

Estimating Models of On-the-Job Search using Record Statistics*

Gadi Barlevy
Economic Research Department
Federal Reserve Bank of Chicago
230 South LaSalle
Chicago, IL 60604
e-mail: gbarlevy@frbchi.org

November 13, 2003

Abstract

This paper proposes a methodology for estimating job search models that does not require either functional form assumptions or ruling out the presence of unobserved variation in worker ability. In particular, building on existing results from record-value theory, a branch of statistics that deals with the timing and magnitude of extreme values in sequences of random variables, I show how we can use wage data to identify the distribution from which workers search. Applying this insight to wage data in the NLSY dataset, I show that the data supports the hypothesis that the wage offer distribution is Pareto, but not that it is lognormal.

Key Words: Job Search, Record Statistics,
Hausdorff Moment Problem, Non-Parametric Estimation

*I would like to thank Joe Altonji, Susan Athey, Marco Bassetto, Jeff Campbell, Kim-Sau Chung, Chuck Manski, Francesca Molinari, Chris Taber, Nick Williams, and especially Eric French and H. N. Nagaraja for helpful discussions. I also benefitted from comments at seminars in Arizona State, Michigan State, the Federal Reserve Bank of Minneapolis, the Federal Reserve Bank of Chicago, the Society of Economic Dynamics, and the NBER Summer Institute. The views expressed here do not necessarily reflect the position of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Introduction

In recent years, economists have made significant progress in estimating models of job search using data on wages and job durations. These estimates can be applied to analyze a host of questions that are of interest to economists. For example, we can use them to identify the sources of wage growth among labor market entrants, which in turn can be used to determine why wage growth varies by race and education, as was done in Wolpin (1992) and Bowlus, Kiefer, and Neumann (2001); or we can use these estimates to infer the degree to which current wage inequality among a given group of workers is likely to translate into inequality over lifetime earnings, as was done in Flinn (2002) and Bowlus and Robin (2003); or we can use them to predict the effects of changes in the minimum wage and other government policies on equilibrium wages and employment, as was done in van den Berg and Ridder (1998); or we can use them to calibrate macroeconomic models in order to gauge the effects of cyclical fluctuations on the labor market, as was done in Barlevy (2002). Clearly, the ability to estimate job search models is an important tool for economists who wish to better understand the operation of labor markets.

While previous authors have made significant progress in estimating such models, the procedures they suggest have some important shortcomings. For example, most of the aforementioned papers employ maximum likelihood using particular functional forms for the wage offer distribution (or alternatively for the distribution of firm productivities that gives rise to the wage offer distribution in equilibrium). As such, these papers assume rather than identify the shape of the wage offer distribution. But since there is no consensus as to what is an appropriate functional form, this seems unsatisfactory without some *a priori* evidence on the shapes of this distribution. Indeed, identifying the exact shape of the wage offer distribution will be important for some of the questions above. For example, the effects of changes in the minimum wage depend on how many firms set their wage close to the minimum wage level, and a functional form that fits the data well along some dimensions may do poorly in matching this region of the distribution. Although some nonparametric procedures to estimating the wage offer distribution have been proposed, most notably in Bontemps, Robin, and van den Berg (2000), these procedures do not allow for unobserved earnings heterogeneity, i.e. they require that differences in wages across workers be driven entirely by differences in wages paid across firms. In practice, though, even within narrowly defined groups, differences in wages are likely to reflect unobserved differences in ability beyond just differences in pay scales across firms. This paper proposes an alternative approach that allows us to estimate these models in the presence of unobserved and time-varying ability, even without having to specify the distribution of the unobserved ability across agents or over time.

The innovation of my approach is that rather than using data on wages and *job duration* to estimate the model, as previous work has done, I use data on wages and *individual work histories*, specifically the number of voluntary and involuntary job changes a worker experienced before settling into each of his jobs. Work history data allows us to treat jobs as records, where a record corresponds to an observation in a sequence that exceeds all of the observations that preceded it in the sequence. Statisticians have studied record processes as a particular branch of extreme value theory, and have applied them to study various phenomena such as record temperatures, record athletic performances, road congestion, optimal tolerance testing, and ruin probabilities.¹ A similar connection exists between record theory and models of on-the-job search: a worker's job at any point in time can be viewed as the record most attractive offer he received since his last involuntary job change. It turns out that we can exploit the implicit record structure of search models to make inference about the search problem workers face.

In what follows, I follow many of the above cited papers in focusing on the Burdett and Mortensen (1998) model of on-the-job search. That model implies an equilibrium wage offer distribution that depends on the distribution of productivity across all firms. I show that we can identify the shape of this wage offer distribution even when there is unobserved variation in workers' ability. Exact identification requires us to estimate an infinite list of record moments. Although we can obtain a consistent estimator for the offer distribution based on a finite number of moments, this estimator may not be very precise when we use a small number of moments. Nevertheless, even a small number of moments can yield sharp tests for particular hypotheses about the shape of the offer distribution. Using data from the National Longitudinal Survey of Youth (NLSY), I argue that the offer distribution for this sample is consistent with a Pareto distribution, which coincides with the functional-form assumptions in some of the papers above, e.g. Flinn (2002). At the same time, I can reject the hypothesis that the wage offer distribution is lognormal, which has also been cited as a plausible form for the wage offer distribution.

The paper is organized as follows. Section 1 introduces the concept of record statistics. Section 2 describes the model and shows how to use record statistics to estimate it non-parametrically. Section 3 describes data from the NLSY that can be used to implement this approach. Section 4 reports the results. Section 5 comments on the applicability of my approach to more general search models in which wages may not correspond to record statistics but where there is still an underlying record structure inherent to the model. Section 6 concludes.

¹An entertaining survey on the various applications of record statistics is provided in Glick (1978).

1. Record Statistics

Although statisticians have written extensively on record theory, it has attracted scant attention from economists.² I therefore begin with a quick overview of record statistics. More comprehensive reviews are available in Arnold, Balakrishnan, and Nagaraja (1992, 1998) and Nevzorov and Balakrishnan (1998).

Consider a sequence of real numbers $\{X_m\}_{m=1}^M$. An element in the sequence is a *record* if it exceeds the realized value of all observations that preceded it in the sequence. Formally, let $M_1 = 1$, and for any integer $n > 1$ define the observation number of the n -th record M_n recursively as

$$M_n = \min \{m : X_m > X_{M_{n-1}}\} \quad (1.1)$$

The n -th record, denoted R_n , is just the value of X_m at the n -th record time, i.e.

$$R_n = X_{M_n} \quad (1.2)$$

As an illustration, suppose we recorded the daily average temperatures in a particular city on the same date each year, and obtained the following sequence:

$$\{65, 61, 68, 69, 63, 67, 71, 66, \dots\} \quad (1.3)$$

The first observation is trivially a record, so $M_1 = 1$ and $R_1 = 65$. The next observation that exceeds this value is the third one, so $M_2 = 3$ and $R_2 = 68$. The very next observation exceeds this value, so $M_3 = 4$, and $R_3 = 69$. Similarly, $M_4 = 7$ and $R_4 = 71$. Thus, we can construct a sequence of records $\{R_n\}$ from the original sequence $\{X_m\}$ in (1.3):

$$\{65, 68, 69, 71, \dots\}$$

Note that $\{R_n\}$ is a subsequence of $\{X_m\}$, and as such will be less informative. For example, we cannot use the sequence $\{R_n\}$ to infer the number of observations between consecutive records, i.e. how many years transpired between when any two consecutive record temperatures were set, whereas we could infer this information from the original sequence $\{X_m\}$. Formally, we cannot deduce M_n from the sequence $\{R_n\}$, an observation I return to below.

²Exceptions are Kortum (1997) and Munasinghe, O’Flaherty, and Danninger (2001). Kortum remarks on the connection between his model of innovation and record theory. However, most of his analysis does not make use of the underlying record structure, since he conditions on time elapsed rather than the number of previously successful innovations. Munasinghe *et al* analyze the number of track and field records in national and international competitions to gauge the effects of globalization, and remark on the likely applicability of record theory in economics.

In applications, the sequence $\{X_m\}_{m=1}^M$ is assumed to follow some stochastic process, so the occurrence of records and the values they assume are well-defined probabilistic events. For example, suppose $M = \infty$ and the individual observations X_m are i.i.d. random variables. This case was examined by Chandler (1952), who first introduced the notion of record statistics, and has come to be known as the classical record model. Various results for this special case have been derived, including the distribution of record times M_n ; the distribution of the number of records within a given sample size; the distribution of the n -th record value given the distribution of any individual observation in the sequence (which is known as the *parent distribution*); the distribution of the original parent distribution given various properties of the record statistics; and the asymptotic distribution of the n -th record value as n tends to ∞ (if it exists). If X_m are not i.i.d., the analysis becomes much more difficult, although some results have been developed for special cases; see Arnold, Balakrishnan, and Nagaraja (1998) for a summary of recent developments. As we shall see below, the conventional search model does not quite correspond to the classical record model, which introduces some complications in the analysis.

Finally, given the frequent occurrence of order statistics in economic applications, it is worth commenting on the connection between order statistics and record statistics. The n -th maximal order statistic, denoted $X_{n:n}$, corresponds to the maximum of n random variables, $\max\{X_1, \dots, X_n\}$. By contrast, the n -th record statistic R_n corresponds to the maximum of a random number M_n of random variables, $\max\{X_1, \dots, X_{M_n}\}$. Thus, if we had to characterize the distribution of the n -th record R_n , we would not be able to describe it as an order statistic, since the record sample comes from a sample whose size is not known in advance and cannot be recovered from the sequence of records $\{R_n\}$. Note that we could still express a record as a *mixture* of order statistics. Specifically, since the number of observations M_n must be equal to an integer greater than or equal to n , the probability that the n -th record is equal to x can be expressed as

$$\text{Prob}(R_n = x) = \sum_{m=n}^{\infty} \text{Prob}(M_n = m) \times \text{Prob}(X_{m:m} = x) \quad (1.4)$$

Since we can often compute the distribution of the n -th record directly without computing the mixing probabilities $\text{Prob}(M_n = m)$, appealing to the order statistics structure that underlies record data may unnecessarily complicate the analysis.

2. Job Search and Record Statistics

We can now turn our attention to the task of estimating job search models. In line with most of the previous literature on structural estimation of search models, I focus on the Burdett and

Mortensen (1998) model for my theoretical framework. I begin with a brief summary of the model. I then illustrate why previous approaches to estimating the model either require us to make explicit functional-form assumptions, or, if not, are incapable of dealing with unobserved heterogeneity. I then show that appealing to the implicit record structure of the model allows us to identify the shape of the wage offer distribution without any assumptions on the distribution of unobserved heterogeneity. Although I focus exclusively on the Burdett and Mortensen model in my discussion, I briefly comment on alternative models below in Section 5.

2.1. The Burdett and Mortensen Model

The original Burdett and Mortensen (1998) model is an equilibrium model in which firms post wages and workers sample these wages at random, choosing optimally among the offers they receive. Given a distribution of productivity across firms, the model makes a precise prediction as to the wage offer distribution that will arise in equilibrium. Among the papers cited above, some, such as van den Berg and Ridder (1998), Bontemps, Robin, and van den Berg (2000) and Bowlus, Kiefer, and Neumann (2001), explicitly estimate the distribution of firm productivities. Others, such as Flinn (2002), estimate the wage offer distribution without directly relating them to an underlying distribution of firm productivity. I will pursue the latter strategy, i.e. I take the distribution of wages posted as a primitive. As I observe below, we will still be able to apply the analysis of Bontemps, Robin, and van den Berg (2000) to back out the original productivity distribution of firms from the wage offer distribution if we are so inclined.

Given a wage offer distribution, the Burdett and Mortensen model can be summarized as follows. At any point in time, individuals can be either employed or unemployed. An unemployed worker receives a utility payoff of b per unit time, and faces a constant hazard λ_0 per unit time of receiving a job offer when he is unemployed. An offer, when it does arrive, specifies a fixed wage drawn from the wage offer distribution denoted by $F(\cdot)$. Employed workers face a constant hazard λ_1 of receiving an offer, which is also drawn from the same offer distribution $F(\cdot)$, and each draw is independent of all previous draws. In addition, employed workers face a probability δ per unit time of losing their job. This rate is assumed constant and independent of the wage on a worker's current job.

Each worker has to decide whether to accept a new offer when one arrives or to stay on his previous job (or remain unemployed if he doesn't yet have a job). The optimal policy for an unemployed worker is to set a reservation wage W^* that depends on b , δ , λ_0 , λ_1 , and $F(\cdot)$, and to accept job offers if and only if they offer a wage of at least W^* . For an employed worker, the

optimal policy dictates accepting any wage offer that exceeds the worker's current wage W and turning down any offer below it. The optimal behavior of firms imposes at least two restrictions that the equilibrium offer distribution $F(\cdot)$, both of which I make use of in my analysis. First, the equilibrium offer distribution $F(\cdot)$ must be continuous, i.e. it cannot exhibit any mass points. Second, $F(W^*) = 0$, i.e. no firm will offer a wage below the common reservation wage W^* .

Note that model makes a clear distinction between voluntary job changes that the worker initiates upon receiving a better offer and involuntary job changes in which the worker is forced to leave his job to unemployment and resume searching from scratch. Following Wolpin (1992), we can break down each worker's employment history in the model into distinct *employment cycles*, where a cycle is defined as the time between *involuntary* job changes. That is, a cycle begins the instant the worker leaves his job involuntarily, continues on through his unemployment spell and each job he subsequently takes on, and ends the next time he is laid off. Of course, in empirical applications it will be important to classify job changes in a way in accordance with the model so that we correctly identify distinct employment cycles, a point I return to in my data analysis. In principle, we should index observations by the employment cycle they are associated with. For ease of notation, though, I will omit this subscript in what follows.

Let M denote the (random) number of job offers a worker receives on a given employment cycle, and let $m \in \{1, 2, \dots, M\}$ index these offers. Let $\{X_m\}_{m=1}^M$ denote the list of wages on the respective job offers that the worker encounters over his employment cycle. The fact that $F(W^*) = 0$ implies that workers will always accept the first offer they receive out of unemployment. Thus, X_1 represents both the wage on first job offer the worker receives and the the wage on the first job the worker is employed on. This will not be true more generally; for example, by a simple exchangeability argument, half of the the time when a worker draws a second wage offer it will be lower than his first offer, so the second offer will often not be the second job the worker is employed on. As such, let us define $N \leq M$ as the number of *actual jobs* the worker is employed on in a given cycle, and let $n \in \{1, 2, \dots, N\}$ index these jobs. Let $\{W_n, t_n\}_{n=1}^N$ denote the wages and durations of all the jobs the worker is employed on over the cycle. Using the definition of M_n in (1.1), the optimal search strategy for a worker implies that

$$W_n = X_{M_n}$$

i.e. the wage on the n -th job in the cycle must be the n -th record from the sequence $\{X_m\}_{m=1}^M$, and N is the total number of records set within this sequence. The fact that wages correspond to record statistics will eventually figure prominently in my analysis.

2.2. Previous Approaches to Estimation

Having described the model, I now turn to the question of how to use data to identify its key parameters. One approach is to assume that the offer distribution $F(\cdot)$ lies in a class of functions parameterized by some finite-dimensional vector ξ , i.e. $F(\cdot) = F(\cdot; \xi)$. Thus, identifying the model amounts to estimating a finite vector $\{\lambda_0, \lambda_1, \delta, \xi\}$. In the full-fledged version of the Burdett and Mortensen model, ξ corresponds to the ratio λ_1/δ and any parameters that characterize the distribution of productivity across firms. For example, if we assume all firms in a given labor market are equally productive, as in van den Berg and Ridder (1998), estimating ξ amounts to estimating a single productivity parameter for each market.³ Alternatively, if we assume the distribution of productivity across firms in each market has finite support, as in Bowlus, Kiefer, and Neumann (2001), ξ amounts to the different levels of productivity and their respective probabilities. In the reduced-form version of the model where we treat the wage offer distribution as a given, ξ corresponds to the parameters that characterize the particular functional form imposed on the wage offer distribution. For example, Flinn (2002) assumes $F(\cdot)$ is Pareto, and then estimates the curvature parameter associated with this distribution.

Given an expression for $F(\cdot; \xi)$, we can write down the likelihood of a given cycle, i.e. the likelihood of there being exactly N jobs in the cycle and that the particular sequence of wages and job durations for these N jobs are given by $\{W_1, t_1, \dots, W_N, t_N\}$. As I show below, this likelihood is a function of λ_1 , δ , and ξ , so we can estimate these parameters using maximum likelihood. Armed with these estimates, we can proceed to estimate λ_0 using data on unemployment durations, although I ignore this step in my discussion since I have nothing novel to say about it.⁴ To obtain the likelihood function $\mathcal{L}(W_1, t_1, \dots, W_N, t_N, N)$, note that the likelihood for any sequence that fails to satisfy the condition that

$$W_1 \leq W_2 \leq \dots \leq W_N$$

must be zero. For sequences that satisfy this requirement, we can compute the likelihood of $\{W_1, t_1, \dots, W_N, t_N, N\}$ as follows. First, we compute the likelihood of the wage and duration on each job and whether it ends voluntarily or involuntarily conditional on the same information for all of the jobs that preceded it in the cycle. Then, we multiply all of these terms to obtain the joint

³More precisely, van den Berg and Ridder assume a worker can participate in only one market. Different workers (as distinguished by observable and unobservable characteristics) operate in different markets, and the productivity of any worker is assumed the same across all firms in the market in which he sells his labor services. Thus, there is only one parameter to estimate per market, but the number of markets is large.

⁴As previous authors have noted, estimating λ_0 in this model is simplified by the fact that $F(W^*) = 0$ in the Burdett and Mortensen model, which avoids the question of recoverability raised in Flinn and Heckman (1982).

likelihood of the entire cycle. Formally, for each job except the last job in the cycle, we compute the conditional likelihood

$$\mathcal{L}(W_n, t_n, N \geq n + 1 \mid W_{n-1}, t_{n-1}, \dots, W_1, t_1, N \geq n) \quad (2.1)$$

For the last job on the cycle, we compute the conditional likelihood

$$\mathcal{L}(W_n, t_n, N = n \mid W_{n-1}, t_{n-1}, \dots, W_1, t_1, N \geq n) \quad (2.2)$$

To compute (2.1) and (2.2), we again break down these likelihoods into products of conditional likelihoods, i.e.

$$\mathcal{L}(W_n, t_n, N \mid W_{n-1}, \dots) = \mathcal{L}(W_n \mid W_{n-1}, \dots) \times \mathcal{L}(t_n \mid W_n, W_{n-1}, \dots) \times \mathcal{L}(N \mid W_n, t_n, W_{n-1}, \dots)$$

Turning to the first term, since the wage on the n -th job is a random draw from the distribution truncated at W_{n-1} , the likelihood of observing W_n on the n -th job given W_{n-1} (as well as all other past data, which are independent of W_n given W_{n-1}) corresponds to

$$\mathcal{L}(W_n \mid W_{n-1}, \dots) = \frac{f(W_n)}{\bar{F}(W_{n-1})} \quad (2.3)$$

where $\bar{F}(\cdot) = 1 - F(\cdot)$. Adopting the convention that $W_0 \equiv W^*$, this expression also characterizes the distribution of the wage on the very first job in the cycle. Next, the duration of the n -th job given the wage W_n will be exponential with an arrival rate equal to $\delta + \lambda \bar{F}(W_n)$. Thus, the conditional likelihood of t_n given W_n and all other past data corresponds to

$$\mathcal{L}(t_n \mid W_n, W_{n-1}, \dots) = (\delta + \lambda \bar{F}(W_n)) e^{-[\delta + \lambda \bar{F}(W_n)]t_n} \quad (2.4)$$

Finally, the conditional probabilities that the n -th job ends because of a quit and a layoff, respectively, correspond to

$$\begin{aligned} \mathcal{L}(N \geq n + 1 \mid W_n, t_n, W_{n-1}, \dots) &= \frac{\lambda_1 \bar{F}(W_n)}{\delta + \lambda_1 \bar{F}(W_n)} \\ \mathcal{L}(N = n \mid W_n, t_n, W_{n-1}, \dots) &= \frac{\delta}{\delta + \lambda_1 \bar{F}(W_n)} \end{aligned} \quad (2.5)$$

Multiplying (2.3) through (2.5), we get

$$\begin{aligned} \mathcal{L}(W_n, t_n, N \geq n + 1 \mid W_{n-1}, \dots) &= \lambda_1 \bar{F}(W_n) \frac{f(W_n)}{\bar{F}(W_{n-1})} e^{-[\delta + \lambda \bar{F}(W_n)]t_n} \\ \mathcal{L}(W_n, t_n, N = n \mid W_{n-1}, \dots) &= \delta \frac{f(W_n)}{\bar{F}(W_{n-1})} e^{-[\delta + \lambda \bar{F}(W_n)]t_n} \end{aligned}$$

Multiplying the above for all jobs $n = 1, \dots, N$ and cancelling redundant terms yields the following expression for the likelihood of a given employment cycle:

$$\mathcal{L}(W_1, t_1, \dots, W_N, t_N, N) = \frac{\delta}{\lambda_1} \left[\prod_{n=1}^N \lambda_1 f(W_n) e^{-[\delta + \lambda_1 \bar{F}(W_n)] t_n} \right] \quad (2.6)$$

Maximum likelihood then amounts to choosing the values of $\{\lambda_1, \delta, \xi\}$ that maximize (2.6) evaluated at the sample data from different employment cycles. As long as we correctly specified the function $F(\cdot; \xi)$, these values are consistent estimates of the true $\{\lambda_1, \delta, \xi\}$.

We can easily amend the model above to allow for heterogeneity in ability across workers, a feature most economists would agree is important for explaining actual data. Suppose the offers $\{X_m\}_{m=1}^M$ reflect price offers of how much a firm is willing to pay *per unit of effective labor* rather than wage offers, and that workers vary in the amount of effective labor they can supply.⁵ All workers, regardless of their productivity, face the same distribution of prices and the same arrival rate λ_1 and δ . Let ℓ_{it} denote the amount of effective labor per hour that worker i can supply at date t , and let \widehat{W}_{it}^n denote the wage we observe for worker i at date t on the n -th job in his employment cycle. Then the wages we would observe for worker i on his various jobs in an employment cycle correspond to

$$\widehat{W}_{it}^n = W_n \ell_{it} \quad (2.7)$$

A virtue of maximum likelihood estimation is that it does not require worker ability to be observable. For example, if we assume that there is a finite number of productivity levels ℓ_{it} a worker could operate at, we can add these values and their respective frequency in the population as parameters we need to estimate.⁶

The main drawback of the approach outlined above is that the choice of the offer distribution $F(\cdot, \xi)$ is arbitrary. Since the answers to the questions raised in the Introduction often hinge on the exact shape of the relevant distribution, assuming rather than estimating the shape of this function runs counter to the original spirit of this undertaking. Although we can conceptually overcome this concern by allowing for a sufficiently large parametric class, in practice the papers cited in the Introduction focus on mutually exclusive parametric families; the wage offer distribution is sometimes assumed to be lognormal, sometimes Pareto, sometimes a power distribution, and so

⁵If the production of output is linear in effective labor, this is equivalent to assuming firms post piece rates.

⁶To ensure workers of different ability choose the same cutoff price W^* in the distribution of prices, we would have to further assume that the value of leisure is proportional to productivity, i.e. $b_{it} = b\ell_{it}$. This interpretation could reflect the fact that the value of leisure is really the value of home production, and individuals are just as productive at home as they are in the market sector.

on. At the very least, we would like some procedure to determine what shape restrictions on the relevant distributions are valid before we carry out maximum likelihood.

In recent work, Bontemps, Robin, and van den Berg (2000) attempt to do just that by devising a non-parametric estimator for the wage offer distribution. They suggest the following two-step procedure in the case where there is no unobserved ability. First, since the first job in the cycle is a random draw from $F(\cdot)$, they advocate using the empirical distribution of wages across all workers on their first job to estimate $F(\cdot)$. Next, given this estimate of the offer distribution, they use data on job duration to identify λ_1 and δ . Recall from (2.4) that the duration of a job is distributed as an exponential with rate $\delta + \lambda_1 \bar{F}(W)$. One can therefore identify λ_1 and δ from the way the duration of a job t varies with the wage paid on the job W . Formally, they choose λ_1 and δ to maximize the likelihood in (2.6) where they substitute in the estimate for $F(\cdot)$ from the first stage.⁷ This approach has the virtue that it does not impose any shape restrictions on $F(\cdot)$, or alternatively on distribution of productivity across firms that gives rise to this distribution. However, this approach cannot accommodate unobserved differences in worker ability. For suppose once again that the wages we observe for worker i on his various jobs over the cycle are given by

$$\widehat{W}_{it}^n = W_n \ell_{it} \tag{2.8}$$

where ℓ_{it} denotes the productivity of worker i at date t . The empirical distribution of wages on the first job workers accept out of unemployment, \widehat{W}_{it}^1 , is a convolution of the true offer distribution $F(\cdot)$ and the distribution of ℓ_{it} across workers. Improperly attributing dispersion in ℓ_{it} to dispersion in offers would overstate the dispersion of the true offer distribution $F(\cdot)$, and this misspecification could then further contaminate estimates for λ_1 and δ .

While we could in principle try to estimate ℓ_{it} from observed worker characteristics, e.g. age and education, it is unlikely that we could ever capture all of the unobserved variation in wages. In particular, an important feature of the data is that a non-negligible fraction of voluntary job changes report a lower wage on their new job than on their previous job. Thus, we need to allow for the possibility that $\widehat{W}_{it}^{n+1} < \widehat{W}_{it}^n$ even when $W_{n+1} \geq W_n$, and it is hard to come up with observable characteristics that can account for such fluctuations in individual earnings over time. Bontemps, Robin, and van den Berg (2000) avoid this issue by only using wage data for the first

⁷Since Bontemps *et al* cannot observe when a worker began his employment cycle in their data, they cannot isolate workers who are on their first job. Instead, they assume that any given worker in their sample is drawn from the steady-state distribution of wages across all workers, denoted by $G(\cdot)$. The distribution $G(\cdot)$ can then be directly related to $F(\cdot)$ given a ratio λ_1/δ . They estimate $G(\cdot)$ non-parametrically, and then maximize a variant of (2.6) in which $f(\cdot)$ is replaced with the appropriate expression in terms of $g(\cdot)$ and the ratio λ_1/δ . They then choose λ_1 and δ to maximize the implied likelihood function.

job they observe for a worker, so they never have to consider wage changes across jobs. But both van den Berg and Ridder (1998) and Flinn (2002), who use multiple jobs from each employment cycle, need to introduce an ℓ_{it} term to reconcile the model with the data. They interpret ℓ_{it} as measurement error in wages, but as the discussion above suggests it could just as well reflect variability in a worker's productivity over time. Either way, unless we knew the distribution of ℓ_{it} , we could not recover the distribution of W_n from the distribution of observed wages \widehat{W}_{it}^n .

2.3. Non-Parametric Estimation using Record Moments

I now propose a way to estimate the wage offer distribution non-parametrically even in the presence of unobserved heterogeneity ℓ_{it} . My approach exploits the implicit record structure of the model. As noted briefly in Section 1, given an i.i.d. sequence $\{X_m\}_{m=1}^\infty$, we can sometimes characterize the distribution of X_m from observations on the records $\{R_n\}_{n=1}^\infty$ in the sequence $\{X_m\}_{m=1}^\infty$. For example, Kirmani and Beg (1984) show that the list of record *moments* $\{E(R_n)\}_{n=1}^\infty$, assuming these moments exist, uniquely characterize the distribution of X_m within the set of continuous distributions. To appreciate why this result might be useful, consider the case where ℓ_{it} reflects pure measurement error, i.e. ℓ_{it} is i.i.d. across workers and over time and is independent of W_n . In this case, the average wage across all workers on their n -th job converges to

$$\begin{aligned} E(\widehat{W}_{it}^n) &= E(W_n) E(\ell_{it}) \\ &= E(W_n) \times \text{constant} \end{aligned}$$

Thus, we can empirically estimate the sequence $\{E(W_n)\}_{n=1}^\infty$ up to a scalar. Since $\{W_n\}$ corresponds to the list of records in a sequence of i.i.d. draws from the underlying wage offer distribution, the list $\{E(W_n)\}_{n=1}^\infty$ should allow us to identify the underlying wage offer distribution without having to characterize the distribution of ℓ_{it} . The approach outlined below adapts this argument for richer specifications of ℓ_{it} that capture differences in productivity as well as just pure measurement error. Note that this approach is conceptually different from the approaches above, since it relies on different data to identify the fundamental parameters of the model. Whereas the approaches above rely on wage and duration data, i.e. on $\{\widehat{W}_{it}^n, t_n\}_{n=1}^N$, my approach relies on wage data and the position in the cycle of each job, i.e. $\{\widehat{W}_{it}^n, n\}_{n=1}^N$.

While the discussion above conveys the basic intuition, it is also somewhat imprecise. Kirmani and Beg (1984) assume records come from an infinite sequence $\{X_m\}_{m=1}^\infty$, whereas in the model wages correspond to records from a sequence $\{X_m\}_{m=1}^M$ in which M is a random variable. Why is this relevant? Since the occurrence of a record is a random event, the expected value of the

n -th record is conditional on the event that there are at least n records in the original sequence, and should more properly be denoted as $E(R_n | N \geq n)$ where N denotes the total number of records in the sequence $\{X_m\}$. If the distribution for each X_m is continuous, the infinite sequence $\{X_m\}_{m=1}^{\infty}$ will contain infinitely many records almost surely, i.e. $\text{Prob}(N \geq n) = 1$ for any finite n , and it is common to simply omit the conditioning event and write $E(R_n)$ for $E(R_n | N \geq n)$. But when we only get to observe $\{X_m\}_{m=1}^M$ where M can be finite, it will no longer be the case that every sequence we observe will contain n records. Thus, when we compute an average value for the n -th record R_n , we can only average over those employment cycles in which the worker makes it to his n -th job. But this is hardly a random sample; quite to the contrary, on those cycles where the worker makes it to his n -th job before being laid off it is more likely that the first few offers the worker received were low enough that the worker could still find a better offer before being laid off. As a result, the moments $E(R_n | N \geq n)$ for records drawn from the sequence $\{X_m\}_{m=1}^M$ will typically be smaller than the corresponding moments $E(R_n | N \geq n)$ from the infinite sequence $\{X_m\}_{m=1}^{\infty}$. We will therefore not be able to rely on Kirmani and Beg's results, and must independently confirm that record moments for records from a sequence of random length still uniquely identify the parent distribution of the individual observations.

I begin by laying out my assumptions on ℓ_{it} . My key assumption is that ℓ_{it} is independent of both the job worker i is employed on at date t and the price per unit labor that it pays. This assumption preserves the basic search structure of the model, since a worker should seek out the jobs offering the highest price per unit labor. The prices per unit labor on the jobs the worker is employed on $\{W_n\}_{n=1}^N$ thus still correspond to record values among the price offers $\{X_m\}_{m=1}^M$ the worker encounters over the cycle. Note that this assumption rules out match-specific wage dynamics, since any change in the wage a worker earns on his current job must also affect the wages he would earn on all other jobs. This is hardly innocuous; for example, my framework does not allow for the possibility that a worker becomes more skilled at only one particular job, i.e. any human capital a worker accumulates must be general in nature. Nevertheless, in the next section I argue that this assumption is reasonable for the sample of workers I consider.

Following Flinn (1986), I proceed to impose the following form for ℓ_{it} :

$$\ell_{it} = \exp(\phi_i + \beta Z_{it} + \varepsilon_{it})$$

The first term, ϕ_i , is fixed over time and serves to capture variations in innate ability that make some workers consistently more productive than others, all other things equal. The next term, Z_{it} , represents observable time-varying characteristics for individual i that affect his productivity. In my application, this will correspond to the time since the worker first entered the market,

i.e. the worker's potential experience. Lastly, ε_{it} denotes unobserved productivity shocks. It serves to capture a combination of any shocks to productivity not accounted for by Z_{it} as well as multiplicative measurement error that may appear in reported wages. I assume ε_{it} follows a stochastic process with the sole restriction that

$$E(\Delta\varepsilon_{it}) = 0 \tag{2.9}$$

i.e. the unconditional mean change in ε_{it} is zero. Note that (2.9) does not restrict the shape of the distribution of ε_{it} or its autocorrelation; in particular, my approach allows for the presence of serial correlation in earnings over the duration of a job.

It will be easier to work from now on with log wages than with wages. Let $w_n = \ln W_n$ denote the log price per unit labor on the worker's n -th job. It follows that w_n represents the n -th record in the sequence of log price offers $\{x_m\}_{m=1}^M$ where $x_m = \ln X_m$, so we can still hope to exploit the record structure of w_n to identify the distribution of x_m (from which we can deduce the distribution of X_m). After substituting in for ℓ_{it} , we obtain the following equation for the log wage:

$$\ln \widehat{W}_{it}^n = w_n + \phi_i + \beta Z_{it} + \varepsilon_{it} \tag{2.10}$$

We next first-difference equation (2.10) to get rid of the fixed effect term ϕ_i . Let Δ denote the difference in a particular variable between two distinct points in time. Then we have

$$\Delta \ln \widehat{W}_{it} = \Delta w + \beta \Delta Z_{it} + \Delta \varepsilon_{it} \tag{2.11}$$

For a worker who is employed on the same job at these two points in time, $\Delta w = 0$, implying wage growth on the job is given by

$$\Delta \ln \widehat{W}_{it} = \beta \Delta Z_{it} + \Delta \varepsilon_{it} \tag{2.12}$$

It follows that we can use ordinary least squares on (2.12) to estimate β , i.e. to estimate the contribution of observable characteristics to productivity growth. This estimate will be important in what follows.

Next, consider the wage growth of workers across jobs. Using our estimate for β , we can net out the role of observable productivity growth in these cases. Thus, for a worker who moves from his n -th job to his $n + 1$ -th job, the net wage gain from changing jobs is given by

$$\Delta \ln \widehat{W}_{it} - \beta \Delta Z_{it} = (w_{n+1} - w_n) + \Delta \varepsilon_{it} \tag{2.13}$$

In other words, the net wage gain for a voluntary job changer who leaves his n -th job is the sum of the gap between the n -th record and the $n + 1$ -th record from a sequence of i.i.d. draws from

the log price offer distribution and a noise term $\Delta\varepsilon_{it}$. Averaging these net wage gains across all such workers, we have

$$\begin{aligned} E(\Delta \ln \widehat{W}_{it} - \beta \Delta Z_{it} \mid N > n) &= E(w_{n+1} - w_n \mid N > n) + E(\Delta\varepsilon_{it} \mid N > n) \\ &= E(w_{n+1} - w_n \mid N > n) \end{aligned}$$

where we use the fact that $\Delta\varepsilon_{it}$ is independent of N . Thus, using observations on wage growth $\Delta \ln \widehat{W}_{it}$ both within jobs and across jobs, we can estimate the expected record gaps

$$\{E(R_{n+1} - R_n \mid N > n)\}_{n=1}^{\infty} \tag{2.14}$$

for records that are drawn from the sequence $\{x_m\}_{m=1}^M$ where the x_m are i.i.d. random variables with the same distribution as the log wage offer distribution. As anticipated by the previous discussion, these record moments may be enough to identify the parent distribution of each x_m . In the next subsection, I show that the sequence of moments in (2.14) is indeed enough to characterize the shape of the parent distribution of x_m within the set of continuous distribution functions, and show how to recover the parent distribution from this list.

2.4. Recovering the Distribution from Record Moments

I begin by establishing that the sequence of expected record gaps in (2.14) uniquely identifies the shape of the parent distribution. First, though, I need to characterize the distribution of the number of offers M in each employment cycle. The proof of the next lemma, as well as all remaining propositions, are contained in an appendix.

Lemma 1: The unconditional number of offers on a cycle M is distributed as a geometric, i.e.

$$\text{Prob}(M = m) = q^{m-1}p$$

where $q = \frac{\lambda_1}{\lambda_1 + \delta}$ and $p = \frac{\delta}{\lambda_1 + \delta}$. ■

The next lemma provides a sufficient condition for the moments in (2.14) to be well-defined:

Lemma 2: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^M$ where M is independent of the realizations of $\{X_m\}$ and $\text{Prob}(M = m) = q^{m-1}p$, for $p+q = 1$. Let $\{R_n\}_{n=1}^N$ denote the records of this sequence. If $E(|X_m|) < \infty$, then the conditional expectation $E(R_{n+1} - R_n \mid N > n)$ is finite for $n = 1, 2, 3, \dots$

Assuming the moments in (2.14) exist, we have the following result:⁸

Proposition 1: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^M$ where $\text{Prob}(M = m) = q^{m-1}p$. If $E(|X_m|) < \infty$, the sequence

$$\{E(R_{n+1} - R_n \mid N > n)\}_{n=1}^{\infty}$$

uniquely characterizes the distribution of X_m in the set of continuous distributions up to a location shift. That is, if $\widehat{F}(\cdot)$ is a continuous function that gives rise to the same sequence $\{E(R_{n+1} - R_n \mid N > n)\}_{n=1}^{\infty}$ as $F(\cdot)$, then $\widehat{F}^{-1}(\cdot) = F^{-1}(\cdot) + \text{constant}$.

Remark: Following up on the present paper, Nagaraja and Barlevy (2003) analyze record moments when the number of observations M is geometric in more detail. Interestingly, they show that characterization results based on record moments from a geometric number of observations are stronger than those that are based on record moments from an infinite number of observations, i.e. moment sequences that are not enough to uniquely identify the parent distribution in the classical model can identify the parent distribution when M is geometric.

To gain some insight as to why record moments allow us to identify the parent distribution, consider the expression $E(X \mid X > x) - x$, i.e. the amount we expect to rise above a given number x when we sample at random from the parent distribution truncated at x . The n -th average record gap $E(R_{n+1} - R_n \mid N > n)$ is just a weighted average of all these expected gains, where the weight on the expected gain over a particular x corresponds to the probability that the n -th record is equal to x . At low values of n , more weight is put on the expected gain starting at low values of x , while at higher values of n , more weight is put on the expected gain starting at high values of x . Taking all of these moments together, we can essentially deduce how much a worker would expect to gain from moving to a better job starting at any initial wage; for low initial wages, these gains will depend more on the average record gap for low values of n , while for high initial wages, these gains will depend on the average record gap for higher values of n . Given two different wage offer distributions, the expected gain from mobility will have to differ at some starting wage, implying at least some of the expected record gaps will be different.

To make practical use of Proposition 1, we need to be able to invert the set of moments to obtain the underlying parent distribution that gives rise to these moments. Since the moments in (2.14) are all functions of the parameter p in Lemma 1, we first need to estimate this parameter.

⁸I am indebted to H. N. Nagaraja for suggesting the proof of this proposition.

Since p is just a function of the ratio $\frac{\lambda_1}{\delta}$, it would seem natural to appeal to existing methods for estimating these parameters. Unfortunately, the way previous authors have estimated these arrival rates is not applicable. Recall that previous work estimated these parameters using the fact that job duration is exponential with rate $\delta + \lambda_1 \bar{F}(W)$, so the way job duration varies with wages allows us to identify these two parameters. However, this approach requires that we know the shape of the function $F(\cdot)$, which we are still in the progress of trying to determine. Fortunately, p can be estimated directly from data on employment history. In particular, Bunge and Nagaraja (1991) show that when M is geometric, the number of records N in an i.i.d. sequence $\{X_m\}_{m=1}^M$ will be distributed according to the truncated Poisson

$$\Pr(N = n) = \frac{p(-\ln p)^n}{qn!} \quad (2.15)$$

Since the number of records N on each employment cycle corresponds to the number of jobs a worker was employed on before he was laid off, we can estimate p and q directly from mobility data, without using information on job duration. Formally, we have

Proposition 2: Let Q_n denote the number of employment cycles with exactly n records. A consistent estimator for p is given by

$$\hat{p} = \arg \max_p \prod_{n=1}^{\infty} \left[\frac{p}{1-p} \frac{(-\ln p)^n}{n!} \right]^{Q_n}$$

Armed with this estimate, I now tackle the problem of constructing a function from the list of expected record gaps in (2.14). To do this, note from the proof of Proposition 1 in the Appendix that the n -th element of the list $E(R_{n+1} - R_n \mid N > n)$ can be expressed as

$$E(R_{n+1} - R_n \mid N > n) = \frac{(-\ln p)^n}{(n-1)! \Pr(N > n)} \int_0^1 g(x) x^{n-1} dx \quad (2.16)$$

where $g(x)$ is a function that depends on p and the parent distribution for each observation in the sequence. To recover the parent distribution $F(\cdot)$, we proceed in two steps. First, we use the moments $E(R_{n+1} - R_n \mid N > n)$ and our estimate for p to determine the function $g(x)$. Second, we invert the function $g(x)$ to recover the distribution $F(\cdot)$.

To determine the function $g(x)$, rewrite (2.16) as

$$\int_0^1 g(x) x^{n-1} dx = E(R_{n+1} - R_n \mid N > n) \frac{(n-1)! \Pr(N > n)}{(-\ln p)^n} \equiv \mu_{n-1} \quad (2.17)$$

Given estimates for p and $E(R_{n+1} - R_n \mid N > n)$ for all n , we can determine each of the μ_{n-1} . The task of finding a function $g(x)$ that solves the infinite system of equations above is known as the Hausdorff moment problem. That is, in a variety of applications, we will often want to find a function $g(x)$ defined over the unit interval whose moments (i.e. integrals of the function multiplied by different powers of x) are equal to a particular sequence of values. Shohat and Tamarkin (1943) offer a set of sufficient conditions for this problem to have a solution (which proves to be unique) and provide an analytical representation for this solution $g(x)$ in terms of μ_n . Essentially, we use the expressions for μ_n to construct coefficients for an infinite polynomial expansion, which allows us to reconstruct any continuous function. A little algebra then allows us to recover the distribution function of interest $F(\cdot)$ from the function $g(x)$ we just constructed. This procedure is summarized in the next proposition:

Proposition 3: Given the complete sequence of moments $\{E(R_{n+1} - R_n \mid N > n)\}_{n=1}^{\infty}$, the inverse parent distribution $F^{-1}(x)$ can be constructed according to the following procedure:

1. Let $\{P_n(x)\}_{n=0}^{\infty}$ denote the set of Legendre polynomials defined on $(-1, 1)$. Define a new set of polynomials $\{\mathcal{P}_n(x)\}_{n=0}^{\infty}$ on $(0, 1)$ as

$$\mathcal{P}_n(x) = \frac{P_n(2x-1)}{\int_0^1 P_n^2(2x-1) dx} \equiv \sum_{j=0}^n c_{nj} x^j \quad (2.18)$$

2. Define a sequence $\{\mu_n\}_{n=0}^{\infty}$ where

$$\mu_{n-1} = E(R_{n+1} - R_n \mid N > n) \times \frac{(n-1)! \Pr(N > n)}{(-\ln p)^n}$$

Using the coefficients c_{nj} from (2.18), construct a new sequence $\{\lambda_n\}_{n=1}^{\infty}$ where

$$\lambda_n = \sum_{j=0}^n c_{nj} \mu_k$$

Define a function $g(x)$ over $(0, 1)$ as the sum of polynomials $\mathcal{P}_n(x)$ with coefficients λ_n , i.e.

$$g(x) = \sum_{j=0}^{\infty} \lambda_n \mathcal{P}_n(x) \quad (2.19)$$

3. The inverse parent distribution function $F^{-1}(x)$ over $(0, 1)$ can be constructed from $g(x)$ as follows:

$$F^{-1}(x) = \int_0^x \frac{qg' \left(\frac{\ln(1-qz)}{\ln p} \right)}{(1-z)(1-qz)\ln p} dz + \text{constant}$$

where the constant of integration denotes the unidentified location parameter.

Remark: Bontemps, Robin, and van den Berg (2000) offer an explicit way to obtain the distribution of productivity in the Burdett and Mortensen model given a wage offer distribution. The inversion requires knowing the ratio $\frac{\lambda_1}{\delta}$. Since Proposition 2 establishes that we can identify p , one can extend Proposition 3 to recover the implied distribution of productivities across firms from the distribution $F(\cdot)$.

In practice, we can only estimate finitely many moments μ_n required to implement the method laid out in Proposition 3, and even those moments are only imprecisely estimated due to sampling error. Talenti (1987) examines a variant of the Hausdorff moment problem with finitely many moments measured with error. He suggests replacing the infinite sum in (2.19) with the finite sum

$$g(x) = \sum_{j=0}^J \lambda_n \mathcal{P}_n(x) \tag{2.20}$$

where J denotes the number of moments we can observe. Talenti shows that the problem is stable, in the sense that the inaccuracy from using a finite set of moments is bounded by the number of moments J and the magnitude of the sampling error, and this bound converges to zero as the number of moments goes to infinity and the error term both go to zero. Thus, the estimator for $F(\cdot)$ in Proposition 3 is consistent as long as in the limit we can precisely estimate all of the relevant moments. But with only a small number of moments, the approximation in (2.20) is likely to be quite poor. Thus, while Proposition 3 yields a consistent non-parametric estimator for the wage offer distribution, this estimator may not be reliable if we can only estimate a small number of moments in the list (2.14).

While a precise estimate for $F(\cdot)$ may be hard to come by in practice, a small number of moments may still suffice to check if a particular functional form is consistent with data. As an illustration, Figure 1 displays the expected record gaps $E(R_{n+1} - R_n \mid N > n)$ for two different log wage offer distributions, an exponential and a normal (which correspond to Pareto and lognormal wage offers, respectively, both of which have been suggested in the previous literature). The moments are computed for a geometric M with the probability p that is estimated in Section 4, and both distributions are normalized to yield the same average wage gain across voluntary job changers as in the data when we use the empirical distribution of job changes across n . As Figure 1 reveals, the two distributions can be easily distinguished from one another even with only a small number moments. In particular, the average net wage gain does not depend on n for the exponential distribution, reflecting the memoryless property of this distribution, while the average wage gain declines rapidly with n for the normal distribution. I will return to this observation below, where I argue that we can in fact reject the null hypothesis that the wage offer distribution is lognormal.

3. Data

To implement the above methodology, I need a dataset with detailed work-history data that can be used to assign n for each job. Since job mobility is highest when workers first enter the labor market, it also seems wise to focus on young workers. Not only will this allow for larger samples of job changers, but the fact that young workers are so mobile makes them less likely to invest in match-specific human capital, in line with my assumption that human capital is general. These considerations lead me to data from the early part of the National Longitudinal Survey of Youth (NLSY) dataset. The NLSY follows a single cohort of individuals who were between 14 and 22 years old in 1979. To avoid using observations where workers are already well established in their careers, I only use data through 1993, at which point the oldest worker in the sample is 36. Each year, respondents were asked questions about all jobs they worked on since their previous interview, including starting and stopping dates, the wage paid, and the reason for leaving the job. To mitigate the influence of mobility due to non-wage considerations, e.g. pregnancy or child-care, I restrict attention to male workers.

Most of the variables that I use are standard. For the wage, I use the hourly wage as reported by the worker for each job, divided by the GDP deflator (with base year 1992). I also experimented with the CPI, but the results were similar. To minimize the effect of extreme outliers on means, I removed observations for which the reported hourly wage was less than or equal to \$0.10 or greater than or equal to \$1000. This eliminated 0.1% of all wage observations. Many of these outliers appear to be coding errors, since these wages appear to be out of line with what these same workers report at other times, including on the same job. For my measure of potential experience, I follow previous literature in dating entry into the labor market at the worker's birthyear plus 6 plus his reported years of schooling (highest grade completed). However, if an individual reported working on a job prior to that year, I date his entry at the year in which he reports his first job. Table 1 provides summary statistics for the original sample of all jobs.

The more unconventional variable in my analysis is the position n each job represents in its respective cycle. To construct this statistic, we first need to distinguish between voluntary job-to-job changes from involuntary job changes in order to delineate employment cycles. One approach is to use individual self-reports on the reason they left each of their jobs, i.e. whether they quit voluntarily or were laid off. Alternatively, we can use the time lapsed between jobs to gauge whether a move was voluntary or not, since a voluntary job changer would immediately move into a new job while a worker who changed jobs involuntarily would spend some time unemployed. According to the model, these approaches should coincide. But in practice, the two agree only 60%

of the time. More precisely, it is true that the vast majority of workers who report an involuntary job loss spend at least one week unemployed. Moreover, the majority of workers who do not spend any time unemployed between jobs indeed report leaving their previous job voluntarily. The main discrepancy is that nearly half of all workers who report leaving their job voluntarily do not start their next job for weeks or even months later. Although some of these cases are probably due to planned delays, it seems as if workers often report leaving a job voluntarily without having another job already lined up. This could be because workers are embarrassed to admit they were laid off, or because they decided to quit for reasons that are not captured by the model. I assume these reasons are independent of the wage, and so can be considered as an involuntary job loss in the model, i.e. workers who quit without lining up another job must resume searching from scratch. Thus, if a worker reports leaving a job voluntarily, I classify that job as having ended involuntarily if his next job began more than two months (8 weeks) after his previous job ended. But if a worker reports leaving a job involuntarily, I continue to classify the job as having ended involuntarily regardless of how long it took him to start a new job. If the worker offers no reason for leaving his job, I classify his job change as voluntary if he starts his next job immediately and involuntary if he starts it over two months later, but otherwise do not classify the job.⁹

With this classification complete, we can assign n as follows. After the first involuntary job change, we set $n = 1$. From then on, we either increment n by 1 if the worker leaves his job voluntarily, or reset n to 1 if he did so involuntarily. One complication is that a non-trivial fraction of workers simultaneously hold more than one job. To deal with this, I draw on Paxson and Sicherman (1996), who argue that the primary reason workers hold multiple jobs is that they are constrained to work a maximum number of hours on each job. Suppose then that workers are constrained and can work on only one job full time. However, workers can receive additional draws from the distribution $F(\cdot)$ and work on those jobs on a part time basis. If we observe a worker employed in job A take on a second job B, we treat job B as a second draw from $F(\cdot)$ that is available for part-time work. If the worker leaves job B before he leaves his original job A, job B provides us with no information on the price of labor on job A. I therefore ignore job B in my analysis, i.e. it is as if the worker had never taken on a secondary job. Alternatively, if the worker leaves job A and remains in job B, then a full-time position must have opened up on job B. Since the wages on these jobs are assumed to be drawn from the same distribution $F(\cdot)$, we can treat it in the same way as a new job that started only after job A ended.

⁹I experimented with cutoffs other than two weeks. These had very little impact on the first few record moments (i.e. $n = 1, 2,$ and 3) that can be estimated quite precisely, although they did affect the less precisely estimated moments for higher values of n .

Out of the 52,827 distinct jobs in my original sample, the procedure above throws out 8,232 as secondary jobs that the worker was always employed on in addition to some other job. As a check, we can use the fact that the NLSY asks workers to rank their jobs each year in terms of which is their primary job. Of the 8,234 jobs I identify as secondary jobs, 72% are never ranked as the primary job in any year, and only 9% are ranked as the primary job each year the job is reported.

Figure 2 reports the distribution of n across the remaining 44,593 jobs. Figure 2a shows the fraction of all jobs each year for which a value for n cannot be assigned. Since we can only assign n following the first involuntary job change, we will have to wait some time before we can assign n for any one worker. Thus, in the first few years of the sample, we can assign n to only a small fraction of jobs. However, by 1993, n was assigned to 87% of all the jobs respondents reported working on in that year. Figure 2b shows the distribution of n where a value for n could be assigned. Not surprisingly, most jobs early on in the sample that can be classified are associated with $n = 1$. But over time, a larger share of workers is observed on higher levels of n . The distribution of n appears to settle down after about 10 years, with roughly half of all jobs associated with $n = 1$, a quarter with $n = 2$, 12% with $n = 3$, 6% with $n = 4$, and 3% with $n = 5$. Note that very high values of n are uncommon, in line with the known result that records from a sequence of i.i.d. draws are relatively rare.

Before I use this data to estimate the average wage for leaving the n -th job net of growth in observable characteristics ΔZ_{it} , a few issues remain to be settled. First, we need to decide the horizon at which to compute the differences in equation (2.13). Since the NLSY provides one wage on each job per interview, we can only measure within-job wage growth at one year differences. However, when Topel and Ward (1992) study a similar sample of young workers using quarterly data, they report a “strong tendency for within-job earnings changes to occur at annual intervals.” Thus, it seems that little is lost by focusing on annual wage growth. Since my estimates involve the difference between wage growth across jobs and within jobs, consistency would suggest restricting attention to wage growth across jobs that is also computed at one year horizons. To ensure this, I only use wage data for jobs the worker reported working on within two weeks of the interview date, which is carried out on a yearly basis. My approach thus abstracts from wage growth for jobs that fall between interviews. Although I ignore intervening jobs over the year in estimating average wage growth, I do use these jobs for constructing n . My ultimate sample consists of 40,370 observations in which a wage is reported in both the current year and previous year. Of these, 28,015 observations involve the same job in both the current year and the previous year, and 12,355 observations involve a change in jobs between the previous interview and the current one.

Next, I need to specify the vector of observable characteristics Z_{it} . I assume Z_{it} is quadratic in potential experience, i.e.

$$Z_{it} = \beta_1 EXP_{it} + \beta_2 EXP_{it}^2 \quad (3.1)$$

Since at annual horizons $EXP_{it} = EXP_{i,t-1} + 1$, it follows that

$$\Delta Z_{it} = \beta_1 + \beta_2 (2EXP_{it} - 1)$$

As noted above, the fact that (3.1) depends on the time the worker spent on *all* jobs rather on any one job rules out the possibility of match-specific human capital. To gauge the plausibility of this assumption, let TEN_{it} denote the tenure of worker i on the job he holds at date t . Suppose we amended (3.1) to include TEN_{it} , i.e.

$$Z_{it} = \beta_1 EXP_{it} + \beta_2 EXP_{it}^2 + \gamma TEN_{it} \quad (3.2)$$

Under this alternative specification, a part of the wage growth the worker achieves on a given job would disappear if he were to move to a new job where $TEN_{it} = 0$. This would invalidate my analysis, since a worker would no longer find it optimal to accept an offer if and only if it pays a higher price per unit labor than his previous job. Thus, it is important to confirm that the coefficient γ in (3.2) is negligible for my sample.

In what follows, I use the approach advocated by Topel (1991) for estimating γ . This approach tends to produce the largest estimates for returns to tenure in other samples, so finding small returns to tenure using this methodology would be more compelling. Topel's approach uses the fact that $EXP_{it} = EXP_{0it} + TEN_{it}$, where EXP_{0it} is the experience of the worker when he first started working on the job he holds at date t . Thus, the observed log wage can be written as

$$\ln \widehat{W}_{it}^n = w_n + \phi_i + \beta_1 EXP_{0it} + \beta_2 EXP_{it}^2 + (\beta_1 + \gamma) TEN_{it} + \varepsilon_{it} \quad (3.3)$$

To estimate γ , we use the following two-step procedure. First, wage growth over a one-year interval on a given job will equal

$$\Delta \ln \widehat{W}_{it} = (\beta_1 + \gamma) + \beta_2 (2EXP_{it} - 1) + \Delta \varepsilon_{it}$$

Hence, we can estimate $(\beta_1 + \gamma)$ and β_2 by ordinary least squares. Next, we use these estimates to construct the difference

$$\ln \widehat{W}_{it}^n - (\beta_1 + \gamma) TEN_{it} - \beta_2 EXP_{it}^2 = \ln w_n + \phi_i + \beta_1 EXP_{0it} + \varepsilon_{it}$$

We then regress this difference on EXP_{0it} and individual fixed effects to estimate β_1 and ϕ_i . The estimate for γ is the difference between the estimates for $\beta_1 + \gamma$ and β_1 . Table 2 reports the

results of this two-step procedure. The point estimate for $\beta_1 + \gamma$ and β_1 are 0.0794 and 0.0740, respectively, implying $\gamma = 0.0054$. The implied point estimate for γ is significantly different from zero at the 5% level, but its magnitude is so small relative to the wage gains from job changing I estimate in the next section that it seems safe to ignore it.¹⁰

In comparing my results to those of Topel (1991), the returns to experience in the two papers are quite consistent. In fact, his point estimate for β_1 of 0.0713 is quite close to mine, implying that returns to experience will be nearly identical at short horizons. The main difference is that wage growth on the job is considerably smaller in my sample than in Topel's sample. In particular, he estimates $\beta_1 + \gamma$ at 0.1258, compared to my estimate of 0.0794. Thus, wages in my sample appear to grow on-the-job at nearly the same rate as wages rise with initial experience across jobs, while wages in his sample grow on-the-job by much more than wages rise with initial experience. This finding appears to be robust to variations in the functional form for the returns to tenure. The bottom panel of Table 2 allows for tenure to enter as a quadratic, i.e. $\gamma_1 TEN_{it} + \gamma_2 TEN_{it}^2$. The estimated returns remain small; although not reported, returns to tenure attain a maximum of only 0.0433 log points at 5 years. The fact that returns to tenure are so much smaller in my sample may reflect the fact that young workers have little incentive to invest in match-specific skills given their high degrees of mobility, in contrast with older workers in Topel's sample for whom this incentive is presumably greater.

4. Empirical Results

In implementing the methodology above, we should keep in mind that different worker groups, e.g. high school graduates and college graduates, might sample from different offer distributions and draw offers at different rates. We could carry out the estimation separately for each such group, but since the number of observations in my sample is already relatively small, I resort to grouping together all workers and assuming they all face identical search problems.

Since relating the moment sequence $\{E(R_{n+1} - R_n \mid N > n)\}_{n=1}^{\infty}$ back to an underlying parent distribution requires knowing the parameter $p = (1 + \lambda_1/\delta)^{-1}$, I begin with estimates for this

¹⁰As noted in Topel (1991), this two-step procedure is likely to overstate the true β_1 , since initial experience EXP_{0it} is likely to be positively correlated with n and hence w_n . As a result, the estimated returns to tenure should really be viewed as a lower bound for the true γ . However, with involuntary job changes, EXP_{0it} is likely to be weakly correlated with n , since workers with high experience can still be on the first job in their employment cycle. Hence, the lower bound we estimate is likely to be close to the true estimate. Indeed, when Altonji and Williams (1998) estimate returns to tenure in the same NLSY sample using an alternative instrumental variables approach, they find returns to tenure that are only slightly larger than those reported here.

parameter. Recall that Proposition 2 offers a maximum likelihood estimator for p based on the distribution of n across jobs that ended in an involuntary job change. Of the 44,593 jobs in my sample, 22,135 are classified as ending involuntarily. Among these, the distribution of n is heavily skewed towards $n = 1$. This would suggest a relatively high probability of involuntary job termination, i.e. a relatively large value of p . The exact point estimates are reported in Table 3. Grouping all workers together, p is estimated at 0.48, implying that the ratio $\lambda_1/\delta \approx 1$. To check whether grouping workers together overlooks important differences in p across identifiable subgroups, I also estimated p separately for different education groups. The point estimates do not seem to differ much from one another, confirming a similar result in van den Berg and Ridder (1998). The implied ratio for λ_1/δ of 1 is smaller than the ratio of 10 reported in several previous papers using job duration data as opposed to employment history data (including some that use the same NLSY dataset). However, not all papers find this ratio to be as large. In particular, Bowlus, Keifer, and Neumann (2001), who use job duration data in the NLSY data to estimate these parameters, also report a ratio $\lambda_1/\delta \approx 1$.

Next, I turn to estimating (2.13), which should yield estimates of the average record gaps. Once again, to mitigate the effect of outliers, I eliminated the most extreme 0.1% of my sample, specifically those observations where $|\Delta \ln \widehat{W}_{it}| > 2$. Most of these deletions appear to be due to coding errors, since these wage changes are typically followed by equally large wage changes in the opposite direction. Since there are very few observations for high values of n , I also confine my analysis to job changers who leave their n -th job for $n \leq 5$. Let $D_{it}^{n,n+1}$ represent a dummy variable which equals 1 if worker i moved from his n -th job in date $t - 1$ to his $n + 1$ -th in date t . Rather than estimating returns to experience from a separate first-stage regression as suggested in Section 2, I combine job stayers and job changers to run a single regression of the form

$$\Delta \ln \widehat{W}_{it} = \beta_1 \Delta EXP_{it} + \beta_2 \Delta EXP_{it}^2 + \sum_{n=1}^{\infty} \pi_n D_{it}^{n,n+1} + \Delta \varepsilon_{it} \quad (4.1)$$

The coefficients π_n in this regression correspond to the estimates of the expected moment gaps $E(R_{n+1} - R_n \mid N > n)$. Combining the two allows the wage growth of job changers to help in identifying the coefficient β_2 , and should therefore be more efficient.

The results of this regression are reported in Table 4. The number of workers who are observed to change from the n -th job to the $n + 1$ -th job in the previous year for each n is reported next to the corresponding dummy variable. The estimated coefficients for (4.1) are reported in the second column of Table 4. The first column in the table reports the estimated coefficients β_1 and β_2 omitting job changers, confirming that estimating β_1 and β_2 from job stayers alone would have negligible effects on my point estimates. The estimated coefficients π_n are all clustered around

8%, with the exception of π_4 . However, this coefficient (as well as those with even higher values of n) is rather imprecisely estimated given the small number of job changers for this value of n .

With only three moments that are estimated with any reasonable degree of precision, using the coefficients π_n to construct the estimator in Proposition 3 is likely to provide a poor approximation of the true log wage offer distribution. However, as noted earlier, we can still use a small number of moments to test particular functional form restrictions. Recall from Figure 1 that if the wage offer is Pareto, in which case the log wage offer distribution is exponential, the coefficients π_n should be constant for all n . Thus, testing this particular functional form restriction amounts to testing a set of linear restrictions on the coefficients π_n in (4.1), namely that they are all equal. Note that this is a test of an entire family of distributions rather than a test of one particular distribution. To the extent that we fail to reject that the π_n are equal, we can then estimate the exact exponential distribution (specifically, the inverse of the hazard rate that characterizes this distribution) from the implied common value of the π_n .

The first row in the bottom panel of Table 4 reports the results for the test that all of the coefficients π_n are equal. The probability of observing these average wage gains under the null that the log wage offer distribution is exponential equals 0.264. Thus, we fail to reject the null that the wage offer distribution is Pareto at conventional significance levels. The third column of Table 4 then reports the estimate for π_n assuming these coefficients are equal. The average net wage growth from voluntarily moving jobs is 0.0806, which is consistent with the estimated wage growth for young workers reported in Topel and Ward (1992). This estimate implies that the underlying log wage offer distribution is exponential with hazard $0.0806^{-1} = 12.41$. This is larger than the estimate of 4.17 Flinn (2002) reports for this hazard using the same NLSY dataset (Table 4, p633). But Flinn abstracts from on-the-job wage growth out of concern for sample selection, and instead attributes all of the growth between the starting wage on the n -th job in the cycle to the starting wage on the $n + 1$ -th job in the cycle to a better price from the underlying offer distribution. Since the average wage gain from the start of one job to the start of the next job is 0.2400 in his sample, the implied hazard rate of the underlying offer distribution, $0.2400^{-1} = 4.17$, will be smaller.

The fact that the estimated coefficients $\pi_n = E(R_{n+1} - R_n \mid N > n)$ are roughly constant naturally suggests the exponential distribution as a candidate for the log wage offer distribution. But we could similarly test other functional forms that have been proposed in the literature. The second row in the bottom panel of Table 4 reports the results for an analogous test of whether the log wage offer distribution is normal. Once again, we can devise a test against the entire class of normal distributions as opposed to a single candidate distribution, and then proceed to estimate

the exact normal distribution to the extent that we fail to reject this hypothesis. To see why, suppose log wage offers are distributed as $N(\mu, \sigma^2)$. One can verify that the average net wage growth among workers who move from their n -th job to their $n + 1$ -th job would equal

$$\sigma E(R'_{n+1} - R'_n \mid N > n) \tag{4.2}$$

where $E(R'_{n+1} - R'_n \mid N > n)$ denotes the average moment gap from the sequence $\{X'_m\}_{m=1}^M$ in which the X'_m are i.i.d. standard normals, i.e. $X'_m \sim N(0, 1)$. Thus, the sequence $\{\pi_n\}_{n=1}^\infty$ is uniquely determined for any normal distribution up to a scalar. To the extent that we fail to reject the hypothesis of a normal distribution, we can estimate σ from the exact values of π_n .¹¹ As can be seen from the last line of Table 4, the probability that we would observe these average wage gains under the null of any lognormal distribution is only 0.014, so we can reject this null at conventional levels of significance. This is not quite precise, since the null hypothesis depends on p , but p is an estimated parameter. However, since the moment sequence in (4.2) is sharply declining for a variety of different p , and since p is relatively tightly estimated, the rejection of the normal is likely to be robust to incorporating sampling error. Interestingly, the expected record gaps for the exponential distribution are independent of p , so the confidence interval in Table 4 does not need to be adjusted to reflect sampling error in the estimate of p .

Since the small sizes involved in my estimation make it hard to estimate the exact offer distribution with great confidence, it would be useful to look for additional tests that can be used to further gauge whether the wage offer distribution is indeed Pareto, i.e. whether the log wage offer distribution is indeed exponential. Since my estimation relies only on the wage growth of voluntary job changers, a natural overidentifying restriction to consider would involve the wage losses of involuntary job changers, data that I have so far ignored in my analysis. If a worker leaves his n -th job involuntarily, it follows that the total number of records in the employment cycle he just ended is exactly n , i.e. $N = n$ (recall that employment cycles are defined as the time between involuntary job changes). Thus, the average log price per unit labor on the job the worker left involuntarily is given by $E(w_n \mid N = n)$, i.e. it is the expected value of the n -th record conditional on the event that there are exactly n records in the sequence $\{X_m\}_{m=1}^M$. On his next job, all we know is that he started a new cycle, so the number of records in his next employment cycle is at least one. Hence, the average log price per unit labor on his new job is given by $E(w_1 \mid N \geq 1)$, which we can denote $E(w_1)$ since the event $N \geq 1$ is true by default. Hence, the average net wage

¹¹By contrast, the parameter μ is not identified, since recall from Proposition 1 that we can only identify a distribution up to a location parameter. Similarly, the exponential distribution is really defined by two parameters, its hazard and its lower support, but only the hazard is identified.

loss suffered by the n -th work can again be expressed as a difference of record moments, namely

$$E\left(|\Delta \ln \widehat{W}_{it} - \beta \Delta Z_{it}| \mid N_{t-1} = n\right) = E(R_n \mid N = n) - E(R_1)$$

Now, suppose the wage offer distribution were exponential with a hazard rate λ . Then these average losses can be written as

$$\lambda^{-1} [E(R'_n \mid N = n) - E(R'_1)] \quad (4.3)$$

where R'_n denotes the n -th record the sequence $\{X'_m\}_{m=1}^M$ in which the X'_m are i.i.d. standard exponentials, i.e. with hazard rate 1. Let $D_{it}^{n,1}$ denote a dummy which equals 1 if worker i moved from his n -th job in date $t-1$ to a job where n is reset to 1 by date t . Then if the log wage offer distribution were indeed exponential, the coefficients π_n in the regression

$$\Delta \ln \widehat{W}_{it} = \beta_1 \Delta EXP_{it} + \beta_2 \Delta EXP_{it}^2 - \sum_{n=1}^{\infty} \pi_n D_{it}^{n,1} + \Delta \varepsilon_{it} \quad (4.4)$$

should line up with (4.3) up to a scale parameter. More precisely, the scaling parameter λ^{-1} should be the inverse of the hazard for the exponential distribution we estimated from voluntary job changers, i.e. λ^{-1} should equal 0.0806.

The first column in Table 5 reports the estimates for the coefficients π_n . Using the estimate for p of 0.48 from Table 3, one can compute the moments $E(R'_n \mid N = n) - E(R'_1)$ numerically, yielding

$$\{0.197, 0.762, 1.127, 1.396, 1.616, \dots\} \quad (4.5)$$

The first row in the bottom panel of Table 5 reports the results of the test that the coefficients π_n are equal to a this sequence up to a scale parameter. The probability of observing these values under the null of an exponential distribution is given by 0.291, so we indeed fail to reject the null hypothesis. The second column of Table 5 reports the implied value of the constant λ^{-1} if we impose the restriction that the wage offer distribution is exponential. That is, we replace $\sum_{n=1}^{\infty} \pi_n D_{it}^{n,1}$ in (4.4) with the expression

$$\lambda^{-1} \left(0.197 D_{it}^{1,1} + 0.762 D_{it}^{2,1} + 1.127 D_{it}^{3,1} + 1.396 D_{it}^{4,1} + 1.616 D_{it}^{5,1} \right)$$

and estimate λ^{-1} by ordinary least squares. The estimated coefficient is 0.0816, nearly identical to the value of 0.0806 implied by the wage gains of voluntary job changers. In other words, the relationship between the wage gains of voluntary job changers and the wage losses of involuntary job changers implied by the model appears to be satisfied in the data.

As a final remark, note that so far I have only used the fact that if the log wage offer distribution is exponential, the average wage losses would have to be proportional to the sequence in (4.5). However, a simple adaptation of an argument in Nagaraja and Barlevy (2003) shows that the converse is also true, i.e. if wage losses are proportional to (4.5), then the offer distribution has to be exponential. Formally, the sequence $\{E(R_n | N = n) - E(R_1)\}_{n=1}^{\infty}$ uniquely characterizes the parent distribution within the class of continuous distribution functions up to a location parameter. As such, we could potentially use wage losses to similarly discriminate between different candidate wage offer distributions. Unfortunately, this approach proves to be less powerful than using the wage gains of voluntary job changers. For example, once again we can test whether the wage change data is consistent with a normal distribution by computing the sequence $\{E(R_n | N = n) - E(R_1)\}_{n=1}^{\infty}$ for the normal distribution and testing whether the coefficients π_n are consistent with this series. As reported in the second row of the bottom panel of Table 5, we cannot reject the null hypothesis that the log wage offer distribution is normal as we could from wage data for voluntary job changers. The reason for this is illustrated in Figure 3, which shows the actual estimated net wage loss and the best-fitting moment differences $E(R_n | N = n) - E(R_1)$ for the normal and the exponential distribution. Although the two sequences are distinguishable – the average wage losses decay more rapidly for the normal distribution – it is more difficult to tell them apart given that both sequences are necessarily increasing, in contrast the sequence of moments for the difference between consecutive records which need not be monotonic. Thus, the fact that wage losses for involuntary job changers can be reconciled with the exponential distribution serves more to affirm the internal consistency of the model than to further pin down the exact functional form of the wage offer distribution.

5. Alternative Models of Search

The estimation strategy described in this paper relies heavily on the fact that according to the model, observed wages over an employment cycle $\{\widehat{W}_{it}^n\}$ correspond to a sequence of records $\{W_n\}_{n=1}^N$ from an i.i.d. sequence that are contaminated by a noise term ℓ_{it} . In more general search models, this may no longer be the case. Nevertheless, I now argue that even when the underlying sequence $\{W_n\}_{n=1}^N$ does not correspond to a list of record values, it will often be the case that a model with on-the-job search has some implicit record structure, and in some cases we might still be able to exploit this fact for purposes of identification. This section offers two suggestive examples. A more comprehensive treatment is clearly necessary to deal with each of these variations, but this is beyond the scope of this paper.

As my first example, suppose a job offer specifies both a wage W and a number of hours H that the worker is required to work. Workers draw job offers from a fixed distribution over (W, H) and choose the job that maximizes their utility. Thus, on a job offering the pair (W, H) , an individual would earn an hourly wage of $\widehat{W}_{it} = W\ell_{it}$, and an income $\widehat{I}_{it} = WH\ell_{it}$. Once again, define an employment cycle as the time between the beginning of adjacent unemployment spells, and let $\{W_n, H_n\}_{n=1}^N$ denote the wages and hours on the different jobs over each such cycle. The sequence $\{W_n\}_{n=1}^N$ will no longer correspond to a sequence of records; in fact, it need not even be monotonic, since a worker might voluntarily move to a job that offers lower W if it is more attractive in terms of the hours it offers. Nevertheless, the n -th job in the cycle still corresponds to the n -th record in *utility space*, i.e. it is the n -th time the worker encounters a job he prefers to all of his previous job offers. Formally, the sequence $\{U(W_n, H_n)\}_{n=1}^\infty$ corresponds to records from the set $\{U_m\}_{m=1}^M$ where U_m denotes the utility the worker derives from the m -th job offer. If we know the function $U(\cdot, \cdot)$, or could estimate it from observed choices, we might be able to use data on wages and hours to identify the distribution of utility across job offers, and possibly even use this to back out the joint distribution of (W, H) . As an illustration, consider the case where agents do not care about leisure. Then they would always choose the job that offers the greatest income, i.e. $U(W, H) = WH$. In this case, the income on the n -th job corresponds to a record from i.i.d. draws from the implied distribution for income across all offers, and we can easily adapt the argument above to identify the distribution of income levels across job offers from observations on income $\widehat{I}_{it}^n = (W_n H_n) \times \ell_{it}$. One might be able to extend a similar argument to other utility functions.

As another example, suppose once again that a job offer only specifies a wage W , and workers draw offers from a fixed offer distribution. However, the productivity of the worker is in part match-specific. Formally, the log wage on the n -th job is given by

$$\ln \widehat{W}_{it}^n = w_n + \phi_i + \beta_1 EXP + \beta_2 EXP^2 + \gamma(TEN_{it}) + \varepsilon_{it}$$

where $\gamma(\cdot)$ is a strictly increasing function. Although we could rule out this case for the NLSY sample, returns to tenure might be more relevant for workers who are at a more advanced stage of their career. In this case, an individual might deliberately choose a job that pays a low wage today because of the promise of higher growth in the future. That is, a worker will move from his n -th job to his $n + 1$ -th job even though $\widehat{W}_{it}^n > \widehat{W}_{it}^{n+1}$ because he anticipates that the wage on the $n + 1$ -th job will be higher at some future date, i.e. that $\widehat{W}_{i\tau}^n < \widehat{W}_{i\tau}^{n+1}$ for some $\tau > t$.^{12,13}

¹²This would not be true in the Burdett and Mortensen model, i.e. it would never be the case that $\widehat{W}_{it}^n > \widehat{W}_{it}^{n+1}$. However, it is possible in that model that $\widehat{W}_{i,t-1}^n > \widehat{W}_{it}^{n+1}$, i.e. the worker might voluntarily move to a job that pays a lower wage than he earned in the past.

¹³A closely related model is developed by Postel-Vinay and Robin (2002). In their model, wages rise on the job not

At first glance, the fact that workers accept wage cuts would suggest the record structure of the model breaks down. But this turns out to be incorrect. In particular, a worker would never accept a job that offers a lower log price w than his current match, since the monotonicity of $\gamma(\cdot)$ implies the wage on such a job would always lag behind his current job. Hence, the sequence $\{w_n\}$ is still a monotonically increasing sequence. More precisely, the set w_n forms a sequence of records in which the threshold for setting a record evolves over time. That is, an observation forms a record if it exceeds the previous record by at least some (time-varying) amount, in much the same way that in athletic competitions a record is set only if it beats the previous record by more than the degree of precision by which performances are measured (which conceivably changes over time as technology makes it possible to measure time at higher precision). More generally, the different jobs in a worker’s employment cycle correspond to records in value function space as opposed to wage space or instantaneous utility space. Using this insight for identification involves a non-trivial modification to the record model explored here, and it is not obvious whether one can still obtain as strong of a characterization result. But to the extent that such a result can be established, the fact that we can directly estimate the function $\gamma(\cdot)$ using Topel’s method above suggests we should still be able to estimate the average underlying record gaps Δw_n .

6. Conclusion

In many applications that use search models, we would like to estimate the structural parameters of the model to learn about the underlying search process. Previous authors have indeed obtained important insights on various interesting questions on labor markets by proceeding to estimate such models. However, the methods they used require fairly stringent assumptions, either on the functional form of the wage offer distribution or on the presence of unobserved earnings heterogeneity. This paper proposes a way to estimate this distribution that can avoid these assumptions by exploiting the underlying record structure of the standard search model. While the number of observations in my dataset is too small to provide very precise estimates of the underlying wage offer distribution, my approach can still rule out certain functional forms, including some that have been used in applications such as the lognormal distribution. At a first pass, the evidence appears consistent with a Pareto wage offer distribution, i.e. the wages a worker could expect to earn in the various jobs available to him have a Pareto distribution. Note that this observation is distinct

because the worker becomes more productive, but because his employer is forced to match outside offers the worker receives. Since outside offers arrive at random, the wage would no longer be a deterministic function of tenure on the job. Ironically, in this case the wage on the job forms a record process, since the current wage reflects the record outside offer the worker encountered since he started working for his current employer.

from the oft-noted fact that the cross-sectional distribution of wages exhibits a Pareto tail.¹⁴ For one thing, the cross-sectional distribution is a convolution of the distribution of prices firms pay and the distribution of ability across agents. In addition, selection from workers moving to higher wage jobs would tend to put more mass on higher values of this distribution. In the simple search model explored here, the record structure is reflected in wages and can be used to uncover the underlying wage offer distribution. But more generally, the jobs on each employment cycle correspond to records in utility space, not wage space. Whether this still yields useful restrictions for observable data (e.g. wages, hours, etc) remains as an open question.

While this paper only examines search applications, record theory is potentially applicable in a variety of contexts. Records statistics arise whenever we get to observe the extremes from an unknown number of observations. This structure characterizes a variety of economic environments. For example, in the Postel-Vinay and Robin (2002) model, the wage a worker earns on his job is the maximum of the outside offers the worker receives, but most panel surveys do not record the number of times the worker receives a matching outside offer. A related example is the problem of optimal contracting with one-sided commitment in Beaudry and DiNardo (1991) in which the optimal strategy for the firm is to pay its worker a wage that reflects the record economic conditions since the employment relationship began, which may be only imperfectly observable to the econometrician. Another example involves auctions in which we do not observe how many potential bidders there are (e.g. internet auctions where we don't know whether those who visit the site are seriously intent on bidding), and all we get to observe are those bids that already exceed previous bids without knowing how many bidders had the opportunity to offer a higher bid but chose not to. Yet another application that is discussed at some length in Arnold, Balakrishnan, and Nagaraja (1998) involves optimal stopping problems, since the event that we reach a point at which we exceed some threshold can be translated into the statement that the record value exceeds some cutoff. Record statistics could thus serve as an important tool for economists in both empirical and theoretical applications.

¹⁴On the presence of a Pareto tail in cross-sectional earnings distributions, see Neal and Rosen (2000).

7. Appendix

Proof of Lemma 1: To derive the expression for $\text{Prob}(M = m)$, let condition on the time between the first offer and the end of the cycle, which is distributed as an exponential with rate δ . Then the probability that there are exactly m offers on an employment cycle can be expressed as

$$\text{Prob}(M = m) = \int_0^\infty \text{Prob}(m - 1 \text{ offers arrive by date } t) \delta e^{-\delta t} dt$$

Since offers arrive at rate λ_1 , the number of offers that arrive within t units of time is Poisson with parameter $\lambda_1 t$, so that

$$\begin{aligned} \text{Prob}(M = m) &= \int_0^\infty \frac{e^{-\lambda_1 t} (\lambda_1 t)^{m-1}}{(m-1)!} \delta e^{-\delta t} dt \\ &= \left(\frac{\lambda_1}{\lambda_1 + \delta} \right)^{m-1} \frac{\delta}{\lambda_1 + \delta} \end{aligned}$$

To solve for these integrals, we use an induction argument together with the fact that for any positive integer k

$$\begin{aligned} \lim_{t \rightarrow 0} t^k e^{-(\lambda_1 + \delta)t} &= 0 \\ \lim_{t \rightarrow \infty} t^k e^{-(\lambda_1 + \delta)t} &= 0 \end{aligned}$$

This establishes the claim. ■

Proof of Lemma 2: Given an i.i.d. sequence $\{X_m\}_{m=1}^M$ where $X_m \sim F(\cdot)$ and where $\text{Prob}(M = m) = q^{m-1}p$, Theorem 4.1 in Bunge and Nagaraja (1991) implies that the probability density for the first $n + 1$ records when there are at least $n + 1$ records in the sequence is given by

$$h(r_1, r_2, \dots, r_{n+1} \cap N > n) = f(r_{n+1}) \prod_{i=1}^n \frac{q f(r_i)}{1 - q F(r_i)} \quad (7.1)$$

where $f(\cdot) = dF(\cdot)$. Integrating out r_1 through r_n in (7.1) and using an induction argument, we can show that the marginal density for r_{n+1} where there are at least $n + 1$ records is given by

$$h(r_{n+1} \cap N > n) = \frac{[-\ln(1 - qF(r_{n+1}))]^n}{n!} f(r_{n+1})$$

Define the inverse cdf $F^{-1}(x)$ for $x \in (0, 1)$ as $\sup\{y : F(y) \leq x\}$. Then using the change of variables $u = F(r_{n+1})$ and $du = f(r_{n+1}) dr_{n+1}$, the expected value of $|R_{n+1}|$ conditional on $N > n$ is given by

$$\begin{aligned} E\left(|R_{n+1}| \mid N > n\right) &= \int_0^1 |F^{-1}(u)| \frac{[-\ln(1 - qu)]^n}{n! \Pr(N > n)} du \\ &\leq \frac{[-\ln(1 - q)]^n}{n! \Pr(N > n)} \int_0^1 |F^{-1}(u)| du \\ &= \frac{[-\ln(1 - q)]^n}{n!} E(|X_m|) < \infty \end{aligned}$$

Since $E(|R_n| \mid N > n) < E(|R_{n+1}| \mid N > n)$, the former is also finite. Lastly, since $E(|R_n|) < \infty$ implies $E(R_n) < \infty$, the lemma follows. ■

Proof of Proposition 1: Integrating out (7.1) yields the following densities:

$$\begin{aligned} h(r_{n+1}, r_n \mid N > n) &= f(r_{n+1}) \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)!} \frac{qf(r_n)}{1 - qF(r_n)} \\ h(r_n \mid N > n) &= \frac{q - qF(r_n)}{1 - qF(r_n)} \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)!} f(r_n) \end{aligned}$$

Define $\Delta = r_{n+1} - r_n$. By construction, $\Delta \geq 0$. Using the law of iterated expectations, we have

$$\begin{aligned} E(\Delta \mid N \geq n) &= E(E(\Delta \mid r_n, N > n)) \\ &= E\left(\int_0^\infty \Delta h(\Delta \mid r_n, N \geq n) d\Delta\right) \end{aligned}$$

where $h(\Delta \mid r_n, N \geq n)$ is the density of the difference between the n -th record and the $n + 1$ -th record conditional on r_n , and is given by

$$h(\Delta \mid r_n, N \geq n) = \frac{f(r_n + \Delta)}{1 - F(r_n)}$$

Hence, the conditional expectation of Δ is given by

$$\begin{aligned} E(\Delta \mid r_n, N \geq n) &= \int_0^\infty \Delta \frac{f(r_n + \Delta)}{1 - F(r_n)} d\Delta \\ &\equiv \mathcal{F}(r_n) \end{aligned}$$

If we integrate the above expression over r_n , we have

$$\begin{aligned} E(\Delta \mid N > n) &= E(\mathcal{F}(r_n) \mid N > n) \\ &= \int_{-\infty}^\infty \mathcal{F}(r_n) \frac{h(r_n \mid N > n)}{\Pr(N > n)} dr_n \\ &= \int_{-\infty}^\infty \mathcal{F}(r_n) \frac{q - qF(r_n)}{1 - qF(r_n)} \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)! \Pr(N > n)} f(r_n) dr_n \\ &= \int_{-\infty}^\infty \left[\int_0^\infty [1 - F(r_n + \Delta)] d\Delta \right] \frac{q}{1 - qF(r_n)} \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)! \Pr(N > n)} f(r_n) dr_n \end{aligned}$$

Now, suppose we have two functions F_1 and F_2 such that

$$E\left(R_{n+1}^{(1)} - R_n^{(1)} \mid N > n\right) = E\left(R_{n+1}^{(2)} - R_n^{(2)} \mid N > n\right)$$

for $n = 1, 2, 3, \dots$. Then we have

$$\begin{aligned} & \int_{-\infty}^{\infty} \left[\int_0^{\infty} [1 - F_1(r_n + \Delta)] d\Delta \right] \frac{(-\ln(1 - qF_1(r_n)))^{n-1}}{(n-1)!(1 - F_1(r_n))} \frac{qf_1(r_n)}{1 - qF_1(r_n)} dr_n = \\ & \int_{-\infty}^{\infty} \left[\int_0^{\infty} [1 - F_2(r_n + \Delta)] d\Delta \right] \frac{(-\ln(1 - qF_2(r_n)))^{n-1}}{(n-1)!(1 - F_2(r_n))} \frac{qf_2(r_n)}{1 - qF_2(r_n)} dr_n \end{aligned}$$

Rewrite both integrals using the change of variables $u = F(r_n)$ to get

$$\begin{aligned} & \int_0^1 \left[\int_0^{\infty} [1 - F_1(F_1^{-1}(u) + \Delta)] d\Delta \right] \frac{(-\ln(1 - qu))^{n-1}}{(n-1)!(1 - u)} \frac{q}{1 - qu} du = \\ & \int_0^1 \left[\int_0^{\infty} [1 - F_2(F_2^{-1}(u) + \Delta)] d\Delta \right] \frac{(-\ln(1 - qu))^{n-1}}{(n-1)!(1 - u)} \frac{q}{1 - qu} du \end{aligned}$$

Applying Lemma 3 in Lin (1987), we know that given a function $\psi(\cdot)$,

$$\int_0^1 \psi(x) (-\ln(1 - x))^n dx = 0$$

for all $n = 1, 2, 3, \dots$ if and only if $\psi(x) = 0$ almost surely. By a simple contradiction argument, one can show that this implies that $\psi(x) = 0$ almost surely if and only if

$$\int_0^1 \psi(x) (-\ln(1 - qx))^n dx = 0$$

Hence, for any u , it follows that

$$\int_0^{\infty} [1 - F_1(F_1^{-1}(u) + \Delta)] d\Delta = \int_0^{\infty} [1 - F_2(F_2^{-1}(u) + \Delta)] d\Delta$$

Let $t = F_1^{-1}(u) + \Delta$. Then it follows that for any u ,

$$\left[\int_{F_1^{-1}(u)}^{\infty} [1 - F_1(t)] dt \right] = \left[\int_{F_2^{-1}(u)}^{\infty} [1 - F_2(t)] dt \right]$$

Since $F_1(\cdot)$ and $F_2(\cdot)$ are continuous, nondecreasing, and bounded, it follows that they are both differentiable almost everywhere. This, in turn, implies that $F_1^{-1}(u)$ and $F_2^{-1}(u)$ are differentiable for almost every $u \in (0, 1)$. Differentiating with respect to such u yields

$$[1 - F_1(F_1^{-1}(u))] \frac{d}{du} F_1^{-1}(u) = [1 - F_2(F_2^{-1}(u))] \frac{d}{du} F_2^{-1}(u)$$

Since $F_1(F_1^{-1}(u)) = F_2(F_2^{-1}(u)) = u$, it follows that for almost all $u \in (0, 1)$,

$$\frac{d}{du} F_1^{-1}(u) = \frac{d}{du} F_2^{-1}(u)$$

Integrating out yields

$$F_1^{-1}(u) = F_2^{-1}(u) + c$$

for some constant c , which establishes the claim. ■

Proof of Proposition 2: the Proposition is an immediate implication of the consistency of the maximum likelihood estimator. ■

Proof of Proposition 3: From equation (7.2), we have

$$\begin{aligned} E(R_{n+1} - R_n \mid N > n) &= \int_{-\infty}^{\infty} \left(\int_0^{\infty} (1 - F(r_n + \Delta)) d\Delta \right) \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)! \Pr(N > n)} \frac{qf(r_n)}{1 - qF(r_n)} dr_n \\ &\equiv \int_{-\infty}^{\infty} G(r_n) \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)! \Pr(N > n)} \frac{qf(r_n)}{1 - qF(r_n)} dr_n \end{aligned}$$

Note that $G(r_n) \geq 0$. Using the change in variables

$$1 - qF(r_n) = p^x$$

we can rewrite the above as

$$\begin{aligned} E(R_{n+1} - R_n \mid N > n) &= \int_{-\infty}^{\infty} G\left(F^{-1}\left(\frac{1 - p^x}{q}\right)\right) \frac{(-\ln p)^n x^{n-1}}{(n-1)! \Pr(N > n)} dx \\ &\equiv \frac{(-\ln p)^n}{(n-1)! \Pr(N > n)} \int_{-\infty}^{\infty} g(x) x^{n-1} dx \end{aligned}$$

where again $g(x) \geq 0$. Set

$$\mu_{n-1} = E(R_{n+1} - R_n \mid N > n) \times \frac{(n-1)! \Pr(N > n)}{(-\ln p)^n}$$

so that $0 < \mu_n < \infty$ for all n . The task of recovering $g(\cdot)$ from the system of equations

$$\int_0^1 g(x) x^n = \mu_n \tag{7.3}$$

for $n = 0, 1, 2, \dots$ is just a case of the Hausdorff moment problem, which asks for an arbitrary sequence $\{\mu_n\}_{n=0}^{\infty}$ with $\mu_0 = 1$ if (1) there exists a function $g(\cdot) \geq 0$ that satisfies (7.3); (2) if the solution is unique; and (3) a closed form expression for any solution $g(\cdot)$. Shohat and Tamarkin (1943) provide a rigorous treatment of this and related moment problems. They prove that if a solution exists to the Hausdorff moment problem, it is unique (as it would indeed have to be from Proposition 1). Moreover, by Theorem 3.7 of Shohat and Tamarkin (p91), they show that the solution is given by

$$g(x) = \sum_{n=0}^{\infty} \lambda_n \mathcal{P}_n(x)$$

where

$$\lambda_n = \int_0^1 g(x) \mathcal{P}_n(x) dx$$

But we let $\mathcal{P}_n(x) = \sum_{j=0}^n c_{nj}x^j$

$$\begin{aligned} \int_0^1 g(x) \mathcal{P}_n(x) dx &= \int_0^1 \left(g(x) \sum_{j=0}^n c_{nj}x^j \right) dx \\ &= \sum_{j=0}^n c_{nj} \int_0^1 g(x) x^j dx \\ &= \sum_{j=0}^n c_{nj} \mu_j \end{aligned}$$

Hence, steps (1) and (2) allow us to recover the function $g(x)$ where

$$\begin{aligned} g(x) &= G \left(F^{-1} \left(\frac{1-p^x}{q} \right) \right) \\ &= \int_0^\infty \left(1 - F \left(F^{-1} \left(\frac{1-p^x}{q} \right) + \Delta \right) \right) d\Delta \\ &= \int_{F^{-1} \left(\frac{1-p^x}{q} \right)}^\infty (1 - F(t)) dt \end{aligned}$$

where the last step uses the change in variables $t = F^{-1} \left(\frac{1-p^x}{q} \right) + \Delta$. Integration by parts and a little algebra reveals that we can rewrite the above integral as

$$g(x) = \int_{\frac{1-p^x}{q}}^1 \left[F^{-1}(u) - F^{-1} \left(\frac{1-p^x}{q} \right) \right] du$$

Differentiating both sides with respect to x yields

$$g'(x) = \left(1 - \frac{1-p^x}{q} \right) F^{-1'} \left(\frac{1-p^x}{q} \right) \frac{p^x}{q} \ln p$$

Using the change of variables

$$z = \frac{1-p^x}{q}$$

and integrating, we have

$$\int_0^1 \frac{qg' \left(\frac{\ln(1-qz)}{\ln p} \right)}{(1-z)(1-qz) \ln p} dz = \int_0^x F^{-1'}(z) dz$$

or

$$F^{-1}(x) = \int_0^x \frac{qg' \left(\frac{\ln(1-qz)}{\ln p} \right)}{(1-z)(1-qz)\ln p} dz + \text{constant}$$

which completes the proof. ■

Table 1: Summary Statistics
for Entire Sample

# of individuals	6,284		
individual characteristics:			
		mean	median
age		24.6	25.0
years of potential experience		8.3	9.0
years of education		12.7	12.0
# of jobs	44,593		
job characteristics:			
% jobs ending voluntarily			0.35
% jobs ending involuntarily			0.50
% jobs censored/not classified			0.15
average job tenure (uncensored)			1.05
average wage (1992 dollars)			\$7.00
median wage (1992 dollars)			\$5.40

Source: National Longitudinal Survey of Youth, author tabulations. Statistics above are for the original sample, i.e. for all jobs reported in each year.

Table 2: Estimating Returns to Tenure γ

linear returns to tenure					
	within-job wage growth $\beta_1 + \gamma$		experience effect β_1		tenure effect γ
	0.0794 0.0065		0.0740 0.0061		0.0054 0.0024
	1 year	2 years	5 years	7 years	10 years
implied returns to tenure	0.0054 0.0024	0.0108 0.0049	0.0271 0.0122	0.0380 0.0171	0.0542 0.0245
implied returns to experience	0.0723 0.0058	0.1411 0.0109	0.3270 0.0226	0.4337 0.0274	0.5680 0.0300
quadratic returns to tenure					
	within-job wage growth $\beta_1 + \gamma_1$		experience effect β_1		tenure effect γ_1
	0.0826 0.0065		0.0661 0.0067		0.0165 0.0024
					tenure squared γ_2
					-0.0016 0.00048

The regressions above follow the two-step method outlined in Topel (1991). The first stage regresses annual within-job real wage growth (in 1992 dollars using the implicit GDP deflator) on a ΔEXP (= constant) and ΔEXP^2 . This is the same regression in column (1) of Table 4, where $\beta_1 + \gamma$ corresponds to the coefficient on ΔEXP . The second stage regresses the log real wage net of the estimated $(\beta_1 + \gamma)TEN + \beta_2 EXP^2$ on initial experience and individual fixed-effects. The coefficient on initial experience corresponds to the estimate of β_1 , and the difference corresponds to the estimate of γ above. Standard errors for β_1 and γ are adjusted to reflect estimation error in the first-stage regressor, using the stacking and weighting procedure in Altonji and Williams (1998). Returns to tenure and experience in the middle of the table are based on estimates for γ , β , and β_2 . In the bottom panel, the first stage regression is amended to allow for a ΔTEN^2 term, which is then subtracted from the log real wage at the second stage.

Table 3: Estimates for p

	Sample size	p	Standard error	Implied λ_1/δ
All	22,135	0.4823	0.0031	1.074
Educ < 12	6,515	0.5008	0.0055	0.997
Educ = 12	6,648	0.4797	0.0058	1.085
Educ \in (13,15)	5,436	0.4504	0.0062	1.220
Educ \geq 16	3,536	0.5049	0.0082	0.981

Estimates for p are derived using maximum likelihood in accordance with Proposition 2 in the text. Sample size corresponds to the number of jobs that end in an involuntary job change used to estimate p . The standard error is the asymptotic standard error. The implied ratio in the last column is computed according to the formula $p = (1 + \lambda_1/\delta)^{-1}$.

Table 4: The Wage Gains of Voluntary
Job Changers, by n

	sample size	(1)	(2)	(3) exponential
ΔEXP	--	0.0794	0.0809	0.0816
		0.0065	0.0050	0.0050
ΔEXP^2	--	-0.0017	-0.0018	-0.0018
		0.0003	0.0002	0.0002
D^{12}	2,473		0.0900	} 0.0806 0.0072
			0.0094	
D^{23}	993		0.0711	
			0.0137	
D^{34}	459		0.0799	
			0.0200	
D^{45}	206		0.0168	
			0.0331	
D^{56}	78		0.0799	
			0.0520	
# obs		28,015	31,868	31,868
stayers		28,015	28,015	28,015
changers		0	3,853	3,853

Test of particular functional forms:		
Exponential	$F(4, 31861) = 1.31$	Prob > $F = 0.2639$
Normal	$F(4, 31861) = 3.12$	Prob > $F = 0.0140$

The dependent variable is the annual growth rate of real wages. The independent variables are the growth ΔEXP , which is identically equal to 1, ΔEXP^2 , which is equal to $2 \text{EXP} - 1$, and a set of dummy variables $D^{n,n+1}$ equal to 1 if the worker moved from his n -th job to his $n+1$ -th job. The column labeled sample size denotes the number of workers in my sample who voluntarily left their n -th job for each value of n . Column (1) estimates the coefficients on ΔEXP and ΔEXP^2 using job stayers only. Column (2) adds job changers and estimates the coefficients on the dummy variables as well. Column (3) estimates the same regression as in column (2) assuming the coefficients on all the dummy variables are equal, which from the text is true if and only if the log wage offer distribution is exponential. The numbers below the coefficient denote robust standard errors. The F -statistics in the bottom panel are the robust Wald-statistics that test constraints on the coefficients on the dummy variables in column (2). The exponential case compares column (3) to column (2), while the normal case involves an alternative set of linear restrictions on the coefficients on the dummy variables.

Table 5: The Wage Losses of Involuntary Job Changers, by n

	sample size	(1)	(2) exponential
ΔEXP	--	0.0837 0.0062	0.0849 0.0050
ΔEXP^2	--	-0.0020 0.0002	-0.0020 0.0002
D^{11}	2,767	0.0029 0.0094	} 0.0816 0.0130
D^{21}	873	0.0843 0.0153	
D^{31}	305	0.0904 0.0278	
D^{41}	137	0.0942 0.0432	
D^{51}	50	0.0754 0.0726	
# obs		31,844	31,844
stayers		28,015	28,015
changers		3,829	3,829
Test of particular functional forms:			
Exponential		$F(4, 31837) = 1.24$	Prob > $F = 0.2905$
Normal		$F(4, 31837) = 1.08$	Prob > $F = 0.3629$

The dependent variable is the annual growth rate of real wages. The independent variables are ΔEXP and ΔEXP^2 as in Table 4, and a set of dummy variables $D^{n,n+1}$ equal to 1 if the worker moved from his n -th job to his $n+1$ -th job. The column labeled sample size denotes the number of workers who involuntarily left their n -th job for each value of n . Column (1) reports the results of this regression, while column (2) estimates the same regression as in column (1) with a particular set of linear restrictions on the coefficients of the dummy variables that are true if and only if the log wage offer distribution is exponential. The numbers below the coefficient denote robust standard errors. The F -statistics in the bottom panel are the robust Wald-statistics that test constraints on the coefficients on the dummy variables in column (2). The exponential case compares column (2) to column (1), while the normal case involves an alternative set of linear restrictions on the coefficients on the dummy variables.

Figure 1: Expected Record Gaps
for Different Parent Distributions

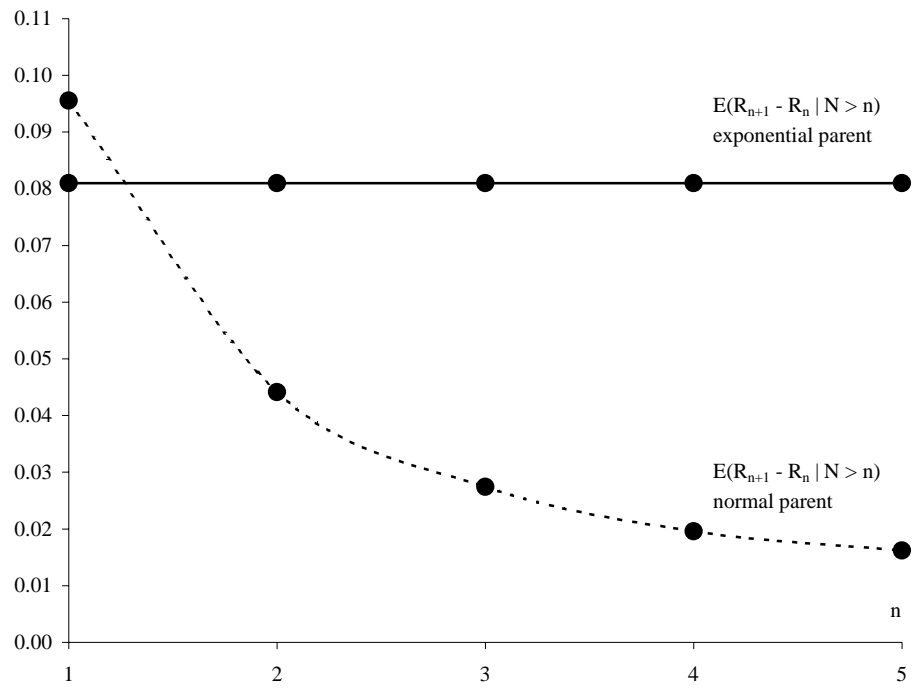


Figure 2: Summary Statistics for n

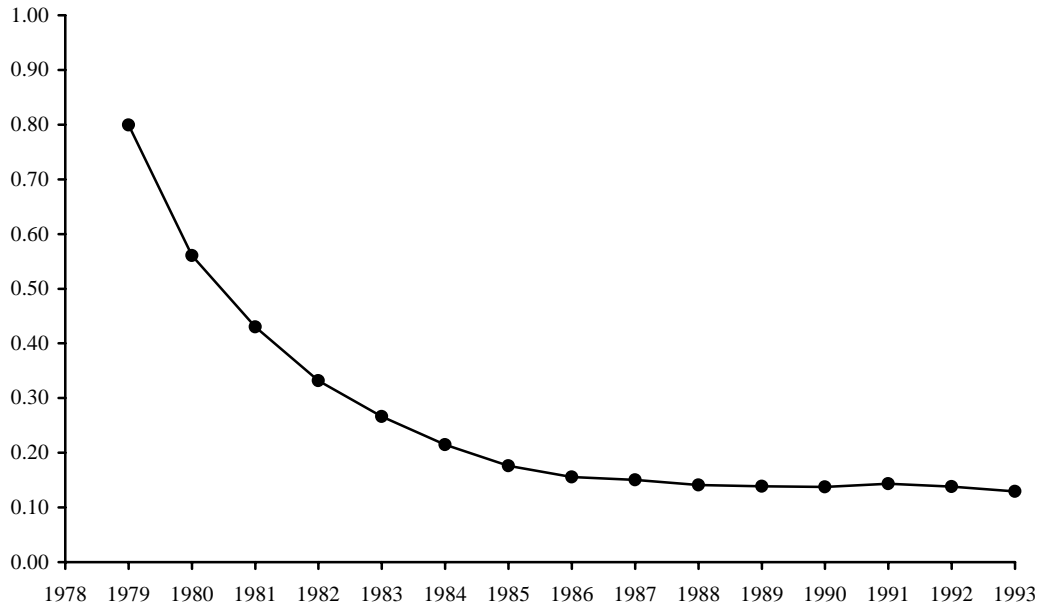


Figure 1a: Proportion of observations where no value for n was assigned

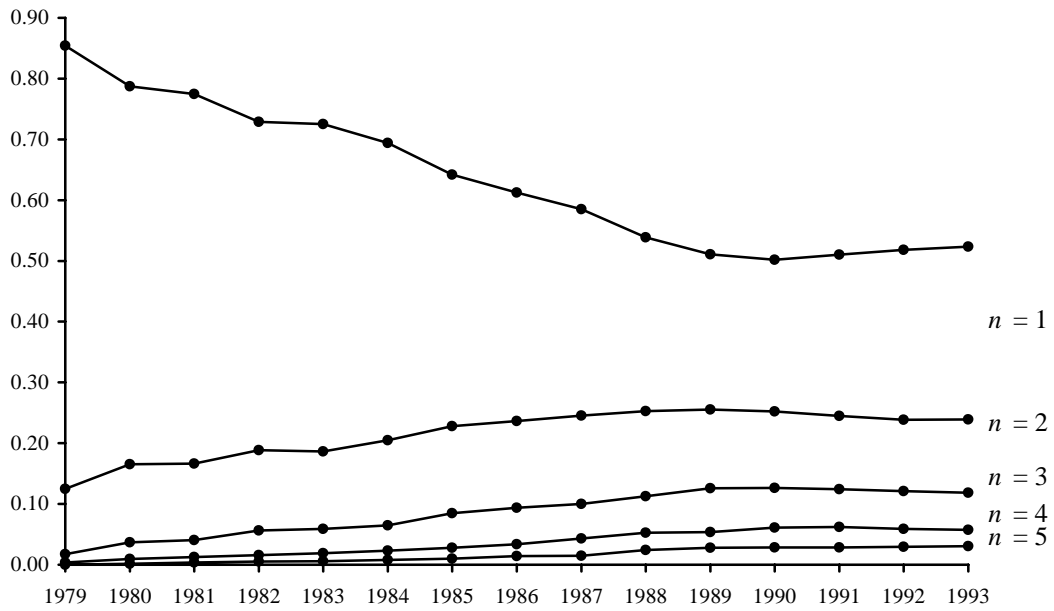
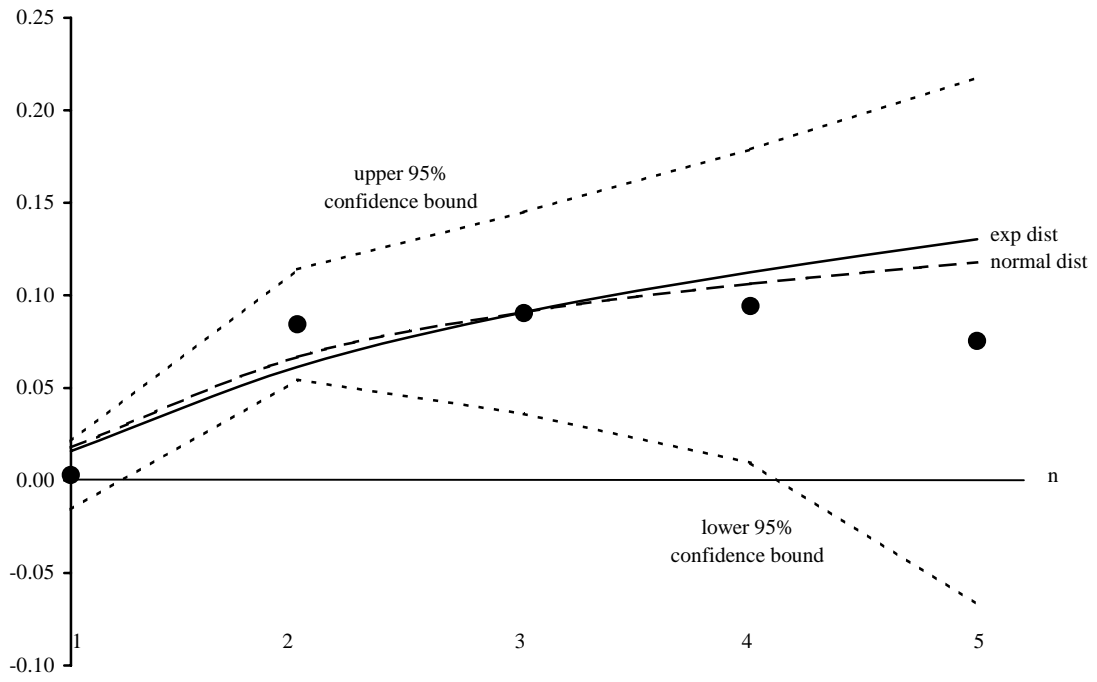


Figure 1b: Share of all observations with $n \geq 1$ for each level of n

Figure 3: Actual vs. Predicted Wage Loss for Involuntary Job Changers



References

- [1] Altonji, Joseph and Nicolas Williams, 1997. "Do Wages Rise with Job Seniority? A Reassessment" Mimeo, Northwestern University.
- [2] Altonji, Joseph and Nicolas Williams, 1998. "The Effects of Labor Market Experience, Job Seniority, and Job Mobility on Wage Growth" *Research in Labor Economics*, 17, p233-276.
- [3] Arnold, Barry, N. Balakrishnan and H. Nagaraja, 1992. *A First Course in Order Statistics*. New York: John Wiley and Sons.
- [4] Arnold, Barry, N. Balakrishnan and H. Nagaraja, 1998. *Records*. New York: John Wiley and Sons.
- [5] Barlevy, Gadi, 2002. "The Sullyng Effect of Recessions" *Review of Economic Studies*, 69 (1), January, p65-96.
- [6] Beaudry, Paul and John DiNardo, 1991. "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data" *Journal of Political Economy*, August, 99(4), p665-88.
- [7] Bontemps, Christian, Jean-Marc Robin, and Gerard van den Berg, 2000. "Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation" *International Economic Review*, May, 41(2), p305-358.
- [8] Bowlus, Audra, Nicholas Kiefer, and George Neumann, 2001. "Equilibrium Search Models and the Transition from School to Work" *International Economic Review*, May, 42(2), p317-43.
- [9] Bowlus, Audra and Jean-Marc Robin, 2003. "Twenty Years of Rising Inequality in US Lifetime Labor Income Values" *Review of Economic Studies*, forthcoming.
- [10] Bunge, John and H. Nagaraja, 1991. "The Distributions of Certain Record Statistics from a Random Number of Observations" *Stochastic Processes and Their Applications*, 38, p167-83.
- [11] Burdett, Kenneth and Dale Mortensen, 1998. "Wage Differentials, Employer Size, and Unemployment" *International Economic Review*, 39, p257-273.
- [12] Chandler, K. N., 1952. "The Distribution and Frequency of Record Values" *Journal of the Royal Statistical Society, Series B*, 14, p220-8.
- [13] Flinn, Christopher, 1986. "Wages and Job Mobility of Young Workers" *Journal of Political Economy* 94(3, Part 2), pS88-S110.

- [14] Flinn, Christopher, 2002. "Labour Market Structure and Inequality: a Comparison of Italy and the U.S." *Review of Economic Studies*, July, 69 (3), p611-45.
- [15] Flinn, Christopher and James Heckman, 1982. "New Methods for Analyzing Structural Models of Labor Force Dynamics" *Journal of Econometrics*, January, 18(1), p115-68.
- [16] Glick, Ned, 1978. "Breaking Records and Breaking Boards" *American Mathematical Monthly*, 85(1), p2-26.
- [17] Kirmani, S. N. U. A. and Beg, M. I., 1984. "On characterization of distributions by expected records" *Sankhyā A*, 46(3), p463-465.
- [18] Kortum, Samuel, 1997. "Research, Patenting, and Technological Change" *Econometrica*, 65(6), November, p1389-1419.
- [19] Lin, G. D., 1987. "On Characterizations of Distributions via Moments on Record Values" *Probability Theory and Related Fields*, 74, p479-83.
- [20] Munasinghe, Lalith, Brendan O'Flaherty, and Stephan Danninger, 2001. "Globalization and the Rate of Technological Progress: What Track and Field Records Show" *Journal of Political Economy*, 109(5), October, p1132-49.
- [21] Nagaraja, H. N. and Gadi Barlevy, 2003. "Characterizations Using Record Moments in a Random Record Model and Applications" *Journal of Applied Probability*, September, 40(3), p826-33.
- [22] Neal, Derek and Sherwin Rosen, 2000. "Theories of the Distribution of Earnings" *Handbook of Income Distribution*, 1, Amsterdam: Elsevier Science, North-Holland, p379-427.
- [23] Nevzorov, V. B. and N. Balakrishnan, 1998. "A Record of Records" *Handbook of Statistics: Order Statistics, Theory and Methods*, v16, eds. N. Balakrishnan and C. R. Rao, Amsterdam: North-Holland, p515-70.
- [24] Paxson, Christina and Nachum Sicherman, 1996. "The Dynamics of Dual-Job Holding and Job Mobility" *Journal of Labor Economics*, July, 14(3), p357-93.
- [25] Postel-Vinay, Fabien and Jean-Marc Robin, 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity" *Econometrica*, November, 70(6), p2295-2350.
- [26] Shohat J. A. and J. D. Tamarkin, 1943. *The Problem of Moments*. New York: American Mathematical Society.
- [27] Talenti, Giorgio, 1987. "Recovering a Function from a Finite Number of Moments" *Inverse Problems*, August, 3(3), p501-17.

- [28] Topel, Robert, 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority" *Journal of Political Economy*, February, 99(1), p145-76.
- [29] Topel, Robert and Michael Ward, 1992. "Job Mobility and the Careers of Young Men" *Quarterly Journal of Economics*, May, 107(2), p439-79.
- [30] van den Berg, Gerard and Geert Ridder 1998. "An Empirical Equilibrium Search Model of the Labor Market" *Econometrica*, September, 66(5), p1183-1221.
- [31] Wolpin, Kenneth, 1992. "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth" *Journal of Political Economy*, June, 100(3), p535-60.