NBER WORKING PAPER SERIES

THE BOSS IS WATCHING:
HOW MONITORING DECISIONS HURT BLACK WORKERS

Costas Cavounidis
Kevin Lang
Russell Weinstein

Working Paper 26319
http://www.nber.org/papers/w26319

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2019

This research was funded in part by NSF grants SES-1260917 and SES-1851636. It draws on similar intuition to Cavounidis and Lang (2015) but greatly simplifies the theory and adds an empirical section testing the main new prediction. We are grateful to Cheonghum Park for excellent research assistance and to Gautam Bose, Sambuddha Ghosh, Larry Katz, Ed Lazear, Calvin Luscombe, Ioana Marinescu, Derek Neal, Marina Soley-Bori, Bill Spriggs and participants in seminars and conferences at Boston University, Cornell, IZA/OECD, NBER, Tufts, the University of California, Riverside and West Point Military Academy for helpful comments and suggestions on this and the earlier paper. The usual caveat applies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Boss is Watching: How Monitoring Decisions Hurt Black Workers
Costas Cavounidis, Kevin Lang, and Russell Weinstein
NBER Working Paper No. 26319
September 2019
JEL No. J01,J3,J71

ABSTRACT

African Americans face shorter employment durations than apparently similar whites. We hypothesize that employers discriminate in either acquiring or acting on ability-relevant information. We construct a model in which firms may "monitor" workers. Monitoring black but not white workers is self-sustaining: new black hires are more likely to have been screened by a previous employer, causing firms to discriminate in monitoring. We confirm the model's prediction that the unemployment hazard is initially higher for blacks but converges to that for whites. Two additional predictions, lower lifetime incomes and longer unemployment durations for blacks, are known to be strongly empirically supported.

Costas Cavounidis
Department of Economics
Warwick University
Coventry, CV4 7AL
Tel: +44 2476528420
United Kingdom
C.Cavounidis@warwick.ac.uk

Kevin Lang
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
lang@bu.edu

Russell Weinstein
School of Labor and Employment Relations
and Department of Economics
University of Illinois at Urbana-Champaign
504 E. Armory Avenue
Champaign, IL 61820
weinst@illinois.edu

1 Introduction

Many Americans, especially African Americans, believe black workers 'don't get second chances'[1] or that they face additional scrutiny in the workplace. Similarly, black workers are admonished to be 'twice as good'[2] in order to succeed. If black workers are subject to higher standards or scrutinized more heavily, we expect this to be reflected in more separations.

Indeed, the data support the idea of shorter employment duration[3] for black workers. Bowlus, Kiefer and Neumann (2001) detect and ponder the disparity in job destruction rates; Bowlus and Eckstein (2002) estimate[4] that young black male high school graduates had roughly 2/3 the job spell duration of their white counterparts. In addition, more of their job spells end in unemployment, suggesting that blacks have much shorter employment spells. Both papers assume an exogenously higher separation rate for black workers to fit their models to the data. Lang and Lehmann (2012) show that differences in unemployment duration alone are insufficient to account for the black/white unemployment rate gap and therefore that black workers' employment stints are shorter. This aspect of labor discrimination has thus far eluded theoretical explication.

In this paper, our proposed explanation for differential employment durations is, in its broadest sense and consistent with the aforementioned observations, that firms discriminate in the acquisition or use of productivity-relevant information. That is, firms either learn differently about black workers or, when information regarding ability is received, they condition how they act on it on workers' race. Crucially, we establish that such discrimination can be self-perpetuating.

---

[1]This assertion can be found in a range of occupations including football coaching (Reid, 2015), music and films (*The Guardian*, 2014) as well as more generally (Spencer, 2014).

[2]Coates, Ta-Nehisi (2012) and Mabry, Marcus (2012)

[3]Throughout this paper we distinguish between employment duration by which we mean the length of an employment spell and job duration by which we mean the time a worker spends with a particular employer. Job duration depends on, among other factors, the arrival rate of outside offers. Our model abstracts from job-to-job transitions, but can incorporate them without trouble, as shown in Section 4.5.8.

[4]Using the NLSY data for 1985 and 1988.

The essence of our model is that, because black workers are more closely scrutinized, a larger share of low-performance workers will separate into unemployment. As a result, since productivity is correlated across jobs, the black unemployment pool is 'churned' and therefore weaker than the white unemployment pool. Since workers can, at least to some extent, hide their employment histories, race serves as an indicator of expected worker productivity. This in turn makes monitoring newly hired black (but not white) workers optimal for firms. Figure 1 illustrates employment in the two labor markets. The churning mechanism is shared with Masters (2014), where information acquisition takes the form of exogenous pre-employment signals rather than endogenous monitoring on the job.

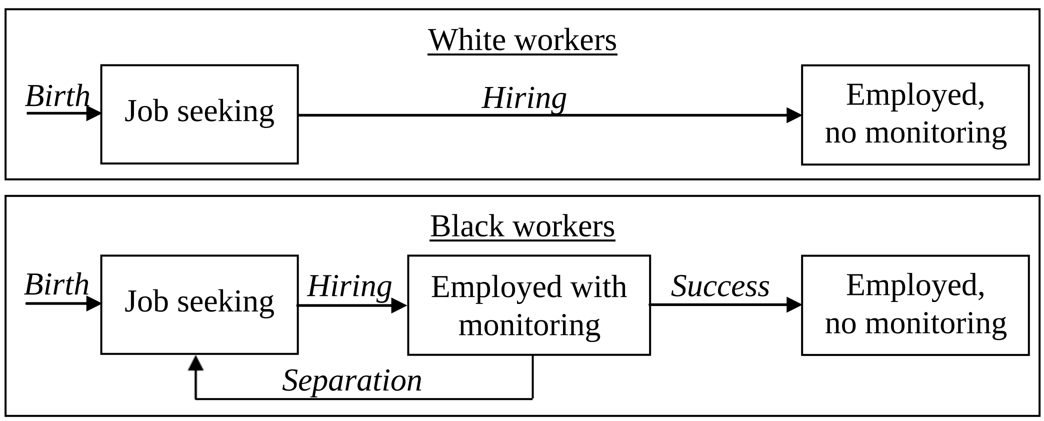


Figure 1: White workers' perpetual employment, black workers' churning cycle

Importantly, our model has excess empirical content. It predicts that involuntary separations from employment will initially be higher for blacks than for whites but that these hazards will converge with seniority. As seniority increases it is more likely that workers have passed monitoring, and are good matches with the firm. We test and largely confirm this previously untested prediction using the National Longitudinal Survey of Youth 1979 (NLSY79). The finding is robust to a variety of sample selection decisions, approaches to smoothing the hazards, measures of seniority and proxies for involuntary separation, and strengthens with the inclusion of controls.

There are multiple equilibria in our model, a property it shares with models of rational stereotyping or self-confirming expectations (Coate and Loury, 1993). However, in our model discrimination is not simply a product of coordination failure; instead, history matters. A group that begins with a low level of skills for which only the bad (monitoring) equilibrium exists will remain in that equilibrium even if its skill level rises to a level consistent with the existence of both the good and bad equilibria. Even if blacks are, on average, more skilled than whites, whites can be in the good steady-state and blacks in the bad steady-state because of a history of lower access to schooling and other human capital investments. Equalizing the human capital that blacks and whites bring to the labor market may be insufficient to equalize labor market outcomes. In contrast, in self-confirming expectations models, if we could just convince blacks to invest in themselves and employers that blacks have invested, we would immediately jump to the good equilibrium

There is an abundance of evidence that black workers face lower wages and longer unemployment duration than white workers. Moreover, these disparities are less prevalent and, perhaps, in some cases nonexistent for the most skilled workers as measured by education or performance on the Armed Forces Qualifying Test. While there are a plethora of models intended to explain wage or unemployment differentials, none addresses both and their relation to skill.[5] Since in our model newly hired black workers are on average less productive than white ones, their wages are lower and firms that expect to hire blacks anticipate less profit from a vacancy and therefore offer fewer jobs. Consequently, blacks have longer unemployment durations.

We derive additional implications from informal extensions to the model.

[5] Many models (e.g. Aigner and Cain, 1977; Becker, 1971; Bjerk, 2008; Charles and Guryan, 2011; Coate and Loury, 1993; Fryer, 2007; Lang, 1986; Lang and Manove, 2011; Lundberg and Startz, 1983; Moro and Norman, 2004) assume market clearing and therefore cannot address unemployment patterns. Search models (e.g. Black, 1995; Bowlus and Eckstein, 2002; Lang and Manove, 2003; Lang, Manove and Dickens, 2005; Rosen, 1997) can explain unemployment differentials, but assume otherwise homogeneous workers and thus cannot address wage differentials at different skill levels. Peski and Szentes (2013) treat wages as exogenous. In general, discrimination models have not addressed employment duration. See the review in Lang and Lehmann (2012).

The higher level of scrutiny increases the return to skill for blacks, consistent with evidence that blacks invest more in schooling compared with apparently equivalent whites. In addition, if unemployment history is partially observable, black job seekers who have experienced enough turnover may be permanently relegated to low-skill, low-wage jobs. Although we do not wish to overstate the predictive power of the model, we note that until around 1940, blacks and whites had similar unemployment rates (Fairlie and Sundstrom, 1999), while blacks faced lower wages. This is consistent with a setting in which, due to low human capital investments, blacks were assumed to have low productivity at most jobs and therefore not monitored for quality. 'Churning' of the black labor market would not begin until human capital investments were sufficiently high.

We believe that the broad implications of our model can be derived through a variety of formalizations. The key elements common to these are:

i. that a worker's productivity at different firms is correlated,

ii. that workers cannot or do not signal their ability and that they can, at least imperfectly, hide their employment histories,[6]

iii. that firms must therefore, to some degree, statistically infer worker ability,

iv. that further information about match productivity arrives during production, and is either costly, imperfect, or both, and

v. that this information, if obtained, may affect retention, so that firm behavior affects the average unemployed worker's ability.

The details of our formal model are driven by our desire for a theoretically rigorous model of wage-setting in a dynamic framework with asymmetric information. Firms and workers bargain over wages and use a costly monitoring technology to assess the quality of the match, which is correlated with the worker's underlying type.

[6] In particular, they must sometimes be able to omit or mischaracterize prior bad matches.

Therefore, use of the monitoring technology depends on the firm's prior: if the belief that a worker is well-matched is sufficiently high or sufficiently low, it will not be worth investing resources to determine match quality. However, if the cost of determining the match quality is not too high, there will be an intermediate range at which this investment is worthwhile. Firm beliefs about black, but not white, workers fall in this region. Consequently, they are subject to heightened scrutiny and are more likely to be found to be a poor match and fired. The increased scrutiny ensures that the pool of unemployed black workers has a higher proportion of workers who have been found to be a poor match at one or more prior jobs. And therefore employers' expectations that black workers are more likely to be poor matches is correct in equilibrium. This, nested in a search model, generates the empirical predictions discussed above.[7]

This churning equilibrium is hard to escape. This is disheartening since *policy succeeding at convergence of group characteristics may fail to equate labor market outcomes*. Only if the skill level of blacks is raised sufficiently above that of whites (technically the proportion of good workers is sufficiently high), does the bad equilibrium cease to exist and white and black workers receive similar treatment.

2 The Model

2.1 Setup

There are two worker groups, 'blacks' and 'whites'. Race is observable by the worker and employers but does not have any direct impact on production.

At all times a steady flow of new workers is born into each population group.[8] A proportion $g \in (0, 1)$ of new workers are type α, for whom every job

[7] Note that our model abstracts from moral hazard and that performance is observed objectively. MacLeod (2003) develops an interesting model in which biased subjective assessments interact with moral hazard concerns.

[8] We do not allow for death but could do so at the cost of a little added complexity.

is a good match.[9] The rest, referred to as type β, have probability $\beta \in (0,1)$ of being a good match at any particular job. The probability of a worker being good at a job, conditional on her type, is independent across jobs. Worker type is private to the worker. Workers begin their lives unemployed. The probability a new worker is good at a particular job is

$$\theta_0 = g + (1-g)\,\beta. \tag{1}$$

Employers cannot directly observe worker type or employment history,[10] but can instead draw statistical inferences from race.

2.2 Match Quality

Production, the payment of wages and the use of the monitoring technology occur in continuous time using a common discount rate r.

Workers can be either well-suited to a task (a 'good' match), producing q per unit time; or ill-suited (a 'bad' match), producing expected output $q - \lambda c$ per unit time. We can interpret the lower productivity of bad workers as errors or missed opportunities, each costing the firm c, that arrive at a constant rate λ. Under this interpretation, opportunities for error are also opportunities to learn the quality of the match as well-matched workers are observed to avoid errors.[11]

The employer does not know the match quality without monitoring. During production, the firm may use a technology that may produce a fully informative signal about match quality. If the signal shows the match to be bad, the firm may terminate it immediately, receiving 0.

In keeping with the opportunities-for-errors interpretation, we assume the signal arrives at a constant hazard rate λ. The monitoring technology costs

[9] Having type α workers perform well at every job is not essential to the argument but simplifies presentation significantly.

[10] At a more informal level, we believe that workers have some ability to hide their employment history and that they will not report information speaking to their own low ability. We show the model is robust to imperfect history revelation in Section 4.5.6.

[11] Alternatively, we could assume that the flows are $q - d$ and q with $d \equiv \lambda c$ and that λ is the arrival rate of opportunities to measure the flows.

b per unit time, so that the expected cost of information is $\int_0^\infty be^{-\lambda t}dt = b/\lambda$ and its expected discounted cost is $\int_0^\infty (e^{-rt}b)e^{-\lambda t}dt = b/(\lambda + r)$. The principal benefit of a signal whose arrival is exponentially distributed, rather than one that arrives deterministically, is that it makes the employment survival function more realistic. In addition, it allows for a certain stationarity in the model: so long as no signal has arrived, the underlying incentives do not change.

For monitoring to ever be useful, matches revealed to be bad must separate. A sufficient condition for this is that $q - \lambda c < 0$. Additionally, we intend that β workers will not be willing to reveal their type in bargaining. To this end, we make the sufficient and simple assumption that such a match is unproductive, regardless of the monitoring choice:

$$\text{(C1)} \qquad \max\left\{q - (1-\beta)\lambda c,\ \beta\frac{q}{r} + (1-\beta)\frac{(q-\lambda c)}{\lambda + r} - \frac{b}{\lambda + r}\right\} \leq 0.$$

It is *much* stronger than necessary. In general, it is sufficient that any wage at which a firm would knowingly hire a β worker is low enough that the worker would rather reject it in order to rematch at a higher (pooling) wage. Assumption (C1) ensures that such separation in search of a new match is beneficial regardless of the expected duration of unemployment.

2.3 Job Search

When a worker is born or her match is terminated, she becomes unemployed. Unemployed workers are stochastically matched to firms, which occurs at a constant hazard μ. For the moment, we treat this rate as exogenous; it will be endogenized in Section 4.4 to address unemployment duration. When a match dissolves, transfers cease and the worker becomes unemployed. A firm does not recoup a vacancy and therefore receives a payoff of 0 on termination.[12]

In the unemployed state, workers merely search for new jobs; we normalize the flow utility from this state to 0. The value from unemployment is thus

[12] This occurs naturally due to free entry when vacancy creation is endogenized; see Section 4.4.

simply the appropriately discounted expected utility from job-finding and is invariant to history. The expected discount on job-finding is $\int_0^\infty e^{-rt}\mu e^{-\mu t}dt = \mu/(\mu + r)$; the value of a new job will depend on the equilibrium. We denote the value of the job-finding state in a market where the average worker quality is θ as U_θ^α for type α workers and U_θ^β for type β workers. These will be constant in steady-state. Furthermore, in order to have simple dynamics outside steady-state, we assume that workers are *myopic*: they bargain as though the future wage distribution is identical to the current one.

2.4 Wage Setting

Given the asymmetry in information between the worker, who knows her type, and the firm, the Nash bargaining model is unusable and the Rubinstein (1982) one suffers from a multiplicity of equilibria. If a β worker does not want to reveal her type, as follows from our assumptions, then the β worker will have to bargain as if she were an α worker. Since in this case the firm cannot distinguish with which type it is bargaining, it should act as if it were bargaining with a random draw from the unemployment pool. Thus an intuitively appealing solution is the outcome that would be reached in Nash bargaining between a firm calculating its rents on the assumption of a random draw from the pool and a worker calculating her rents as if she were an α worker.

We posit a simple wage bargaining model akin to Lauermann and Wolinsky (2016) that produces this outcome albeit only in expectation. When a worker and firm meet, a wage offer w is randomly drawn from some distribution F. They then simultaneously choose whether to accept or reject the offer. If either rejects the offer, the match is dissolved; the firm receives 0 while the worker searches for the next firm. If both parties accept the offer, production proceeds at that wage. We are looking for Perfect Bayesian Equilibria in which neither party uses a (weakly) dominated strategy.[13] Using a randomly drawn take-it-or-leave-it offer allows us to escape both the multiplicity of equilibria caused by off-path beliefs if the worker can make offers, and the Diamond paradox if

[13]This is needed to rule out equilibria where mutually acceptable wages are rejected by both parties.

only the firm makes offers.[14]

To ensure that the wage process does not end in disagreement in equilibrium, we assume only agreeable wages are proposed. Specifically, we assume that F is a uniform distribution on the set of wages the firm and worker would both accept; thus, wage negotiation is as though an arbitrator proposes any wage on the contract curve with equal likelihood. Crucially, this assumption guarantees that, once we endogenize the job-finding rate, disparities in new match formation rates are solely caused by lower demand for black workers, and not the bargaining process.[15] Fortuitously, the assumption also results in simple solutions.

Jointly, our assumptions will ensure that every match will find a mutually acceptable wage, that equilibrium in steady-state will be unique, that wages are uniformly distributed over the contract curve, and that they are on average equal to the equal-weights Nash bargaining solution (between a firm with beliefs given by θ and an α worker), despite the asymmetric information.

2.5 Steady State

A steady state of a labor market is a mass of α job seekers, a mass of β job seekers and a mass of monitored β workers along with equilibrium firm and worker wage acceptance and monitoring strategies that make these populations constant over time. There are three kinds of steady states: those in which all employees are monitored until match quality is revealed, those in which no monitoring occurs, and those in which the monitoring choice depends on the

[14]Earlier versions of this paper used an alternating-offers bargaining model with off-path belief restrictions and derived an equivalent set of theoretical results. The somewhat artificial nature of the current wage-setting structure dramatically simplifies the presentation without fundamentally changing the results.

[15]This of course makes F an equilibrium object, as the acceptability of wages in turn depends on F; but the solution is unique given our myopia assumption. We could instead assume F is uniform on $[0, q]$ but then unacceptable wages would be encountered, and the probability of disagreeable wages would vary between matches with black and white workers. Notice that F does not depend on worker type as due to assumption (C1) there are no wages the firm and β workers would accept but α workers would not. As equilibrium acceptable wages will form an interval, we could, instead of a uniform, use any distribution with connected, compact support by scaling it to the acceptable wage interval.

wage draw. As the latter are unstable to small perturbations in the parameters, we focus on the former two.

Consider the case where no employees are monitored: the white labor market. Matches never deteriorate and therefore the only source of job seekers is newly born workers. In this scenario, a firm just matched with a worker infers his probability of being of type α is the population prevalence g; the chance of a white job-seeker being good at a job to which he is matched is therefore

$$\theta_W = \theta_0 = g + (1-g)\beta.$$

Now suppose that all newly hired black employees are monitored and all bad matches are terminated. Newly matched black workers will be worse than average.

Lemma 1 *The probability a newly hired black worker is in a good match is*

$$\theta_B = \frac{\beta}{\beta g + (1-g)} < \theta_W. \tag{2}$$

Proof. See A.1 ■

Therefore, although monitoring may be individually prudent for each firm, it creates a negative externality by feeding a stream of workers who are worse than the population average (i.e. containing more β types) back into the job-seeker pool. Surprisingly, the steady state θ_B of this process does not depend on the rate of information λ, the worker matching rate μ, or the rate at which new workers enter the market.[16]

2.6 Parametric Assumptions

Now we impose certain restrictions on the joint values of parameters sufficient to ensure the existence of both steady states.

For an equilibrium with no monitoring to exist for white workers, we want to assume that monitoring costs are not too low. For monitoring not to be

[16]This is an artifact of the assumption that workers are infinitely lived.

optimal, the instantaneous monitoring cost must not be worth paying to detect bad matches, accounting for the fact that the cost must be recouped on the surviving fraction of workers.

$$(C2) \qquad \underbrace{\frac{b}{\lambda}}_{\text{Monitoring cost}} > \underbrace{(1-\theta_W)\frac{\lambda c}{r}}_{\text{Reduction in losses to errors}} \cdot \underbrace{\theta_W}_{\text{Proportion of remaining workers}}$$

Our second condition, antisymmetrically to (C2), posits that "monitoring costs must not be too high" and ensures that all new black employees will be monitored in equilibrium. As the monitoring decision is increasing in the wage, for there to be no monitoring in the black labor market, a condition is needed at the lowest wage in that market. As the lowest wage in the market depends on the speed at which workers match rather than simply the firm's break-even point, the relevant expression is a bit more complex.

$$(C3) \qquad \frac{b}{\lambda} < (1-\theta_B)\,\frac{\lambda c\,(\theta_B\mu + 2r) - 2rq}{r\,(\mu + 2r)}$$

In other words, $\theta_{B,}$ the belief about the average ability in the black unemployed pool, must be sufficiently low that acceptably high wages are only possible if the firm monitors. Strictness of the inequality ensures that switching to an unchurned market is not simply a matter of switching equilibria (as the non-monitoring one will not exist here). Notice that as the matching frictions vanish (μ increases), (C3) becomes the opposite inequality of (C2) for θ_B rather than θ_W.

Finally, for both labor markets to exist, it must be that workers can be in expectation gainfully employed; a sufficient condition is that the expected non-monitoring product of workers drawn from the black unemployed pool is positive:

$$(C4) \qquad q - \lambda c(1-\theta_B) > 0.$$

3 Solution

First, we will use the model's properties to characterize the firm's and worker's actions. The main intuition behind the following result is that the firm is more willing to monitor if the bad matches terminated by monitoring are costlier, due to higher wages.

Lemma 2 *The firm's monitoring decision is increasing in the wage w and decreasing in its belief about match quality θ.*

Proof. Fix a wage w and a belief θ for the firm. If the firm doesn't monitor the worker, it receives the production net of the wage and the expected cost of errors, forever:

$$V^w_{\theta,N} = \frac{q - w - (1-\theta)\lambda c}{r}. \tag{3}$$

On the other hand, the firm can choose to monitor the worker. If it is optimal to do so at any instant, it is optimal to do so until the signal arrives as the problem doesn't otherwise change. With probability θ, the match is good and the production net of wages $q - w$ is received by the firm forever as no separation will occur; with the complementary probability the match is bad so $q - w - \lambda c$ is received by the firm only until the signal arrives and the match ends; in either case, the monitoring cost of b is paid until revelation. The firm's expected lifetime payoff if it monitors is therefore

$$V^w_{\theta,M} = \theta \frac{q-w}{r} + (1-\theta)\frac{q - w - \lambda c}{\lambda + r} - \frac{b}{\lambda + r}. \tag{4}$$

For monitoring to be optimal for the firm, we need

$$V^w_{\theta,M} \geq V^w_{\theta,N} \tag{5}$$

which reduces to

$$w \geq q - \lambda c + \frac{rb}{\lambda(1-\theta)}, \tag{6}$$

or equivalently

$$\theta \leq 1 - \frac{rb}{\lambda(w - q + \lambda c)}. \tag{7}$$

■

Our next result shows that the wages acceptable to both the firm and the workers form an interval.

Lemma 3 *For a labor market with steady state expected match quality* θ*, there is an interval of wages,* $[\underline{w}_\theta, \overline{w}_\theta]$*, the worker and firm both accept.*[17]

Proof. see A.2 ■

An intervalic structure for the wages in each market will simplify analysis significantly. From Lemma 3 we have that the mutually acceptable wages are an interval $[rU_\theta^\alpha, \overline{w}_\theta] = [\underline{w}_\theta, \overline{w}_\theta]$. As α workers never separate once they find a job, we have that the lowest wage is equal to the expected wage they'd get at another firm, adjusted for search time: $\underline{w}_\theta = \frac{\mu}{\mu+r}\int_0^q wdF = \frac{\mu}{\mu+r}[.5\underline{w}_\theta + .5\overline{w}_\theta]$, so that

$$\underline{w}_\theta = \frac{\mu}{\mu + 2r}\overline{w}_\theta. \tag{8}$$

We now present the main theoretical results of the paper: existence and uniqueness of equilibria in the two markets that perpetuate their associated steady states.

3.1 The Non-Monitored Market

Proposition 1 *Assuming (C1)-(C4), the white (non-churned) labor market has a unique solution where the monitoring technology is not used. The average wage in this market is*

$$w_{\theta_W}^{avg} = \frac{\mu + r}{\mu + 2r}\left[q - (1 - \theta_W)\lambda c\right]. \tag{9}$$

Proof. see A.3 ■

The main intuition for the proposition comes from Lemma 2. Since the value of monitoring is increasing in w, for a non-monitoring solution we need

[17] Incentives are weak at the interval's endpoints, but this is immaterial as F will put zero probability on them.

only check whether the firm chooses to monitor at the break-even wage $\overline{w}_{\theta_W}$. And (C2) ensures that monitoring does not occur at that wage.

Interestingly, since the firm cannot learn the worker's type in this non-churned equilibrium, type has no effect on wages.

3.2 The Monitored Market

Here, as workers are monitored, β workers sometimes face separation and therefore have a low outside option. However, they cannot accept low wages at which monitoring would not occur at beliefs θ_B without revealing their type; thus, such wages are not accepted by the firm. Therefore this equilibrium is effectively a pooling one as well, despite the fact β workers receive significantly lower utility than α workers.

Proposition 2 *Assuming (C1)-(C4), the black (churned) labor market has a solution where the monitoring technology is used in every match. The average wage in this market is*

$$w_{\theta_B}^{avg} = \frac{\mu + r}{\mu + 2r}\left[q - \frac{r(\lambda c(1-\theta_B) + b)}{\lambda\theta_B + r}\right]. \tag{10}$$

Proof. see A.4 ■

The intuition here again comes from (2), which tells us that the monitoring decision is increasing in w and therefore if monitoring occurs at $\underline{w}_\theta$ it occurs at all matches, and (C3) which ensures this condition holds. As the equilibrium strategies induce monitoring at every equilibrium wage, employees who are revealed to be in bad matches separate from the firm. This sends only β workers back into the job-seeking pool, churning the market quality to θ_B.

4 Implications for Labor Markets

The previous sections establish conditions under which there are two distinct steady-states of the labor market. In this section, we compare labor market outcomes for workers in these steady states. We first discuss a prediction that has not previously been tested and then discuss the relation of our other predictions to known labor market regularities.

4.1 Employment Duration

Absent monitoring, there is no new information to dissolve the match. Therefore, taken literally, the model implies no turnover in the white equilibrium. In contrast, with monitoring, some workers prove ill-suited for the job and return to the unemployment pool. We interpret this as predicting that black workers will have lower average employment duration. Recall that workers who return to the unemployment pool are all type β. Therefore, turnover is even higher than if only new entrants were monitored. The model, again taken literally, implies that the separation hazard for blacks is

$$h_t = \frac{(1-\beta)(1-g)\,\lambda e^{-\lambda t}}{1-(1-\beta)(1-g)\,e^{-\lambda t}} \tag{11}$$

which is decreasing in t.

Importantly, h declines with t and asymptotes to 0, the hazard rate for whites. We expect this prediction to be robust to consideration of important real world elements not addressed by the model. Whether the hazard rates, in fact, converge is not something we are aware of the literature addressing and is the subject of our empirical investigation later in this paper.

As our model abstracts from firm-to-firm hiring, we have no prediction with regard to it. Although it may seem that firms would be out to poach black workers with high seniority (that are likely to have passed monitoring), adverse selection effects (with the worst workers more willing to leave) could unravel such effects, depending on the ability of outside employers to commit. Still, our predictions are in terms of employer-initiated separations, not moves to better jobs. Therefore, in the empirical section, we treat spells that end in a job-to-job transition as censored.

4.2 Wages

As $w^{avg}_{\theta_B} < w^{avg}_{\theta_W}$, blacks, on average, earn less than whites. The highest wage firms are willing to pay is lower for blacks than for whites since the average quality of new hires is lower, and the lowest wage blacks are willing to accept is lower because they expect other employers to pay less, as well. Interestingly, because whites are not monitored, their lifetime utility does not depend on type, and both types have higher utility than black αs who, in turn, have higher lifetime utility than black βs:

$$U^{\alpha}_{\theta_W} = U^{\beta}_{\theta_W} > U^{\alpha}_{\theta_B} > U^{\beta}_{\theta_B}.$$

It is less obvious whether the wage distributions of blacks and whites will overlap, and for some parameter values they do not. Figure 2 illustrates an example of the model where the equilibrium wages for white and black workers significantly overlap, despite the fact only the latter are monitored.

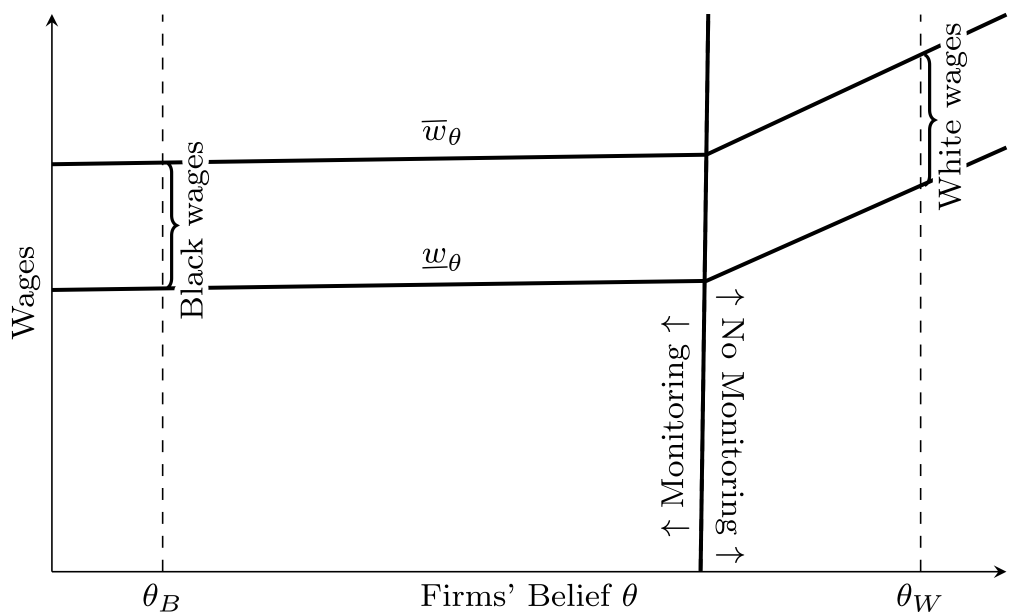


Figure 2: An example where the black and white wage ranges overlap. $r = .05,\ g = .99,\ \mu = 2,\ \beta = .1,\ q = 1,\ b = 2,\ \lambda = 2,\ c = 4$

4.3 Persistence of Discrimination

A key result of the churning mechanism in this paper is that deleterious steady states are persistent. In this section we show just how hard it is to transition to a good steady state. We regard this as illustrating the difficulty of addressing labor market discrimination in the context of policy, particularly policy aimed at improving the skills of black workers. The existence of a range of g values for which both steady states exist allows us to talk about *persistence* of the deleterious equilibrium.

Heretofore we have assumed that average skill levels for the two population groups are identical. Suppose instead that skill levels are $g_B \neq g_W$ and that the steady-state of each market provides monitoring for black but not white workers. Monitoring will persist as the equilibrium in the black labor market until g_B rises above some critical level while the no monitoring equilibrium will persist in the white market provided that g_W remains above a lower critical level. In principle, we can have the black workers in the bad equilibrium and the white workers in the good equilibrium provided that (C2) and (C3) hold for θ_W and θ_B calculated using g_W and g_B respectively. Put simply, this means that discrimination in wages and monitoring (and therefore also separations) can continue even if black workers are significantly better, on average, than white workers.

4.4 Unemployment Duration

We have so far treated the workers' matching rate, μ, as exogenous. Making the standard assumption of free entry, we now allow firms to post and maintain vacancies at a cost k per unit time. When a firm creates a vacancy, it can direct its search. This can take several forms, most notably locating production operations in an area with specific population characteristics or advertising the vacancy in different areas and through different media. In general, a firm can target markets indexed by i where a proportion ρ_i of unemployed workers are white. The open vacancy cost k is invariant to this target choice. We assume that in each market i the bargaining equilibria and population group steady states break down along the discriminatory lines described so far.

Define ϕ as market tightness and let the worker job-finding rate function follow the commonly assumed form

$$\mu(\phi) = m\phi^{\gamma} \tag{12}$$

for constants $m > 0$ and $0 < \gamma < 1$. Note that if firms expect a match to be worth V, the free-entry level of ϕ in such a market sets

$$\frac{\mu(\phi)}{\phi}V - k = 0 \tag{13}$$

so that

$$\phi = (\frac{Vm}{k})^{\frac{1}{1-\gamma}}. \tag{14}$$

Therefore, ϕ is an increasing function of V.

Assuming the parametric assumptions hold for the entire breadth of derived matching rates, we can now derive the free-entry equilibrium level of μ_{ρ_i} for each market i. The payoff to a firm for matching is the same as for an α worker, that is, when hiring from pool i, the firm expects a successful match to pay

$$V_i = \rho_i \frac{1}{\mu + 2r}\left[q - \lambda c(1 - \theta_W)\right] + (1 - \rho_i)\frac{1}{\mu + 2r}\left[q - \frac{r(\lambda c(1 - \theta_B) + b)}{\lambda\theta_B + r}\right] \tag{15}$$

The above expression is increasing in ρ_i. Therefore, for the same μ, markets with more black workers will have a lower expected payoff for a filled vacancy. Therefore, the free-entry $\phi(\rho_i)$ and $\mu(\phi(\rho_i))$ are increasing in ρ_i, so that workers searching for jobs in markets with a higher proportion of black workers take longer, on average, to find employment. A first-order stochastic dominance argument can then show that black workers take longer, on average, to find employment, and receive lower average wages.

4.5 Extensions

4.5.1 Eventual revelation in all matches

We have assumed unrealistically that the match quality of workers who are not monitored is never revealed. More plausibly, heightened scrutiny speeds the rate at which match quality is revealed. In a model in which workers live forever, this change considered in isolation would eliminate our results because the composition of the jobless pool is independent of the rate at which bad matches are revealed. However, if workers do not live forever, then reducing the rate at which match quality is revealed does affect the quality of the unemployment pool, and our basic results go through.

4.5.2 Skill level and discrimination

Further, we can allow for observable heterogeneity among workers. If there are groups of workers for whom g is high, only the no-monitoring equilibrium will exist for these groups, regardless of race. This is also true at very low g and very low β (although we have assumed away this case to simplify the proofs). The first result is consistent with similar outcomes for blacks and whites with high levels of skill as measured by education or the Armed Forces Qualifying Test (Neal and Johnson, 1996; Lang and Manove, 2011). The latter is consistent with some evidence that the bottom of the labor market is similarly bad for blacks and whites. On the other hand, Lang and Manove find that the market learns the productivity of white but not black high school dropouts. This is consistent with an equilibrium in which white unemployed dropouts are, on average, more skilled than black unemployed dropouts and therefore in which white but not black dropouts are monitored. Nevertheless, without additional, largely *ad hoc* assumptions, this story cannot account for the very high unemployment rate among black dropouts.

4.5.3 Investment in unobservable skills

We have heretofore postulated that the proportion of α types is exogenous. Assume instead that some fraction of workers are innately of type α. Others can transform themselves from βs into αs at some cost ω with cdf $F(\omega)$. Provided that the fraction of natural αs satisfies (C2) and (C3), both equilibria will continue to exist for the right choice of F. However, since in the no-monitoring equilibrium αs and βs receive the same wage, there is no incentive to invest in becoming an α. In contrast, in the monitoring equilibrium, lifetime earnings are strictly higher for αs than for βs. Thus, some individuals will have an incentive to make the investment.[18] This prediction contrasts with Coate and Loury (1993), where black workers are less willing to invest in skills.

4.5.4 Education

Suppose now that there exists a signal,[19] which we identify with education, that α workers can purchase at some personal cost $\kappa \sim F$. Assume doing so ensures that any employer will be immediately aware that the worker is indeed type α. An educated α worker of either population with cost κ will then anticipate a lifetime utility of $V_{Educ}(\kappa) = \frac{\mu}{\mu+2r}q - \kappa$. In Section 4.2 we showed white α workers receive a higher lifetime payoff than their black counterparts; therefore, the incentive for the latter to invest in education is greater. As this implies that $\bar{\kappa}_W \equiv max\{\kappa : V_{Educ}(\kappa) \geq U^{\alpha}_{\theta_W}\} < max\{\kappa : V_{Educ}(\kappa) \geq U^{\alpha}_{\theta_B}\} \equiv \bar{\kappa}_B$, we must have that $F(\bar{\kappa}_W) < F(\bar{\kappa}_B)$ and therefore more black workers will purchase education. In particular, there exists some range of idiosyncratic costs for which black workers will purchase education but white workers will not. This is consistent with the finding in Lang and

[18]It might appear that the incentive to undertake such investments would unravel the monitored equilibrium. However, if this were the case, no worker would have an incentive to invest. This raises messy dynamic issues which we sidestep by assuming that the fraction of additional workers who would choose to invest is insufficient to overturn (C3).

[19]We analyze the case of a pure signal. If education can also turn a β into an α, the analysis is a combination of the analysis in this and the prior subsection since productive investment increases the fraction of workers who are α but investment that reveals workers to be α reduces the fraction of unrevealed workers who are α.

Manove (2011) that, conditional on past test scores, blacks get more education than whites do. The intuition here is simple; if a worker of high skill is treated as if she has the average hire's skill for her group, she has a greater incentive to reveal her high skill if that average is lower.[20]

Perhaps equally importantly, this extension suggests that blacks and whites with high observable skills will have similar outcomes as discussed in the previous subsubsection.

4.5.5 Imperfect monitoring

We show in Cavounidis and Lang (2015) that one can write a very similar model in which β workers always match badly but monitoring can result in false positive good matches. Much of the analysis would remain unchanged in such a model. Parameters would exist that would force monitoring on blacks but not whites, the black labor market would churn, and it would produce lower employment duration, higher unemployment duration and lower lifetime wages for blacks. In this formulation, black workers succeeding at monitoring would only be as good as whites who had never been monitored; therefore a churned market does not necessarily produce better long-run matches for the successfully monitored.

However, this alternate model would imply that some workers are purely parasitic and cannot be matched well, but rather aspire simply to find a job where their lack of productivity is undiscovered.

4.5.6 Stigma and degeneration into lower-skilled jobs

Our model assumes unrealistically that employers have no information regarding the time that workers have been in the labor market or the number of jobs they have held. If the other aspects of our model were a rough representation of reality, it is implausible that firms would not recognize that some workers were unlikely to be new entrants and therefore very likely to be β types. Then

[20] Strictly speaking, this creates a feedback loop from lower wages for the uneducated to a greater measure of education. The right assumptions on F rule out associated complexities.

workers who do not find a good match sufficiently quickly will be permanently barred from the monitoring sector.

Somewhat more formally, as an extension to the model, we can relax the assumption that past history is entirely unobservable. Assume instead that each separation has a probability ζ of becoming public common knowledge. Any worker who has a revealed separation is known to be of type β in any new match. Thus, a newly hired worker who does not have such a stigma will be of average quality $\theta'_B = [\beta + g\zeta(1-\beta)] \,/\, [g\beta + (1-g) + g\zeta(1-\beta)]$. If we assume θ'_B satisfies (C3), churning can persist but will be primarily a phenomenon for relatively young workers.

But what will happen to workers revealed to be βs? It is straightforward to extend the model to allow for a second occupation type (q', c') lacking monitoring technology[21] that is less skill intensive than the task described so far, i.e. $q > q'$ and $q - \lambda c < q' - \lambda c'$. Unrevealed β types can be strictly better off than revealed ones in a new match of the first task, but revealed βs are forced to enter the market for the second occupation.

In this scenario, a fraction of black workers are relegated to low-wage jobs while white workers with similar skills can always get better jobs.

4.5.7 Changing screening and monitoring technology

Autor and Scarborough (2008) examine the effect of bringing in a new screening process. They find that the screening process raised the employment duration of both black and white workers with no noticeable effect on minority hiring. In our model, we can think of this technology as allowing the firm to screen for job match quality prior to employment. This increases the proportion of hired blacks who become permanent workers since some bad matches are not hired. If the screening mechanism is good enough, the firm will choose not to monitor the black workers it hires, and all black workers will be permanent. Formally, since all white workers are permanent in the absence of the screen, the screen does not affect this proportion. Informally, if poor matches

[21] Or, more palatably, the same technology but without the incentives to use it, as in the case of a small enough c'.

are more likely to depart even without monitoring, then there will also be positive effects on white employment duration.[22] Similarly, Wozniak (2015) shows that drug testing increases black employment and reduces the wage gap; we interpret this as confirming evidence for the notion that employers are more uncertain about the quality of black workers, and therefore that black workers benefit more from early resolution of such uncertainty.[23]

We note that improved technology appears to have reduced monitoring costs. This is unambiguously good for blacks who share the cost of being monitored. Unless the reduction shifts whites into the monitoring equilibrium, they are unaffected by the cost reduction. However, if firms begin monitoring, α workers and firms will initially be better off. Firms will be able to better screen their workers, and as a consequence can offer higher wages, which should make α workers better off as the monitoring does not put them at risk. On the other hand, β workers will generally be worse off. In a collective bargaining setting, the union might resist monitoring. The more interesting point is that since monitoring creates an externality, it is easy to develop an example in which monitoring makes both types of workers and capital worse off in the long run.[24]

4.5.8 Job-to-job transitions

There are many ways to incorporate job-to-job transitions in a search model. Some would not alter our results qualitatively, and others would, while yet others would produce a multiplicity of equilibria. Of particular import is the

[22] Formally, the model would have to be modified to ensure that some β workers are never perfectly matched and/or that some β workers are still in bad matches when they exit the labor force.

[23] Wozniak (2015) is not to be interpreted as evidence that monitoring is good for black workers on the aggregate. As in the present paper, it can beneficial on an individual level (as it allows good workers to get higher wages than otherwise); our model, however, shows it can also create a worse externality.

[24] Suppose that g_0 is just sufficient to sustain a no-monitoring equilibrium. A small reduction in b puts the labor market into a monitoring equilibrium. Initially, α workers and firms would experience a slight gain, but the churning will wipe this out and more. Firms always make zero profit on vacancies, but if we allow for a distribution of vacancy costs, then the rents earned by firms with low costs of creating vacancies will also fall.

information accessible to outside firms. Under the simplest assumption, that of information symmetry across firms, an outside firm that meets an employed worker learns whether monitoring has concluded successfully (or, alternatively, the worker could disclose this). Keeping with the wage offer structure, such an outside firm and worker pair would draw a wage uniformly among all mutually acceptable wages. Wages would jump on transition, but our predictions regarding average wages by race would be preserved, conditional on any of age, experience, or job spell duration. The monitoring state of the worker would persist from firm to firm (as incentives to monitor increase in the wage, but white workers are not monitored even at firms' break-even wage) so that our predictions regarding employment spell duration go through.

5 Empirics

We test the model's prediction that the hazard into nonemployment is initially higher for blacks but converges to that for whites. To our knowledge this prediction has not previously been tested. The model also predicts longer unemployment durations and lower lifetime incomes for blacks relative to whites. These are known to be strongly empirically supported (see Lang and Lehmann 2012).

5.1 Methods

We estimate the hazard of entering a period of nonemployment using the first fulltime spell of each individual at each employer. Because the model is about separation into nonemployment, not job to job transitions, we censor spells ending in a job to job transition. For these spells, we assert we do not know when the spell would have ended in nonemployment. We use both nonparametric and semiparametric survival analysis methods to estimate the hazard.

First, using standard techniques, we calculate the hazard over time intervals with the intervals large enough not to require further smoothing. For each nonoverlapping time interval, $(t_{j-1}, t_j], j = 1...k+1$, we obtain the number of employment spells at the start, the number of spells ending in nonemployment (failures) over the interval, and the number of spells ending but not in nonemployment (censored). A conventional way of calculating the hazard in this setting is to assume that censoring and death times are uniformly distributed within each interval. The hazard at the midpoint m for each nonoverlapping interval is then :

$$\hat{h}(t_{mj}) = \frac{d_j}{[(t_j - t_{j-1})(Y_j - \frac{d_j}{2})]} \tag{16}$$

The variable Y_j is the number of spells at the start of the interval minus half of the spells censored over the interval, and d_j is the number of failures over the interval (Klein and Moeschberger 2003).[25]

Each interval must have at least one failure in order to calculate the hazard, and so we use smaller intervals at shorter durations when there is more data. We use intervals of 13 weeks until this size interval no longer includes a failure. Starting at week 546 we use intervals of 26 weeks, until this no longer includes a failure. Starting at week 832 we use intervals of 52 weeks. We calculate the hazard separately over these intervals for black and white workers. We obtain confidence intervals based on the estimated standard deviation of the hazard function at the midpoint of interval j, using that the number of failures in the interval is a binomial random variable.[26]

This nonparametric method does not allow controlling for covariates, and so we additionally plot baseline hazard functions for blacks and whites from a Cox proportional hazards model stratified by race. The stratified Cox model allows for different baseline hazard functions for blacks and whites, rather than assuming the baseline hazards for blacks and whites are proportional. As in the traditional Cox model, we constrain the coefficients on the covariates to be the same for blacks and whites.

We use the time intervals defined above as our measures of time, so that the baseline estimates do not require further smoothing. Using time inter-

[25] This is referred to in the literature as the life table method.

[26] The formula for the estimated standard deviation of the hazard is found in Klein and Moeschberger (2003) and used in STATA. A similar formula is derived in Gehan (1969).

vals rather than week as a measure of time creates more instances of failures occurring at the same "time", since time is now a larger unit.[27]

The baseline contributions we obtain from this model are the same as the Nelson-Aalen contributions in the case of no covariates, using the week intervals as a measure of time.[28]

Specifically, we estimate

$$h(t|W, \mathbf{Z}) = h_W(t) exp(\mathbf{Z}\gamma) \tag{17}$$

The variable W is an indicator for whether the individual is white and $\mathbf{Z}$ includes whether the individual had completed any college, indicators for geographic region (Northeast; North Central; South; West is omitted), whether the individual lived in an urban area, age, blue collar, non-blue and non-white collar occupation (white collar is omitted), and year fixed effects, all measured at the start of the spell, and the standardized AFQT score.

While we are able to use the default STATA program for our main hazard estimates, there is no built-in program to obtain bootstrapped standard errors for the baseline hazard contributions and so we describe the estimation in detail. Following Kalbfleisch and Prentice (2002), we obtain the baseline hazard contributions at each failure time by maximizing the likelihood function

$$\Pi_{i=1}^{k} \left[\Pi_{j \in D_i} \left(1 - \alpha_i^{\exp[Z_j(t_i)'\widehat{\gamma}]} \right) \Pi_{l \in R(t_i) - D_i} \alpha_i^{\exp[Z_l(t_i)\widehat{\gamma}]} \right] \tag{18}$$

where $1 - \alpha_i$ is the baseline hazard at each failure time $t_i, i = 1, ...l$, D_i is the set of individuals who fail at time t_i, and $R(t_i)$ is the set of individuals at risk of failing just prior to time t_i. Maximizing (18) with respect to α_i implies the maximum likelihood estimate of α_i is the solution to

[27] There are several methods for dealing with these ties, all requiring assumptions about the timing of these failures. We present results using the Breslow approximation, one of the conventional methods and the STATA default. This is based on the assumption that the subjects failed at different times, but we do not know the order.

[28] There are several estimators of the baseline hazard rate in a proportional hazards model. We use the estimator from Kalbfleisch and Prentice (2002), which is also the default in STATA.

$$\sum_{j \in D_i} exp[Z_j(t_i)\hat{\gamma}][1 - \alpha_i^{exp[Z_j(t_i)\hat{\gamma}]}]^{-1} = \sum_{l \in R(t_i)} exp[Z_l(t_i)\hat{\gamma}] \tag{19}$$

This is solved using an iterative procedure. Because these baseline contributions are functions of the estimated coefficients γ on the covariates, to obtain standard errors, we use a parametric bootstrap in which we draw 10,000 sets of coefficients using the variance-covariance matrix of our estimated coefficients and calculate the estimated baseline hazard contribution at each time using (19).[29] We use the standard deviation of these estimated contributions to form the confidence intervals.

For robustness, we determine the hazards at each week, rather than for an interval of weeks, and then smooth using a kernel-smoother and local linear smoothing. These methods require choices of kernels and bandwidths, and, in the former case, an approach to addressing bias in the boundary regions.

5.2 Data

We test the model's prediction using the NLSY79, a nationally representative sample of 12,686 individuals, 14-22 years old when first surveyed in 1979, with oversamples of blacks, Hispanics and poor whites. These individuals are surveyed annually through 1994, and biennially afterwards.

We construct employment spells using the Employer History Roster which greatly facilitates linking job spells across survey years, by assigning each job a unique identification number consistent across surveys.[30] We define employment spells as the first fulltime spell with each employer, defining fulltime as at least 30 hours per week. For each survey year in which an individual reported employment at a given employer, we collect the start and end week

[29]For the iterative procedure to solve (19), we use the initial value suggested by Kalbfleisch and Prentice (2002): $1-\alpha_{i0} = d_i\{\sum_{l \in R(t_i)} exp[Z_l(t_i)\hat{\gamma}]\}^{-1}$, where d_i is the number of failures over the interval.

[30]Information for jobs six through 10 reported in some of the early survey years may not have been added by NLSY to the roster due to difficulty recovering these data. This is unlikely to have a large impact on the results given these jobs are a small proportion of those ever reported, for a small proportion of individuals (National Longitudinal Survey of Youth 2019).

of employment with that employer reported in the survey. We construct the total length of the employment spell by grouping all consecutive fulltime spells at the employer across survey years.

We treat an employment spell as ending in nonfulltime employment if there is more than one week between the spell's end and the start of the next fulltime employment spell, reported at any other employer in any survey year. We treat all exits into nonfulltime employment as involuntary, regardless of the reason the spell ended (including the worker being fired, quitting, or losing a job because of a layoff). For robustness, we also show results from treating quits into nonfulltime employment as censoring the spell.

If a spell with an employer ends in a job-to-job transition we treat the employment spell as censored. We treat spells for which individuals report employment at the time of the interview but do not respond to the next survey as censored. Thus, if the individual reports working at the time she was last surveyed, that spell is censored.

We restrict the sample to non-Hispanic males and exclude spells in which the worker ever reports self employment or working for a family business. We further exclude individuals with missing start or end weeks for any fulltime spell, and individuals with fulltime spells that end before they begin.[31] These missing or unclear start and end weeks make it difficult to know whether any of the individual's spells end in nonfulltime employment, and so we exclude all spells for these individuals.

The Employer History Roster also shows weeks within a span in which the individual reported not working at the employer for reasons such as pregnancy, health issues, quits, and temporary layoffs. Since we do not know whether the spell is at risk of ending during the gap, our main specifications include only spells without within-job gaps. This sample restriction excludes about 5000 spells.

Because the survey is conducted every two years starting in 1994, we do

[31] Due to rounding week numbers, it is possible that the start week of the span is greater than the end week. In cases when the start week is up to two weeks greater than the end week, we replace the start week equal to the end week (National Longitudinal Survey of Youth 2019).

not know the values of some of the control variables in some years and must impute their values from adjacent years. The online appendix describes these imputations in detail. In order to avoid excluding individuals with missing values of the covariates, we include an indicator for whether the individual is missing the value of the covariate, and set the value of the covariate to zero. We define occupations as white, blue collar or neither based on Gibbons and Katz (1991).[32]

As Table 1 shows there are nearly 34,700 employment spells in the sample, which are the first fulltime employment spells at each employer for individuals in the sample. There are over 23,100 employment spells for white workers and nearly 11,600 employment spells for black workers. For both, the average spell duration is approximately 80 weeks. On average there are nearly seven employment spells for each white worker in the sample, and nearly eight employment spells for each black worker in the sample. While 55% of white employment spells end in nonfulltime employment, 63% of black employment spells end in nonfulltime employment. The table shows other differences between the average white and black employment spells, including age, education, AFQT, urban location, blue-collar versus white-collar occupations, and region. Importantly, the stratified Cox proportional hazard models will include these as covariates.

5.3 Results

Figure 3 shows the hazard estimates using bins. We plot the hazard function only at durations less than or equal to 1000 weeks because of small sample sizes with longer durations. We see evidence that a gap opens up by 39 to 52 weeks, with blacks substantially more likely than whites to separate. Although the estimates become imprecise at longer durations, after ten years there is little evidence of a gap in the separation rate.

Appendix Figure A2 shows nonparametric plots with hazards by week rather than larger bins, smoothed using kernel smoothers with various band-

[32]Agricultural and private household workers are coded as neither blue nor white collar. Details in the appendix.

Table 1: Summary Statistics

	White	Black
End in Nonemployment	0.55	0.63
	[.5]	[.48]
Duration	78.13	76.96
	[159.7]	[146.37]
Age at Spell Start	25.21	27.33
	[7.45]	[8.07]
Any College Completed at Spell Start	0.32	0.25
	[.47]	[.43]
AFQT (standardized)	0.3	-0.63
	[1.]	[.76]
Urban Location at Spell Start	0.74	0.85
	[.44]	[.36]
Region at spell start: Northeast	0.18	0.16
	[.39]	[.37]
Region at spell start: North Central	0.28	0.16
	[.45]	[.37]
Region at spell start: South	0.34	0.59
	[.47]	[.49]
Region at spell start: West	0.2	0.08
	[.4]	[.28]
Spells per Person	6.69	7.9
	[5.1]	[5.25]
Blue Collar Occupation at Spell Start	0.55	0.66
	[.5]	[.47]
White Collar Occupation at Spell Start	0.26	0.2
	[.44]	[.4]
Spells at Risk of Ending in Nonemployment at		
200 weeks	2220	1077
400 weeks	992	435
600 weeks	522	214
800 weeks	289	114
1000 weeks	159	66
Total Spells	23112	11580

Notes: Standard deviations in brackets.

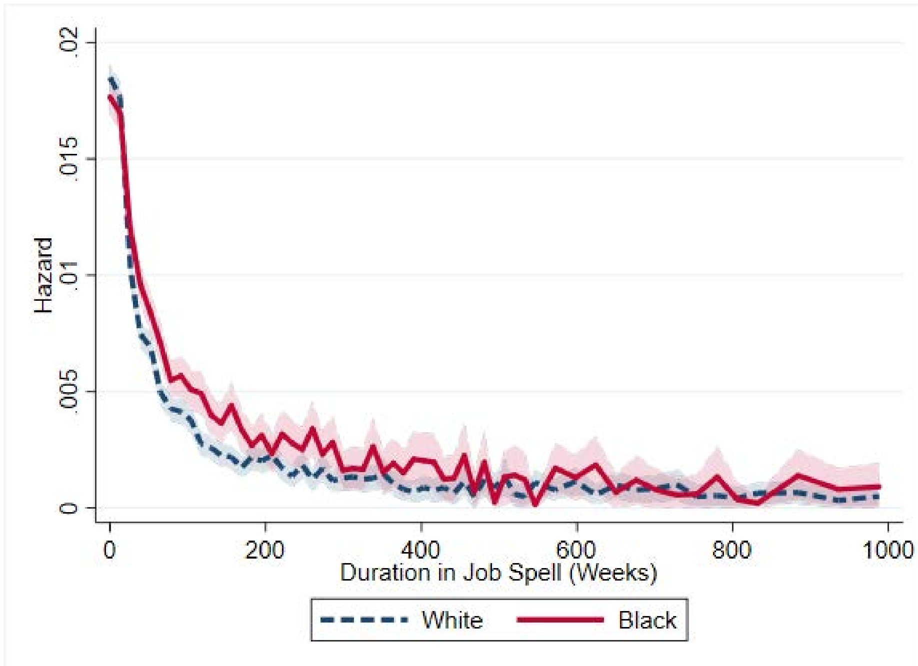


Figure 3: Nonparametric Estimates of the Hazard into Nonfulltime Employment by Week Bins: first fulltime spell of non-Hispanic males at each employer. Confidence intervals based on the estimated standard deviation of the hazard function at the midpoint of the interval, using that the number of failures in the interval is a binomial random variable.

widths and kernels, as well as local linear smoothing. Most, but not all, of these plots show a smaller or nonexistent gap in the first year followed by an opening of the gap, which closes by roughly year 10.

One potential reason for the initially higher hazard for white workers in Figure 3 is that white workers are more likely to quit in the early stages of their job due to better outside options. While this is not a part of our model, we estimate an additional specification in which we treat quits as censored rather than as failures.[33] As expected, Appendix Figure A1(a) shows this reduces the hazard rate and also suggests slightly higher hazards for black workers starting from the beginning of their employment spell. These results also suggest a widening of the gap between black and white workers followed

[33]See appendix for details on coding quits.

by convergence. We are inclined to believe that our model fits well past a normal probationary period of six months to a year but probably does not fully describe what is going on during the probationary period when all workers are being scrutinized and workers are trying to figure out whether they want to stay at the job.

In Figure 4, we present Cox proportional hazard estimates using the multi-week intervals as units of time. While the pattern is quite similar to that in Figure 3, the gap in the hazard rate opens up earlier when adjusting for individual-level covariates.[34]

The online appendix shows additional plots in which we estimate Cox regressions using week rather than larger bins, and smoothing hazard contributions using kernel smoothers with various bandwidth and kernels, as well as local linear smoothing. The results also show an initial gap in the hazards of black and white workers, which eventually becomes nonexistent.

We also estimated a Cox model with the same covariates, but modeled the percentage gap between the black and white hazards to be a cubic in seniority. To allow the effect of race on the hazard to vary over time, we include an observation for each job spell at each failure time in the data (as Cox models are only estimated when failures occur). We find that the gap at week 1 is a statistically significant 6.7 percent but initially grows (Appendix Figure A3). Over the range from 1 to 1000 weeks, the gap hits its maximum of 74 percent at 294 weeks (roughly five years and eight months). It then falls, ceasing to be significant at the 5% level at approximately twelve years and six months, and continues falling for durations through 872 weeks (nearly 17 years). At 872 weeks, the point estimate suggests a gap of 8.3%.

[34] Workers in blue-collar occupations at the start of the spell have higher hazards while those with at least some college, higher AFQTs, and who are older have lower hazards (details in the online appendix). The larger initial gap between blacks and whites in the Cox estimation than in the nonparametric estimation is importantly affected by controlling for age at start of spell and fixed effects for year of spell start. Because they have more spells, blacks on average are older at spell start. Controlling for year of spell start, older workers have lower hazards. As a result, controlling for the fact that blacks are on average older at spell start shifts up their baseline hazard relative to not controlling for this difference.

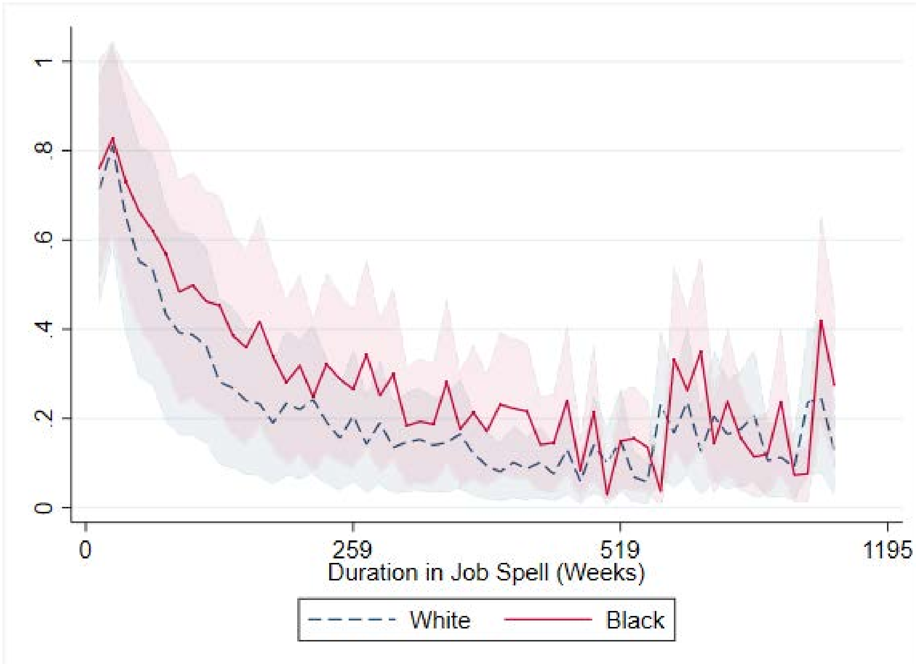


Figure 4: Estimates of the Hazard into Nonfulltime Employment by Week Bins, Based on a Cox Model Stratified by Race. Confidence intervals based on a parametric bootstrap of 10,000 sets of coefficients.

5.4 Robustness

The incentive to monitor new workers is relevant mainly for individuals hired out of nonemployment. For robustness, we identify employment spells for which the individual entered from nonemployment, and restrict to those spells. We continue to see that a gap in black-white hazards disappears over time with some suggestive evidence that convergence takes a longer time, more than ten years.

In the main specification, if there are any gaps in employment at an employer, the employment spells are treated separately, and we include only the first of these spells. As discussed above, there are two ways in which an individual can report gaps at an employer. The individual could report multiple employment spells at the same employer, reporting a start and end week for each span. Additionally, for any given spell the individual can report a within-job gap at the employer and the start and end week of that gap. For robustness, we treat gaps of less than or equal to 26 weeks as continuations of the same employment spell at the employer, but subtract the length of the gap from the duration. The appendix shows the hazard functions for this robustness specification look quite similar to the principal results.

Results also look similar when restricting to workers who are no more than 30 at the start of the spell. Finally, for robustness we treat transitions to part-time employment as censoring the spell rather than as spells ending in nonemployment. This has little effect on the results (see appendix).

6 Conclusion

We develop a model that predicts known disparities between blacks and whites: blacks earn lower wages, have longer unemployment duration, and obtain more education conditional on measured ability. It also predicts one previously unstudied disparity: the hazard for separation from employment into unemployment is higher for blacks at low tenure but the hazard rates converge as tenure increases. As we have argued previously, while the model, of necessity, relies on some special assumptions, the key elements are 1) that worker productivity is correlated across jobs, 2) that ability is neither perfectly observed or signalled and workers can to some extent hide past firings, 3) that firms therefore use race to statistically infer worker ability, 4) that additional information arrives during employment and is either imperfect, costly or both so that a worker's productivity can never be known perfectly at zero cost, and 5) that firms can and do act on new information by firing some workers.

The prediction is largely confirmed. In our stratified Cox models, conditional on observables, at the beginning of a job black workers are more likely to exit the job into unemployment. There is little evidence of a gap after roughly 10 years of seniority. Contrary to the model's prediction, many but not all of our results show that conditional on observables the gap is much smaller over roughly the first year on the job before growing. Obviously, it is up to the reader to decide how problematic this is. Our interpretation is that the first year is a period when all worker/firm pairs are discovering whether they are grossly mismatched, information that both the worker and firm receive for free. The mechanism we underline becomes increasingly important with tenure and dominates after this initial probationary period.

Our message is in some ways depressing. Simply addressing education or human capital disparities between blacks and whites need not eliminate labor market disparities. The 'bad equilibrium' in which many blacks find themselves is difficult to escape.

References

Autor, David H. and David Scarborough, 2008. "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics,* 123(1):219-277.

Becker, Gary S. 1971. *The Economics of Discrimination*, 2nd edition. Chicago, IL: Chicago University Press.

Bjerk, David. 2008. "Glass Ceilings or Sticky Floors? Statistical Discrimination in a Dynamic Model of Hiring and Promotion." *Economic Journal,* 118(530): 961-982.

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." *Journal of Labor Economics*, 13(2): 309-333.

Bowlus, Audra J., and Zvi Eckstein. 2002. "Discrimination and Skill Differences in an Equilibrium Search Model." *International Economic Review*, 43(2): 1309-1345.

Bowlus, A. J., Kiefer, N. M. and Neumann, G. R. 2001. "Equilibrium Search Models and the Transition from School to Work." *International Economic Review*, 42(2), 317-343.

Cameron, A. Colin and Pravin K. Trivedi. 2009. *Microeconometrics: Methods and Applications*, 8th edition. Cambridge, England: Cambridge University Press.

Cavounidis, Costas and Kevin Lang. 2015. "Discrimination and Worker Evaluation." *National Bureau of Economic Research Working Paper No. 21612.*

Charles, Kerwin, and Jonathan Guryan. 2008. "Prejudice and The Economics of Discrimination." *Journal of Political Economy*, 116(5): 773-809.

Coate, Stephen, and Glenn C. Loury. 1993. "Will affirmative-action Policies Eliminate Negative Stereotypes?" *American Economic Review,* 83(5): 1220–1240.

Coates, Ta-Nehisi. 2012. "Fear of a Black President." *The Atlantic.* Retrieved on January 12th, 2016 from: http://www.theatlantic.com/magazine/archive/2012/09/fear-of-a-black-president/309064/#

Epanechnikov, V. A. 1969. "Non-Parametric Estimation of a Multivariate Probability Density." *Theory of Probability and its Applications,* 14(1): 153-158.

Fairlie, Robert W. and William A. Sundstrom. 1999. "The Emergence, Persistence, and Recent Widening of the Racial Unemployment Gap." *Industrial and Labor Relations Review.* 52(2): 252-70.

Fryer Jr., Roland G. 2007. "Belief Flipping in a Dynamic Model of Statistical Discrimination. *Journal of Public Economics*, 91(5-6): 1151-1166.

Gehan, Edmund A. 1969. "Estimating Survival Functions From The Life Table." *Journal of Chronic Diseases.* 21: 629-644.

Guardian, 2014. "Prince: Black People Don't Get Second Chances," February 25, 2014, http://www.theguardian.com/music/2014/feb/25/prince-black-people-dont-get-second-chances, accessed January 24, 2015.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data.* 2nd ed. New York: Wiley.

Klein, J. P., and Ml. L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data.* 2nd ed. New York: Springer.

Lang, Kevin. 1986. "A Language Theory of Discrimination." *Quarterly Journal of Economics*, 101(2): 363-382.

Lang, Kevin and Jee-Yeon K. Lehmann. 2012. "Racial Discrimination in the Labor Market: Theory and Empirics." *Journal of Economic Literature.* 50(4): 959-1006.

Lang, Kevin, and Michael Manove. 2003. "Wage Announcements with a Continuum of Types." *Annales d'Economie et de Statistique.* 71-72: 223-244.

Lang, Kevin, and Michael Manove. 2011. "Education and Labor Market Discrimination." *American Economic Review*, 101(4): 1467-1496.

Lang, Kevin, Michael Manove, and William T. Dickens. 2005. "Racial Discrimination in Markets with Announced Wages." *American Economic Review*, 95(4):1327-40.

Lauermann, Stephan and Wolinsky, Asher. 2016. "Search with adverse selection. "*Econometrica*, 84(1):243-315.

Lundberg, Shelly J., and Richard Startz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review*, 73(3): 340-7.

Mabry, Marcus. 2007. *Twice as Good: Condoleezza Rice and Her Path to Power.* New York: Modern Times. Print.

MacLeod, W. Bentley, 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review,* 93(1):216-240.

Masters, Adrian. "Statistical discrimination from composition effects in the market for low-skilled workers." *Labour Economics,* 26 (2014): 72-80.

Moro, Andrea, and Peter Norman. 2004. "A General Equilibrium Model of Statistical Discrimination." *Journal of Economic Theory*, 114(1): 1-30.

National Longitudianl Survey of Youth. 2019. "NLSY 1979 Appendix 18: Work History Data. " https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-appendix-18-work-history-data, accessed, April 15, 2019.

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy,* 104(5): 869-95.

Peski, Marcin, and Balazs Szentes. 2013. "Spontaneous Discrimination." *American Economic Review*, 103(6): 2412-36.

Reid, Alvin. 2015. "Black Head Coaches Don't Get Second Chances," *St. Louis American,* Aug 31, 2015, http://newpittsburghcourieronline.com/2015/08/31/black-head-coaches-dont-get-second-chances, accessed, January 24, 2016.

Rosen, Asa. 1997. "An Equilibrium Search-Matching Model of Discrimination." *European Economic Review,* 41(8): 1589-1613.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50(1): 97-109.

Spencer, DeShuna. 2014. "For Black Youth in America, There Are No Second Chances," *emPower*, February 5, 2014, http://www.empowermagazine.com/black-youth-america-second-chances, accessed January 24, 2016.

Wozniak, Abigail. 2015. "Discrimination and the Effects of Drug Testing on Black Employment." *Review of Economics and Statistics*, 97(3): 548-566.

A Appendix

A.1 Proof of Lemma 1

Proof. Define the quantities

ξ	Flow mass of workers born per unit time
A	Mass of unemployed black type α workers
B	Mass of unemployed black type β workers
Λ	Mass of currently monitored black type β workers

As g is the fraction of new workers that is type α and unemployed α workers are becoming employed each at a Poisson rate μ and never separate, A obeys

$$\frac{dA}{dt} = \xi g - \mu A$$

Similarly, a proportion $(1-g)$ of new workers is type β and such unemployed workers are also being hired at a Poisson rate μ each. However, as Λ workers who are of type β are being monitored, a flow mass $\Lambda\lambda(1-\beta)$ of black β workers are separating after monitoring reveals a bad match are also coming in to the black unemployed pool. Hence, B obeys

$$\frac{dB}{dt} = \xi(1-g) - \mu B + \Lambda\lambda(1-\beta).$$

Finally, unemployed β workers are becoming employed with monitoring at a Poisson rate μ and once they are employed they cease being monitored when match quality is revealed, which occurs at a rate λ. Thus the mass of monitored black β workers Λ must satisfy

$$\frac{d\Lambda}{dt} = \mu B - \Lambda\lambda.$$

Steady state implies that

$$\frac{dA}{dt} = \frac{dB}{dt} = \frac{d\Lambda}{dt} = 0.$$

Solving, we obtain

$$A = \frac{\xi g}{\mu}$$

$$B = \frac{\xi(1-g)}{\mu\beta}$$

and therefore the proportion of α workers in the unemployed pool is

$$\frac{A}{A+B} = \frac{\frac{\xi g}{\mu}}{\frac{\xi g}{\mu} + \frac{\xi(1-g)}{\mu\beta}} = \frac{g}{g + \frac{1}{\beta}(1-g)}.$$

Thus, a new match from the black job-seeker pool is of average quality

$$\frac{g}{g + \frac{1}{\beta}(1-g)} \cdot 1 + \left(1 - \frac{g}{g + \frac{1}{\beta}(1-g)}\right) \cdot \beta = \frac{\beta}{\beta g + (1-g)} \equiv \theta_B.$$

As $\beta < 1$ this is less than θ_W. ■

A.2 Proof of Lemma 3

Proof. As we require that strategies are not weakly dominated and form a PBE, the worker must at every wage draw be using an undominated action. As the worker's action doesn't affect his payoff if the firm rejects the wage, but does if the firm accepts it, she must always act as though the firm will accept the wage. Then, it follows that α workers will accept the wage offer if

$$w \geq rU_\theta^\alpha. \tag{20}$$

On the other hand, β workers will accept wages of

$$w \geq rU_\theta^\beta \tag{21}$$

regardless of whether they think the firm is likely to monitor and possibly fire them.[35] As α workers can mimic the β acceptance rule and not suffer separation, we have $U_\theta^\alpha \geq U_\theta^\beta$.

There are thus no wages α workers accept that β workers do not. If $U_\theta^\alpha > U_\theta^\beta$ there are wages that β workers accept that α workers do not; but then the firm will assign probability 1 to a worker accepting such a wage being type β and by $(C1)$ refuse such a wage. Thus no wage below rU_θ^α is ever accepted by both parties. On the other hand, since all workers accept higher wages, accepting such a wage does not shift beliefs; therefore, firms will accept all wages higher than rU_θ^α at which they make profits when they believe θ. The requirement that the firm makes nonnegative expected profit corresponds to at least one of $V_{\theta,M}^w$ and $V_{\theta,N}^w$ being positive; as they both are decreasing and continuous in w, we have that there is a single upper cutoff

$$\overline{w}_\theta = \max\{w | \max\{V_{\theta,M}^w, V_{\theta,N}^w\} \geq 0\}. \tag{22}$$

The fact that w_{θ_B} and w_{θ_W} are positive follows from C4 and the fact that both $V_{\theta,M}^w$ and $V_{\theta,N}^w$ are increasing in θ. Accordingly the lower cutoff is $\underline{w}_\theta = rU_\theta^\alpha$. ■

A.3 Proof of Proposition 1

Proof. First, we show that in the white labor market, at the maximal wage $\overline{w}_{\theta_W}$, the employer does not monitor the worker. Suppose $w = q - (1-\theta_W)\lambda c$. Then we have

$$V_{\theta_W,M}^w = \theta_W \frac{q-w}{r} + (1-\theta_W)\frac{q-w-\lambda c}{\lambda + r} - \frac{b}{\lambda + r} \tag{23}$$

$$V_{\theta_W,M}^w = \frac{1}{r(\lambda+r)}\left[\theta_W(1-\theta_W)\lambda c - \frac{b}{\lambda}\right] \tag{24}$$

[35] Considerations about monitoring by other firms enter U_θ^β, but the worker's decision to accept a match does not depend on whether the firm will monitor.

From C2 this expression is negative. As $V^w_{\theta_W,M}$ is decreasing in w, we must have that $\overline{w}_{\theta_W} = q-(1-\theta_W)\lambda c$. Therefore, no monitoring occurs at the upper end of the equilibrium wage interval. From Lemma 2 we have that the employer's monitoring decision is increasing in w; therefore, there is no monitoring in the market with belief θ_W. Given that, $\theta_W = \theta_0$ is the resulting steady state belief about newly hired workers.

The average wage in the market is therefore

$$w^{avg}_{\theta_W} = .5\overline{w}_{\theta_W} + .5\underline{w}_{\theta_W} = .5\overline{w}_{\theta_W} + .5\frac{\mu}{\mu+2r}\overline{w}_{\theta_W} = \frac{\mu+r}{\mu+2r}[q-(1-\theta_W)\lambda c]. \quad (25)$$

■

A.4 Proof of Proposition 2

Proof. Consider the wage at which the firm would break even if it does not monitor, $\overline{w}^n_{\theta_B}$. It is given by

$$q-(1-\theta_B)-\overline{w}^n_{\theta_B} = 0 \quad (26)$$

$$\overline{w}^n_{\theta_B} = q-(1-\theta_B)\lambda c. \quad (27)$$

Assume for contradiction that this is the highest equilibrium wage. As the monitoring decision is increasing in w, it must also not monitor at the lowest equilibrium wage, which by (8) is $\underline{w}^n_{\theta_B} = \frac{\mu}{\mu+2r}\overline{w}^n_{\theta_B}$. Furthermore, we have from (6) that such a non-monitoring wage must satisfy

$$\overline{w}^n_{\theta_B} \geq q-\lambda c+\frac{rb}{\lambda(1-\theta)}. \quad (28)$$

But that implies

$$\frac{b}{\lambda} \geq (1-\theta_B)\frac{\lambda c(\theta_B\mu+2r)-2rq}{r(\mu+2r)}, \quad (29)$$

a direct contradiction to (C3). Therefore, we conclude that the firm would monitor at a wage of $\frac{\mu}{\mu+2r}\overline{w}^n_{\theta_B}$, that the highest equilibrium wage is the break-even monitoring wage $\overline{w}^m_{\theta_B}$, and that $\overline{w}^m_{\theta_B} > \overline{w}^n_{\theta_B}$. It then follows from the fact that $\frac{\mu}{\mu+2r}\overline{w}^n_{\theta_B}$ was a wage at which the firm would monitor and the increasing monitoring decision that the firm would monitor at $\frac{\mu}{\mu+2r}\overline{w}^m_{\theta_B}$, as well. Therefore, all workers are monitored and the average unemployed worker's quality is in steady state at θ_B.

The average wage in the back labor market is thus

$$w^{avg}_{\theta_B} = .5\overline{w}_{\theta_B} + .5\underline{w}_{\theta_B} = .5\overline{w}^m_{\theta_B} + .5\frac{\mu}{\mu+2r}\overline{w}^m_{\theta_B} = \frac{\mu+r}{\mu+2r}\left[q - \frac{r(\lambda c(1-\theta_B)+b)}{\lambda\theta_B + r}\right]. \tag{30}$$

■

A.5 Empirical Appendix

A.5.1 Constructing Individual Covariates

The stratified Cox estimation adjusts for whether the individual has completed at least some college at the time the employment spell begins. If the individual is surveyed in the year the employment spell begins, this is simply an indicator for whether highest grade completed that year is at least 13. If the individual is not surveyed in that year, we impute the value of this variable as described below.

If the highest grade completed is the same in the previous and subsequent surveys, we impute the value for highest grade completed in the survey of interest. We first impute the values for survey years based on previous and subsequent surveys. After imputing these, we impute the value for nonsurvey years based on previous and subsequent surveys.[36]

The indicator for completion of any college at the time the spell began will still be missing for individuals whose spell begins in nonsurvey years,

[36] This addresses the problem arising from missing highest grade completed in 1998, but nonmissing in 1996 and 2000. Imputing the nonsurvey years and the survey years together, the value in 1997 would still be missing since the next survey year is 1998. Imputing the survey years first implies that 1998 becomes nonmissing, and so then 1997 becomes nonmissing as well because both 1996 and 1998 are nonmissing.

and whose highest grade completed changes between the previous and the subsequent surveys. For these individuals, we impute the value of any college at the beginning of the spell using highest grade completed at the last survey before the start of the spell, and years since this survey. For example, if the highest grade completed at the last survey was at least 13, then we impute a value of one for any college at the beginning of the spell. If the highest grade completed at the last survey was 11, and there was less than or equal to one year between the last survey and the beginning of the spell, then the individual has not completed at least 13 years of education at the time the spell begins. We impute a value of zero.

If the individual is not surveyed in the year the spell begins, we determine the individual's geographic region at the beginning of the spell, and whether they live in an urban or rural location, using the value of these variables in surrounding years. We impute only if the value of these variables is the same in the previous and subsequent surveys.

We standardize AFQT scores so they have a mean of zero and standard deviation of one.

A.5.2 Kernel- and Local-Linear Smoothing to Obtain Hazard Estimates: Methods

The principal results use intervals of weeks to smooth the hazard estimates. For robustness, we use week as a unit of time, the smallest unit of time for which we know employment status. We obtain the steps (hazard contributions) of the Nelson-Aalen cumulative hazard, and smooth them using a kernel smoother.[37] We obtain the cumulative hazard separately for black and white workers.

The cumulative hazard at time t_j is denoted $H(t_j)$. Then, the hazard contributions at each time t_j in which some individual moves from employment

[37] The Nelson-Aalen cumulative hazard is $\hat{H}(t) = \sum_{j|t_j \leq t} \frac{d_j}{n_j}$, where d_j is the number of failures at time t_j and n_j is the number at risk of failure at time t_j. The steps of this function are equal to $\frac{d_j}{n_j}$.

to nonemployment are defined as:

$$\Delta\hat{H}(t_j) = \hat{H}(t_j) - \hat{H}(t_{j-1}) \tag{31}$$

We plot the smoothed hazard function separately for black and white workers

$$\hat{h}(t) = b^{-1}\sum_{j=1}^{D} K_t(\frac{t-t_j}{b})\Delta\hat{H}(t_j) \tag{32}$$

The hazard estimate at time t, $h(t)$, is based on the the hazard contributions $\hat{H}(t_j)$ at all failure times j, where each contribution is weighted by the kernel function K and bandwidth b. We use the Epanechnikov kernel, with the bandwidth equal to .5*Silverman's plug-in estimate.[38] This yields a bandwidth of 75 for black workers and 87 for white workers.

We use a boundary-adjusted Epanechnikov kernel (based on Müller and Wang (1994)) to address bias in the boundary regions ($t_{min} \leq t < b$; $t_{max} - b < t \leq t_{max}$) from using a symmetric kernel. As a further alternative to using a boundary-adjusted kernel, we show hazard estimates only outside the boundary region. For those results, we smooth the hazard contributions using the version of the Epanechnikov kernel described in Epanechnikov (1969).[39] This yields equivalent results to the version of the Epanechnikov kernel described in the paper if the bandwidth in the principal specifications in the paper is divided by $\sqrt{5}$. This implies the boundary region is smaller when using this kernel, which is helpful given we plot only outside the boundary. These plots show a clear gap at the boundary of roughly 35-40 weeks which closes over time.

We also smooth the hazards using kernel-weighted local linear regression

[38] The version of the Epanechnikov kernel we refer to here is $K[z] = .75(1 - z^2)$ if $|z| < 1$. Silverman's plug-in estimate is given by $b^* = 1.3643\delta N^{-.2}\min(s, iqr/1.349)$ where δ depends on the kernel and is 1.7188 for the Epanechnikov kernel, N is the number of unique failure times, s is the sample standard deviation of the failure times and iqr is the interquartile range of the failure times. Cameron and Trivedi (2005) suggest using Silverman's plug-in estimate, as well as bandwidths half and twice the size. Because Silverman's plug-in estimate for the bandwidth is quite large, we present estimates using bandwidths half the size as well as the actual Silverman plug-in estimate.

[39] $K[z] = .75(1 - \frac{1}{5}z^2)$ if $|z| < \sqrt{5}$

which does not yield biased estimates in the boundary regions (see Nielsen and Tanggaard 2001 and Cameron and Trivedi 2005).

We additionally estimate Cox models stratified by race, and use the same Epanechnikov kernel and bandwidth as in the nonparametric specifications to smooth the baseline hazard contributions from the Cox model. These models include the same covariates as in the principal results.

A.6 Coding Quits

The respondents' choices for why they left their job change with the surveys. For most years the reasons why the respondent left their job include several which are explicitly named "quits" for example quit for pregnacy or family reasons or quit to look for another job, and several which are clearly not quit related. In the first survey year, as well as the surveys starting in 2002, there are several additional categories that do not include the word "quit" that we code as quits as they are arguably voluntary separations.

In 1979, these include "Pregnancy" and family changed jobs or moved, or family reasons. In years after 2002, these include "transportation problems" "no desirable assignments available" "retirement" "job assigned through a temp agency or a contract firm became permanent" and "dissatisfied with job matching service". There are few instances of spells ending for these reasons.

A.7 Blue- and White-Collar Workers

We use data on occupation from the Employer Roster. Up until 2000, the 1970 Census occupation codes are used to define blue- and white-collar workers. We define the following as white collar: managers, officials, and proprietors; professional, technical, and kindred; clerical and kindred; sales workers. The following are defined as blue collar: craftsmen, foremen, and kindred; operatives and kindred; laborers, except farm; service workers, except private household. Starting in 2002, the 2000 Census occupation codes are used. The following codes are defined as white collar: management; business and financial operations; computer and mathematical; architecture and engineering;

legal; education, training, and library; arts, design, entertainment, sports, media (except equipment workers); healthcare practitioners and technical; sales and related; office and administrative support; life, physical, and social sciences; community and social services. The following are defined as blue collar: healthcare support; protective service; food preparation and serving related; building and grounds cleaning and maintenance; personal care and service; construction and extraction; installation, repair, and maintenance; production; transportation and material moving; arts, design, entertainment, sports, and media (only equipment workers).

There are several individuals in the main sample who have employment roster data, but are not in the occupation roster data. Other individuals are in the occupation roster data, but do not have a valid occupation code. Both of these groups of individuals are coded as missing occupation, and have a value of zero for the indicators for whether they are blue- or white-collar workers, or neither blue- nor white-collar workers. Through 2002, invalid occupation codes include those less than or equal to zero, or greater than or equal to 995. Starting in 2004, invalid occupation codes include those less than or equal to zero, or greater than or equal to 9950.

There is one individual in the main sample with the same week intervals at the same employer, reported in two consecutive survey years, and who reports a valid occupation code in one year but not the other. We take the observation with the valid occupation code. For this individual, in one year the reported start week is one week greater than the end week. As discussed in the paper, we set the start week equal to the end week for these observations because of the possibility of rounding week numbers. After making this adjustment, the start and end weeks are the same across the two consecutive survey years for this individual.

References

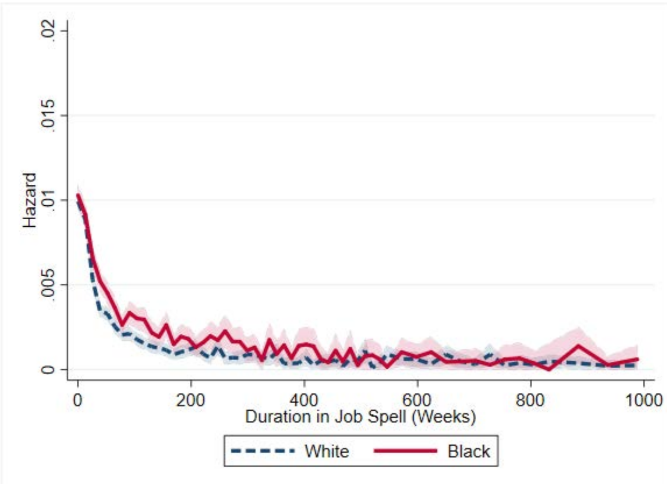
Cameron, A. Colin and Pravin K. Trivedi. 2009. *Microeconometrics: Methods and Applications*, 8th edition. Cambridge, England: Cambridge University Press.

Epanechnikov, V. A. 1969. "Non-Parametric Estimation of a Multivariate Probability Density." *Theory of Probability and its Applications*, 14(1): 153-158.

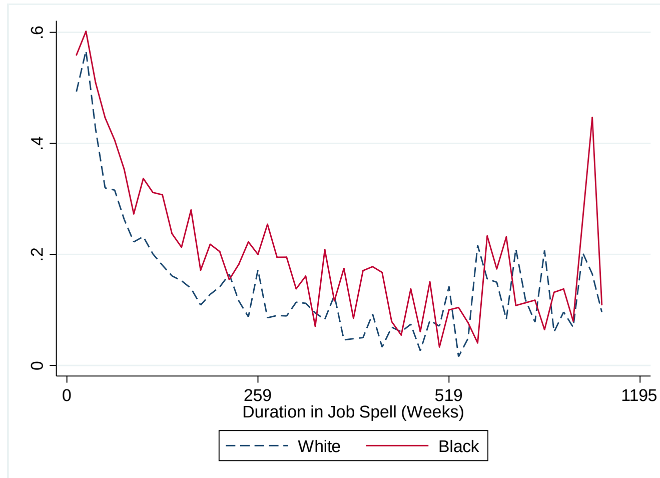
Nielsen, Jens Perch and Carsten Tanggaard. 2001. "Boundary and Bias Correction in Kernel Hazard Estimation." *Scandinavian Journal of Statistics*, 28: 675-698.

Appendix Figure A1: Robustness Hazard Estimates using Week Bins

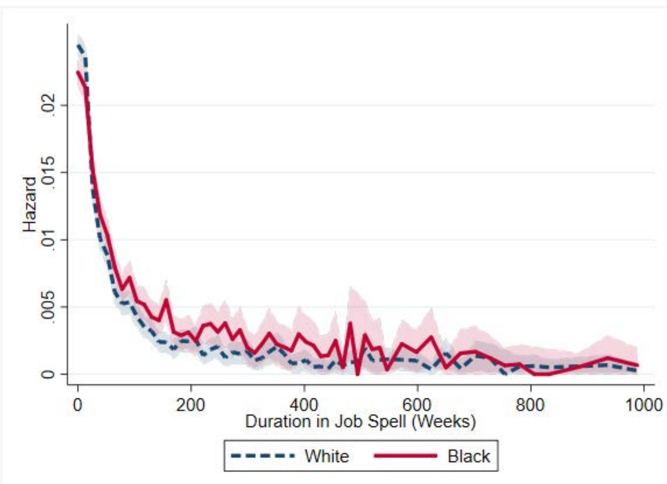
(a) Quits as Censoring the Spell, Nonparametric



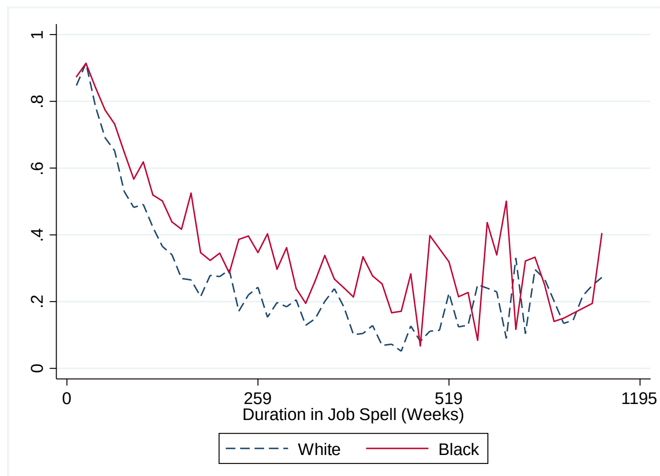
(b) Quits as Censoring the Spell, Cox



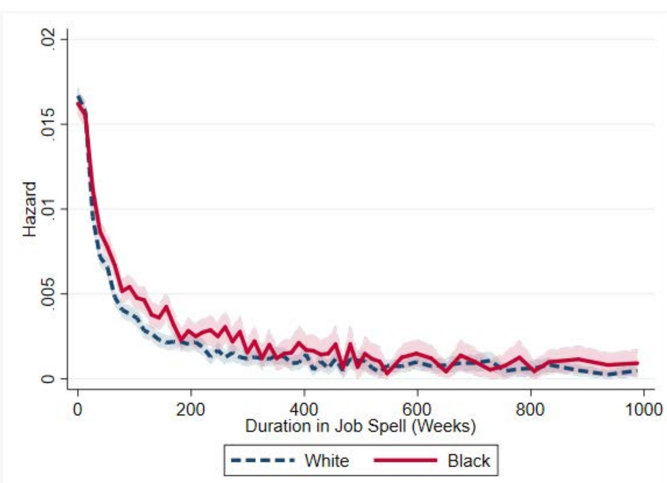
(c) Enter Spell from Nonemployment, Nonparametric



(d) Enter Spell from Nonemployment, Cox

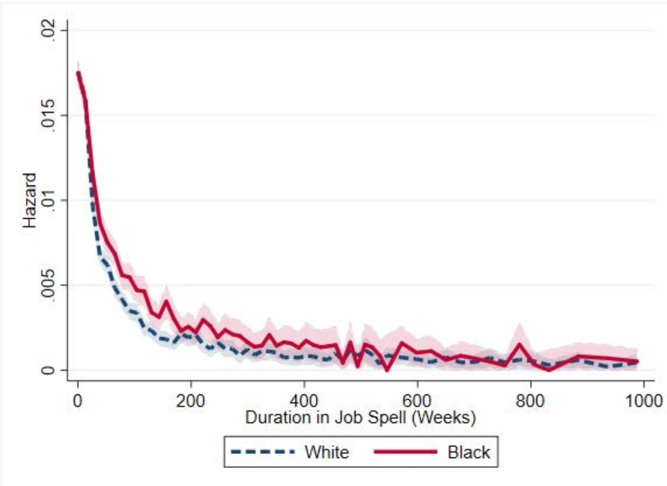


(e) 26 Week Interruptions as Same Spell, Nonparametric

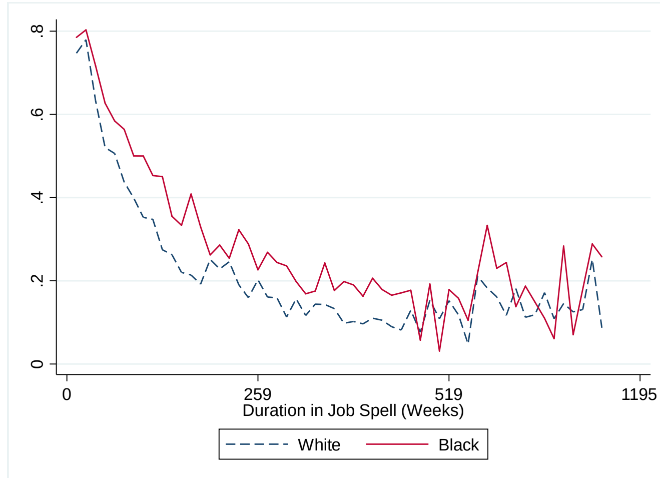


(f) 26 Week Interruptions as Same Spell, Cox

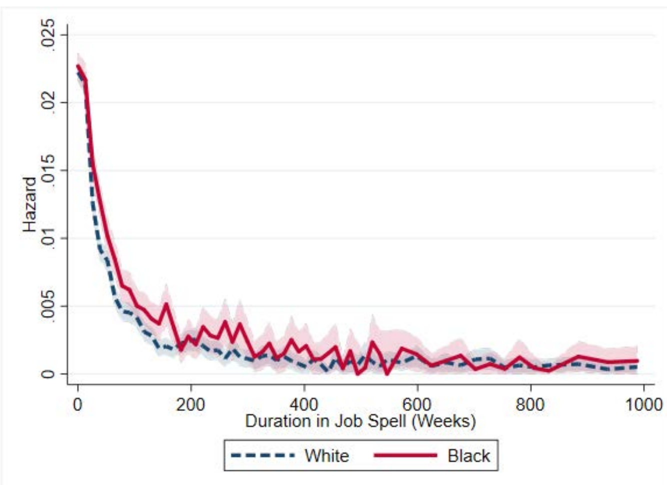
(g) Including Parttime Spells, Nonparametric



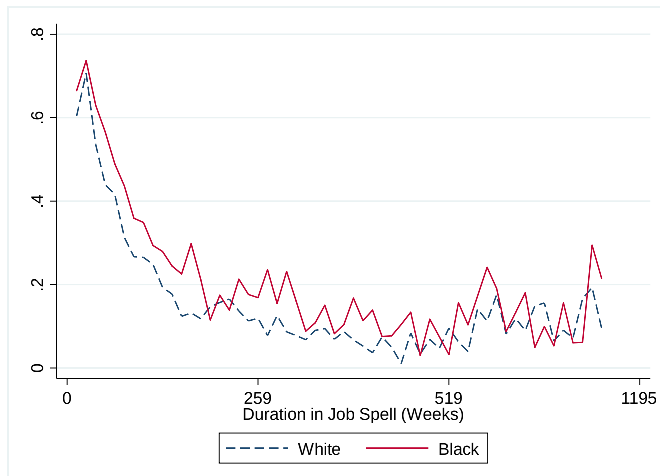
(h) Including Parttime Spells, Cox



(i) Age ≤ 30 at Spell Start, Nonparametric



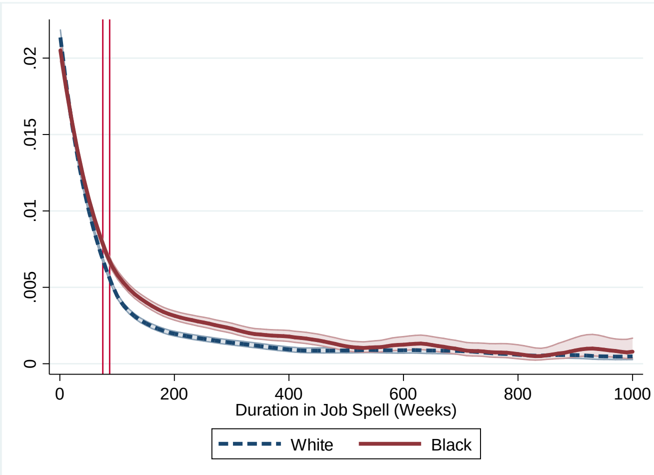
(j) Age ≤ 30 at Spell Start, Cox



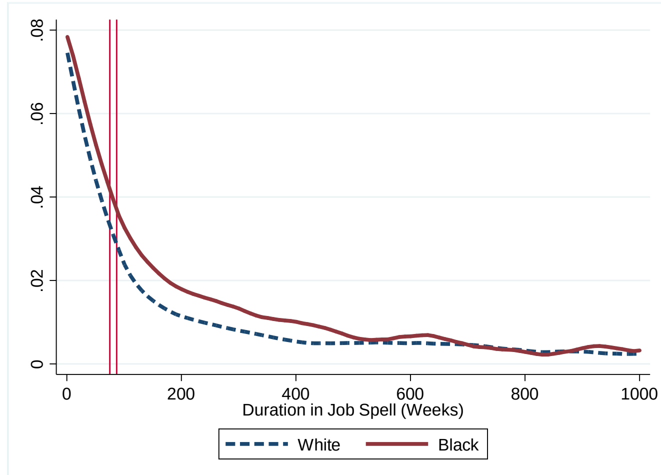
Notes: This figure presents robustness results, using both nonparametric methods and stratified Cox regressions. These are the same methods as used in the principal results described in Figures 3 and 4. The plots in (a) and (b) treat quits into nonemployment as censoring the employment spell, rather than the spell ending in nonemployment. The plots in (c) and (d) restrict to employment spells for which the individual entered from nonemployment. The plots in (e) and (f) treat ≤ 26 week gaps in spells at the same employer as the same spell. The plots in (g) and (h) include parttime spells and treat transitions to parttime employment as censored rather than as failures. The plots in (i) and (j) restrict to spells for which the individual was less than or equal to 30 at the start of the spell. For all plots showing confidence intervals, bands around estimates are 95% confidence intervals.

Appendix Figure A2: Robustness Kernel-Smoothed Hazard Estimates

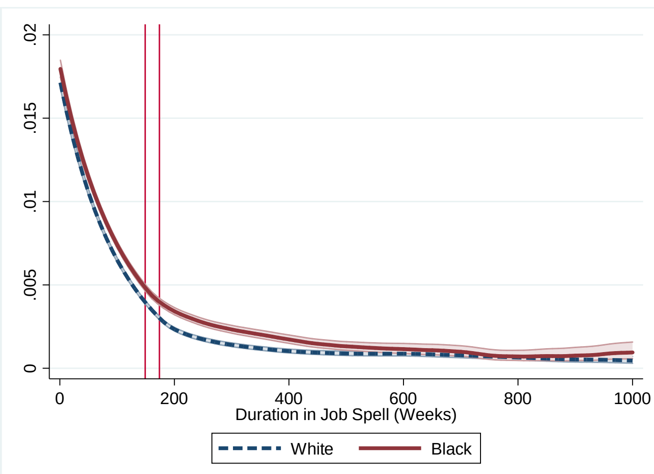
(a) Boundary-adjusted, Smaller Bandwidth, Nonparametric



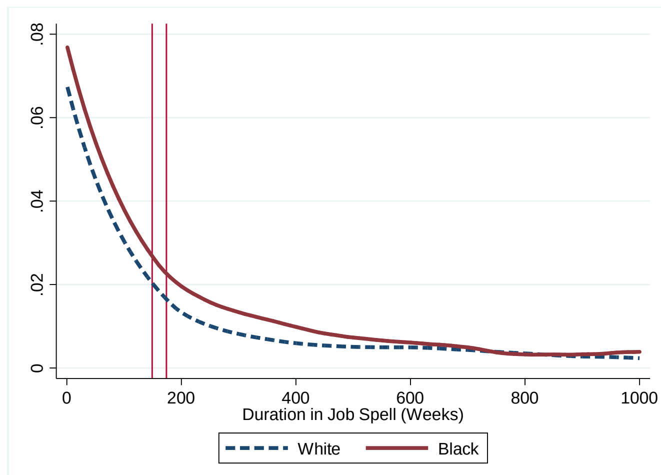
(b) Boundary-adjusted, Smaller Bandwidth, Cox



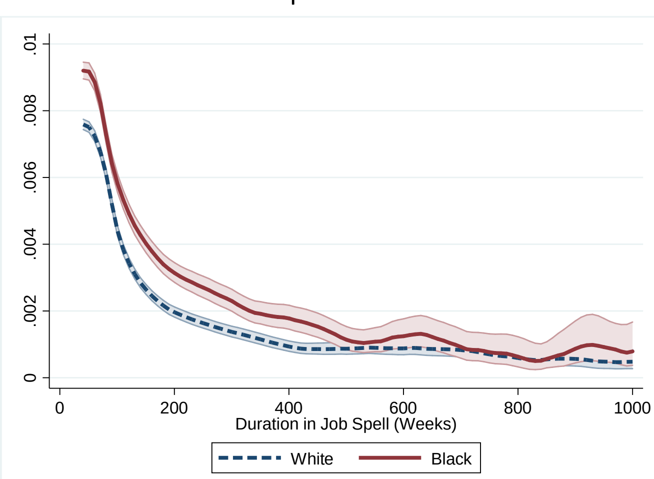
(c) Boundary-adjusted, Larger Bandwidth, Nonparametric



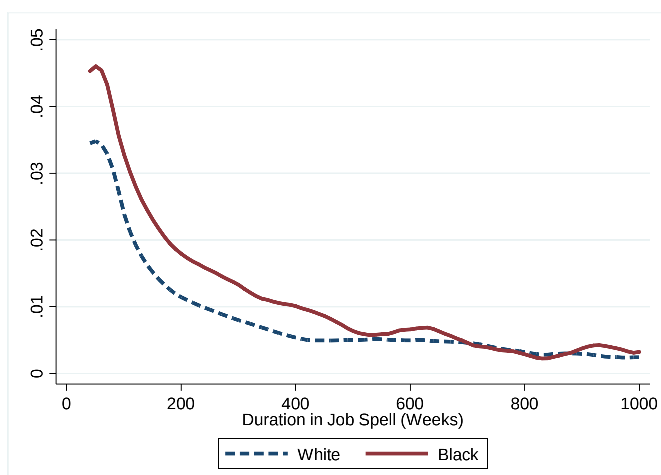
(d) Boundary-adjusted, Larger Bandwidth, Cox



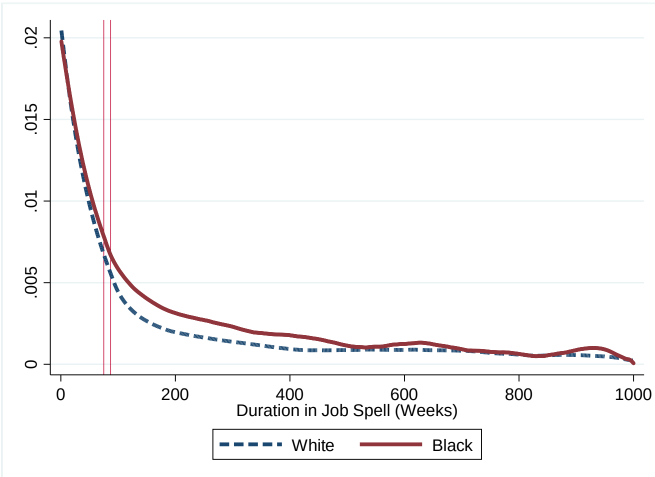
(e) Exclude Boundary Region, Smaller Bandwidth, Nonparametric



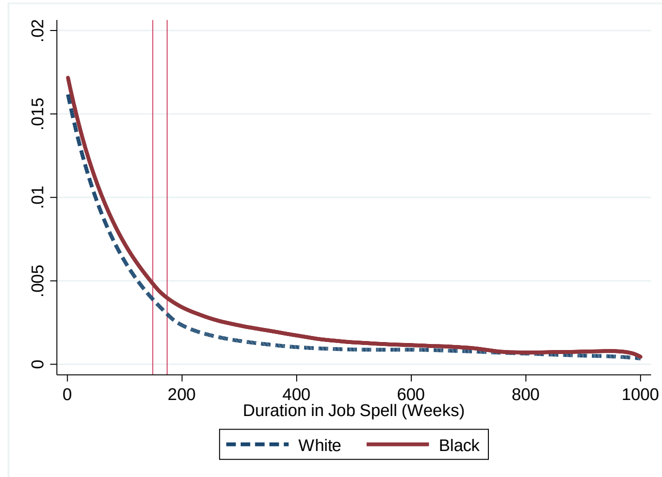
(f) Exclude Boundary Region, Smaller Bandwidth, Cox



(g) Local Linear Smoothing, Smaller Bandwidth, Nonparametric

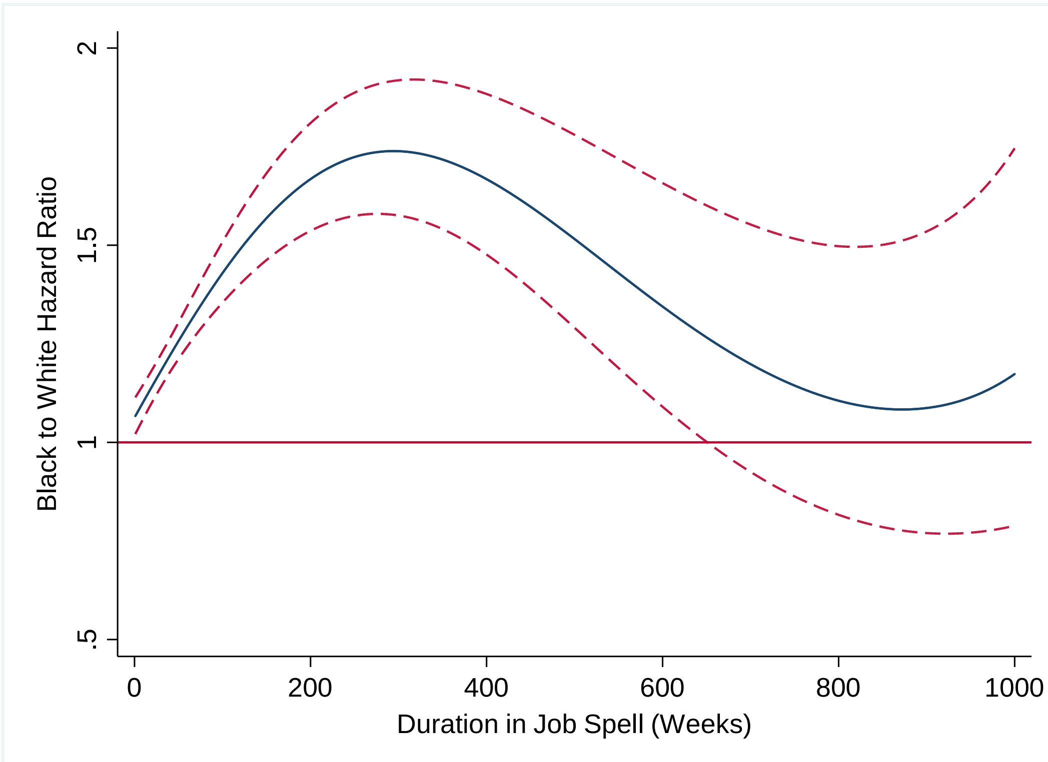


(h) Local Linear Smoothing, Larger Bandwidth, Nonparametric



Notes: The plot in (a) presents kernel-smoothed hazard contributions from the Nelson-Aalen cumulative hazard, using the boundary-adjusted alternative Epanechnikov kernel, with bandwidth equal to 87 weeks for whites and 75 weeks for blacks, these are one half of the Silverman plug in for this kernel (174 weeks for whites, 149 weeks for blacks). The plot in (b) presents kernel-smoothed hazard contributions from a Cox Proportional Hazards model, stratified by race. The kernel and the bandwidth are the same as those in (a). The explanatory variables in the Cox model are the same as those included in Figure 4. The plots in (c) and (d) are analogous to (a) and (b), but use the Silverman plug-in bandwidth rather than half of the Silverman plug-in bandwidth as in (a) and (b). The plots in (e) and (f) are analogous to (a) and (b), but use the Epanechnikov kernel (1969 version), with bandwidth equal to half the Silverman plug in for this kernel, and show results only outside the boundary regions. The plot in (g) uses local linear smoothing of the hazard contributions, using the alternative Epanechnikov kernel with bandwidth equal to half the Silverman plug-in estimate. The plot in (h) is the same as in (g), but uses the Silverman plug-in estimate as the bandwidth. For all plots showing confidence intervals, bands around estimates are 95% confidence intervals. Vertical lines show boundary regions for black and white workers.

Appendix Figure A3: Black to White Hazard Ratio Controlling for Covariates in a Cox Model, Allowing the Percentage Gap in Hazards to be a Cubic in Seniority



Notes: This is a plot of the Black to White Hazard ratio from a Cox model controlling for the same covariates included in Figure 4. Additionally, we include an indicator for black, and interact this with a cubic in seniority (duration in job spell in weeks). In order to allow the effect of race to vary over time, we include an observation for each job spell at each failure time in the data (as Cox models are only estimated when failures occur in the data). We obtain the linear combination of the coefficients on black, for values of week from 1 to 1000. We then exponentiate these to obtain the hazard ratio. Dashed lines are the 95% confidence intervals.

Appendix Table A1: Coefficients from Cox Model Stratified by Race

Any College Completed at Spell Start	-0.0713***
	[0.0200]
Any College Completed at Spell Start Missing	0.101
	[0.171]
AFQT (standardized)	-0.0260***
	[0.00932]
AFQT (standardized) Missing	-0.0461
	[0.0339]
Region at spell start: Northeast	-0.0912***
	[0.0261]
Region at spell start: North Central	-0.0578**
	[0.0245]
Region at spell start: South	-0.164***
	[0.0233]
Region at spell start: Missing	-0.183***
	[0.0524]
Urban Location at Spell Start	-0.00161
	[0.0194]
Urban Location at Spell Start Missing	-0.0321
	[0.0371]
Blue Collar Occupation at Spell Start	0.327***
	[0.0203]
Non-Blue Non-White Collar Occupation at Spell Start	0.543***
	[0.0601]
Occupation at Spell Start Missing	1.413***
	[0.0280]
Age at Spell Start	-0.0553***
	[0.00341]
Observations	34692

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Coefficients are from a Cox Proportional Hazards model stratified by race, using week bin as a unit of time. Each observation is a job spell. We model hazard into nonemployment, and the failure variable is an indicator for whether the job spell ended in nonemployment. See text for details.