

Equilibrium Search Models and The Transition from School to Work

Audra J. Bowlus
Nicholas M. Kiefer
George R. Neumann *

Date: June 4, 1997

Abstract

This paper applies an equilibrium search to study the transition from schooling to work of U.S. high school graduates. We consider the case where there is heterogeneity in firm productivity and the number of firm types is discrete. For this case the estimation problem is non-standard and the likelihood function is non-differentiable. This paper provides a computational method to obtain the MLE and, through several Monte Carlo studies, characterizes the behavior of the estimator. Applying these methods to the transition from school to work, our results show that nonemployed blacks receive fewer offers than whites and employed blacks are more likely to lose their jobs. Importantly, employed blacks and whites receive job offers at the same rate. However, the difference in job destruction rates is so great that it accounts for three-quarters of the black-white wage differential.

Keywords: Labor market transitions; duration models; simulated annealing; discrete non-parametric heterogeneity; black-white earnings differential.

JEL Classification: J64, J3, C41.

Editorial Communications: Prof. George R. Neumann, Dept. of Economics, Univ. of Iowa, Iowa City, IA, 52242-1000, USA

*Department of Economics, University of Western Ontario, London, Ontario Canada N6A 5C2; CLS Aarhus, Department of Economics, University of Aarhus and Department of Economics, Cornell University, Ithaca NY 14853; and Department of Economics, University of Iowa, Iowa City IA 52242. We have benefited from comments made by colleagues and participants in seminars at Aarhus, Berkeley, Cornell, Iowa, the Free University of Amsterdam and the University of Western Ontario. This work was supported in part by NSF grant SES-9122253.

1. Introduction

The essential idea of equilibrium search models of labor market behavior is that wage policy matters. In contrast, the stylized neoclassical competitive model predicts that firms paying a wage above the competitive equilibrium will disappear; those offering less will attract no workers. The search approach introduces "friction" via information asymmetries. Here, firms that offer high wages are more attractive to workers, obtaining and retaining employees more readily than firms offering lower wages. Other things equal high wage firms generate lower profit per worker but make it up on volume. These simple ideas about wage policy have been stated in informal ways by several scholars. Hicks (1932, [1966]) discusses the 'Gospel of High Wages' whereby unusually successful employers pay high wages to have the "pick of the market" (p.36). Kerr (1954) initiated a long-surviving, although never mainstream, line of research on empirical correlates of wage policy that has since become known as "Dual Labor Market" theory. While these early discussions of wage policy are often colorful, formal content has been given to the ideas only recently in equilibrium search models by Albrecht and Axell (1984), Burdett (1990), Burdett and Mortensen (1995), and Mortensen (1990). In competitive models wage policy doesn't matter because by definition in equilibrium the law of one price holds: all workers of a given type receive the same wage. Even in simple monopsony models of the labor market (Card and Krueger (1995)) wage policy does not matter because the law of one price still holds, albeit at lower than the competitive level. In contrast, search models generate dynamic monopsony power for employers due to the presence of trade

frictions such as the length of time it takes to find a new job. A firm's wage policy is important in such models because it directly affects the distribution of income in an economy. Moreover, in such dynamic monopsony models public policy experiments such as introducing or changing a minimum wage can have employment effects that are quite different from those expected in the standard competitive case.

Because equilibrium search models provide a natural interpretation of interesting labor market phenomena, the estimation of such models has attracted considerable attention. Eckstein and Wolpin (1990) estimate a version of Albrecht and Axell's (1984) model; van den Berg and Ridder (1993a,1993b), and Kiefer and Neumann (1993), and Bowlus, Kiefer, and Neumann (1995) estimate a version of Mortensen's (1990) model. Although these models differ in the forces of competition generated by firms, each predicts a disperse price equilibrium to exist. In the Albrecht and Axell model the equilibrium wage distribution is determined by heterogeneity in reservation wages of workers. In Mortensen's approach the equilibrium wage distribution is determined by the technology that matches workers to jobs. This is a pure search equilibrium model, in that heterogeneity in workers or firms is not required for the existence of a dispersed equilibrium. To date, efforts at fitting these models to data have not been completely successful. The tight theoretical structure needed to generate simple estimation strategies results in a poor match between theory and evidence. "Evidence" in this context consists of data on the search time to find employment, the wage accepted, and the duration of the job. While the duration data are in reasonable agreement with the predictions of the theory, the distribution of earnings is not. This is to be expected: when

wage policy matters, higher wages bring forth greater labor supply to firms, and since workers prefer higher wages, the distribution function of wages is predicted to have a thick right tail. However, empirical wage distributions typically have thin right tails, looking much like income distributions for which Pareto distributions are frequently used to characterize the upper tail. One explanation for this mismatch between theory and evidence is that unmeasured differences across workers and jobs can not be ignored. Of course, unobserved heterogeneity, like dummy variables in a regression, can explain any divergence between fact and theory. We adopt the approach of introducing heterogeneity in small doses - so that the essential information basis for wage dispersion is not completely absorbed into "heterogeneity."

In this paper we provide methods for estimating equilibrium search models using a non-parametric estimator of heterogeneity. In Section 2 we describe the basic equilibrium search model due to Mortensen (1990) and Burdett and Mortensen (1995) and we develop the likelihood function appropriate to data generated by this model. For the case that we consider, namely a discrete number of firm types having different productivity levels P_j , $j=1, \dots, J$, the estimation problem is non-standard, and the likelihood function is not differentiable. We characterize the MLE for this problem and provide a means of computing it. The computational method we employ exploits the special structure of the problem. In the first part of the algorithm we calculate the points of support of the wage distribution in the presence of heterogeneity; in the second part the profile likelihood function, conditioned on the points of support of the wage distribution, is maximized with respect to the remaining parameters in the usual manner. Asymptotic independence of

these parameter blocks guarantees that the stepwise maximum is a global maximum.

The third section contains a Monte Carlo study of the properties of the estimators we propose using all the data from the search model - unemployment duration, accepted wages, job spell duration and job transitions. In Sections 4 and 5 we apply this technology to study the transition between school and work for black and white male high school graduates using the National Longitudinal Survey of Youth (NLSY). This application has been of considerable policy interest in recent years, so we use it to examine how well the search framework can explain empirical patterns found in the data. The model provides a satisfactory fit to the nonemployment durations following graduation and the wage offer distribution for the first full-time job for both groups. It has more difficulty matching the job duration data. Longer nonemployment spells and lower wages for blacks are captured in the model through a lower arrival rate of offers while nonemployed and a higher job destruction rate while employed. Three quarters of the black-white wage differential is explained by differences in the search parameters highlighting the important role labor market mechanisms can play in explaining wage differentials. Our conclusions are given in the sixth and final section of the paper.

2. Equilibrium Search Models

To characterize labor market transition data we use the equilibrium search model set out in Mortensen (1990). We summarize the homogeneous version of it briefly here. The primitives of the model are (1) the offer arrival rate for unemployed workers, λ_0 , (2) the arrival rate of offers while employed, λ_1 , (3) the job destruction rate, δ , and (4) the productivity level of firms, P . In the model workers, taking the wage offer distribution of the firms as given, solve the standard search utility maximization problem and adopt a reservation wage strategy. Following Mortensen and Neumann (1988) a worker's reservation wage while unemployed is:

$$r = b + (\kappa_0 - \kappa_1) \int_r^{\infty} \left[\frac{1 - F(w)}{1 + \kappa_1(1 - F(w))} \right] dw \quad (1)$$

where $F(w)$ is the wage offer distribution, $\kappa_0 = \lambda_0/\delta$ and $\kappa_1 = \lambda_1/\delta$, and b is the value of non-market time. Unemployed workers see jobs arrive at rate λ_0 , and they accept the first job that offers more than their reservation wage. While employed at wage w , a worker's reservation wage is also w . Job offers arrive at rate λ_1 for employed workers, and jobs are destroyed at rate δ . In the homogeneous version firms are identical with productivity level P , face constant returns to scale in production, and maximize profits by choosing the wage to pay. The balancing condition which equates supply and demand is that firms will offer higher wages if and only if they can expect to get an additional number of workers to cover the lower per worker profits. Higher wages attract more workers to a firm and allow firms to retain the workers longer. Mortensen (1990) shows that the unique equilibrium

wage offer distribution is:

$$F(w) = \frac{1 + \kappa_1}{\kappa_1} \left[1 - \left(\frac{P - w}{P - r} \right)^{1/2} \right] \quad (2)$$

with density

$$f(w) = \left[\frac{1 + \kappa_1}{2\kappa_1} \right] \frac{1}{\sqrt{(P - w)(P - r)}}. \quad (3)$$

The development so far has considered the case where all firms have identical productivity, P . Bowlus, Kiefer and Neumann (1995) and Koning, Ridder and van den Berg (1995) point out that this implies from (3) that the density of wage offers, is increasing over its support. But, as noted above, empirical wage distributions typically have a thin right tail.

The observed empirical distribution of wages can be made consistent with the equilibrium search model if within a market there are differing levels of productivity across firms and if firms with especially high productivity are relatively rare.¹ Heterogeneity in firm productivity can be introduced in two ways. Following Bontemps, Robin and van den Berg (1996) one could assume P is continuously distributed. Alternatively, following Bowlus, Kiefer and Neumann (1995) heterogeneity can be viewed as arising from a finite number of firm types, Q . Either approach can fit the wage distribution, but there are advantages and disadvantages of the approaches in other respects. The disadvantage of assuming a continuum of firm types is that the assumption creates a direct map

¹ Unobserved between market heterogeneity in productivity levels has also been used to reconcile the shape of observed empirical wage distributions with that implied by the search model (see Koning, Ridder and van den Berg (1995)).

between the distribution chosen to represent productivity and the resulting wage distribution due to the competition faced on either side by any one firm. A one-to-one mapping exists between productivity and wages. The model is then less clearly a search model - in that there is no longer pure search wage dispersion. Thus, it is difficult to assess the role of information in the operation of labor markets. The advantage of assuming a continuous distribution function is that it makes the estimation problem standard.

The advantages of modeling heterogeneity as discrete are twofold. First, while allowing for some flexibility in the shape of the wage offer distribution, it retains the key model element of search induced wage dispersion. Second, the classification of workers into different levels of productivity generates testable economic implications that will be of use when paired worker-firm data are available.² A disadvantage is that discrete heterogeneity results in a likelihood function that is discontinuous in some of the parameters and whose estimation is non-standard. Non-standard estimation issues arise frequently in the application of dynamic programming models and can be handled.

Because we find the advantages of the discrete specification compelling enough to justify a trial application, we pursue the discrete heterogeneity approach in this paper. To this end, assume that there are Q types of firms with productivity $P_1 < P_2 < \dots < P_Q$. The fraction of firms having productivity P_j or less is $\gamma_j = \gamma(P_j)$. The equilibrium wage distribution, following Mortensen (1990), is:

² The Danish IDA dataset currently follows a sample of 1000 firms and reports earnings and labor market transitions for all workers at those firms.

$$F(w) = \phi_j(w) \forall w \quad (4)$$

with ϕ_j defined by

$$\phi_j(w) = \frac{1 + \kappa_1}{\kappa_1} \left[1 - \frac{1 + \kappa_1 [1 - \gamma_{j-1}]}{1 + \kappa_1} \left[\frac{P_j - w}{P_j - w_{Hj-1}} \right]^{1/2} \right], \quad w_{Lj} < w \leq w_{Hj} \quad (5)$$

where w_{Lj} is the lowest wage that will be offered by a firm of type j and w_{Hj} is the highest wage paid by a type- j firm.

Several restrictions are implied by the model. First, $w_{L1} = r$, and $w_{Hj} = w_{Lj+1}$, $j = 1, \dots, Q-1$. Second, $F(w_{Hj}) = \gamma_j$, $j=1, \dots, Q$. Define $B_j = [(1 + \kappa_1(1 - \gamma_j)) / (1 + \kappa_1(1 - \gamma_{j-1}))]^2$ and observe that $0 < B_j < 1 \forall j$. Then, from the condition $F(w_{Hj}) = \gamma_j$ it follows that

$$P_j = \frac{w_{Hj} - B_j w_{Hj-1}}{1 - B_j}. \quad (6)$$

Equation (6) implies that if we know κ_1 , r , w_{Hj} and γ_j , $j = 1, \dots, Q$, we can estimate the unobserved productivity levels, P_j .

The stochastic processes induced by the model described above completely characterize labor market histories, whether complete or incomplete. The data to be explained are the duration of unemployed search, D_1 , the wage received on an accepted job, w , the length of that job, D_2 , and whether the job ends because it was lost ($C=1$) or left ($C=0$). According to the theory unemployed search duration is exponential with intensity parameter λ_0 . The marginal distribution of accepted wages is $f(w)$ given by

$$f(w) = \sum_{j=1}^Q \left[\left[\frac{1 + \kappa_1 [1 - \gamma_{j-1}]}{2\kappa_1} \right] \left[(P_j - w)(P_j - w_{Lj}) \right]^{-1/2} \right] * I(w_{Lj} < w \leq w_{Hj}) \quad (7)$$

where $I(x)$ is the indicator that the event x occurs. The theory also predicts that conditional on the wage rate received, w , the density of job duration, $f(D_2 | w)$, is exponential with intensity parameter $(\delta + \lambda_1(1-F(w)))$ where $F(w)$ is given in equation (4). Finally the probability that a job ends by being lost rather than left, $\Pr(C=1 | w)$, is

$$\Pr(C = 1 | w) = \frac{\delta}{\delta + \lambda_1(1-F(w))}. \quad (8)$$

The likelihood function for this segment of the labor market history is:

$$\ell(\theta) = \lambda_0 \exp(-\lambda_0 D_1) f(w) \exp(-(\delta + \lambda_1[1 - F(w)]) D_2) \delta^C (\lambda_1[1 - F(w)])^{1-C}. \quad (9)$$

We partition the parameters of this model into three groups as: $\theta = \langle \theta_1, \theta_2, \theta_3 \rangle$

where

$$\theta_1 = \langle w_{H1}, w_{H2}, \dots, w_{HQ-1} \rangle$$

$$\theta_2 = \langle r, w_{HQ} \rangle$$

$$\theta_3 = \langle \lambda_0, \lambda_1, \delta \rangle.$$

Observe that the likelihood function is continuous but not differentiable in w_{H1}, \dots, w_{HQ-1} , which can be seen from the distribution function of wages shown in equation (5).

However, the cut points w_{Hj} , $j=1, \dots, Q-1$, are points of discontinuity of the density and the results of Chernoff and Rubin (1956) show that maximum likelihood estimates of discontinuity in density converge to their true value at rate N . Therefore asymptotically, to order $N^{1/2}$, the variability in estimates of w_{Hj} is unimportant and can be ignored. We provide empirical evidence on this conclusion in a small sampling experiment below. In

our application, we use the bootstrap to provide standard errors for these parameters. The likelihood function (9) with w_{Hj} treated as known remains non-standard because the range of the random variable, w , is from r to w_{HQ} and these are parameters of the model. Kiefer and Neumann (1993) propose the estimators

$$\hat{r} = \min\{w_i\}, \quad \hat{s} = \max\{w_i\} \quad (10)$$

for r and w_{HQ} . The estimators \hat{r} and \hat{s} are super-efficient but in finite samples they are not necessarily the MLE's for r and w_{HQ} . However, asymptotically they are the MLE's and the theory of local cuts (Christensen and Kiefer (1994)) justifies conditioning on them. As van den Berg and Ridder (1993a) note, these estimators are sensitive to measurement error, but at least for the case of classical measurement error their performance is actually improved in small samples.³

A characterization of maximum likelihood estimators for w_{Hj} $j=1, \dots, Q-1$ in wage data is given by the following theorem. The result is that the MLEs for these parameters occur at observed wages.

Theorem 1.

Let $\{W_N\}$ be the set of observed wages from a sample of size N drawn from the wage distribution described in equation (7). Denote the parameters corresponding to points of discontinuity of this distribution as w_{Hj} , $j = 1$ to $Q-1$. The maximum likelihood estimator for w_{Hj} , denoted \hat{w}_{Hj} , is a $(Q-1)$ element of W_N .

³ In small samples the extremum estimators are biased - r is overestimated and s is underestimated; this bias disappears asymptotically. However, classical measurement error adds noise to each observation so that the observed sample minimum is likely to have negative measurement error, and conversely for the sample maximum. This partially offsets the bias in the extremum

Proof.

It suffices to consider the case of a single cutpoint, Θ . Write the likelihood for wage data as

$$\begin{aligned} \ell(\Theta) = & \sum_{w \leq \Theta} [\ln(1/2\kappa_1) - 1/2 \ln[(P_1 - w)(P_1 - r)]] \\ & + \sum_{w > \Theta} [\ln((1 + \kappa_1)/2\kappa_1) - 1/2 \ln[(P_2 - w)(P_2 - \Theta)]] \end{aligned}$$

Here P_1 and P_2 are functions of Θ and Θ is a single kink point in the distribution. The argument is as follows: if $\hat{\Theta}$ does not occur at an observed wage, then the first and second order conditions for a maximum must be satisfied at $\hat{\Theta}$. We show this is not possible. It suffices to do this for the case where Θ is the only unknown parameter as we show the second derivative is positive - so in the multiparameter case a diagonal element of the Hessian would be positive, violating the second order condition. Differentiating twice and using $\partial P_1 / \partial \Theta = 1/(1-B_1)$ and $\partial P_2 / \partial \Theta = -B_2/(1-B_2)$ gives

$$\begin{aligned} 2l_{qq} = & \sum_{w \leq \theta} [(P_1 - w)^{-2} + (P_1 - R)^{-2}](1 - B_1)^{-2} \\ & + \sum_{w > \theta} [(P_2 - w)^{-2} + (P_2 - \Theta)^{-2}(-B_2 / (1 - B_2))^2] \\ & + \sum_{w \leq \theta} [(P_2 - \Theta)^{-2}(B_2 / (1 - B_2))] \\ & + \sum_{w > \theta} [(P_2 - \Theta)^{-2}(B_2 / (1 - B_2) + 1)] \\ & > 0 \end{aligned}$$

Theorem 1 tells us that there is no purpose looking for solutions to normal equations for

estimator.

estimates of the points of discontinuity; the estimates are observed values of wages, which are the points of discontinuity of the likelihood function of the sample. Though our theorem and proof focus on the leading case of wage data (the most relevant, and necessary, source of information on the wage distribution), experience has shown us that the addition of duration data, etc. does not affect the main result. Many variations are possible with different data configurations. For Q known there are $N!/(N-Q+1)!(Q-1)!$ potential estimates of w_{Hj} , which can be a large number. Finding the maximum of the likelihood function can be done by a grid search in low dimensions, although for $Q > 3$ this becomes time intensive for samples of the size typically used in panel data - that is, $N \approx 500-1,000$. For example, for $Q=3$ and $N=500$ there are $N(N-1)/2 = 124,750$ distinct combinations of w_{H1} and w_{H2} possible. The $N(N-1)/2$ elements are the upper off-diagonal elements in a matrix array of $W \times W$ where search is restricted by the order relation $w_{H2} > w_{H1}$. This triangular array is replaced in higher dimensions by the suitably restricted sub-matrices with $w_{Hj} > w_{Hi}$ for $j > i \leq Q$. Not all of these ordered points are admissible. To be admissible we must have $P_j > P_i$ whenever $j > i$. Mortensen (1990) shows that $P_j > P_i \Rightarrow w_{Hj} > w_{Hi}$, but the converse is not true as can be seen from manipulation of equation (6). Equation (6) provides the solution to $P = A \cdot W_H$, where P is an $Q \times 1$ vector of productivities of the Q firm types, and W_H is an $(Q+1) \times 1$ vector of wage cut points - $W_H(1) = r$, $W_H(2) = w_{H1}$, etc. When P is not ordered, the estimate of w_{Hj} that generated P is not admissible, and the likelihood function need not be evaluated at that point.

The likelihood function shown in (9) can be maximized using an iterative procedure

with two steps in each iteration. In the first step of the algorithm the function is $l(\theta_1 | \hat{\theta}_2, \hat{\theta}_3)$ maximized using Simulated Annealing (Szu and Hartley (1987), Otten and Ginneken, (1989), Kirkpatrick et. al.(1983)) to obtain an estimate of $\theta_1, \hat{\theta}_1$. In the smooth maximization step $\hat{\theta}_1$ is used to maximize $l(\theta_3 | \hat{\theta}_2, \hat{\theta}_1)$ using a Newton-Raphson procedure. These steps are iterated until convergence occurs. Since the estimator $\hat{\theta}_1$ converges at rate N (Chernoff and Rubin, 1956) it is asymptotically independent of $\hat{\theta}_3$. The estimator $\hat{\theta}_2$ is also superefficient (and a local cut) and asymptotically independent of $\hat{\theta}_3$. Thus, iterative separate maximizations lead to a joint maximum of the likelihood function on convergence. In the Simulated Annealing step we use a maximum number of iterations that depends on the number of points of support of the heterogeneity distribution, Q , with a "wrong" acceptance probability of .05 for a unit step in the wrong direction. On the smooth part of the function we maximize $l(\theta_3 | \hat{\theta}_1, \hat{\theta}_2)$ using Gauss's OPTMUM procedure with a stopping criteria of a change in function value less than 10^{-8} .

So far we have treated Q , the number of points of support of the heterogeneity distribution, as known. The choice of Q in this framework is similar to choosing the points of support in the Heckman-Singer (1984) estimator of a mixing distribution. As yet there is no formal test for choosing the correct level of Q . Certainly one could consider conditional moments tests, choosing the first Q that produces a set of estimated moments within a given tolerance range of the sample moments. In our experience the quasi-

likelihood ratio test - $-2\Delta\log l < \chi_{0.05}$ - appears to work reasonably well as a criterion for choosing Q . We describe our experience with this rule with a small Monte Carlo study in Section 3.

3. Monte Carlo Results

To examine the behavior of the estimator described in section 2 we conducted a small Monte Carlo analysis. Samples of size 500 were generated according to the true model and 500 replications were performed. The true model is specified as $\lambda_0 = 0.03$, $\lambda_1 = 0.01$, $\delta = 0.0035$, $Q=3$, $r = 100$, $w_{H1} = 179$, $w_{H2} = 377$, $w_{H3} = 677.35$, $\gamma_1 = .3$, $\gamma_2 = .7$. The model was estimated for $Q= 1, \dots, 7$, and the number of break points was selected by comparing minus twice the loglikelihood difference associated with increasing Q by one (this statistic is denoted V) with the .05 critical value of $\chi^2(1)$. We denote the Q chosen by this criterion as Q^* . As we noted above, the sampling distribution of V is unknown. It is clear that the likelihood function is nondecreasing in Q . Note that the Neyman-Pearson lemma applies, so V is the right criterion function to use. What remains unknown is the distribution of V and therefore the appropriate critical values.

For 500 replications the marginal distribution of the estimated Q is given in Table 1.

Table 1 Sample Distribution of Q True value of Q = 3					
Q =	1	2	3	4	5+
h(Q) =	.000	.000	.368	.398	.234

Using this criterion Table 1 indicates a tendency towards overfitting the points of support for the wage distribution. Note that this is the opposite of the pattern typically found in using the Heckman-Singer estimator. In fact, Heckman and Singer (1984) find that the estimator typically chooses a small number of points of support. As Table 1 shows, in this Monte Carlo experiment we never underfit Q and 63% of the time we choose a value greater than the true value of Q. This bears further investigation and suggests increasing our critical values.

The obvious question is how does the choice of Q affect the estimates of the parameters of the model, in particular λ_0 , λ_1 , and δ .⁴ Table 2 shows the sampling distribution for these parameters with Q chosen according to the criterion we have given. Thus these are marginal quantities (with respect to Q). As the table shows the mean values of these estimates are dead on. Furthermore the sampling variation is very small. For example, 95% of the sampling distribution of λ_1 lies between .009 and .011, and for δ between .0031 and .0038.⁵

⁴ The estimates of the reservation wage r and the highest wage w_{HQ} are unaffected by the choice of Q.

⁵ The estimate of λ_0 does not depend upon the value of Q.

Table 2 Sample Distribution of λ_0 , λ_1 , and δ			
Parameter	Mean	5th %-tile	95th %-tile
λ_0	0.030	0.028	0.033
λ_1	0.010	0.009	0.011
δ	0.0035	0.0031	0.0038

If we use a value of Q greater than the true value of Q , Q^0 , we find that this has a relatively small effect on the parameters. For example, the root mean square error for λ_1 increases by 2.4% if we compare the estimates implied by Q^* with those produced using Q^0 . However, the root mean square error for δ decreases by 2.2% using the same comparison, which implies that very little is lost on average by overfitting Q . In contrast, although we would never choose to underfit the distribution by following this criterion, had we chosen Q equal to 2, for example, the root mean square errors for λ_1 and δ double compared to Q^* . Thus the "loss" associated with overfitting Q seems minor.

4. The Transition From School to Work

We apply the estimation procedure described above to a sample of wage and duration data from the NLSY for black and white male high school graduates who did not continue on in formal education.⁶ We collected information on: (1) the non-employment

⁶ The sample includes only those individuals who receive a high school diploma. GED recipients are dropped.

spell between graduation and the first full-time job, (2) the wage accepted on the first full-time job, (3) the job spell length, and (4) whether the job was left (a quit) or lost. Two important, practical measurement issues arise in studying the transition between school and work. One involves the timing of when a job is found, and, consequently, measurement of non-employment spells. The second issue is the definition of a job: How should one distinguish full-time employment from casual, stop-gap jobs? Obviously, the definition of a full-time job affects measurement of job search duration, so we begin with our definition of full-time employment.

Nonemployment durations

We define the first full-time job as the first recorded employment opportunity that:

- (a) consists of thirty-five hours or more a week;
- (b) lasts at least one month (four weeks), and
- (c) starts within three years after high school completion.

To make sure we have the first full-time job spell after graduation we include only individuals who finished high school after January 1, 1978.⁷ If the first full-time job spell happens to surround the education completion date, it is used as the first spell only if the individual holds the job longer than two months after the education completion date. This eliminates summer jobs and temporary jobs held while in school. Nonemployment durations then consist of the number of weeks between the start of the first full-time job and the respondent's high school graduation date. Consequently, the spell of job search may actually contain one or more spells of part-time employment. This focus on full-time

⁷ Reconstructing employment histories prior to 1/1/78 is not possible in the NLSY.

employment is standard in the literature.⁸

Even with the restriction of employment to full-time jobs a substantial number of nonemployment spells have weeks of search equal to zero; i.e., the individuals were employed prior to finishing their education. In the data 22% (44/199) of black high school graduates had zero search duration, as did 32% (149/465) of white high school graduates.⁹ Eckstein and Wolpin (1995), who also study the school to work transition using NLSY data, measure spell duration in quarters and treat these spells as having a duration of 1 quarter.¹⁰ This procedure artificially produces the appearance of duration dependence even when none is present, and in fact can produce the appearance of negative duration dependence when the data actually are drawn from a distribution characterized by positive duration dependence. This problem is, of course, the classic problem of initial conditions. When an individual starts searching for a job is unobserved; a search duration is partially observed only if a job is not found by the end of formal education, and in this case only the forward recurrence time, T_F , is observed. Completed spells of search, T_C , are random variables characterized by observed measurements $T_C = T_B + T_F$, where T_B , is a realization of a backward recurrence time. The problem is that T_B is unobserved.¹¹ As is well known (Billingsley (1961), Cox and Lewis (1966), Heckman and Singer (1986)) the distribution of forward recurrence times is not in general

⁸ See for example, Eckstein and Wolpin (1995), p. 269. We use the same data source as Eckstein and Wolpin, although our data come from the 1992 version of the NLSY while they use data through 1986.

⁹ In addition, 70 white and 80 black high school graduates were dropped from the sample because they had not found a full-time job within three years of graduating.

¹⁰ Eckstein -Wolpin do not state how they treated such spells. We infer that they are given the shortest duration of 1 quarter from the text on page 269 and the plots of the Kaplan-Meier survival functions on pages 271-3.

¹¹ In principle the backwards recurrence time could be measured, but to our knowledge this has not been done in any of the data

the same as the distribution of completed spells, the only exception is the exponential distribution. There is no simple fix for this data problem. Discarding the first observation as Billingsley suggests is not an option as there are no repeat spells of transiting from school to work. In what follows we maintain the assumption that search durations are exponentially distributed, but recognize that the observed measurements are left-censored. That is, only individuals with a positive forward recurrence time contribute information on nonemployment durations in the likelihood function. Based on the estimates obtained under this specification we test whether the data can reject the exponential specification and, if so, why.

sets used to study the transition from school to work.

Job Duration

Job durations are recorded as the number of weeks from the starting date of the job to the ending date of the job or survey, whichever comes first. If the latter, then the spell is right-censored. Because we follow individuals through the 1992 wave of the NLSY, the censoring rates are low - less than 10% for both blacks and whites. Individuals are coded as leaving the first job if they leave it for another job without an intervening spell of nonemployment, and as having lost it if they transit into nonemployment.¹² It is possible for a nonemployment/job spell series to be inadmissible. This occurs if either one of the spells has an erroneous start or stop date.¹³ These spell series are excluded as well as those with job spells that are not in the private sector and those with missing wages or hours.

Wages

Wage data in the NLSY are categorized according to time rates: hourly, daily, weekly, bi-weekly, monthly, and annually. Each time a respondent is questioned about a job, wage information is collected. For this study the first wage reported for a job is used as the accepted wage offer. All wages are converted into weekly wages and reported in 1990 constant dollars. Some care needs to be taken with the wage information reported by the NLSY. Several cases can be found where a respondent reports a time rate that does not agree with the pay rate. To identify these problem responses we cross-checked all time and pay rate responses against upper and lower bounds collected for males of

¹² Respondents with less than two weeks between the end date of their first job and the start date of the next are coded as having left to make a job-to-job transition.

the same age and education from the Current Population Surveys (CPS) for 1979-1992.¹⁴ Those respondents with wages that do not fall within the acceptable ranges are dropped from the sample. Wages greater than \$600 per week are also dropped from the sample.

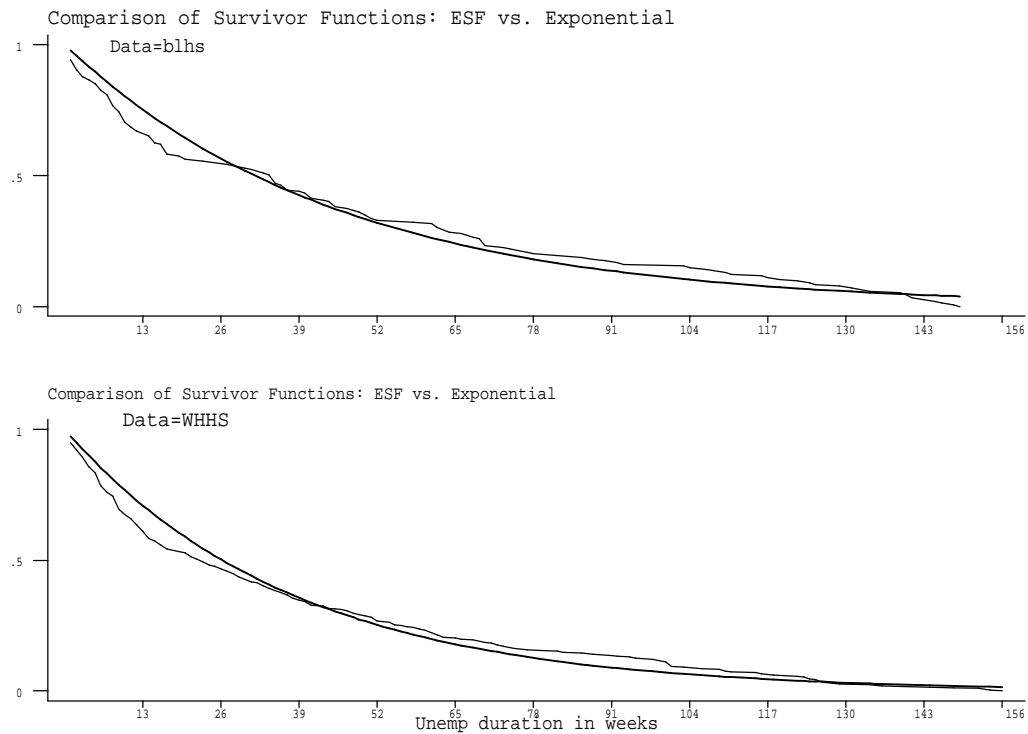


Figure 1
Nonemployment Duration

Sample Characteristics: Durations and Accepted Wages.

The raw data on durations and accepted wages are informative about the underlying search process. In Figure 1 we show the survivor function for nonemployment durations for black and for white high school graduates. In each case we plot the Kaplan-Meier estimate of the survivor function along with the survivor function for an exponential

¹³ A few spells in the data have no start and/or no stop dates and a few have start dates which are greater than their stop dates.

¹⁴ We use the 5th and 95th percentiles from the set of hourly wages for paid hourly workers from the March outgoing rotation groups for

random variable with the same mean duration. The visual impression one gets from these data is that there is slight drop off in the hazard rate in the second quarter of non-participation - from 13 to 26 weeks - but that overall for both groups the exponential specification does not appear at odds with the data. We will return to this point in section 5 below where we formally test this proposition. In comparing blacks and whites we find a substantial difference in mean nonemployment durations. On average, black high school graduates spend 45 weeks searching for full-time employment after graduation while whites spend 37 weeks. Taken together with the higher tendency of whites to already be employed at the time of graduation (32% versus 22%), black males experience a more difficult time making the transition to work.

Figure 2 shows the empirical distribution of accepted wages for high school graduates by race. Along with the histogram of wages we have drawn the frequency function for a log normal distribution with the same mean and variance as the underlying data. In both cases the data suggest that the wage distribution is less skewed to the right than a simple log normal. The mean accepted wage for white high school graduates is \$291.70 and for black graduates is \$255.08. Thus we find a substantial black-white wage differential of 13% even among first jobs. Given the age of the respondents (14 to 21) at the start of the survey in 1979, the majority graduate from high school between 1978 and 1984. This period covers the 1980-1982 recession. Examining wages in the year these first jobs started, we find no significant effect of the recession for either blacks or whites.

each year. This avoids wage conversion problems also present in the CPS. The lower bounds are quite close to the minimum wage.

However, as pointed out by Eckstein and Wolpin (1995), whites who graduate during the recession have significantly lower wages on their first jobs.¹⁵ Further analysis of this effect in our sample reveals this result stems not from a cycle effect but rather from a general decline in wages over the sample period.¹⁶ This decline fits with evidence from Juhn, Murphy and Pierce (1993), although in these data it appears to be too large. Further investigation of this patterns seems warranted but is both beyond the scope of this paper and the NLSY data.¹⁷

One indicator of specification problems in a stationary search model is the presence of correlation between accepted wages and search duration.¹⁸ For white high school graduates the correlation between accepted wage and measured search duration is -0.072, with a marginal significance level of 20%. Similarly, for black high school graduates the correlation is -0.149, with a marginal significance level of 6%. Thus these data do not suggest evidence of a changing reservation wage or of unmeasured worker heterogeneity, both of which would generate correlation between accepted wages and search duration.

¹⁵ Wages for blacks are also lower but not significantly so.

¹⁶ Regressing wages on either the starting year or the graduating year reveals a significant downward trend for both black and white high school graduates over this time period.

¹⁷ Given the panel nature of the NLSY and the original age constraints, 1984 effectively contains the last cohort of high school graduates. Thus it is difficult to distinguish between trend, cycle and other unobservable effects.

¹⁸ The interpretation of the correlation depends upon the alternative one has in mind, but all stationary search models with homogeneous workers predict that the correlation should be zero.

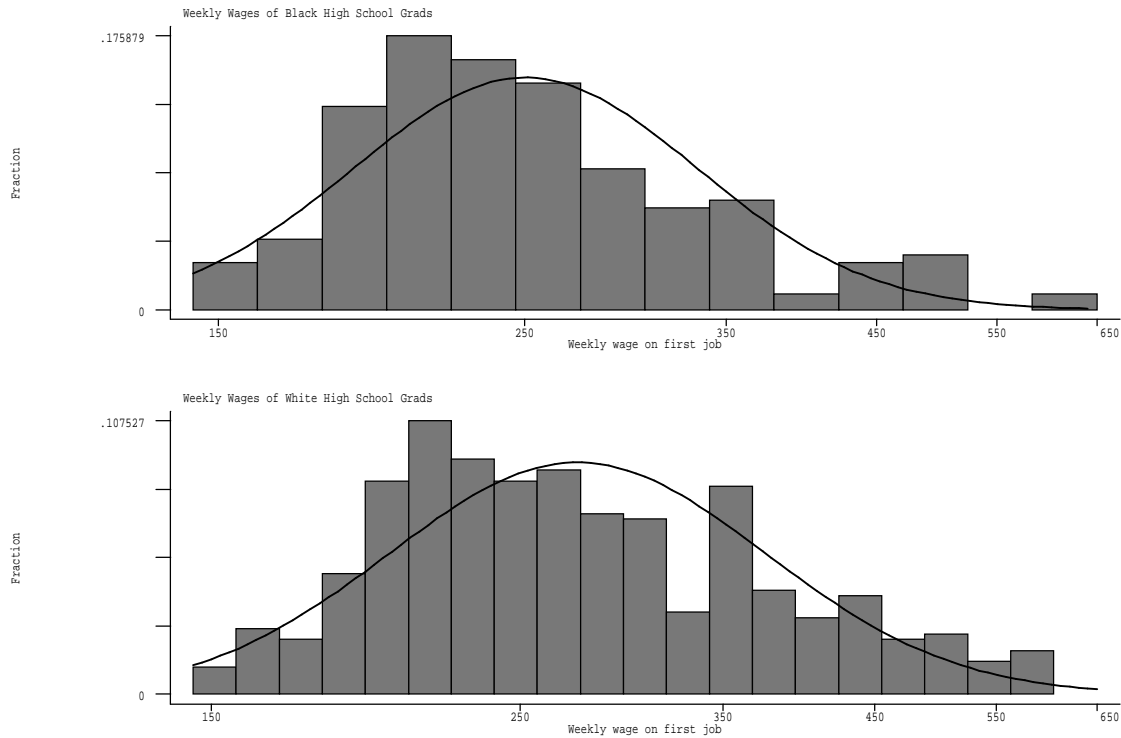


Figure 2
Accepted Wages

In addition to differences across blacks and whites in nonemployment durations and accepted wages, differences also emerge in job spell durations and transition patterns. The mean job spell duration (including censored spells) for white high school graduates is 114 weeks while it is only 85.4 weeks for black high school graduates. Of those job spells that end within the sample period, 71% end in a transition to nonemployment for black graduates. The comparable figure for white graduates is only 55%. A further implication of search models is that the expected duration of a job spell is positively related to the accepted wage. This pattern is also found in these data -

$\rho_{D2,W} = 0.258$ (with $msl=0.0002$) for black high school graduates and $\rho_{D2,W} = 0.107$ (with $msl= 0.021$) for white high school graduates, although the overall levels of correlation (especially for whites) are surprisingly weak. We return to this issue in section 5.

We also examine the employment outcomes of workers who start jobs before their education is completed. In these data 44 of 199 black high school graduates and 149 of 465 white high school graduates have such employment. Accepted wages, on average, are the same whether the job began before schooling ended or not. For blacks the difference between the group means is \$7.15, while for whites the difference is \$5.44. Compared to mean weekly earnings of \$255 for blacks and \$291 for whites, these differences are both statistically and practically small. On average, black high school graduates worked 23.4 weeks prior to graduating. Similarly, white high school graduates worked 34.8 weeks before school ended. This leads to a marginally significant difference in mean job spell durations. For blacks, those who started employment before they completed schooling stayed with their first employer 29 weeks longer ($t=1.33$), while for whites the difference in job durations was 20 weeks ($t=1.40$).

5. Estimation Results

Tables 3 and 4 contain the estimation results for white and black males, respectively. In both cases a large improvement in the log likelihood function occurs when heterogeneity is introduced. As in Bowlus, Kiefer and Neumann (1995), the value

of the log likelihood function stabilizes as Q gets larger. Using our criterion the optimal choice for Q is 4 for whites, and either 3 or 4 for blacks. For ease of comparison of wage distributions we choose a value of Q = 4 in both cases. Examining the behavior of the parameter estimates over Q we find that they converge as well. At levels of Q beyond the optimal choice, the differences in the parameters across Q are insignificant. This is primarily due to the inclusion of duration and transition data, in addition to wage data, in the log likelihood function.

Table 3				
Estimates for White Male High School Graduates				
Q	λ_0	λ_1	δ	LogL
1	0.026878 (.001542)	0.004079 (.000301)	0.005143 (.000302)	-7064.67
2	0.026878 (.001542)	0.006947 (.000497)	0.004612 (.000291)	-6932.55
3	0.026878 (.001542)	0.007877 (.000562)	0.004445 (.000287)	-6905.17
4	0.026878 (.001542)	0.008041 (.000574)	0.004409 (.000287)	-6898.73
5	0.026878 (.001542)	0.008119 (.000581)	0.004395 (.000287)	-6897.58
6	0.026878 (.001542)	0.008321 (.000596)	0.004364 (.000286)	-6895.67

Table 4 Estimates for Black Male High School Graduates				
Q	λ_0	λ_1	δ	LogL
1	0.022258 (.001798)	0.003207 (.000450)	0.008448 (.000694)	-3064.56
2	0.022258 (.001798)	0.006940 (.000923)	0.007730 (.000674)	-2958.37
3	0.022258 (.001798)	0.007472 (.000971)	0.007634 (.000673)	-2939.99
4	0.022258 (.001798)	0.008001 (.001065)	0.007551 (.000671)	-2938.17
5	0.022258 (.001798)	0.008165 (.001078)	0.007519 (.000670)	-2936.45
6	0.022258 (.001798)	0.007978 (.001080)	0.007532 (.000672)	-2935.97

Comparing the estimates across the two groups reveals that blacks face a significantly higher arrival rate of offers while unemployed, λ_0 , and a significantly higher job destruction rate, δ . Interestingly λ_1 , the arrival rate of job offers while employed, is not different for young black and white males. These results imply much lower relative competition levels for blacks than whites.¹⁹ The values of $\kappa_0 = \lambda_0 / \delta$ and $\kappa_1 = \lambda_1 / \delta$ for white males are 6.10 and 1.82, respectively, while for black males they are 2.95 and 1.06. Table 5 shows the estimates of r , w_{Hj} , γ_j and P_j , $j=1, \dots, Q^*$, for both groups. It is

¹⁹ The ratios of the job offer arrival rates to the job destruction rate, κ_0 and κ_1 , can be viewed as the relative competition levels firms face for workers.

interesting to note that black and white men have similar reservation wages²⁰ and similar degrees of heterogeneity in productivity . However, as stated before, blacks earn 87.4% of whites in this sample. This wage differential arises because job destruction is higher for blacks²¹ and because of differences in productivity. The higher outflow of blacks means the effective competition for them is almost half that for whites (compare κ_1 s) giving firms more monopsony power in the black labor market and lowering their wages. In addition, a higher job destruction rate slows the rate at which blacks are able to move up the wage distribution through job-to-job transitions. The difference in the values of κ_1 accounts for approximately three-quarters of the black-white wage differential.²² The remainder stems from differences in the productivity distribution. Even though average productivity is higher in the black labor market (\$520 versus \$486), the white distribution produces higher average wage offers. The higher mean comes primarily from the high value of P_4 for blacks.²³ However, a job offer from this firm type is highly unlikely, and so mean wage offer differences are primarily driven by the higher values for whites at low productivity firms, the lower fraction of P_1 firms for whites, and the higher fraction of P_2 firms for whites. Thus, while

²⁰ This is likely due to the choice of high school graduates, our wage screening procedure, and a binding minimum wage during this period.

²¹ The large difference in the arrival rates of offers while unemployed, λ_0 , only affects the wage offer distribution through the reservation wage. If blacks and whites had the same value of non-market time, b , then a lower value of λ_0 for blacks would imply a lower reservation wage for blacks and would add to the explanation of any wage differential. However, in the presence of a binding minimum wage that avenue is eliminated.

²² There is a \$37 difference in mean wages. Holding reservation wages constant the search model predicts the mean wage offer for blacks would increase by \$25 if blacks faced the whites' value of κ_1 and the mean wage offer of whites would decrease by \$29 if whites faced blacks' κ_1 .

²³ High values of P often result in the estimation when the estimated value of κ_1 is low and relatively high wages are present in the data.

Table 5 Estimated Support Points of the Wage Distribution		
Parameters	White Males Q=4	Black Males Q=4
r	145.54	143.79
w_{H1}	284.29	253.16
γ_1	0.5825	0.6315
P_1	372.66	344.55
w_{H2}	365.63	281.83
γ_2	0.8046	0.7563
P_2	484.18	411.10
w_{H3}	472.29	373.26
γ_3	0.9427	0.9301
P_3	682.18	618.22
w_{H4}	597.16	598.47
γ_4	1	1
P_4	1164.92	2060.84

productivity differences are important, the greater presence of the wage differential is explained in this model by differences in labor market competition across the groups.

Asymptotic standard errors for λ_0 , λ_1 , and δ are shown in parentheses in Tables 3 and 4. Table 6 provides bootstrap standard errors for λ_0 , λ_1 , and δ using the white high school graduate sample, based on 100 replications for the parameters corresponding to Q^* . We note that the standard bootstrap may not provide the correct coverage probabilities in this case because we have estimated the endpoints of the distribution

(Bickel and Freedman (1981)). Beran and Ducharme (1991) and Swanepoel (1986) show that this problem can be avoided by undersampling in the bootstrap. That is, instead of drawing N bootstrap observations with replacement from the sample of N observations, one needs to draw $M \leq N$. Unfortunately, the asymptotic bounds on the sequence needed do not provide any aid in choosing M in application.²⁴ To provide some feel for how to choose M we ran a small Monte Carlo experiment to bootstrap the distribution of the maximum from a sample of N observations. The distribution was $U(0,1)$. We set $N=500$, and drew a sample. We bootstrapped from this sample setting M at 100,200,...,500. We examined the sampling distribution of the maximum value, $X_{(N)}$, and the normed and centered statistic defined by

$$R_{M,N} = M(X_{(N)} - X_{(M)}^*) / X_{(N)} \quad [11]$$

The sample maximum has density function $g(x) = Nx^{N-1}$ and $R_{M,N}$ is asymptotically distributed as extreme value 1 in the case of the unit uniform. Here $X_{(M)}^*$ is the first order-statistic from the resample of M observations and $X_{(N)}$ is the sample maximum observation.

In this experiment we find that the 90 and higher percentiles of $R_{M,N}$ line up very well for the sample maximum, even for $M/N = 1$ for sample sizes of 500. For the centered and normed maximum this agreement fades as the ratio of M/N exceeds 70%.

In Table 6 we report standard errors corresponding to $M=N$; i.e., no under-sampling. Results using a 30% undersampling scheme are similar, although the standard errors

²⁴ Swanepoel's bound requires that the sequence M_N must satisfy $O(N^\epsilon) \leq M_N \leq o((N^{1+\epsilon}/\log(N))^{1/2})$, which is of little use in choosing

are a little larger.

In Table 6 we also compare the standard errors from the Hessian for the regular parameters to those from the bootstrap. They line up well, showing that the asymptotic standard errors are a good approximation. As noted, above the bootstrap standard errors are the only ones available for the change-point parameters. Note that, as expected, these parameters are quite precisely estimated.

Diagnostics

With estimates of the model in hand we can ask how well does this simple model fit the data. For non-employment spells the issue is whether the assumption of exponentiality can be rejected. Figure 1 above suggests no great discrepancy, but we proceed to a formal test. Within groups the duration of recorded search is exponential with intensity parameter λ_0 so a natural test would be White's Information Matrix test, which in this context can be shown (Kennan and Neumann (1988)) to test whether the variance equals the mean squared. There is substantial evidence that the finite sample distribution of the IM test is poorly approximated by its asymptotic distribution for samples of the size used in most economic studies, (i.e., $N < 50,000$). However, Horowitz (1994) has shown that bootstrap-based critical values eliminate essentially all of the size problem with this test. We bootstrapped the IM test for exponential search spells using 1,000

Table 6a
Comparison of Hessian and Bootstrap
Standard Deviations For White Male NLSY Sample

M for a given sample.

Parameter s.d.'s x (10 ⁺³)	Hessian	Bootstrap	Bootstrap CV
λ_0	1.542	1.844	0.068
λ_1	.574	0.736	0.097
δ	.287	0.425	0.095

Table 6b Bootstrap Standard Deviations of Cut Points			
Parameter	S.D.	Parameter	S.D.
r	3.638	P ₁	21.767
W _{H1}	28.782	P ₂	57.484
W _{H2}	43.857	P ₃	97.972
W _{H3}	50.611	P ₄	157.307
W _{H4}	4.941	--	--

replications and obtained marginal significance levels of 0.549 for white high school graduates and 0.492 for black high school graduates. Thus, in both cases we cannot reject the hypothesis that waiting times are exponential. This is not surprising. In the data the mean observed duration is 45.6 weeks for blacks and the standard deviation is 43.7 weeks. For whites, comparable figures are a mean of 37.8 weeks and standard deviation of 38.9 weeks. Therefore, a test that looks at whether the variance equals the mean squared will not find evidence against it in these data. For example, the standard chi-squared based 95% confidence interval (which is based on, and sensitive to, a normality assumption) yields confidence intervals of 38.7 to 52.5 weeks for the standard

deviation of search spells for blacks, and 33.4 to 41.1 weeks for whites. Of course, this conclusion would be reversed if we treated the zero duration spells as having 1 week duration. Using that convention results in a mean “duration” of 35.7 weeks for blacks with a standard deviation of 42.8 , and for whites a mean of 26 weeks and a standard deviation of 36.4 weeks. Both the bootstrapped White IM test and the classical chi-square test for the standard deviation equal to the mean strongly reject the hypothesis.

These tests look only at the relation of the mean and variance. To pursue these diagnostic issues further we examine the slope of the hazard function. To do so, we fit a Weibull model to the duration data and test whether the estimated shape parameter, ρ , differs significantly from 1.²⁵ The results are shown in Table 7 for only duration greater than zero (Dur1) and durations defined as Dur2 = Dur1 + I(Dur1==0)*1. For both groups using only durations recorded as positive, the shape parameter is estimated to be .92, which is fairly close to 1. For blacks the asymptotic confidence interval straddles 1.0, while for whites it is, to two decimal points, right on the line.

Table 7
Specification Tests For Exponential Search Times

Dependent Var/ Group	ρ	95% Upper Limit	95% Upper Limit
Dur1			
Black HS Graduates	.9190	.8089	1.0441
White HS Graduates	.9151	.8387	.9985
Dur2			
Black HS Graduates	.6669	.5964	.7459
White HS Graduates	.6255	.5827	.6714

When durations for which we do not have a forward recurrence time (Dur1 ==0) are

treated as having 1 week of search, as in Dur2, it is obvious from the table that the shape parameter reveals a declining hazard rate; i.e., $\hat{\rho} < 1$.

We conclude from this that the assumption of exponential search times is not rejected by the data once it is realized that the observed times are not completed durations but forward recurrence times from the date of completed education. The claim that “It is a well established empirical regularity that hazard rates are decreasing with the duration of search... “ (Eckstein and Wolpin, 1995, p. 263) is not supported by the data and is based only on a coding convention.

Figure 3 shows the empirical distribution function of wages and the theoretical cumulative distribution function implied by the parameter estimates for each group. The overall fit of the wage data is quite good, although for both groups there is a tendency to over-predict wages in the left tail.

²⁵ See Neumann [1997] for details on diagnostic testing in duration models.

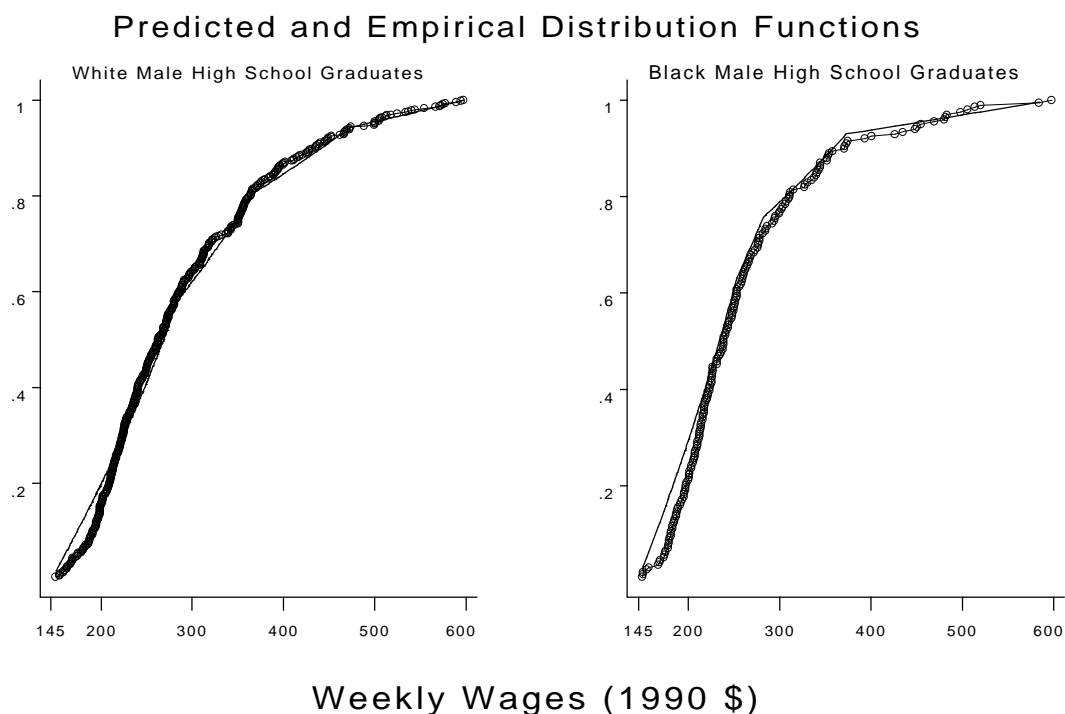
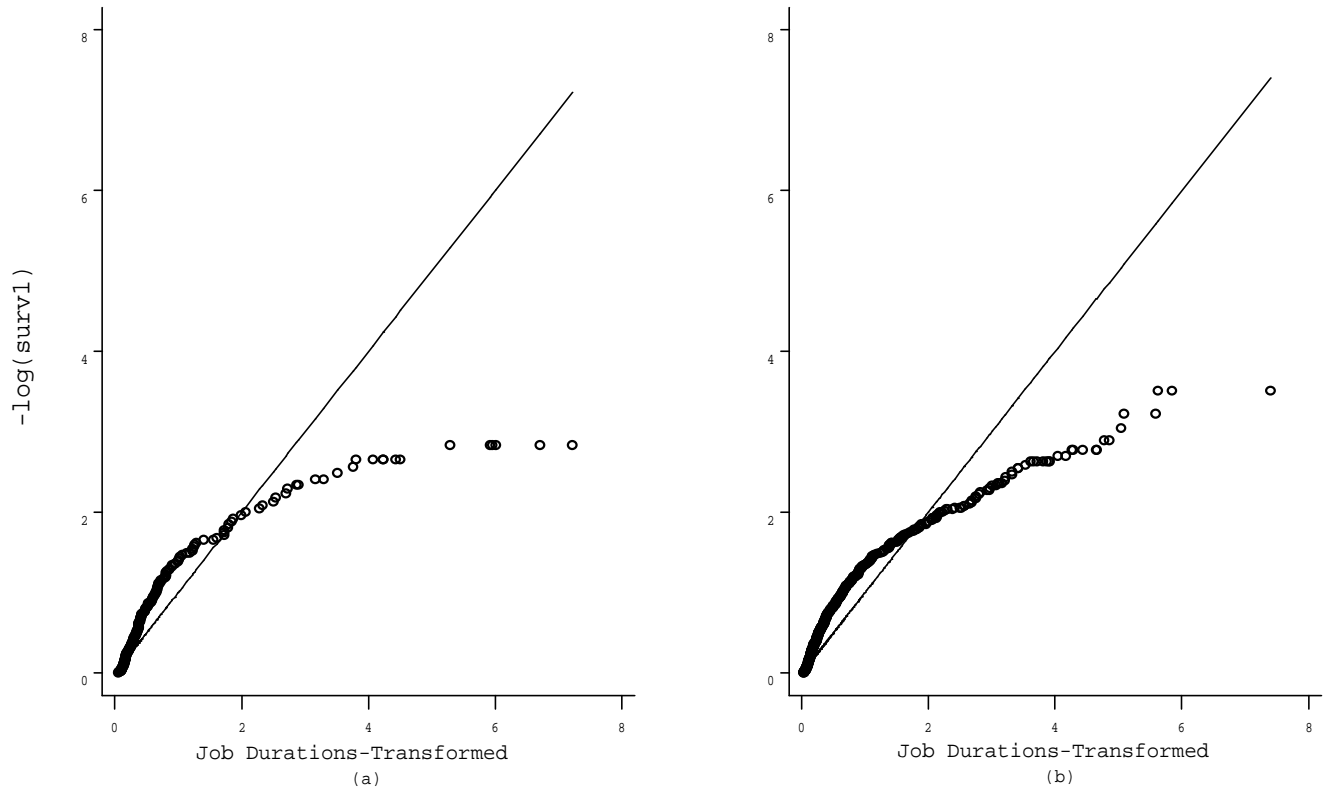


Figure 3

Of course, closeness is measured here in terms of the distribution function. The theoretical probability density function increases monotonically on any sub-interval of w , that is, on any interval where P is constant, and consequently it would not be close to any data that had a declining density. This caveat aside, Figure 3 shows that the wage data can be reconciled using only 4 types of firms.

Given the acceptable fit of the model to the wage and nonemployment data, we turn our attention to the job spell data. Job durations, D_2 , in the model are exponential with parameter $\delta + \lambda_1(1-F(w))$. They are independent but, because of the dependence on the current wage w , not identically distributed across individuals. Thus, we observe a mixture of exponentials. To test whether or not this distributional assumption holds we

note that the following transformation of the data - $z=(\delta + \lambda_1(1-F(w)))D_2$ - is independently and identically distributed as unit exponential if all of the job durations are completed spells. In both high school graduate samples the level of censoring is less than 10%. Therefore, we ignore the complications in testing brought about by the presence of censoring. In Figure 4 we test unit exponentiality of the transformed durations using a graphical test (see Horowitz and Neumann (1989)). If the distributional assumption holds, then the integrated hazard function of the transformed durations should be a straight line with slope 1. Clearly in both cases the test fails. More formal tests, such as estimating a Weibull model, lead to the same conclusion. The graphs indicate that there are too few long spells in the data. To assess the reasons why the job durations reject the exponential specifications, we performed a small Monte Carlo experiment using the estimated values of $\delta + \lambda_1[1-F(w)]$ to generate the transformed unit exponential job durations. Table 8 shows data for the high school graduate sample and from simulations of exponential job durations. Mean duration is lower for both blacks and whites and the quartiles indicate that this holds throughout the mid-range of the duration distributions. In other words, the actual distribution is skewed to the left, indicating that there are many more short duration jobs than would be expected. Interestingly, a simple hypothesis such as “short term jobs are dead-end jobs with low wages” is not supported by the data. White high school graduates have a correlation between wages and job duration that, while positive, is only half that predicted by the model, and black high school graduates job duration is more correlated with the wage than would be expected.



Integrated Hazard Functions For Job Durations
Figure 4

Table 8. Monte Carlo Results of Distribution of Job Spell Durations				
	Whites		Blacks	
	Sample	Monte Carlo	Sample	Monte Carlo
Mean	110.97	128.69	88.69	93.04
25th %-tile	20	35	18	26
50th %-tile	50	85	37	63
75th %-tile	131	174	84	128
Corr(D_{2,w})	.1073	.2696	.2575	.1841

It is possible that the poor fit of the job duration data is due to the estimates of δ

and λ_1 being unduly influenced by the wage data. To estimate the arrival rates one needs unemployment durations, job durations and wages. However, one need not impose the model's implied structural form of the wage offer distributions, as identification only requires an estimate of the empirical distribution function of the wage offer distribution $F(w)$.²⁶ Re-estimating δ and λ_1 using the empirical distribution function of wages yields estimates that are very close to those of $Q=4$ for whites ($\lambda_1 = .007898$ and $\delta = .004420$) and $Q=3$ for blacks ($\lambda_1 = .007588$ and $\delta = .007603$). Tests of the distributional assumptions using these parameters do not overturn the results found above. Thus, high school graduates spend less time in their first job given their wages, than implied by the search model. It may be that the job shopping tendencies of young workers lead to exit patterns inconsistent with the search model's rule of only voluntarily leaving a job if another with a higher wage is waiting.

6. Conclusions

We have demonstrated that the equilibrium search model with heterogeneity in firm productivity can be estimated and provides a dramatic improvement in fit with respect to wage data relative to the homogeneous model. The estimation problem is nonregular, but its special features can be exploited to design an efficient estimation procedure. A limited sampling experiment verifies that the estimates are feasible and have reasonable properties. We provide an application using NLSY data on new entrants in the U.S. labor

²⁶ Of course, the productivity distribution can not be recovered without using the model's implied structure of the wage offer distribution.

market. Since the model is nonlinear and nonregular we provide bootstrap standard errors of our parameter estimates. Diagnostic tests show the model provides a satisfactory fit to nonemployment duration and wage data, but has trouble fitting job durations. Our results show that nonemployed blacks receive fewer offers than whites and employed blacks are more likely to lose their jobs. Importantly, employed blacks and whites receive job offers at the same rate. However, the difference in job destruction rates is so great that it can account for three-quarters of the black-white wage differential. These findings underscore the importance of labor market institutions - affecting matching and job loss - in explaining racial differences in incomes.

References.

- Albrecht, J. W., and B. Axell, [1984] "An Equilibrium Model of Search Unemployment," Journal of Political Economy, 92:824-40.
- Beran, R. and G.R. Ducharme, [1991] "Asymptotic Theory for Bootstrap Methods in Statistics," Les Publications CRM, Centre de Recherches Mathematiques, Universite de Montreal.
- van den Berg, G. J. and G. Ridder, [1993a] "Estimating Equilibrium Search Models from Wage Data," in N. C. Westergaard-Nielsen and P. Jensen, eds., **Proceedings of the Third Symposium On Panel Data and Labour Market Dynamics**, New York: North-Holland, 43-55.
- _____, [1993b] "On the Estimation of Equilibrium Search Models from Panel Data," in J.C. Van Ours, G. A. Pfann, and G. Ridder, eds., **Labor Demand and Equilibrium Wage Formation**, New York: North-Holland, 227-245.
- Bickel, Peter J. and David A. Freedman, [1981] "Some Asymptotic Theory for the Bootstrap," The Annals of Statistics 9: 1196-1217.
- Billingsley, Patrick, [1961], **Statistical Inference for Markov Processes**, University of Chicago Press, Chicago.
- Bontemps, Christian, Jean-Marc Robin, and Gerard J. van den Berg, [1996] "Equilibrium Search with a Continuous Distribution of Productivities: Theory and Estimation," mimeo.
- Bowlus, A., N. M. Kiefer, and G. R. Neumann, [1995], "Estimation of Equilibrium Wage Distributions with Heterogeneity," Journal of Applied Econometrics, 10: S119-S131.
- Burdett, K., [1990], "Search Models: A Survey," University of Essex Discussion Paper.
- Burdett, K., and D. T. Mortensen, [1995], "Wage Differentials, Employer Size, and Unemployment," International Economic Review, forthcoming.
- Card, David, and Alan B. Krueger, [1995], **Myth and Measurement: The New Economics of the Minimum Wage**, Princeton University Press: Princeton, N.J.

- Chernoff, H., and H. Rubin, [1956], "The Estimation of the Location of a Discontinuity in Density", **Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability**, Berkeley: University of California Press, 1, 19-37.
- Christensen, B. J. and N. M. Kiefer, [1994], "Local Cuts and Separate Inference," Scandinavian Journal of Statistics, 20:1-13.
- Cox, D. R., and P. A. Lewis, [1966], **The Statistical Analysis of Series of Events**, Methuen Co.: London.
- Eckstein, Z., and K. I. Wolpin, [1990], "Estimating a Market Equilibrium Search Model From Panel Data on Individuals," Econometrica, 58:783-808.
- _____ [1995] "Duration to First Job and the Return to Schooling: Estimates from a Search-Matching Model," Review of Economic Studies, 62: 263-286.
- Heckman, J. J., and B. Singer, [1984], "Econometric Duration Analysis," Journal of Econometrics, 32:59-84.
- _____, [1986], "Econometric Analysis of Longitudinal Data," Chapter 29 in Z. Griliches and M. D. Intriligator, eds., **Handbook of Econometrics, Volume III**, North-Holland, Amsterdam, 1689-1763.
- Hicks, J. R. [1932], **The Theory of Wages**, MacMillan, London. [2nd edition, 1966].
- Horowitz, J. L., [1994], "Bootstrap-based Critical Values for the Information Matrix Test," Journal of Econometrics, 61:395-411.
- Horowitz, J. L. and G.R. Neumann, [1989], "Specification Testing in Censored Regression Models: Parametric and Semiparametric Methods," Journal of Applied Econometrics, 4: S61-S86.
- Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce, [1993], "Wage Inequality and the Rise in Return to Skill," Journal of Political Economy 101(3): 410-442.
- Kennan, J., and G. R. Neumann, [1988], "Why does the Information Matrix Test Reject So Often?," Working Paper No. 88-4, Department of Economics, University of Iowa, Iowa City, IA.

- Kerr, C. S., [1954], "The Balkanization of Labor Markets," in E. Wright-Bakke et al., eds., **Labor Mobility and Economic Opportunity**, MIT Press, Cambridge, MA.
- Kiefer, N. M., and G. R. Neumann, [1993] "Wage Dispersion with Homogeneity: The Empirical Equilibrium Search Model," in N. C. Westergaard-Nielsen and P. Jensen, eds., **Proceedings of the Third Symposium On Panel Data and Labour Market Dynamics**, New York: North-Holland, 57-74.
- Kirkpatrick, S., C. D. Gelatt, Jr., and M.P. Vecchi, [1983] "Optimization by Simulated annealing," Science, 220:671-80;
- Koning, P., G. Ridder, and G. J. van den Berg, [1995], "Structural and Frictional Unemployment in an Equilibrium Search Model with Heterogeneous Agents," Journal of Applied Econometrics, 10: S133-151.
- Mortensen, D. T., [1990] "Equilibrium Wage Distributions: A Synthesis," in J. Hartog, G. Ridder, and J. Theeuwes, eds., **Panel Data and Labor Market Studies**, New York: North-Holland, 279-96.
- Mortensen, D. T., and G. R. Neumann, [1988], "Estimating Structural Model of Unemployment and Job Duration in Dynamic Econometric Modeling," **Proceedings of the Third International Symposium in Economic Theory and Econometrics**, Cambridge University Press.
- Neumann, G. R., [1997], "Search Models and Duration Data," ch. 3 in H. Pesaran and P. Schmidt, eds., **Handbook of Applied Econometrics: Volume II Microeconometrics**, Basil Blackwell, Oxford.
- Otten, R., and L. van Ginneken, [1989], **The Annealing Algorithm**, Boston: Kluwer.
- Swanepoel, Jan W.H., [1986], "A Note on Proving that the (Modified) Bootstrap Works," Communications in Statistics: Theory and Methods 15: 3193-3203.
- Szu, H., and R. Hartley, [1987] "Fast Simulated Annealing," Physics Letters, A, 12:157-62.