Fitting Equilibrium Search Models to Labour Market Data

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

Draft Date: March 14, 1995
Print Date: March 5, 1996

*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. This work was supported in part by NSF grant SES-9122253.

1. Introduction

The essential idea of equilibrium search models of labour market behaviour 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 Labour 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 labour 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 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 labour 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 (1993,1994), and Kiefer and Neumann (1994) 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. The key insight was the addition of the reasonable assumption that workers search while on the job and change jobs for higher wages. 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 is the data on the search time to find employment, the wage rate received 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 labour supply to firms, and since workers prefer higher wages, the distribution function of wages is predicted to have a thick right tail. On the other hand, 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 be used to explain almost any error 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, $j=1,..,J$, 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 durations, accepted wages, and job transitions. In Section 4 we apply this technology to labour market transitions for young males using the National Longitudinal Survey of Youth (NLSY). Our conclusions are given in the fifth and final section of the paper.

2. Equilibrium Search Models

To characterize labour 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:

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

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 distribution is:

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

with density

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

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, shown in Figure 1(a), is increasing over its support. This is the estimated equilibrium wage distribution from Kiefer and Neumann (1994).

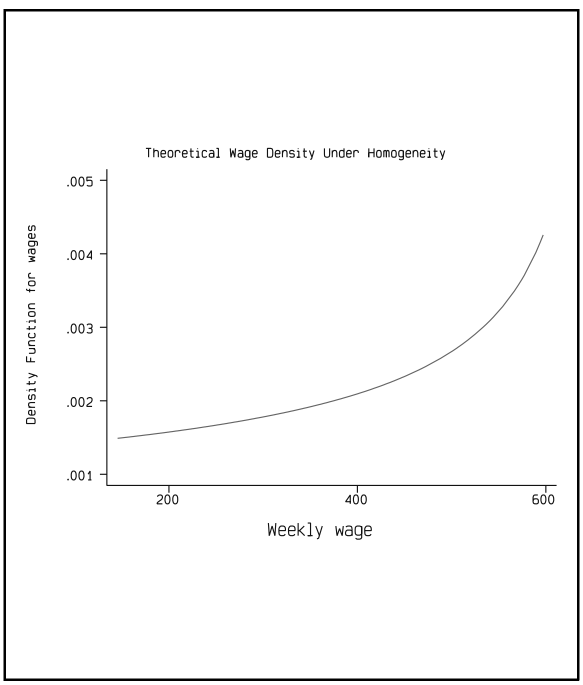

Figure 1(a)

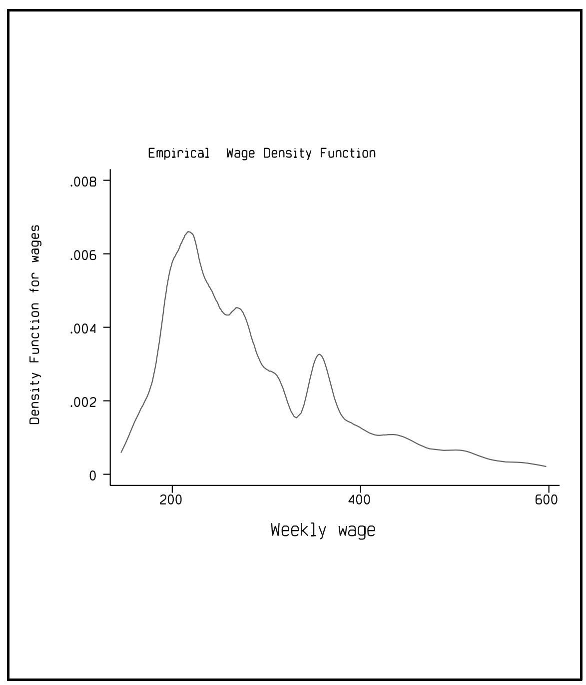

Figure 1(b)

But, as noted

above, empirical wage distributions typically have a thin right tail. Figure 1(b) shows the density function of accepted wage offers for white male high school graduates from the NLSY. The empirical density function shown was obtained using gaussian kernel estimator with a bin width of $10. The long right tail is apparent in these wage data, as is the presence of multiple modes, which provides some hint that the empirical density might be a generated by a mixture of densities. These features are not artifacts of the choice of kernel or binwidth; they are present in a variety of reasonable nonparametric density esimates.

The observed empirical distribution of wages can be made consistent with the equilibrium search model if there are differing levels of productivity across firms and if firms with especially high productivity are relatively rare. Heterogeneity in firm productivity can be introduced in two ways. Following Koning, Ridder and van den Berg (1995) one could assume P is distributed, say, log normally across firms. 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 between the distribution chosen to represent productivity and the resulting wage distribution due to the competition faced on either side by any one firm. Thus if one assumes a log normal distribution, as done by Koning, Ridder and van den Berg, the induced equilibrium wage distribution will be log normal. This leaves little room for the underlying parameters of the model to influence the shape of the distribution. The model is then less clearly a search model - in that there need be no 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 modelling heterogeneity as discrete are twofold. First, it allows the data to describe the shape of the wage offer function. Second, the classification of workers into different levels of productivity generates testable economic implications. The disadvantage is that discrete heterogeneity results in a likelihood function that is discontinuous in some of the parameters and whose estimation is non-standard. Nonstandard estimation issues arise frequently in 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 $P_1 < P_2 < ... < P_Q$. The fraction of firms having productivity P_j or less is $\gamma_j = \gamma(P_j)$. The equilibrium wage distribution, following Mortensen (1990), is:

$$F(w) = \varphi_j(w) \quad \forall \; w \tag{4}$$

with ϕ_j defined by

$$\varphi_j(w) = \frac{1 + \kappa_1}{\kappa_1}[1 - \frac{1+\kappa_1[1-\gamma(P_{j-1})]}{1+\kappa_1}[\frac{P_j - w}{P_j - w_{Hj}} \qquad w_{Lj} < w \le w_{Hj} \tag{5}$$

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,..,Q-1. Second, $F(w_{Hj}) = \gamma(P_j)$, j=1,...,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

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

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

The stochastic processes induced by the model described above completely characterize labour market histories, whether complete or incomplete. The data to be explained are durations 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 durations are exponential with intensity parameter λ_0. The marginal distribution of accepted wages is f(w) given by

$$f(w) = \sum_{j=1}^{Q} [[\frac{1+\kappa_1[1-\gamma_{j-1})]}{2\kappa_1}][(P_j - w)(P_j - w_{Lj})]^{-1/2}] \qquad \textbf{(7)}$$

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 durations, $f(D_2|w)$, is exponential with intensity parameter $(\delta + \lambda_1(1\text{-}F(w))$ 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

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

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

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

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

$$\theta_1 = < w_{H1}, w_{H2}, .., w_{HQ-1} >$$

$$\theta_2 = < r, w_{HQ} >$$

$$\theta_3 = < \lambda_0, \lambda_1, \delta >.$$

Observe that the likelihood function is continuous but not differentiable in $w_{H1},...,w_{HQ-1}$, which can be seen from the distribution function of wages shown in equation (5). However, the cut points w_{Hj}, j=1,..,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_1, is from r to w_{HQ} and these are parameters of the model. Kiefer and Neumann (1994) propose the estimators

$$\hat{r} = \min \{w_i\}, \quad \hat{s} = \max \{w_i\} \qquad \textbf{(9)}$$

for r and w_{HQ}. The estimators $\hat{r}$ and $\hat{s}$ are super-efficient but in finite sample 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.

A characterization of maximum likelihood estimators for w_{Hj} j=1,..,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.

Proof.

Write the likelihood for wage data

$$\ell(\Theta) = \sum_{w c \Theta} [1n(1/2\kappa_1) - 1/2 \ln[(P_1 - w)(P_1 - R)]]$$

$$+ \sum_{w \geq \Theta} [\ln((1 + \kappa_1)/2\kappa_1) - 1/2 \ln[(P_2 - w)(P_2 - \Theta)]]$$

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

$$2\ell_{\theta\theta} = \sum_{w<\theta} [(P_1 - w)^{-2} + (P_1 - R)^{-2}](1 - B_1)^{-2}$$

$$+ \sum_{w \geq \Theta} [(P_2 - w)^{-2} + (P_2 - \theta)^{-2}(-B_2/(1 - B_2))^2$$

$$+ \sum_{w>\theta} [(P_2 - \theta)^{-2}(B_2/(1 - B_2))$$

$$+ \sum_{w \geq \theta} [(P_2 - \theta)^{-2}(B_2/(1 - B_2) + 1)$$

$$> 0$$

Theorem 1 tells us that there is no purpose looking for solutions to normal equations for 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 $\hat{w}_{Hj}$, 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 x 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 $P = A^*W_H$, where P is an Mx1 vector of productivities of the M firm types, and W_H is an (M+1)x1 vector of wage cut points --$W_H(1) = R$, $W_H(2) = w_{H1}$, etc. When P is not ordered, the estimate of $\hat{w}_{Hj}$ 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 $\ell(\theta_1|\theta_2, \theta_3)$ is maximized using Simulated Annealing (Szu and Hartley (1987), Otten and Ginneken, (1989)) to obtain an estimate of θ_1,$\hat{\theta}_1$. In the smooth maximization step $\hat{\theta}_1$ is used to maximize $\ell(\theta_3|\hat{\theta}_2,\hat{\theta}_1)$ 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 indepenent 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 $\ell(\theta_2, \theta_3 | \hat{\theta}_1)$ 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 - $-2*\Delta \log l < \chi^2_{.05}$ - 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 behaviour 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 λ_0 = .03, λ_1 = .01, δ = .0035, Q=3, r = 100, w_{H1} = 179, w_{H2} = 377, w_{H3} = 677.35, γ_1 = .3, γ_2 = .7. The model was estimated for Q= 1,...,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 the 500 replications the marginal distribution of the estimated Q is given in Table 1.

Table 1
Sample Distribution of $\hat{Q}$
True value of Q = 3

$\hat{Q}$ =	1	2	3	4	5+
$h(\hat{Q})$ =	.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 noticed in using the Heckman-Singer estimator. In fact, Heckman and Singer (1985) 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.[1] 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).

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

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.[2]

If we use a value of $\hat{Q}$ greater than the true value of Q, Q^o, 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 $\hat{Q}$ with those produced using Q^o. However, the root mean square error for δ decreases by 2.2% using the same

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

[2] The estimate of λ_0 does not depend upon the value of Q.

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 $\hat{Q}$ 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 wages and durations from the NLSY for male high school graduates who did not continue on in formal education.[3] We collect information on the unemployment spell between graduation and the first full-time job, the wage accepted, the job spell length, and whether the job was left for another or lost. We define the first full-time job as the first one that consists of thirty-five hours or more a week, lasts longer than two weeks, and starts within three years after high school completion. To make sure we have the first full-time spell after graduation we include only those individuals who finished high school after January 1, 1978.[4] If the first full-time job spell happens to surround the education date, it is used as the first spell only if the individual holds the job longer than two months after the education date. This eliminates summer jobs and temporary jobs held while in school. Individuals are coded as having left the first job if they leave it for another and as having lost it if they transit into unemployment. It is possible for an unemployment/employment spell series to be inadmissible. This occurs if either one of the spells has an erroneous start or stop date.[5] 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.

---

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

[4]Reconstructing the NLSY employment history prior to 1/1/78 is not possible.

[5]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.

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 1982 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 check all time and pay rate responses against upper and lower bounds collected for males of the same age and education from the Current Population Surveys (CPS) for 1979-1991.[6] 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.

Tables 3 and 4 contain the estimation results for white and black males, respectively. In both cases a large improvement in the log likelihood occurs when heterogeneity is introduced. As in Bowlus, Kiefer and Neumann (1995) the value of the log likelihood 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 distrinutions we choose Q=4 in both caes. Examining the behaviour 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.

---

[6]We use the 5th and 95th percentiles from the set of hourly wages for paid hourly workers from the March outgoing rotation groups for each year. This avoids wage conversion problems also present in the CPS. The lower bounds are quite close to the minimum wage.

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 enough 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.[7] The values of κ_0 (= λ_0/δ) and κ_1 (= λ_1/δ) for white males are 3.03 and 2.96, respectively, while for black males they are 1.81 and 1.06. Table 5 shows the estimates of r, w_{Hj}, and γ_j and P_j, j=1,..,Q^*, for both groups. It is interesting to note that black and white men have similar reservation wages and similar distributions (white men have an additional cut point). However, blacks earn 87.4% of whites in this sample.[8] This wage differential arise, render this search interpretation because job offers while unemployed arrive sloer than for whites and because job destruction is higher. Because of this higher outflow blacks have less opportunity to advance up the wage distribution. Thus, the wage differential is explained in this model primarily by differences in labour market competition across the groups. Table 6 gives bootstrap standard errors based on 100 replications for the parameters corresponding to Q*. The standard errors from the Hessian are given for the regular parameters. These line up well with the bootstrap, showing that the asymptotic standard errors are a good approximation. For the change-point parameters the bootstrap standard errors are the only ones available. Note that, as expected, these parameters are quite precisely estimated.

---

[7]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.

[8]The mean wage for whites is \$291.70 and for blacks \$255.08.

Table 3
Estimates for White Male High School Graduates

Q	λ_0	λ_1	δ	LogL
1	0.013359 (.000769)	0.004079 (.000302)	0.005143 (.000303)	-7231.67
2	0.013359 (.000769)	0.006660 (.000477)	0.004643 (.000293)	-7100.35
3	0.013359 (.000769)	0.007855 (.000562)	0.004449 (.000288)	-7074.06
4	0.013359 (.000769)	0.008007 (.000572)	0.004415 (.000288)	-7065.83
5	0.013359 (.000769)	0.008224 (.000591)	0.004379 (.000287)	-7064.80
6	0.013359 (.000769)	0.008255 (.000769)	0.004372 (.000287)	-7063.22

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

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}	291.74	253.16
γ_1	0.6118	0.6315
w_{H2}	374.66	281.83
γ_2	0.8244	0.7563
w_{H3}	469.48	373.26
γ_3	.9399	.9301
w_{H4}	597.16	598.47

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

Parameter s.d.'s x (10^{+3})	Hessian	Bootstrap	Bootstrap CV
λ_0	.769	1.844	0.068
λ_1	.572	0.736	0.097
δ	.288	0.425	0.095

Table 6b
Bootstrap Standard Deviations of Cut Points

Paramaeter	S.D.	Parameter	S.D.
R	3.638	P_1	21.767
WH_1	28.782	P_2	57.484
WH_2	43.857	P_3	97.972
WH_3	50.611	P_4	157.307
WH_4	4.941	--	--

5. Conclusions

We have demonstrated that the equilibrium search model with heterogeneity in firms' productivities can be estimated and provides a dramatic improvement in fit 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 market. Since the model is nonlinear and nonregular we provide bootstrap standard errors of our parameter estimates. Our results show that unemployed blacks receive fewer offers than whites and employed blacks are more likely to lose their jobs. Importantly, employed blacks and whites have equal probabilities of advancing up the job ladder. The net result is that the black employment flows lead to higher concentrations in lower wage jobs. 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.

van den Berg, G. J., and G. Ridder, [1993] "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.

________, [1994] " Estimating Equilibrium Search Models from Wage Data," in N. C. Westergård-Nielsen and P. Jensen, eds., **Proceedings of the Third Symposium On Panel Data and Labour Market Dynamics**, New York: North-Holland, 43-55.

Bowlus, A., N. M. Kiefer, and G. R. Neumann, [1995], "Estimation of equilibrium wage distributions with heterogeneity," Journal of Applied Econometrics, forthcoming.

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.

Devine, T. J., and N. M. Kiefer, [1991], **Empirical Labor Economics: The Search Approach**, Oxford: Oxford University Press.

Eckstein, Z., and K. I. Wolpin, [1990], "Estimating a Market Equilibrium Search Model From Panel Data on Individuals," Econometrica, 58:783-808.

Heckman, J. J., and B. Singer, [1984], "Econometric Duration Analysis," Journal of Econometrics, 32:59-84.

Hicks, J. R. [1932], **The Theory of Wages**, MacMillan, London. [2nd edition, 1966].

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, [1991] "Notes On Monopsony Models", mimeo, Princeton University.

________, [1994] "Wage Dispersion with Homogeneity: The Empirical Equilibrium Search Model," in N. C. Westergård-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, forthcoming.

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 modelling," **Proceedings of the Third International Symposium in Economic Theory and Econometrics**, Cambridge University Press.

Otten, R., and L. van Ginneken, [1989], **The Annealing Algorithm**, Boston: Kluwer.

Szu, H., and R. Hartley, [1987] "Fast Simulated Annealing," Physics Letters, A, 12:157-62.