next periods random utility shock, η and next periods state variables $(S_{it+1}, b_{i,t+1})$, conditional on the state variables in period t and the choice in period period t. If the individual chooses employment in one of the j occupations they will receive a noisy signal of their unknown skill which also factors in to their future beliefs. Given the state variables S_{it} this random variable $b_{ijt} \sim N(b_{it}, \sigma_{it}^2)$. Importantly if the worker chooses non-employment in period t they will not receive a signal of their unknown skill.

Given the size of the state space in our model, it is not feasible to solve the future value function analytically. However, in constructing the choice probability, evaluating this value function may not be necessary if when differencing the value functions, $v_j - v_0$, this unknown term cancels out. This is the main insight in Altug and Miller (1998) who show that in models possessing finite state dependence the conditional value functions can be written in an analytically equivalent way such when the conditional value function are differenced in estimation the unknown expected future value terms cancels out, circumventing the need to calculate these very difficult expressions. James (2011) applies this concept to occupational choice and learning models. James (2011) shows that if no learning occurs during non-employment, then these models satisfy the condition of finite state dependence and can be tractably estimated despite their large state space. The estimation is done by drawing individuals beliefs from the estimated beliefs distributions in each period and using this distributions to form expectations and solving the dynamic problem along the relevant paths. Appendix B describes the details of the estimation.

5 Estimation Results

5.1 Wage Equation and Skills

Wages Table 5 presents the wage equation estimates. We allow the returns to all skills, observed and unobserved, to differ by occupation. We normalize the returns to all the skills to be 1 in occupation 6. As in equation 5 the estimates of a_j are presented. Table 6 represents the wage equation coefficients on the composite skill. The returns to observed pre-market skills is not entirely monotonic in occupations, but is highest in occupation 10 (1.38) and is the lowest in occupation 1 (0.98) and 3 (0.95) [10] The occupation specific return to the unobserved skill, however, is higher overall in occupations with higher α.[11] The returns are below 1 for occupations 1-6 and above 1 in occupations 8-10. We allow for the measurement error to vary by occupations in order to account for systematic differences in contract forms in different occupations that might result in differences in the variance of errors in reporting. While the variances differ, there is no systematic patterns of change by our measure of occupational ranking.

The experience variable was constructed to include the occupation-specific experience. Table 6 describes it. The low-Alpha experience is a weighted average of specific experience in occupations 1-5 and the high-Alpha occupation is a weighted average of experience in occupations 1-6. The estimated weights are reported in the Table. The returns to occupation experience in the composite experience skills is increasing in occupational ranking (with the exception of occupation 3 in the low skill experience.) The weights of occupation 9 and 10 are almost double the weight of occupation 8. Therefore, although in the wage equation the occupation-specific returns on the low-occupations-skills the returns to the high-skill-occupation are larger. For example the return in the composite skill on experience in occupation 9 is 1.467 and the returns to occupation 5 in the low skill composite experience is 0.53. The occupational returns to low-skill experience in occupation 5 is 1.22 and the returns to high skill experience in occupation 9 is 0.84. Therefore, an additional year in occupation 5 has a return of 0.64 in that occupation; the returns to a year in occupation 9 is 1.23 in that occupation. With the exception of occupation 3, the returns to ”low alpha” occupational experience is higher in the lower alpha occupations than in the high-alpha occupations. The returns to the ”high alpha” experience are high in all occupations. This suggested that individuals with ”high-alpha” occupational experience will have high returns for their experience even if they move to an occupation with lower alpha. This indicates that occupation specific skills play an important role in wages, which suggest that if there is state dependence in job assignment , then the first occupation can predict wage increase over and above the”pre-market” skills that we find in Table 3. We further explore the role of switching costs and the state dependence source in our counterfactual simulations.

5.2 Beliefs and Learning

Table 7 and Figure 6 presents the mean and variances and distributions of the beliefs by race and actual experience. The estimates are of the yea by year distribution. We did not impose in estimation the updating

[10] The non-monotonicity may be due to the bundling of the different skills.

[11] This is consistent with the construction of the occupation index in Anotonovics and Golan 2012. It was constructed by classifying the skills that are hard to observe or learn about quickly.

as described by equation 3 and 4. However, the estimates are of the distribution: $b_{it} \sim N(b_{i0}, \frac{\sigma_t^4}{\sigma_t^2 + \delta_t^2})$ which maps directly into the fundamental parameters in the above equations. Therefore, we can identify the fundamental parameters of $\sigma_{0,r}^2$ and δ_t^2, using moments from the estimated VAR-COV matrix. Note that these estimates confirm that the variance of the noise δ_t varies over time. The patterns of the variance are not consistent with a standard learning models in which the variance is constsistant ($\delta_t = \delta, \forall t$).

The top panel in Table 7 describes the race gap in the mean beliefs. It shows that the initial beliefs are small initially, but between the first two years and years 8-10 the gap is more than 3 time larger. The gap initially may reflect gap in skills (although we cannot distinguish between skill gap and employers discrimination), but it amplifies with actual experience. Once interpretation is that black workers experience smaller increase in on the job training or unobserved skill accumulation. Another possibility is that they are assigned to task that depend less on the unobserved skill. Both interpretations might be consistent with discrimination.

The variance of beliefs is similar at the first year of employment but increases for white workers more than it does for black workers. Between year 2 and 9 the variance doubles for white workers and is about 1.3 times larger for black workers. and by 42% for white workers These findings are consistent with Phelps notion of discrimination in the sense that employers find it more difficult to interpret signals of minorities; as a result, given a signal, the weight it receives is smaller for minorities. This means that black workers with positive signals receive less reward, but that black workers with low signals are penalized less than white workers with the same signals.

We will further explore the quantitative impact the beliefs distribution on the observed sorting of blacks and white and the pay gap when we perform counterfactual analysis.

5.3 Utility function

Table 8 presents the estimates of the utility function. The utility is linear and separable in consumption and the non-pecuniary benefits/costs of working and occupations choices. The value of not working is normalized to zero. The estimates of the non-pecuniary costs of working in each occupation are substantially larger for the "high α" occupations. The cost at the sixth highest alpha is more than twice the cost at the fifth alpha. The cost at the two highest ranked occupations is about 2.5 times the cost at the third highest ranked occupation. They can represent taste but also entry cost of working in the complex tasks occupations. That is if there are education or skill barriers to entry and education requirement for employment in these occupations, we are not modeling it, and therefore for the lowest education groups the intercept reflect the very low employment rate of low education individuals in the occupation.

Indeed, the non-pecuniary costs of working in these occupations vary substantially by education group. For example with 16 years of education the non-pecuniary costs in the third highest occupation ($\alpha = 8$) is the lowest. With 18 years the non-pecuniary benefit is highest in the highest rank occupation. With 10 years of education the highest non-pecuniary benefit is at the second lowest occupation. Black workers have substantially larger costs of working in the top 5 ranked occupations, and no significant higher costs in the bottom 3. For example, while a white worker with 16 years of education have substantially lower costs of working in occupation 8 than in 5, the non-pecuniary cost for a black worker with 16 years of education is about the same in these two occupations, reflecting the larger non-pecuniary costs for blacks to enter in the third higher occupation.

In addition there are costs to entering into work from non-employment (they are about 65% larger than the non-pecuniary costs of working in the $\alpha = 6, -8$ occupations) and there is a switching costs of occupations. The switching costs are 2.5 times larger than costs of entry from unemployment. This is consistent with the patterns that initial high α occupation choice predicts large wage increases because of large switching costs. These costs may reflect higher costs of recieving offers in different occupations or frictions that we do not model.

6 Decomposing the Black-White Career Gaps

The model provides two channels for blacks and whites to have differences in labor market outcomes conditional on the pre-market observed skills: The first of the differences in utility function parameters, in particular preferences over occupations. As discussed above the higher non-pecuniary costs of working in high-α occupations can mask discrimination or low hiring rates of black workers, as well as taste differences. The second is differences in beliefs, which include the mean differences and the variance differences. In order to quantify the relationships of these parameters to differences in career earnings gaps, we perform a number of counterfactual simulations. A full list of the counterfactuals and the impact of these parameter changes is seen in the lower panel of Table 9 and Figures 7-8. The first two columns in Table 9 report the levels in the data and the model's prediction. The entries in the table report the percentage of the racial gap that is closed in each counterfactual. The levels of the variables for black and white workers are plotted in the figures.

The first set of counterfactuals (column 3 in Table 9) quantifies the role of the pre-market observable skills, that is the Rotter, AFQT and Education, in the observed gap. We simulate the model assigning black workers the pre-market skills of white workers and report the impact on the wage, total earnings, actual experience and α gap. Since the occupation gap has no obvious scale we report the median α gap in each category, which is between $0 - 1$ and compute the mean by race.[12]

The second counterfactual equalized the initial mean beliefs for white and black workers. The results are reported in Column 4. To the extent that these differences reflect mean unobserved skill differentials, over and above AFQT and education, the counterfactual quantifies its importance. In Column 5 we equalizes the mean and variances of the beliefs of black workers to that of white workers. Column 6 reports the counterfactual results in which all the black coefficients are set to zero in the utility, thus there is no difference in preferences of white and black workers. Lastly, the final counterfactual equates the pre-market-skills and the initial mean beliefs coefficients of black and white worker. The results are reported in column 7.

Turning to the first counterfactual, the differences in pre-determined skills account for 36% of the total experience gap after 5 years and 37% after 15; they account for 28% and 24% of the occupation gap 5 an 15 years later (respectively). Thus the majority of the employment and occupation gaps are not explained by the AFQT and education differences. They explain 47% of the wage gap after 5 years of experience but explains much less (35%) of the gap of workers with 15 years of experience. Overall, these factors explain about half of the gap in total earning after 5 years and only 37% of the gap after 15 years.

The differences in initial mean of unobserved skill differences explain 21% and 14% of the earnings gaps after 5 and 15 years, respectively. Equalizing all the beliefs reduces the gap by 27%.

In order to compare our model predictions to the literature that quantifies initial skills, observed and unobserved, we conduct a counterfactual that equates both the observed pre-market skills and unobserved initial means (columns 7 of Table 9). This counterfactual explains about 40% of the experience gap and between $29 - 25\%$ of the occupation gap.

Overall, it explains 70% of the total pay gap after 5 years but only half of the wage and total earnings gap after 15 years. Sorting over time into jobs in which workers have comparative advantage reduces the effect of the differences in the initial means of unobserved ability and observed skills in our model. However, the dynamics of sorting into jobs and accumulating experience is affected by the variances in beliefs and the aggregate increase in the gap of the mean beliefs over time (these are the "wage residual" gaps) may reflects discrimination; furthermore, the preferences differences in entry into working in the different occupation may also partly reflect discrimination and affect the widening of the racial pay gaps.

[12] we divided α the occupations by the levels of the α variabe. $\alpha = 1$ contains occupations with $\alpha \leq 0.1$,, the top category $\alpha = 10$ contains occupaitons with $\alpha > 0.9$.

In order to analyze what affects the remaining gaps (about half of the race gap earnings after 15 years is and 60% and $70-75\%$ of the experience and occupation gaps), Counterfactual 4 sets all the black-specific coefficients in the utility to zero and therefore equates the preferences of black and white workers. The preference differences explains most of the experience and occupation gaps (73% and 75%), respectively. It explains 21% of the total earning gap after 5 years and 27% after 15 years. Thus, the impact of preferences increases over time as occupation-specific experience becomes more important in explaining the total earnings gap between white and black workers. These, preferences differences can reflect taste differences but it can also capture other friction in the labor market such as lower employment opportunities in these occupations which means black works are less likely to be employed or "promoted" into the high-skill occupations.

7 Conclusions

This paper analyzes the sources of the life-cycle pay, employment and occupation racial gaps. We document that white workers are assigned initially to occupations with more complex task requirements. Over time, workers are sorted into occupations with more non-routine and complex tasks. After controlling for the observable gaps, the degree of "task complexity" of initial occupation is correlated with larger rise in earnings (over and above the increase associated with education and AFQT). The accumulated earnings gap after 15 years is about 117K. Empirically, about 53% of this gap can be accounted for after controlling for observable pre-market characteristics (AFQT, and schooling).

To address the question of why the gaps increase over time and in order to account for dynamic selection, we develop a dynamic model of employment and occupation choice which nests discrimination. The model is a dynamic Roy model in which different occupations have different returns to the different skills. In particular, some skills are initially observed, but there is incomplete information about skill. Over time workers and employers receive signals and workers are sorted into occupations with a better match. The mean beliefs about workers unknown skills vary by race, reflecting potential skill gaps in addition to observed gaps. The variance of the noise to signal ratio varies by race as well; if employers receive less informative signals on minority groups (as in Phelps 1972), it can imply that minority workers will be sorted more slowly into matches and some of the gap can be due to a "mismatch" between workers and occupations. We also allow for the non-pecuniary value from an occupation to depend on race and education. This can reflect preference differences but also frictions differences if black workers are less likely to be hired or receive offers from employers in more complex-tasks occupations.

We then develop a two-step estimation method to recover the model's structural parameters, and use them to simulate counterfactual exercises that allows us to decompose the different factors affecting these gaps.

We find that the majority of the total experience and occupational experience gaps is explained by the differences in the preference parameters that capture costs of entry into the different occupations. These costs are substantially larger for blacks in skills with high requirement for complex-skills. About 27% of the total pay gap after 15 years is explained by this factor. While we cannot separately identify taste differences and discrimination in hiring or lack of network and low offer arrivals of high task complexity jobs to black workers, it is suggesting that such discrimination might play a role in the observed widening black-white pay gaps. In addition, equating the initial pre-market skills and the initial mean differences in beliefs explain about half of the total earnings after 15 years. This reflects the importance of sorting of workers into jobs over time in explaining the long-term earnings gaps, and the fact that it is not predicted by initial skill differences. Our estimates indicate that the variance of beliefs is increasing substantially more for white workers, reflecting possibly unobserved promotion gaps or training and increase in the initially unknown skills over time. While the pre-market skill gaps (observed and unobserved) between black and white workers are clearly an important factor in the labor market outcome gaps, other factors and frictions are important to understanding the life-cycle labor market outcome gaps.

References

V. Aguirregabiria and P. Mira. Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica*, 70(4):1519–1543, 2002.

J. G. Altonji and R. M. Blank. Race and gender in the labor market. *Handbook of labor economics*, 3: 3143–3259, 1999.

J. G. Altonji and C. R. Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, 2001.

S. Altug and R. A. Miller. The effect of work experience on female wages and labour supply. *The Review of Economic Studies*, 65(1):45–85, 1998.

H. Antecol and K. Bedard. The racial wage gap the importance of labor force attachment differences across black, mexican, and white men. *Journal of Human Resources*, 39(2):564–583, 2004.

K. Antonovics and L. Golan. Experimentation and job choice. *Journal of Labor Economics*, 30(2):333–366, 2012.

P. Arcidiacono and R. A. Miller. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867, 2011.

P. Arcidiacono, P. Bayer, F. A. Bugni, and J. James. Approximating high-dimensional dynamic models: Sieve value function iteration. In *Structural Econometric Models*, pages 45–95. Emerald Group Publishing Limited, 2013.

D. H. Autor, F. Levy, and R. J. Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333, 2003.

C. Brown. Black-white earnings ratios since the civil rights act of 1964: The importance of labor market dropouts. *The Quarterly Journal of Economics*, 99(1):31–44, 1984.

S. V. Cameron and J. J. Heckman. The dynamics of educational attainment for black, hispanic, and white males. *Journal of political Economy*, 109(3):455–499, 2001.

P. Carneiro, J. J. Heckman, and D. V. Masterov. Labor market discrimination and racial differences in premarket factors. *The Journal of Law and Economics*, 48(1):1–39, 2005.

A. Chandra. Labor-market dropouts and the racial wage gap: 1940-1990. *The American Economic Review: Papers and Proceedings*, 90(2):333–338, 2000.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Z. Eckstein and K. I. Wolpin. Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339, 1999a.

Z. Eckstein and K. I. Wolpin. Estimating the effect of racial discrimination on first job wage offers. *The Review of Economics and Statistics*, 81(3):384–392, 1999b.

R. G. Fryer. Racial inequality in the 21st century: The declining significance of discrimination. *Handbook of Labor Economics*, 4:855–971, 2011.

G.-L. Gayle and L. Golan. Estimating a dynamic adverse-selection model: Labour-force experience and the changing gender earnings gap 1968-1997. *Review of Economic Studies*, 79(1):227–267, 2012.

G.-L. Gayle, L. Golan, and M. Soytas. Family structure and intergenerational income mobility. *Unpublished manuscript, Washington University in St. Louis, Department of Economics*, 2014.

R. Gibbons, L. F. Katz, T. Lemieux, and D. Parent. Comparative advantage, learning, and sectoral wage determination. *Journal of labor economics*, 23(4):681–724, 2005.

A. Hincapie. Entrepreneurship over the life cycle. *Working Paper*, 2017.

V. J. Hotz and R. A. Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

V. J. Hotz, R. A. Miller, S. Sanders, and J. Smith. A simulation estimator for dynamic models of discrete choice. *The Review of Economic Studies*, 61(2):265–289, 1994.

J. James. Ability matching and occupational choice. *FRB of Cleveland Working Paper No. 11-25.*, 2011.

C. Juhn. Labor market dropouts and trends in the wages of black and white men. *ILR Review*, 56(4): 643–662, 2003.

K. Lang and M. Manove. Education and labor market discrimination. *The American Economic Review*, 101 (4):1467–1496, 2011.

R. A. Miller. Job matching and occupational choice. *Journal of Political Economy*, 92(6):1086–1120, 1984.

D. A. Neal and W. R. Johnson. The role of premarket factors in black-white wage differences. *Journal of political Economy*, 104(5):869–895, 1996.

G. S. Oettinger. Statistical discrimination and the early career evolution of the black-white wage gap. *Journal of Labor Economics*, 14(1):52–78, 1996.

J. O'Neill. The role of human capital in earnings differences between black and white men. *The Journal of Economic Perspectives*, 4(4):25–45, 1990.

E. Phelps. Statistical theory of racism and sexism. *American Economic Review*, 62(4):659–661, 1972.

J. Speer. Pre-market skills, occupational choice, and career progression. *Journal of Human Resources*, 52.1: 187–246, 2017.

B. Western and B. Pettit. Black-white wage inequality, employment rates, and incarceration. *American Journal of Sociology*, 111(2):553–578, 2005.

A Tables and Figures

Table 1: Summary Statistics

race	wage	alpha	year of schooling	experience	AFQT
black	8.89	0.36	12.6	8.02	22.56
white	12.59	0.50	13.2	8.33	43.34
Total	10.56	0.46	13.0	7.75	

Table 2: Age-Race Occupation Task Gaps

	White	Black
Age 20	Cashiers, Typist, Carpet Installers (0.33)	Carpenter's Helpers, Janitor (0.25)
Age 25	Proofreaders, Plumbers (0.48)	Cashiers, Typist, Carpet Installers (0.35)
Age 30	Jewelers, AC/Heating Repairs (0.54)	Receptionist, Bus Drivers (0.38)
Age 35	Bill and Account Collectors, Engine Mechanics (0.58)	Stenographers, Auto Body Repair (0.40)
Age 40	Dental Hygenist, Radio/TV Repair, (0.6)	Stenographers, Auto Body Repair (0.4)

Table 3: Earnings Gap

	5 Years in Labor Market		15 Years in Labor Market	
	Average Years Employed	Total Estimated Earnings	Average Years Employed	Total Estimated Earnings
Unconditional				
White	3.76	$96k	13.58	$382k
Black	3.27	$70k	11.82	$265k
Gap	0.49	$26k	1.76	$117k
Conditional on sample mean AFQT, Rotter, and 12 years of schooling				
White	3.70	$83k	13.51	$362k
Black	3.29	$78k	11.97	$307k
Conditional Gap	0.41	$5k	1.54	$55k
% of Unconditional Gap Closed	16%	81%	13%	53%

Table 4: Alphas by Race Regression Results

	(1)	(2)
dependent: wage	white	(black)
initial alpha	2.192***	1.333**
	(0.349)	(0.468)
AFQT score	0.519***	1.028***
	(0.113)	(0.172)
Education category 2	1.273***	-0.0186
	(0.276)	(0.300)
Education category 3	1.064**	0.925*
	(0.336)	(0.369)
Education category 4	2.357***	2.187***
	(0.381)	(0.482)
experience	0.243***	0.254***
	(0.0311)	(0.0404)
Education category 2 * experience	0.0251	0.0414
	(0.0320)	(0.0338)
Education category 3 * experience	0.192***	0.0985*
	(0.0384)	(0.0417)
Education category 4 * experience	0.443***	0.293***
	(0.0443)	(0.0572)
AFQT * exp	0.0584***	0.0921***
	(0.0124)	(0.0198)
initial alpha * exp	0.162***	0.0955
	(0.0391)	(0.0536)
constant	8.388***	9.081***
	(0.266)	(0.355)
	22282	8881

Standard errors in parentheses. Base education category is less than high school. Edu cat 2 is high school. Edu cat 3 is some college. Edu cat 4 is some college and above.

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

Table 5: Wage Coefficients

	Alpha 1	Alpha 2	Alpha 3	Alpha 4	Alpha 5	Alpha 6	Alpha 7	Alpha 8	Alpha 9	Alpha 10
Intercept	10.70	11.75	14.00	10.35	11.67	11.81	11.84	12.45	13.49	12.80
	(0.12)	(0.29)	(0.47)	(0.05)	(0.13)	(0.20)	(0.29)	(0.61)	(0.84)	(0.56)
Loadings										
Pre-market	0.98	1.23	0.95	1.01	1.12	1.00	1.07	1.20	1.05	1.38
	(0.06)	(0.02)	(0.19)	(0.03)	(0.12)	(0.00)	(0.08)	(0.15)	(0.23)	(0.25)
Low Alpha Experience	1.53	1.43	0.33	1.81	1.22	1.00	1.25	0.93	1.38	1.04
	(0.08)	(0.15)	(0.07)	(0.09)	(0.13)	(0.00)	(0.14)	(0.18)	(0.43)	(0.46)
High Alpha Experience	0.99	0.92	0.82	0.94	0.87	1.00	0.89	0.87	0.84	0.86
	(0.09)	(0.18)	(0.03)	(0.12)	(0.13)	(0.00)	(0.06)	(0.10)	(0.06)	(0.09)
Innate Ability	0.81	0.88	0.88	0.88	0.89	1.00	0.93	1.01	1.06	1.04
	(0.04)	(0.05)	(0.03)	(0.06)	(0.01)	(0.00)	(0.02)	(0.04)	(0.01)	(0.06)
Measurement Error Variance	0.02	3.03	1.31	1.28	0.75	1.33	1.00	6.32	2.11	1.89
	(0.00)	(2.04)	(0.92)	(1.11)	(0.45)	(1.74)	(0.21)	(1.01)	(0.17)	(3.47)

Table 6: Skill Coefficients

Pre-Market Variables		Low-Alpha Experience Variables		High-Alpha Experience Variables	
HGC	-0.094	xpr_1	0.271	xpr_6	0.708
	(0.008)		(0.033)		(0.121)
HGC<12	-0.715	xpr_2	0.337	xpr_7	0.766
	(0.195)		(0.030)		(0.136)
HGC>12	0.788	xpr_3	0.761	xpr_8	0.793
	(0.029)		(0.079)		(0.058)
HGC≥16	1.538	xpr_4	0.265	xpr_9	1.467
	(0.377)		(0.076)		(0.184)
AFQT	0.020	xpr_5	0.530	xpr_{10}	1.358
	(0.003)		(0.062)		(0.139)
Rotter	-0.177	xpr_1 squared	0.001	xpr_6 squared	-0.018
	(0.011)		(0.006)		(0.026)
		xpr_2 squared	-0.001	xpr_7 squared	-0.027
			(0.004)		(0.004)
		xpr_3 squared	-0.033	xpr_8 squared	-0.007
			(0.009)		(0.017)
		xpr_4 squared	0.004	xpr_9 squared	-0.016
			(0.013)		(0.046)
		xpr_5 squared	-0.011	xpr_{10} squared	-0.032
			(0.006)		(0.009)
		$(\sum_{j=1}^{5} \text{xpr}_j)^2$	-0.015	$(\sum_{j=6}^{10} \text{xpr}_j)^2$	-0.002
			(0.005)		(0.009)

Table 7: Beliefs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Belief Means (white normalized to zero)																
black / diff.	-0.63	-0.47	-0.86	-0.83	-0.52	-1.09	-1.06	-1.74	-1.82	-1.67	-1.92	-1.96	-1.75	-2.25	-1.51	-0.83
	(0.23)	(0.22)	(0.22)	(0.15)	(0.08)	(0.08)	(0.07)	(0.16)	(0.14)	(0.34)	(0.27)	(0.40)	(0.66)	(0.33)	(0.34)	(0.65)
Belief Standard Deviation																
white	4.74	5.21	5.85	5.95	6.11	6.71	7.06	6.66	7.47	7.36	6.96	7.28	7.36	7.65	8.08	8.08
	(0.05)	(0.11)	(0.12)	(0.29)	(0.34)	(0.37)	(0.21)	(0.16)	(0.26)	(0.26)	(0.07)	(0.03)	(0.26)	(0.19)	(0.12)	(0.58)
black	4.61	4.61	4.37	4.80	5.33	4.96	5.32	5.08	5.53	5.74	5.35	5.77	6.46	7.16	6.68	6.90
	(0.34)	(0.24)	(0.13)	(0.29)	(0.23)	(0.28)	(0.35)	(0.09)	(0.20)	(0.31)	(0.40)	(0.46)	(0.37)	(0.75)	(0.47)	(0.32)
diff.	0.12	0.60	1.47	1.15	0.78	1.75	1.73	1.58	1.94	1.62	1.61	1.51	0.90	0.50	1.40	1.18
	(0.38)	(0.29)	(0.21)	(0.56)	(0.10)	(0.35)	(0.27)	(0.16)	(0.09)	(0.13)	(0.35)	(0.49)	(0.20)	(0.91)	(0.46)	(0.69)
Belief Correlation (lower white / upper black)																
1	1	0.47	0.42	0.31	0.30	0.28	0.23	0.20	0.23	0.18	0.18	0.24	0.16	0.04	0.20	0.01
2	0.60	1	0.66	0.54	0.48	0.46	0.44	0.37	0.37	0.31	0.32	0.32	0.30	0.25	0.27	0.14
3	0.42	0.60	1	0.69	0.53	0.60	0.53	0.43	0.40	0.43	0.36	0.38	0.30	0.26	0.35	0.18
4	0.44	0.58	0.69	1	0.66	0.67	0.56	0.52	0.43	0.47	0.40	0.42	0.34	0.31	0.38	0.23
5	0.37	0.48	0.59	0.68	1	0.69	0.57	0.52	0.47	0.37	0.41	0.33	0.36	0.41	0.36	0.18
6	0.37	0.43	0.55	0.64	0.73	1	0.76	0.65	0.59	0.50	0.53	0.49	0.45	0.44	0.51	0.30
7	0.30	0.39	0.48	0.61	0.65	0.75	1	0.78	0.74	0.60	0.58	0.58	0.49	0.50	0.52	0.48
8	0.32	0.44	0.49	0.56	0.64	0.68	0.72	1	0.77	0.69	0.62	0.64	0.55	0.57	0.66	0.53
9	0.31	0.38	0.43	0.46	0.52	0.57	0.65	0.72	1	0.69	0.68	0.69	0.60	0.54	0.65	0.51
10	0.29	0.34	0.38	0.45	0.50	0.54	0.56	0.68	0.79	1	0.65	0.67	0.63	0.60	0.58	0.70
11	0.29	0.37	0.37	0.43	0.47	0.50	0.54	0.65	0.72	0.81	1	0.80	0.73	0.70	0.77	0.64
12	0.29	0.31	0.33	0.39	0.43	0.44	0.47	0.61	0.66	0.75	0.85	1	0.79	0.62	0.80	0.59
13	0.24	0.29	0.31	0.36	0.43	0.44	0.50	0.62	0.65	0.69	0.78	0.83	1	0.81	0.76	0.70
14	0.21	0.23	0.28	0.36	0.44	0.41	0.50	0.57	0.58	0.67	0.71	0.73	0.85	1	0.81	0.78
15	0.18	0.20	0.29	0.38	0.45	0.40	0.46	0.56	0.56	0.65	0.66	0.71	0.83	0.90	1	0.83
16	0.15	0.18	0.24	0.34	0.49	0.41	0.48	0.54	0.47	0.59	0.66	0.66	0.82	0.75	0.80	1

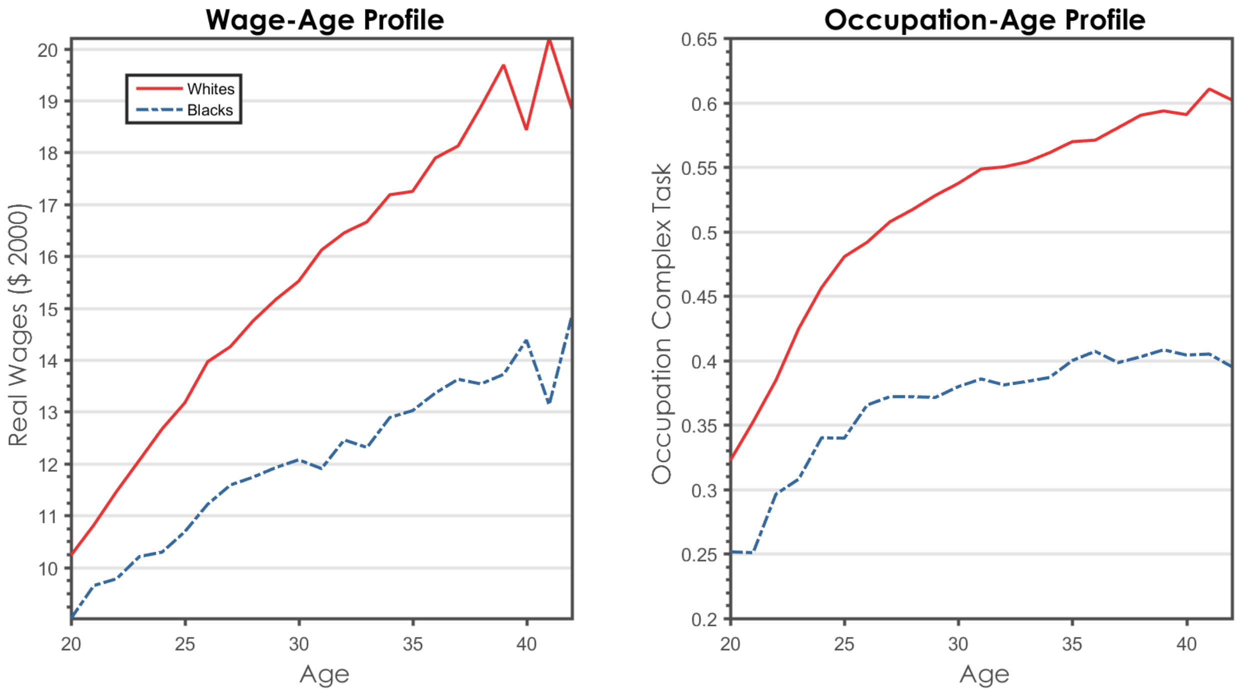
Table 8: Utility Parameters

	Alpha 1	Alpha 2	Alpha 3	Alpha 4	Alpha 5	Alpha 6	Alpha 7	Alpha 8	Alpha 9	Alpha 10
Initial Choice (relative to alpha 1)										
intercept	–	0.04	-0.58	0.02	-0.30	-0.79	-1.59	-1.83	-2.60	-3.20
		(0.06)	(0.10)	(0.08)	(0.04)	(0.15)	(0.06)	(0.04)	(0.13)	(0.10)
HGC-12	–	0.01	0.05	0.17	0.21	0.23	0.43	0.45	0.59	0.74
		(0.01)	(0.04)	(0.04)	(0.03)	(0.00)	(0.01)	(0.02)	(0.06)	(0.02)
Black	–	-0.50	-0.32	-0.59	-0.55	-0.91	-0.72	-0.53	-0.92	-0.75
		(0.11)	(0.05)	(0.17)	(0.12)	(0.19)	(0.05)	(0.11)	(0.53)	(0.09)
Subsequent Choice (relative to unemployment)										
intercept	-0.40	-0.31	-0.58	-0.31	-0.27	-0.58	-0.58	-0.65	-1.76	-1.68
	(0.09)	(0.09)	(0.13)	(0.10)	(0.10)	(0.14)	(0.13)	(0.11)	(0.12)	(0.19)
HGC-12	-0.09	-0.09	-0.07	-0.04	0.00	0.06	0.09	0.15	0.28	0.31
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)
Black	0.09	-0.06	-0.01	-0.13	-0.21	-0.50	-0.50	-0.44	-0.56	-0.53
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.14)	(0.01)	(0.10)	(0.09)	(0.05)
Common Parameters										
Switching Cost	-2.54									
	(0.02)									
Additional Cost if from Unemployment	-1.04									
	(0.04)									
(Age - 18)	0.02									
	(0.00)									
Wage	0.05									
	(0.01)									

Table 9: Simulations

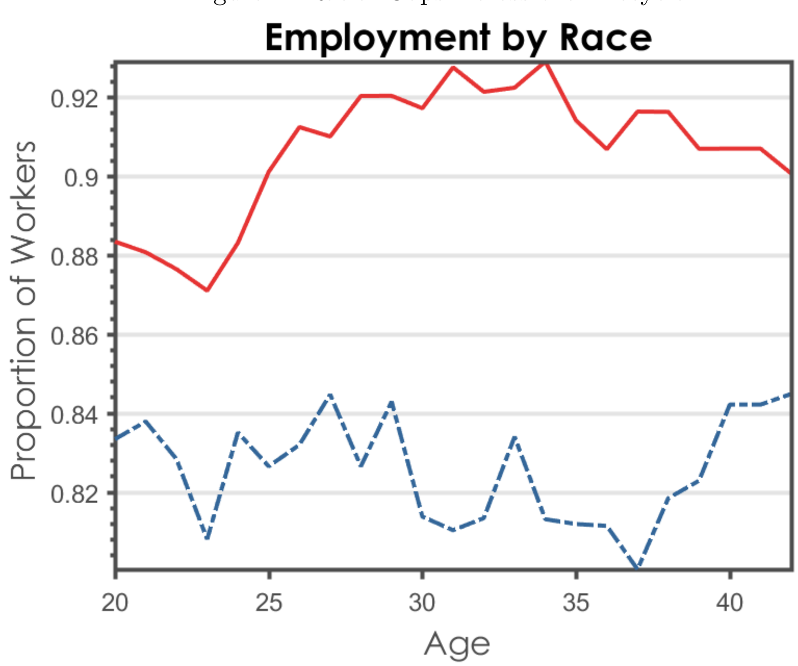
	Nominal		Fraction of Model Gap				
Period	Data	Model	Equalized Observed Pre-Market	Equalize Initial Mean Beliefs	Equalize All Beliefs	Equalize All Utility	Equalize All Initial Conditions (3+4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total Earnings							
5	22.49	24.23	0.49	0.21	0.22	0.21	0.70
	(0.91)	(2.60)	(0.11)	(0.08)	(0.04)	(0.06)	(0.09)
10	58.77	64.24	0.40	0.16	0.24	0.24	0.56
	(1.27)	(3.92)	(0.08)	(0.06)	(0.02)	(0.07)	(0.06)
15	107.92	112.60	0.37	0.14	0.27	0.27	0.51
	(5.28)	(4.46)	(0.06)	(0.05)	(0.02)	(0.07)	(0.04)
Total Experience							
5	0.25	0.24	0.36	0.04	0.04	0.68	0.39
	(0.05)	(0.07)	(0.15)	(0.01)	(0.02)	(0.09)	(0.15)
10	0.59	0.59	0.36	0.04	0.05	0.71	0.39
	(0.02)	(0.10)	(0.12)	(0.01)	(0.02)	(0.09)	(0.12)
15	1.23	0.88	0.37	0.04	0.07	0.73	0.40
	(0.06)	(0.12)	(0.11)	(0.01)	(0.02)	(0.10)	(0.11)
Alpha							
5	0.13	0.14	0.28	0.00	0.01	0.72	0.29
	(0.01)	(0.01)	(0.04)	(0.00)	(0.01)	(0.07)	(0.04)
10	0.14	0.15	0.26	0.00	0.01	0.69	0.26
	(0.01)	(0.01)	(0.04)	(0.01)	(0.02)	(0.08)	(0.03)
15	0.15	0.14	0.24	0.01	0.03	0.75	0.25
	(0.01)	(0.01)	(0.03)	(0.00)	(0.01)	(0.06)	(0.03)

Figure 1: Racial Gaps Across the Lifecycle



Note:

Figure 2: Racial Gaps Across the Lifecycle



Note:

Figure 3: Occupation-Employment Gaps

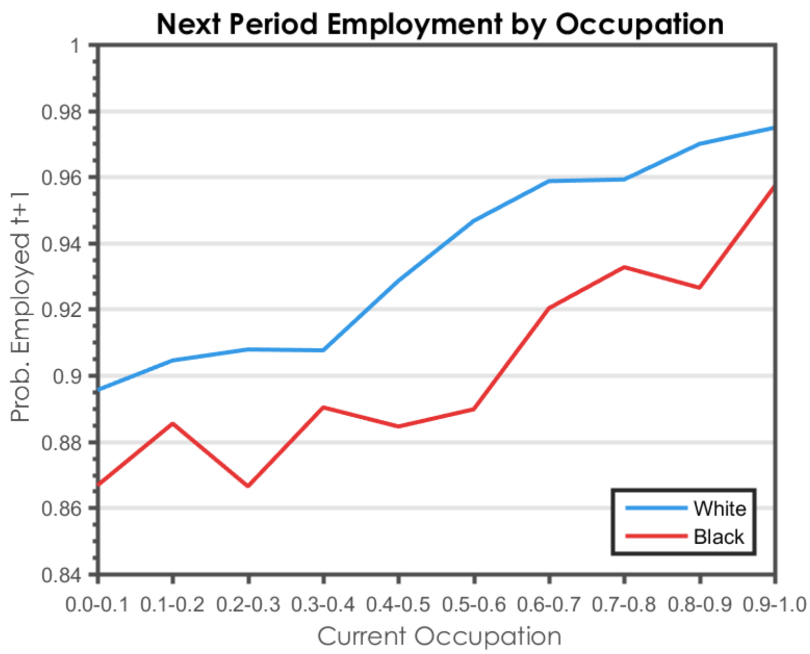
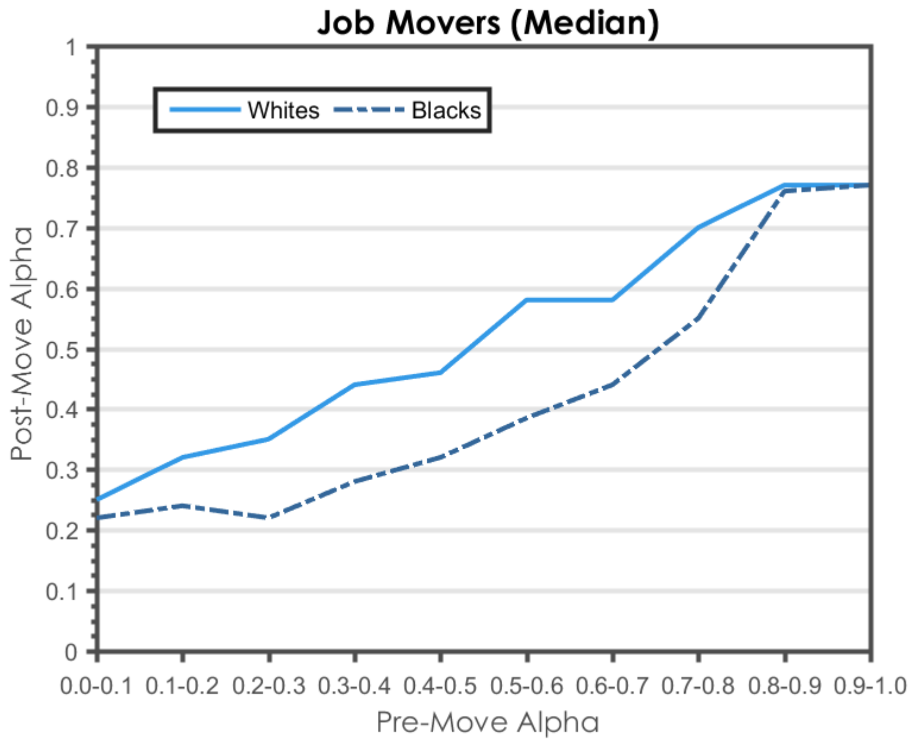
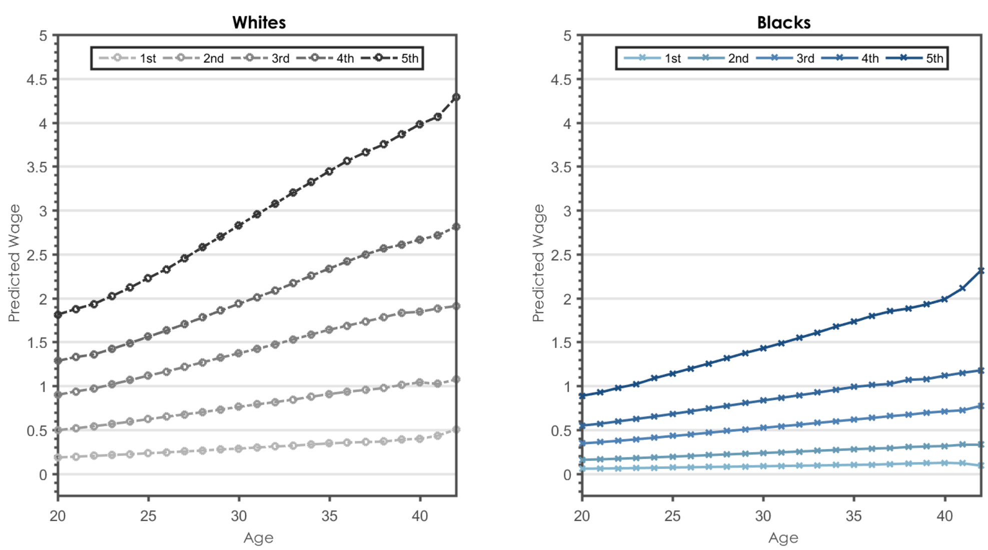


Figure 4: Occupational Transitions by Race



Note:

Figure 5: Alphas by Race



Note: Based on regression in from Table 4

Figure 6: Estimated Beliefs

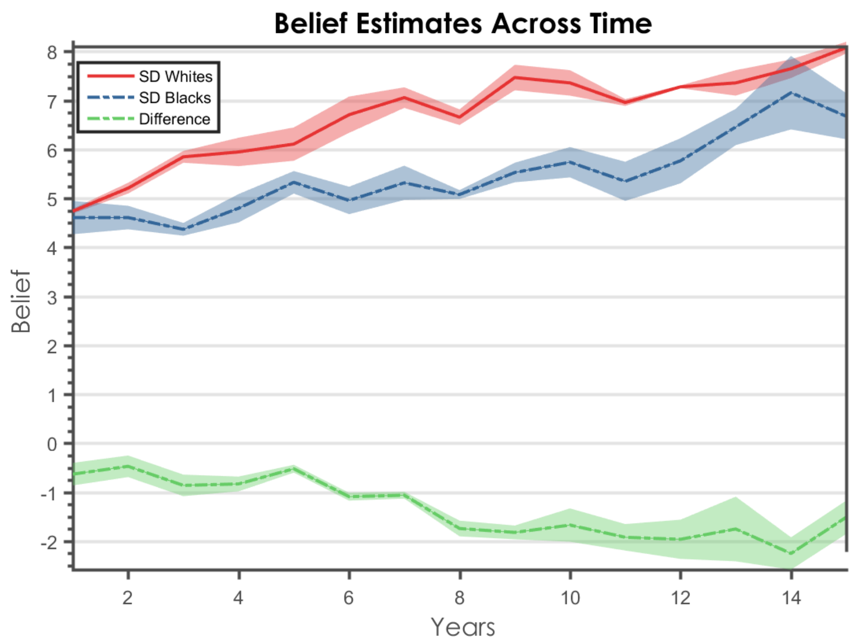


Figure 7: Decomposition of the Earnings Gap

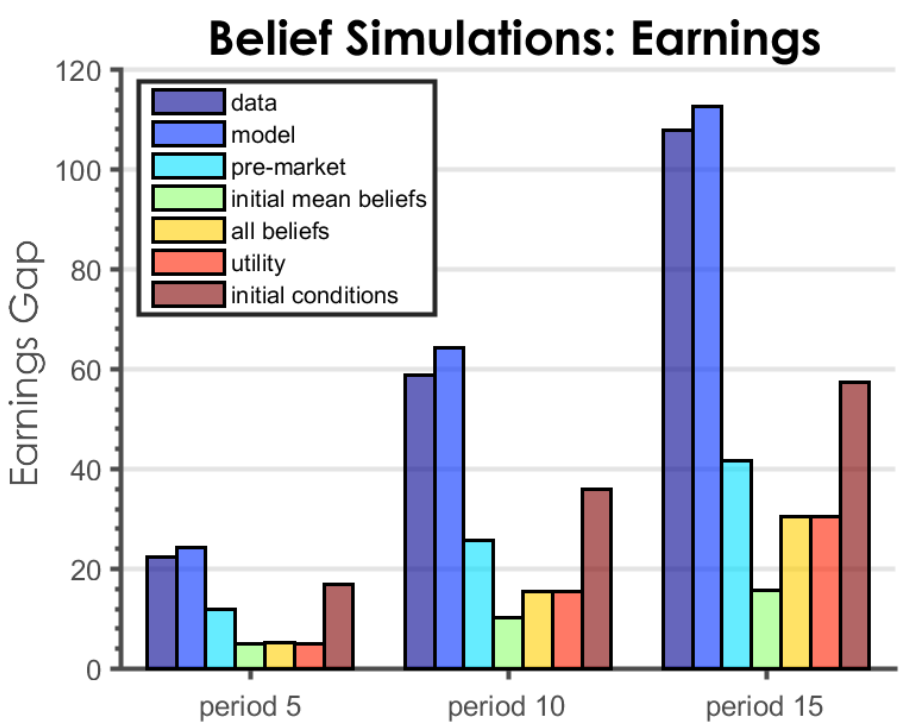


Figure 8: Decomposition of the Experience Gap

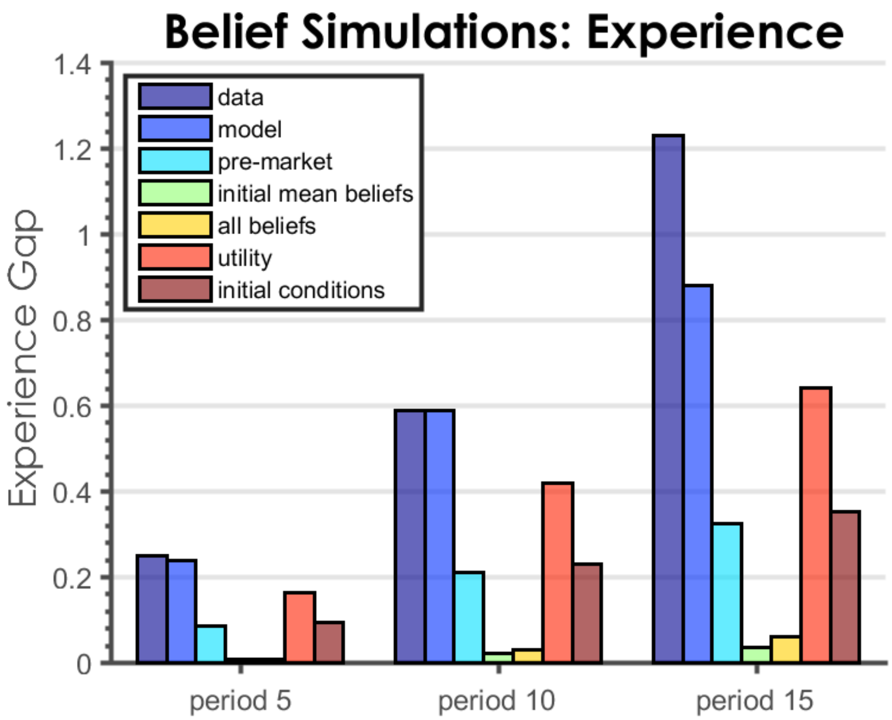
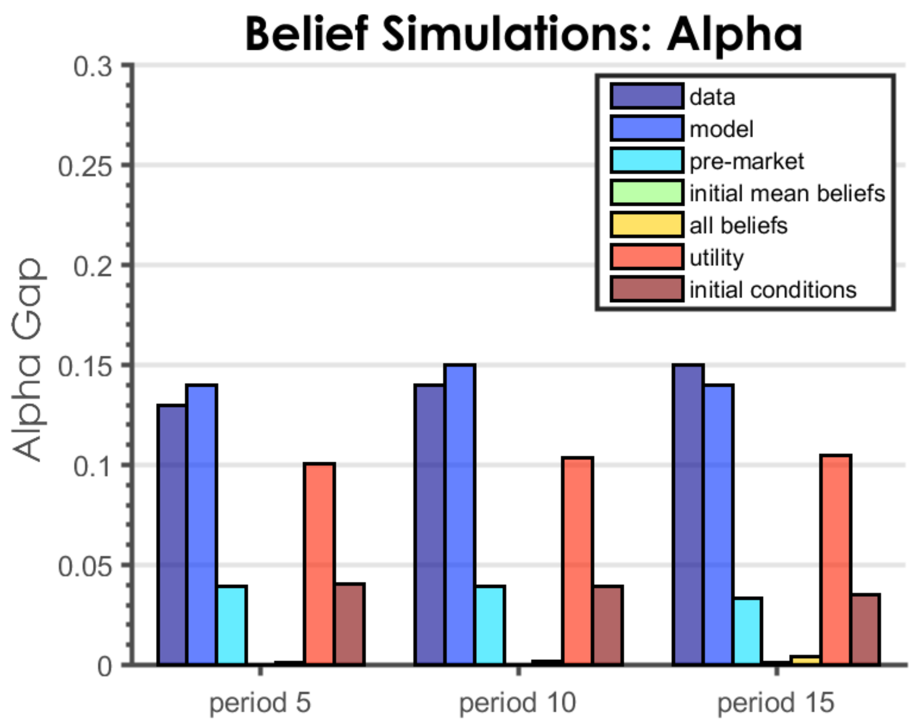


Figure 9: Decomposition of the Occupation Gap



B Estimation-Second Stage

First, we re-write the choice probability in Eq. (12) by differencing each value function by v_0. Define $\tilde{v}_j = v_j - v_0$. Then the choice probabilities can be written as

$$\begin{aligned}
\Pr(I_{ijt}|S_{it},\alpha,\theta,\kappa) &= \frac{\exp(\tilde{v}_j(S_{it},\alpha,\theta,\kappa))}{1+\sum_{j'=1}^{3}\exp(\tilde{v}_j(S_{it},\alpha,\theta,\kappa))} && \text{for } j \in \{1,2,3\} \\
\Pr(I_{ijt}|S_{it},\alpha,\theta,\kappa) &= \frac{1}{1+\sum_{j'=1}^{3}\exp(\tilde{v}_j(S_{it},\alpha,\theta,\kappa))} && \text{for } j = 0
\end{aligned} \tag{13}$$

Where

$$\tilde{v}_j(S_{it},\alpha,\theta,\kappa) = u_1(S_{it},j) + u_2(S_{it},j) + \beta E[V(S_{it+1},\eta_{ijt+1})|S_{it}] \tag{14}$$

$$- \beta E_\eta[V(S_{it+1}, I_{i0t}=1), \eta_{ijt+1}|S_{it}] \tag{15}$$

Our strategy in estimation is to follow Altug and Miller (1998) and re-write the two expectations in a way that when they are differenced they reduce to a set of terms that feasible to calculate. To accomplish this we exploit the Hotz and Miller (1993) mapping of value function and conditional choice probabilities. Hotz and Miller (1993) show that given state variables S, for any j', the expected value function can be written as taken with respect to next periods random utility shock η can be re-written as

$$\begin{aligned}
E_\eta[V(S,\eta)] &= \ln\left(\sum_{k=1}^{3}\exp(v_k(S))\right) + \gamma \\
&= \ln\left(\sum_{k=1}^{3}\exp(v_k(S))\frac{\exp(v_{j'}(S))}{\exp(v_{j'}(S))}\right) + \gamma \\
&= \ln\left(\frac{\sum_{k=1}^{3}\exp(v_k(S))}{\exp(v_{j'}(S))}\right) + v_{j'}(S) + \gamma \\
&= \ln \Pr{}_{j'}^{-1}(S) + v_{j'}(S) + \gamma
\end{aligned}$$

Where γ is Euler's constant. Importantly this mapping can be written in terms of any choice j'. Altug and Miller (1998) suggest intentionally choosing a sequence of choices that lead to a convenient expression of the value function.

Define the following 1, 2, and 3-period ahead state variables

$$\begin{aligned}
S^A_{it+1} &= \{S_{it}, I_{ijt}=1, z_{ijt}\} \\
S^A_{it+2} &= \{S_{it}, I_{ijt}=1, z_{ijt}, I_{i0t+1}=1\} \\
S^A_{it+3} &= \{S_{it}, I_{ijt}=1, z_{ijt}, I_{i0t+1}=1, I_{i0t+2}=1\}
\end{aligned}$$

Using this sequence of state variables and the ability to re-write the future value function in terms of a conditional choice probability and a conditional value function. we can write the first future value term in Eq. (15) conditional on employment j

as

$$
\begin{aligned}
&E_z E_\eta \left[V(S_{it+1}, z_{ijt}, I_{ijt}\!=\!1, \eta_{ijt+1})|S_{it}\right] \\
&\quad = E_z E_\eta \left[V(S^A_{it+1}, \eta_{ijt+1})|S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + v_0(S^A_{it+1}) + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + 0 + \beta E_\eta \left[V(S_{it+2}(S^A_{it+1}, I_{i0t+1}\!=\!1), \eta_{ijt+2})\right] + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + 0 + \beta E_\eta \left[V(S^A_{it+2}, \eta_{ijt+2})\right] + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + 0 + \beta \left[\ln \Pr_0^{-1}(S^A_{it+2}) + v_0(S^A_{it+2}) + \gamma\right] + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + 0 + \beta \left[\ln \Pr_0^{-1}(S^A_{it+2}) + 0 + \beta E_\eta \left[V(S_{it+3}(S^A_{it+2}, I_{i0t+2}\!=\!1), \eta_{ijt+3})\right] + \gamma\right] + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + 0 + \beta \left[\ln \Pr_0^{-1}(S^A_{it+2}) + 0 + \beta E_\eta \left[V(S^A_{it+3}, \eta_{ijt+3})\right] + \gamma\right] + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) + \beta \ln \Pr_0^{-1}(S^A_{it+2}) + \beta^2 E_\eta \left[V(S^A_{it+3}, \eta_{ijt+3})\right] + \beta\gamma + \gamma \Big| S_{it}\right] \\
&\quad = E_z \left[\ln \Pr_0^{-1}(S^A_{it+1}) \Big| S_{it}\right] + \beta E_z \left[\ln \Pr_0^{-1}(S^A_{it+2}) \Big| S_{it}\right] + \beta^2 E_z \left[E_\eta \left[V(S^A_{it+3}, \eta_{ijt+3})\right] \Big| S_{it}\right] + \beta\gamma + \gamma
\end{aligned}
$$

Given estimates of the conditional choice probabilities, the first two terms of this expression can be easily calculated with the data. The third term, which is a three period away expected future value term is too complicated to be calculated. However, as we show below, we can write $E_\eta[V(S_{it+1}, I_{i0t}\!=\!1, \eta_{ijt+1})|S_{it}]$ in Eq. (15) in way so that this term drops out.

Define an alternative 1, 2, and 3-period ahead state variables

$$
\begin{aligned}
S^B_{it+1} &= \{S_{it}, I_{i0t}\!=\!1\} \\
S^B_{it+2} &= \{S_{it}, I_{i0t}\!=\!1, I_{ijt+1}\!=\!1, z_{ijt+1}\} \\
S^B_{it+3} &= \{S_{it}, I_{i0t}\!=\!1, I_{ijt+1}\!=\!1, z_{ijt+1}, I_{i0t+2}\!=\!1\}
\end{aligned}
$$

In this expression the one-period ahead state variables S^B_{it+1} results from the individual choosing non-employment in period t. Furthermore, S^B_{it+2} is the state variable when the individual subsequently chooses employment in occupation j in period $t+1$ and receives the signal from z_{ijt+1}. Notice in period $t+1$, the beliefs $b_{it+1} = b_{it}$, since the worker did not receive any signals of their unknown skill while not employed in period t, Thus $z_{ijt+1} \sim N(b_{it}, \sigma^2_{it})$ comes from the same distribution of skill signals that would have resulted if they were employed in period t.

Using this sequence of state variables and the ability to re-write the future value function in terms of a conditional choice probability and a conditional value function. we can write the second future value term in Eq. (15) conditional on non-employment as

$$
\begin{aligned}
&E_\eta[V(S_{it+1}I_{i0t}\!=\!1,\eta_{ijt+1})|S_{it}]\\
&\quad= E_\eta[V(S^B_{it+1},\eta_{ijt+1})|S_{it}]\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + v_j(S^B_{it+1}) + \gamma\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z E_\eta\Big[V(S_{it+2}(S^B_{it+1},z_{ijt+1},I_{ijt+1}\!=\!1),\eta_{ijt+2})\Big|S_{it}\Big] + \gamma\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z E_\eta\Big[V(S^B_{it+2},\eta_{ijt+2})\Big|S_{it}\Big] + \gamma\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2}) + v_0(S^B_{it+2}) + \gamma\Big|S_{it}\Big] + \gamma\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2})\\
&\qquad + 0 + \beta E_\eta\Big[V(S_{it+3}(S^B_{it+2},I_{i0t+2}\!=\!1),\eta_{ijt+3})\Big] + \gamma\Big|S_{it}\Big] + \gamma\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2}) + \beta E_\eta\Big[V(S^B_{it+3},\eta_{ijt+3})\Big] + \gamma\Big|S_{it}\Big] + \gamma\\
&\quad= \ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2})\Big|S_{it}\Big]\\
&\qquad + \beta^2 E_z\Big[E_\eta\Big[V(S^B_{it+3},\eta_{ijt+3})\Big]\Big|S_{it}\Big] + \beta\gamma + \gamma
\end{aligned}
$$

Using these two expressions for the expected future value term, we can write the differenced value function in Eq. (15) as

$$
\begin{aligned}
\tilde v_j(S_{it},\alpha,\theta,\kappa) =&u_1(S_{it},j) + u_2(S_{it},j) + \beta E_z E_\eta[V(S_{it+1},z_{ijt},I_{ijt}\!=\!1,\eta_{ijt+1})|S_{it}]\\
&- \beta E_\eta[V(S_{it+1}I_{i0t}\!=\!1,\eta_{ijt+1})|S_{it}]\\
=&u_1(S_{it},j) + u_2(S_{it},j) + \beta\Big\{E_z\Big[\ln \Pr_0^{-1}(S^A_{it+1})\Big|S_{it}\Big] + \beta E_z\Big[\ln \Pr_0^{-1}(S^A_{it+2})\Big|S_{it}\Big]\\
&+ \beta^2 E_z\Big[E_\eta\Big[V(S^A_{it+3},\eta_{ijt+3})\Big]\Big|S_{it}\Big] + \beta\gamma + \gamma\Big\}\\
&- \beta\Big\{\ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2})\Big|S_{it}\Big]\\
&+ \beta^2 E_z\Big[E_\eta\Big[V(S^B_{it+3},\eta_{ijt+3})\Big]\Big|S_{it}\Big] + \beta\gamma + \gamma\Big\}\\
=&u_1(S_{it},j) + u_2(S_{it},j) + \beta\Big\{E_z\Big[\ln \Pr_0^{-1}(S^A_{it+1})\Big|S_{it}\Big] + \beta E_z\Big[\ln \Pr_0^{-1}(S^A_{it+2})\Big|S_{it}\Big]\Big\}\\
&- \beta\Big\{\ln \Pr_j^{-1}(S^B_{it+1}) + u_1(S^B_{it+1},j) + u_2(S^B_{it+1},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2})\Big|S_{it}\Big]\Big\}\\
&+ \beta^3 E_z E_\eta\Big[V(S^A_{it+3},\eta_{ijt+3}) - V(S^B_{it+3},\eta_{ijt+3})\Big|S_{it}\Big]\\
=&u_1(S_{it},j) + u_2(S_{it},j) + \beta E_z\Big[\ln \Pr_0^{-1}(S^A_{it+1})\Big|S_{it}\Big] + \beta^2 E_z\Big[\ln \Pr_0^{-1}(S^A_{it+2})\Big|S_{it}\Big]\\
&- \beta \ln \Pr_j^{-1}(S^B_{it+1}) - \beta u_1(S^B_{it+1},j) - \beta u_2(S^B_{it+1},j) - \beta^2 E_z\Big[\ln \Pr_0^{-1}(S^B_{it+2})\Big|S_{it}\Big] \qquad (16)
\end{aligned}
$$

The results above stem from tow important facts. First, the distribution of z_{ijt} and z_{ijt+1} are the same, i.e. $F(z_{ijt}) = F(z_{ijt+1})$ where $F(\cdot)$ denotes the conditional distribution function of the unknown skill signal. In addition to the fact that z and η are independent, this allows us to write

$$
\begin{aligned}
&E_z E_\eta\Big[V(S^A_{it+3},\eta_{ijt+3})\Big|S_{it}\Big] - E_z E_\eta\Big[V(S^B_{it+3},\eta_{ijt+3})\Big|S_{it}\Big]\\
&\quad= E_z E_\eta\Big[V(S^A_{it+3},\eta_{ijt+3}) - V(S^B_{it+3},\eta_{ijt+3})\Big|S_{it}\Big]
\end{aligned}
$$

Second, conditional on a given signal z, the state variables $S^A_{it+3} = S^B_{it+3}$. This is because in the utility function only the prior choice impacts current utility an in both cases S^A_{it+3} and S^B_{it+3}, $I_{i0t+2} = 1$. Therefore, for a given z and η $V(S^A_{it+3}, \eta_{ijt+3}) = V(S^B_{it+3}, \eta_{ijt+3})$ and thus

$$\begin{aligned} E_z E_\eta \Big[V(S^A_{it+3}, \eta_{ijt+3}) - V(S^B_{it+3}, \eta_{ijt+3}) \Big| S_{it} \Big] =& E_z E_\eta \Big[0 \Big| S_{it} \Big] \\ =& 0 \end{aligned}$$

B.1 Likelihood

Equation (16) provides a representation of the conditional value function that can be used to estimate the structural parameters. Let Γ denote the worker's policy function, where $\Gamma_j(S_{it}(b), \alpha, \theta, \kappa)$ denotes the probability that this worker given the state variables and structural parameters makes choice j. These policy functions are defined as

$$\Gamma_j(S_{it}(b), \alpha, \theta, \kappa, \Gamma) = \frac{[\exp(\tilde{v}_j(S_{it}, \alpha, \theta, \kappa, \Gamma)]^{(j>0)}}{1 + \sum_{j'=1}^{J} \exp(\tilde{v}_{j'}(S_{it}, \alpha, \theta, \kappa, \Gamma))}$$

In the second stage, we estimate the structural parameters by maximizing the second stage log-likelihood

$$LL_{second-stage} = \sum_{i=1}^{n} \int \ln \left[\prod_{t=1}^{T_i} \Gamma_j(S_{it}(b), \alpha, \theta, \kappa, \Gamma)^{I_{ijt}} \right] h_i(b) db \tag{17}$$

Where $h_i(b)$ is the posterior distribution of unobserved beliefs for individual i given the estimates from the first stage.

$$h_i(b|W_i, I_i, \hat{\sigma}, \hat{\delta}, \hat{\lambda}, \hat{\omega}) = \frac{L(I_i|\Omega_i(b), \hat{\omega}) L(W_i, I_i|\Omega_i(b), \hat{\lambda}) f(b|I_i, r_i, \hat{\sigma}, \hat{\delta})}{\int L(I_i|\Omega_i(b'), \hat{\omega}) L(W_i, I_i|\Omega_i(b'), \hat{\lambda}) f(b'|I_i, r_i, \hat{\sigma}, \hat{\delta}) db'} \tag{18}$$

One approach to estimating the structural utility parameters, would be following Hotz and Miller (1993), which is non-parametrically estimate $\widehat{\Gamma}(S_{it}(b))$ with the data, and then choose the parameters as

$$\underset{\hat{\alpha}, \hat{\theta}, \hat{\kappa}}{\text{argmax}} \sum_{i=1}^{n} \int \ln \left[\prod_{t=1}^{T_i} \Gamma_j(S_{it}(b), \hat{\alpha}, \hat{\theta}, \hat{\kappa}, \widehat{\Gamma})^{I_{ijt}} \right] h_i(b) db$$

This expression is only a function of contemporaneous utility functions and conditional choice probabilities that where computed from the data.

Instead we pursue a full solution method to this problem and solve for

$$\underset{\hat{\alpha}, \hat{\theta}, \hat{\kappa}, \widehat{\Gamma}}{\text{argmax}} \sum_{i=1}^{n} \int \ln \left[\prod_{t=1}^{T_i} \Gamma_j(S_{it}(b), \hat{\alpha}, \hat{\theta}, \hat{\kappa}, \widehat{\Gamma})^{I_{ijt}} \right] h_i(b) db$$

Where the estimate the policy functions along with the parameters, this imposes a degree of internal consistency with the utility parameters and the contraction mapping for the policy function.

The integral in the likelihood does not have a closed form, so we simulate it taking R draws of b, labeled, b_{ir} for each individual and maximize

$$\underset{\hat{\alpha}, \hat{\theta}, \hat{\kappa}, \widehat{\Gamma}}{\text{argmax}} \sum_{i=1}^{n} \sum_{r=1}^{R} \ln \left[\prod_{t=1}^{T_i} \Gamma_j(S_{it}(b_{ir}), \hat{\alpha}, \hat{\theta}, \hat{\kappa}, \widehat{\Gamma})^{I_{ijt}} \right]$$

Since the policy functions are approximated using flexible polynomials given the size of the state space, We solve the optimization problem above using Arcidiacono et al. (2013), which is similar to the algorithm in Aguirregabiria and Mira (2002).