Title
Identification of Search Models Using Record Statistics

Author
Barlevy, Gadi

Publication Date
2005-04-14

Identification of Search Models using Record Statistics*

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

March 8, 2005

Abstract

This paper shows how record-value theory, a branch of statistics that deals with the timing and magnitude of extreme values in sequences of random variables, can be used to non-parametrically identify the offer distribution of wages workers face. Using NLSY wage data, I show that the data supports the hypothesis that the wage offer distribution is Pareto but rejects that it is lognormal. In addition, I show that my approach can be used to construct a bound on the return to job-specific human capital. Using the same NLSY data, I find that job-specific human capital plays only a minor role in the wage growth of the workers in my sample. Instead, wage growth among the young workers in my sample appears to be driven primarily by the accumulation of general human capital as well as on-the-job search.

Key Words: On-the-Job Search, Non-Parametric Identification, Wage Growth, Specific Human Capital

*I would like to thank Joe Altonji, Susan Athey, Marco Bassetto, Jeff Campbell, Kim-Sau Chung, Zvi Eckstein, Chuck Manski, Francesca Molinari, Chris Taber, Nick Williams, and especially Eric French and H. N. Nagaraja for helpful discussions. I also wish to acknowledge the help of my research assistant Merritt Lyon. Finally, I have benefitted from comments at seminars in Arizona State, the Federal Reserve Bank of Minneapolis, Tel Aviv University, Michigan State, NYU, Penn State, University of Pennsylvania, University of British Columbia, SUNY Albany, Berkeley, 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 increasingly turned to job-search models to examine key questions in labor economics. For example, search models allow us to distinguish among competing hypotheses for why wage growth varies by race: are blacks and whites paid differently, do they encounter job offers at different rates, or do they accumulate human capital at different rates? Examples of papers that explore these issues include Wolpin (1992) and Bowlus, Kiefer, and Neumann (2001). As another example, search models can be used to estimate the effects of changes in the wage offer distribution on long-run earnings inequality; papers in this vein include Flinn (2002) and Bowlus and Robin (2004). Finally, search models can be useful for predicting the effects of various events on labor markets. This includes the effects of increasing the minimum wage, as in van den Berg and Ridder (1998) and Flinn (2004), or of a macroeconomic shock, as in Barlevy (2002).

Unfortunately, the answers we obtain using this approach can be sensitive to functional form assumptions, particularly those assumptions that govern the offer distribution workers face. For example, we may erroneously conclude blacks and whites face a common offer distribution if the same distribution happens to provide the best overall fit for both groups within the limited family of distributions we consider. As another example, the effect of an increase in the minimum wage depends on how many employers choose to offer a wage just above the original minimum wage, and a functional form that provides a good general fit may do poorly in matching this particular part of the distribution. Thus, it is important to ask whether job search models can be identified non-parametrically, at the very least so that we can verify candidate functional forms before proceeding with parametric estimation. Moreover, since differences in wages across workers likely reflect both differences in what employers offer as well as differences in worker quality, we need an approach to identification that is robust to the presence of worker heterogeneity.

This paper examines whether search models are non-parametrically identified. My approach is related to work by Athey and Haile (2002) on non-parametric identification of auction models. Appealing to the auction literature is only natural; after all, search models also involve multiple bidders competing over a common object, namely the worker's time. One key difference, though, is that auction data typically include the number of active bidders, while worker surveys seldom ask workers how many employers have previously bid to hire them. This is important, since many of Athey and Haile's results require that we know the number of bidders in the auction.[1]

[1] Song (2004) examines identification in auction models when the number of bidders is unknown. Her approach can be applied to search models, but it is not robust to the presence of unobserved heteroegneity.

While we cannot appeal to results in Athey and Haile (2002) to identify search models, it turns out that the job offers a worker accepts – the analog of winning bids in auction models – have a particular structure that we can still exploit. Specifically, the jobs the worker accepts form a sequence of *records*, in the sense that the worker must value each job he accepts more than the offers that preceded it. Statisticians have studied the behavior of record values from random sequences, and have applied their findings to study various phenomena such as global warming, record athletic performances, road congestion, and tolerance testing.[2] Analogous results to those that have already been established in the statistics literature reveal that if the jobs a worker accepts correspond to records, we can identify the offer distribution workers face even with only limited information on how worker quality varies across individuals or over a worker's lifetime. By contrast, previous work on identification of search models has had to assume that worker heterogeneity is either absent or perfectly observable.

In what follows, I describe a model of on-the-job search in which workers of varying ability draw offers from a fixed offer distribution. I show that this distribution is identified, although exact identification turns out to be impractical in the presence of unobserved heterogeneity. Nevertheless, we can test particular hypotheses about the shape of the offer distribution and narrow down the set of possible functional forms. For example, using data on young men from the National Longitudinal Survey of Youth (NLSY), I find that the offer distribution is consistent with a Pareto shape, a functional-form used by Flinn (2002), but not with a lognormal distribution as has been assumed in some of the other aforementioned works.

While I mostly focus on one particular approach to identifying the wage offer distribution, this distribution is actually overidentified: the offer distribution uniquely determines both the average wage gains of voluntary job changers and the average wage losses of involuntary job changers. Much of my discussion focuses on the former, although I confirm that both approaches consistently identify the same offer distribution. As I explain below, this consistency is itself informative, since it suggests job-specific human capital cannot be an important source of wage growth for the young workers in my sample; if it were, the average wage losses for involuntary job changers would have seemed too large relative to the average wage gains of voluntary job changers. This suggests an alternative approach for identifying the contribution of job-specific human capital to the ones proposed by Altonji and Shaktoko (1987) and Topel (1991).

The paper is organized as follows. Section 1 introduces the concept of record statistics. Section

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

2 describes the model and shows how it can be identified non-parametrically. Section 3 describes data from the NLSY that can be used to implement this approach. Section 4 reports the results. Section 5 discusses the wage losses of involuntary job changers. Section 6 considers search models in which wages do not correspond to record statistics but where there is still an underlying record structure inherent to the model. Section 7 concludes.

1. Record Statistics

Although statisticians have written extensively on record processes, their work has attracted scant attention from economists.[3] 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 all observations that preceded it in the sequence. Formally, let $L_1 = 1$, and for any integer $n > 1$ define the n-th record time L_n recursively as

$$L_n = \min\left\{m : X_m > X_{L_{n-1}}\right\} \tag{1.1}$$

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

$$R_n = X_{L_n} \tag{1.2}$$

As an illustration, suppose we recorded the daily average temperature in a given location on the same date each year, and obtained the following sequence:

$$\{65, 61, 68, 69, 63, 67, 64, 66, ...\} \tag{1.3}$$

The first observation is trivially a record, so $L_1 = 1$ and $R_1 = 65$. The next observation that exceeds this value is the third one, so $L_2 = 3$ and $R_2 = 68$. The very next observation exceeds this value, so $L_3 = 4$, and $R_3 = 69$. Thus, we can construct a sequence of records $\{R_n\}$ from the original sequence $\{X_m\}$ in (1.3):

$$\{65, 68, 69, ...\}$$

[3]Exceptions 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.

Note that $\{R_n\}$ is a subsequence of $\{X_m\}$, and as such is less informative. For example, we cannot infer how many years transpired between when any two consecutive record temperatures were set, i.e. we cannot deduce L_n from the sequence $\{R_n\}$.

Next, suppose the sequence $\{X_m\}_{m=1}^{M}$ represents some stochastic process. In this case, the number of records in the sequence $\{X_m\}_{m=1}^{M}$ and their values are well-defined probabilistic events. One case that has been analyzed extensively, enough that it is referred to as the *classical* record model, is where $M = \infty$ and X_m are independent and identically distributed with a given distribution that is referred to as the *parent distribution*. This case was first analyzed by Chandler (1952). Various results for this case have since been derived: formulae for the distribution of record times L_n and the number of records within a given sample size; the distribution of various attributes of the record process $\{R_n\}_{n=1}^{\infty}$ for a given parent distribution; and, conversely, characterizations for the parent distribution given information on the record process $\{R_n\}_{n=1}^{\infty}$. Record processes are more difficult to characterize when X_m are not i.i.d., 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 standard search model does not quite reduce to the classical record model, so we will not be able to rely on existing results for our analysis.

As a final note, it is worth commenting on the connection between record statistics and order statistics. The n-th maximal order statistic, denoted $X_{n:n}$, is the maximum of n random variables, $\max\{X_1, ..., X_n\}$. By contrast, the n-th record statistic R_n is the maximum of a *random* number L_n observations, $\max\{X_1, ..., X_{L_n}\}$. Given a value for L_n, the n-th record can certainly be viewed as an order statistic, i.e. $R_n = X_{L_n:L_n}$, and the fact that the highest recorded number in the series changed $n-1$ times can be ignored. But without conditioning on the value of L_n, the n-th record R_n is a *mixture* of order statistics, whose mixing probabilities depend on n. Formally, the probability that the n-th record value equals x can be expressed as

$$\Pr(R_n = x) = \sum_{m=n}^{\infty} \Pr(L_n = m) \times \Pr(X_{m:m} = x) \tag{1.4}$$

Since mixtures of distributions do not necessarily inherit the properties of the underlying distributions, results that are true for order statistics may not be true for record statistics. For example, the average value of the n-th record value may not exist even though the average value of the corresponding order statistic exists for any finite sample size. Thus, although order statistics and record statistics are closely related, results on order statistics that have proven so useful for analyzing auction models cannot be directly applied to studying record processes.

2. Job Search and Record Statistics

Having introduced the concept of records, I can turn my attention to job search. This section describes a model in which workers search from a fixed offer distribution and shows how insights from record statistics can be used to identify this distribution. Note that I treat the offer distribution as a primitive rather than deriving it from economic fundamentals. However, I show below that my model can be viewed as a reduced form of richer models in which the equilibrium offer distribution is uniquely determined by economic fundamentals. For these models, identifying the offer distribution is equivalent to identifying the fundamentals we might ultimately care about.

This section is organized as follows. I first describe the economic environment. Next, I discuss identification in the benchmark case where worker productivity is perfectly observable. I then turn to the case where worker productivity is imperfectly observable.

2.1. A Model of Job Search

Consider an economy populated by employers and workers. Workers supply a homogeneous labor input, although they may each supply different amounts of labor. Let ℓ_{it} denote the amount of labor worker i can supply per hour at date t. This amount – which is essentially the worker's productivity – is observable to both the employer and the worker, but need not be observable to the econometrician who collects data on this market. Later on I will be more precise as to what the econometrician observes and what assumptions I impose on the unobservable part.

A worker can be either unemployed or working for an employer. While unemployed, a worker can produce $b\ell_{it}$ units of output per hour, where b is the productivity of the technology in the home sector. Alternatively, b can be viewed as the marginal value of leisure, and $b\ell_{it}$ is the amount of leisure he gets to enjoy. All workers are assumed to share the same value of b. The reason for imposing this structure will become clear momentarily.

When a worker is unemployed, he encounters potential employers at rate λ_0 per unit time. When an employer meets a worker, he offers to employ him at a fixed price w per unit of effective labor. As I explain below, in various models where employers choose their wages, they will in fact offer a fixed price per unit of effective labor in equilibrium. Alternatively, one can view this as an assumption that employers pay a piece rate, so workers earn in direct proportion to what they produce. If the worker accepts a job offer, his hourly wage would be

$$W_{it} = w\ell_{it} \tag{2.1}$$

I use an upper-case W to denote the hourly wage and a lower case w to denote the price per unit of effective labor. In the data, we will only get to observe hourly wages W_{it}.

Let $F_i(\cdot)$ denote the distribution of the price per unit labor w across all potential employers worker i could meet. Since the worker is assumed to search haphazardly, each new offer is an independent draw from $F_i(\cdot)$. All workers face the same distribution, i.e. $F_i(\cdot) = F(\cdot)$ for all i.[4] This assumption is not unreasonable if we limit attention to workers searching in broadly similar labor markets. However, since practical considerations ultimately require me to group together workers with different educational, racial, and geographic backgrounds, this assumption may be questionable in my actual application. Still, there is nothing that conceptually precludes us from implementing the approach outlined below separately for distinct groups.

Employed workers face a constant hazard λ_1, possibly different from λ_0, of encountering potential employers. Once again, they are offered a fixed price per unit of labor drawn from $F(\cdot)$.

Finally, employed workers face a constant hazard δ of losing their job. This rate is assumed to be independent of the wage on a worker's current job. Workers cannot recall offers they already turned down, so a worker who loses his job must resume searching from scratch.

Assuming the worker seeks to maximize the present discounted value of his earnings, his search problem is fairly simple. While unemployed, he should set a reservation price w^* and accept offers of at least w^* per unit of labor. While he is employed, he should trivially accept any offer that exceeds the price on his current job and turn down any offer below it. The optimal cutoff w^* depends on $F(\cdot)$ as well the parameters b, δ, λ_0, and λ_1. By assuming these parameters are the same for all workers, I ensure the cutoff w^* will be as well. All workers thus face the same essential search problem. Differences in ability only scale the price w paid by an employer, but the distribution of w a worker will accept is the same for all workers.

The focus of this paper is whether we can identify the offer distribution $F(\cdot)$ non-parametrically from hourly wage data $\{W_{it}\}$. As we shall see below, identification is trivial when ℓ_{it} is perfectly

[4] Assuming that all workers face the same offer distribution need not require that they prefer the same employers. For example, Marimon and Zilibotti (1999) and Barlevy (2002) consider models in which workers have a comparative advantage for certain jobs. Under the symmetry assumptions they impose, $F_i(\cdot)$ is the same for all i, but each worker prefers the particular job where his own comparative advantage lies.

observable (which implies w is perfectly observable as well). The more interesting question is whether we can identify $F(\cdot)$ when ℓ_{it} is imperfectly observable.

Before I turn to the question of identification, let me digress briefly on whether $F(\cdot)$ is really the object we ought to be interested in. In particular, certain applications in which we wish to estimate a search model require us to identify not an offer distribution but the economic fundamentals that shape it. For example, to simulate the effects of changes in policy, we need to know the underlying fundamentals to derive the equilibrium offer distribution under the new policy. However, I shall now argue that the environment above represents a reduced form of several equilibrium search models in which there is a one-to-one mapping between the equilibrium offer distribution $F(\cdot)$ and the fundamentals that shape it. Thus, identifying $F(\cdot)$ is equivalent to identifying the deeper structural parameters of interest for particular equilibrium models.

Consider the case where the economic fundamental is the distribution of productivity across employers. That is, employer j can produce z_j units of output per unit of effective labor, and $\Gamma(\cdot)$ is the cumulative distribution of z_j across all employers. One model of this type is Lucas and Prescott (1974). They assume workers search across locations, where each location contains many employers using the same technology. In equilibrium the worker must be paid his productivity, or else another firm in the same location would hire the worker away. Thus, the wage in location j is given by $W_{it} = z_j \ell_{it}$, confirming employers will offer a constant price per unit of effective labor in equilibrium. Since $\Gamma(z) = F(z)$ for all z, identification of one implies identification of the other.

Subsequent researchers have questioned Lucas and Prescott's assumption that workers can immediately take a job from other equally productive employers but must wait for offers from more productive employers. Instead, they assume workers must wait for any offer. In this case, it matters what the employer can promise the worker when they meet. One possibility is that employers cannot commit, so no matter what the employer promises, he can always renegotiate with the worker. Mortensen (1986) suggested resolving this negotiation through Nash bargaining. Shimer (2004) solves the Nash bargaining problem when on-the-job search is possible. His solution is consistent with employers offering to pay a fixed price per unit of effective labor. It also implies a one-to-one mapping from $\Gamma(\cdot)$ to $F(\cdot)$, so $\Gamma(\cdot)$ can be identified from $F(\cdot)$. Alternatively, employers might be able to commit to a fixed price per unit of labor, as in Burdett and Mortensen (1998). In this case, we again obtain a one-to-one mapping from $\Gamma(\cdot)$ to $F(\cdot)$, which Bontemps, Robin, and van den Berg (2000) derive explicitly. Under all of these alternative assumptions, then, we can recover the structural parameters of interest by applying the relevant inversion formula on the offer distribution $F(\cdot)$ we identify from wage data.

2.2. Identification with Observed Heterogeneity

Let us return to the question of whether it is possible to identify the offer distribution $F(\cdot)$ from hourly wage data $\{W_{it}\}$. I begin with the case in which labor productivity ℓ_{it} is perfectly observable. In this case, we can recover the price w from the hourly wage W_{it}. Not surprisingly, the fact that we can observe w allows us to easily identify $F(\cdot)$, a point previous authors have already noted. However, analyzing this case sheds some insight as to why we might still be able to identify the wage offer distribution even when we cannot recover w from W_{it}.

I begin my analysis by describing the data available for identification. The most common sources of wage data are surveys that follow workers over time and keep track of their work histories: the hourly wage paid on each job a worker held, how long each job lasted, why the job ended if it did, and so on. As noted earlier, these surveys typically ask about the jobs workers accept, not the offers they encounter. Hence, the only available data are the hourly wages W_{it} workers earned on jobs they were employed on, as well as data on how many jobs a worker held and the reason each job ended. Most of the papers cited in the Introduction also use data on how long workers were employed on each job, but we don't need this data to identify the offer distribution $F(\cdot)$.

Following Wolpin (1992), I partition the data for each worker into distinct *employment cycles*, where a cycle is defined as the time between forced layoffs. That is, a cycle begins when the worker is forced to leave a job, continues on through his unemployment and subsequent employment, and ends the next time he is forced out of a job. It is therefore important to distinguish between instances in which a worker is forced out of a job, i.e. involuntary job changes, and those in which the worker chooses to move upon meeting a higher paying employer, i.e. voluntary job changes. While voluntary and involuntary job changes have precise meanings in the model, distinguishing between them empirically raises some issues that I discuss in more detail later. We should index observations by their respective employment cycle, but I omit this subscript in what follows.

Within each employment cycle, the worker first spends some time unemployed, followed by a period of uninterrupted employment in one or more jobs. Let M_u denote the (random) number of offers he receives *before* the first offer he accepts. Thus, if the worker accepts his very first job offer, $M_u = 0$. It is easy to show that the number of offers until he accepts his first offer has a geometric distribution, namely $\text{Prob}(M_u = m) = F(w^*)^m (1 - F(w^*))$.

Similarly, let M denote the (random) number of offers he receives before he is laid off, *starting from the first offer he accepts*. Thus, $M = 1$ implies that the worker was laid off from the first job he accepted. As the next lemma illustrates, M is also geometrically distributed.

Lemma 1: The unconditional number of offers on a cycle M has a geometric distribution, i.e. $\text{Prob}(M = m) = (1-p)^{m-1}p$, where $p = \delta/(\lambda_1 + \delta)$. $\blacksquare$

The proof of this lemma and other results are contained in an Appendix. Let $m \in \{1, 2, ..., M\}$ index the offers the worker receives, and $\{X_m\}_{m=1}^{M}$ denote the list of prices per unit labor the worker encounters over an employment cycle, starting with the first offer he accepts. Define N as the (random) number of *actual jobs* the worker is employed on in a given cycle, so that $N \leq M$, and let $n \in \{1, 2, ..., N\}$ index these jobs. Finally, let $\{w_n\}_{n=1}^{N}$ denote the price per unit of labor on each of these jobs. The optimal search strategy for a worker implies that

$$w_n = X_{L_n}$$

i.e. the price per unit labor on the n-th job in the cycle is the n-th record from the sequence $\{X_m\}_{m=1}^{M}$, and N is the number of records in this sequence.

Let W_{it}^{n} denote the hourly wage of the i-th worker at time t who is on the n-th job in his cycle. Since ℓ_{it} is observable, we can divide W_{it}^{n} by ℓ_{it} to recover the price per unit labor w_n on his n-th job. Since the latter are just record statistics, identifying the wage offer distribution reduces to recovering the parent distribution from information on the record values in the sequence $\{X_m\}_{m=1}^{M}$. Recall that this is one of the problems statisticians have analyzed for the classical record model in which the number of observations M is infinite. By contrast, here the number of observations M is itself random, a case that has received less attention in the statistics literature.

Before proceeding, I should point out that since we never observe data below w^*, we couldn't possibly identify $F(\cdot)$ non-parametrically below this threshold.[5] All we can hope to identify is

$$F(w \mid w \geq w^*) = \frac{F(w) - F(w^*)}{1 - F(w^*)} \tag{2.2}$$

I therefore focus on the truncated distribution above. For some applications, this distribution suffices. Moreover, in some models, economic theory implies $F(w^*) = 0$, so the truncated distribution *is* the true offer distribution. In a slight abuse of terminology, I will interchangeably refer to identifying $F(\cdot)$ when I mean identifying $F(\cdot \mid w \geq w^*)$.

[5] One can potentially identify $F(\cdot)$ below w^* by imposing parametric assumptions. Flinn and Heckman (1982) derive conditions for when a given parametric functional form for $F(\cdot)$ is recoverable from data on $w \geq w^*$.

So, can we identify $F(\cdot \mid w \geq w^*)$ from hourly wage data? Since we can recover w, identification is straightforward. As Bontemps, Robin, and van den Berg (2000) observe, since the first job in an employment cycle is a random draw from $F(\cdot \mid w \geq w^*)$, wages on these jobs will be distributed as the offer distribution.[6] Translated into the language of records, this implies that the *first* record identifies the offer distribution. However, there is no need to appeal to the implicit record structure of wages to recognize the potential of wages on a first job for identification.

Nevertheless, acknowledging the record structure of wages, while not essential for identification when worker heterogeneity is observable, does provide insights that prove useful when we allow for unobserved heterogeneity. Using only the wages from the first job out of unemployment ignores potentially useful data. In particular, the wages on jobs beyond the first job in an employment cycle – which correspond to subsequent record statistics – are also useful for identifying the offer distribution. The next proposition implies that for *any* integer n, knowing the distribution of wages on just the n-th job in the cycle can allow us to identify the offer distribution:

Proposition 1: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^{M}$ where $\Pr(M = m) = (1-p)^{m-1}p$ for some $p \in (0,1)$. Let $\{R_n\}_{n=1}^{N}$ denote the records in this sequence. For any integer n, the distribution of X_m is uniquely determined in the class of continuous distribution functions by (1) the distribution of R_n given $N \geq n$; and (2) the distribution of the number of records N.

For $n > 1$, we need data not only on wages but also on the number of jobs workers hold in a typical employment cycle (the analog of the number of records N). To appreciate why we need this additional information, consider the distribution of wages on the second job in an employment cycle. We only observe these wages if a worker managed to switch into a second job before being forced out of a job. But if a worker was lucky enough to get a high offer on his first job, he is unlikely to find an even better job in time. Thus, workers who make it to a second job are more likely to be those who drew low offers on their first job. To correct for this selection, we need to know something about how many jobs workers pass through on a typical employment cycle.

More precisely, we need to know λ_1/δ, the rate at which workers meet employers relative to the rate at which they lose contact with them. How can we recover this ratio? Some of the papers cited in the Introduction identify λ_1 and δ from job duration data. They use the fact that the duration of a job that pays w is exponential with hazard $\lambda_1(1 - F(w)) + \delta$, so data on duration and wages

[6] More accurately, Bontemps *et al* argue that the wage of a worker on the first job we observe him on provides a non-parametric estimator of the steady-state wage distribution $G(\cdot)$, from which we can back out $F(\cdot)$. But the logic for using the wage on the first job we observe the worker on out of unemployment is identical.

can separately identify λ_1 and δ. But this hinges on a functional form for $F(\cdot)$, whereas we need the ratio λ_1/δ to identify $F(\cdot)$. Thus, the way previous authors estimated these parameters is not relevant for our purposes. However, if $F(\cdot)$ is continuous, the distribution of N allows us to determine p, which from Lemma 1 is monotonic in λ_1/δ:

Lemma 2: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^{M}$ where $\Pr(M = m) = (1-p)^{m-1} p$ for some $p \in (0,1)$. Let N denote the number of records in $\{X_m\}_{m=1}^{M}$. Then $\Pr(N = n) = \dfrac{p}{1-p}\dfrac{(-\ln p)^n}{n!}$, which implies that the distribution of N identifies p.

To summarize, for any integer n, we can uncover the offer distribution from the distribution of wages on the n-th job of their employment cycle in two steps. First, we use mobility data to identify p. Given p, we can then recover the offer distribution from wage data.

The implication of Proposition 1 is that when worker quality is observable, the wage offer distribution is overidentified. This suggests a way to test the validity of the underlying search model. That is, we can always construct an empirical distribution of wages on the first job out of unemployment, even if workers are not searching optimally from a fixed offer distribution. But to the extent that the model is true, wages on different jobs in an employment cycle should consistently reveal the same offer distribution. More importantly, though, the proposition reveals that the offer distribution uniquely determines the evolution of wages over an employment cycle. Thus, the wage growth of workers over the employment cycle contains revealing information about the wage offer distribution. This additional information is redundant when worker quality is observable, but as I show in the next section, it proves essential for identification when worker quality is unobservable.

2.3. Identification with Unobserved Heterogeneity

Given the inherent difficulty of measuring a worker's true productivity, I now allow ℓ_{it} to be unobservable. In this case, the wages of workers on their first job no longer uniquely identify the offer distribution; without any information on ℓ_{it}, it is impossible to tell if variation in the wages on the first job out of unemployment W_{it}^1 is due to variation in the prices w_1 across employers or ability ℓ_{it} across workers. Formally, the distribution of hourly wages is a convolution of prices and ability, and without further restrictions there is no unique way to deconvolute these terms and identify $F(\cdot \mid w \geq w^*)$.

Of course, if we impose enough structure on the unobservable component, we might be able to uncover ℓ_{it} even when it is not directly observable. For example, suppose each worker's productivity were constant over time, i.e. $\ell_{it} = \ell_i$ for all t. Observing workers over multiple employment cycles would allow us to infer their relative productivities, since more productive workers consistently earn higher wages. Once we know which workers are more productive, we can back out w and use wages on the first job out of unemployment to recover the offer distribution. In fact, using a result from Kotlarski (1967), we only need to observe each worker on just two employment cycles. Let w_1' and w_1'' denote the price per unit labor on the first job in the first and second employment cycles, respectively. Kotlarski's theorem states that under certain regularity conditions, given any three independent variables w_1', w_1'', and ℓ_i, the joint distribution of $(w_1'\ell_i, w_1''\ell_i)$ identifies the distribution of all three variables up to a scale parameter.

The problem with imposing assumptions on unobserved ability this way is that such assumptions are impossible to verify (although one might be able to rule them out; e.g. the assumption that ability is fixed over time is inconsistent with the fact that wages vary over the duration of a job). A more satisfying approach would be to determine whether the offer distribution can be identified even under minimal assumptions on ℓ_{it}. This is the approach I pursue.

Consider the following specification for ℓ_{it}, based on Flinn (1986), which allows for both observable and unobservable variation in worker ability:

$$\ell_{it} = \exp\left(\beta Z_{it} + \phi_i + \varepsilon_{it}\right) \tag{2.3}$$

The first term, Z_{it}, represents observable characteristics for individual i that affect his productivity, and β represents the returns to these characteristics. The next term, ϕ_i, is fixed over time, reflecting variations in innate ability that make some workers consistently more productive than others. I do not require this term to be observable. The last term, ε_{it}, denotes unobserved variation in productivity, as well as multiplicative measurement error in reported wages.

In what follows, I consider changes in wages at regularly-spaced intervals, e.g. one year apart, denoted $\Delta \ln W_{it}$. Differencing wages has the virtue of eliminating the fixed effect term ϕ_i. Some of my assumptions on ℓ_{it} involve differences of variables rather than the variables themselves. In particular, I impose the following three assumptions:

Assumption 1: ℓ_{it} is independent of job-specific characteristics

Assumption 2: ΔZ_{it} is independent of $\Delta\varepsilon_{it}$

Assumption 3: $E\left[\Delta\varepsilon_{it}\right] = 0$

The first assumption states that the choice of employer has no effect on the worker's productivity. This insures a worker will accept any offer that pays more per unit of labor than his current job. It also implies any human capital the worker accumulates must be general in nature, since it cannot be specific to any one employer. An important part of my empirical work will be to confirm that job-specific human capital is indeed negligible in my sample.

The second assumption states that growth in observable and unobservable worker productivity are independent. This insures we can consistently estimate the returns to observable characteristics from wage data. Since the only observable characteristic in my empirical application is potential experience, which evolves deterministically, this assumption seems plausible.

The final assumption states that ε_{it} should not grow on average over the time interval we consider. This assumption is essentially without loss of generality, since we can always include intercepts in ΔZ_{it} to capture growth in ε_{it}. Intuitively, if ε_{it} grows systematically over time, we could infer this from workers who remain on the same job, so such growth is essentially observable. The fact that ε_{it} is a martingale at the relevant time horizon imposes very minimal restrictions on earnings. For example, it allows for serial correlation in wages over the duration of a job, including the case where ε_{it} is non-stationary. Likewise, the variance of ε_{it} can vary arbitrarily over time and across individuals, and each individual's productivity can follow a different stochastic process.

Given such weak assumptions, it will be impossible to uncover ℓ_{it} from data on W_{it}. Nevertheless, I now show that by appealing to the underlying record structure of the search model, we can still identify the distribution of prices w workers face. Define $\omega_n = \ln w_n$ as the log price per unit labor on the worker's n-th job, so that ω_n represents the n-th record in the sequence of log price offers $\{x_m\}_{m=1}^{M}$ where $x_m = \ln X_m$. After substituting in for ℓ_{it}, we obtain the following equation for the log hourly wage:

$$\ln W_{it}^n = \omega_n + \beta Z_{it} + \phi_i + \varepsilon_{it} \tag{2.4}$$

We next first-difference equation (2.4) to get rid of the fixed effect term ϕ_i:

$$\Delta \ln W_{it}^n = \Delta\omega + \beta\Delta Z_{it} + \Delta\varepsilon_{it} \tag{2.5}$$

For a worker employed on the same job at these two points in time, $\Delta\omega = 0$, so that

$$\Delta \ln W_{it}^n = \beta\Delta Z_{it} + \Delta\varepsilon_{it} \tag{2.6}$$

Since ΔZ_{it} and $\Delta\varepsilon_{it}$ are assumed to be independent, we can estimate (2.6) by ordinary least squares, i.e. we can estimate the contribution of observable characteristics to productivity growth.

Next, using our estimate for β, we can net out the role of observable productivity growth for workers who change jobs voluntarily. 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

$$\Delta \ln W_{it}^{n} - \beta \Delta Z_{it} = (\omega_n - \omega_{n-1}) + \Delta \varepsilon_{it} \tag{2.7}$$

The net wage gain for a voluntary job changer who leaves his n-th job is thus the sum of a noise term $\Delta\varepsilon_{it}$ and the gap between the n-th and $n+1$-th records among i.i.d. draws from the log offer distribution. Since I imposed no assumptions on $\Delta\varepsilon_{it}$ other than that its mean, we still face a deconvolution problem in recovering the distribution of the record gap $\Delta\omega$ from data on hourly wages. However, since $\Delta\varepsilon_{it}$ has zero mean, we can recover *expected* record gaps. That is, averaging the net wage gains for workers who move from their $n-1$-th to their n-th job, we obtain

$$\begin{aligned} E(\Delta \ln W_{it}^{n} - \beta \Delta Z_{it} \mid N \geq n) &= E\left(\omega_n - \omega_{n-1} \mid N \geq n\right) + E(\Delta\varepsilon_{it} \mid N \geq n) \\ &= E\left(\omega_n - \omega_{n-1} \mid N \geq n\right) \end{aligned}$$

where the fact that $E(\Delta\varepsilon_{it} \mid N \geq n) = 0$ follows from the assumption that $\Delta\varepsilon_{it}$ is independent of job characteristics. I shall now argue that the sequence of expected record gaps

$$\left\{E\left(R_{n+1} - R_n \mid N \geq n+1\right)\right\}_{n=1}^{\infty} \tag{2.8}$$

from an i.i.d. sequence $\{X_m\}_{m=1}^{M}$ uniquely characterizes the parent distribution of each X_m. I first need to provide conditions under which this sequence of moments exists.

Lemma 3: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^{M}$ where $\Pr(M=m) = (1-p)^{m-1} p$ for some $p \in (0,1)$. Let $\{R_n\}_{n=1}^{N}$ denote the records of this sequence. If $E\left(|X_m|\right) < \infty$, then the conditional expectation $E\left(R_{n+1} - R_n \mid N \geq n+1\right)$ is finite for $n = 1, 2, 3, ...$

Thus, we need to assume that the offer distribution has a finite mean. Under this assumption, given a value of p, which recall we can back out from the distribution of N, we can identify the shape of the wage offer distribution from the sequence in (2.8):[7]

Proposition 2: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^{M}$ where $\Pr(M=m) = (1-p)^{m-1} p$ for some $p \in (0,1)$. If $E\left(|X_m|\right) < \infty$, the sequence

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

characterizes the distribution of X_m in the set of continuous distributions, up to a location shift.

Remark: Gupta (1984), building on Kirmani and Beg (1984), shows that when $\Pr(M=\infty) = 1$, the sequence $\left\{E\left(R_{n+1} - R_n\right)\right\}_{n=1}^{\infty}$ uniquely characterizes the parent distribution up to a location

[7] I am indebted to H. N. Nagaraja for his assistance with the proof of this proposition.

shift. Since in the classical record model $N = \infty$ with probability one, there is no need to condition on there being at least n records. But when $\Pr(M = \infty) < 1$, as in our case, the expected value of the n-th record must be conditioned on there being at least n records among the sequence of offers the worker encounters. This conditioning is non-trivial, and we cannot simply extend Gupta's result to the present setting. In a related paper to the present one, Nagaraja and Barlevy (2003) provide a more rigorous analysis of record moments when the number of observations M has a geometric distribution. They show that characterization results based on record moments from a geometric number of observations are stronger than those from an infinite number of observations, i.e. moment sequences that are not enough to uniquely identify the parent distribution when M is infinite can identify the parent distribution when M has a geometric distribution.

Proposition 2 implies that the average wage gains of voluntary job changers (net of returns to experience) identify the distribution of log wage offers $x_m = \ln X_m$ up to a location parameter. We can then recover the offer distribution in levels up to a scale parameter. The average wage growth of voluntary job changers provides us enough information to identify the shape of the offer distribution.

What is the intuition for the above identification result? Recall from my discussion of the case of observable heterogeneity that the offer distribution uniquely determines the evolution of wages over an employment cycle. Thus, looking at the extent to which wages grow with job mobility yields a great deal of information on the underlying offer distribution over which workers are searching. A more technical explanation is that for any random variable X, we can always uncover its distribution by tracing out $E(X \mid X > x) - x$ for all values of x. The n-th average record gap $E(R_{n+1} - R_n \mid N > n)$ is just a weighted average of $E(X \mid X > x) - x$ over all values of x, which puts more weight on low values of x for low values of n and more weight on high values of x for high values of n. Tracing the way in which the average record gap varies with n provides as much information as tracing the way $E(X \mid X > x) - x$ varies with x.

2.4. Applying Identification to Estimate the Offer Distribution

Proposition 2 establishes that the average net wage gains of voluntary job changers at different points in an employment cycle can identify the wage offer distribution. But to make practical use of this result, i.e. to obtain an actual estimate of the offer distribution, we need to be able to map these averages into a distribution function. In a previous version of this paper, I explicitly solve the inversion problem of how to construct the parent distribution from the infinite set of expected record gaps $\{E(R_{n+1} - R_n \mid N \geq n + 1)\}_{n=1}^{\infty}$. I also derive a consistent estimator for this distribution based on a finite number of moments that converges to the true parent distribution as the number of employment cycles goes to infinity and with it the number of moments one can estimate. However, this estimator is too noisy for the sample sizes I use, since I can estimate only a small number of moments with great precision.

Nevertheless, even a small number of moments allows us to test whether certain functional forms are consistent with the data. As an illustration, Figure 1 displays the expected record gaps $E\left(R_{n+1} - R_n \mid N \geq n+1\right)$ for two different distributions, an exponential and a normal (which in inverse logs correspond to Pareto and lognormal distributions, respectively). The moments are computed assuming M has a geometric distribution consistent with my estimates in Section 4, and both distributions are normalized to yield the same average log wage gain across voluntary job changers as we observe in the data. 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, reflecting its logconcave shape. In my empirical work, I will focus on this implication rather than try to construct a non-parametric estimator for $F(\cdot)$.

3. Data

To apply the insights above, I need a dataset with detailed work-history data to assign n to jobs. Moreover, since job mobility is highest when workers enter the labor market, it seems wise to focus on young workers. In addition, my assumption that all human capital is general is more likely to be true for younger workers, whose high mobility should make investment in job-specific skills less attractive. These considerations led me to the National Longitudinal Survey of Youth (NLSY) dataset. The NLSY tracks a 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, when the oldest worker in the sample was 36. Each year, respondents were asked questions about the jobs they held since their previous interview, including starting and stopping dates, the wage paid, and the reason for leaving. 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 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 outliers, 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 they are far out of line with what the same workers report at other dates, including for the same job. For my measure of potential experience, I follow previous work in dating entry into the labor market at the worker's birthyear plus 6 plus his reported years of schooling (highest grade completed). If an individual reported working prior to that year, I date his entry at the year in which he reports his first job. Table 1 provides summary statistics across all jobs.

The one new variable I use is the position n of each job in its respective employment cycle. First, I need to partition the data into employment cycles, using the occurrence of involuntary job changes as break points. To identify these occurrences, I could use the worker's response on whether he quit voluntarily or was laid off. Alternatively, the model implies involuntary job changes will be followed by an unemployment spell, so I could classify job changes in which the worker spent some time not working between jobs as involuntary changes. In the model, these measures coincide. But in the data they agree only 60% of the time. More precisely, workers who report a layoff do seem to spend at least one week without a job, and workers who directly move into their next job without a spell of unemployment do often report quitting their job. However, nearly half of all workers who reported quitting did not start their next job until weeks or even months later. Some of these delays may be planned; for example, a teacher who leaves to work for another school would likely spend two months in the summer not working; likewise, a worker may use up vacation days when he leaves an employer, but report leaving his job on the day he started his vacation. Yet in many of these instances the worker probably resumed searching from scratch after quitting, e.g. because he quit to avoid being laid off or he was embarrassed to admit he was laid off. As a compromise, I use the worker's stated reason for leaving his job as long as he starts his next job within 8 weeks of when his previous job ended, but treat him as an involuntarily job changer regardless of his stated reason if he does not start his next job until more than 8 weeks later. 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 is he starts it after two months, but otherwise do not classify the job. I experimented with cutoffs other than eight weeks. These had very little impact on the first few record moments (i.e. $n = 1$, 2, and 3), although they did affect my estimates for higher values of n where sample sizes were already small.

Next, I assign all jobs within each employment cycle a value of n. That is, I set $n = 1$ after the first involuntary job change I observe for a person, so a worker must experience at least one involuntary job change before I can start assigning values for n. From then on, I increment n by 1 whenever the worker changes jobs voluntarily, until the employment cycle ends and n is reset to 1 at the start of the next cycle. 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 can work on only one job full time, but they can receive additional draws from the distribution $F(\cdot)$ and work on those part-time. Thus, 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 he then 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, so we can ignore it. Alternatively, if the worker leaves job A and remains in job B, 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 offer distribution, we can treat it the same way as a new job that started only after job A ended, whether job A ended voluntarily or not.

Out of the 52,827 distinct jobs in my original sample, the procedure above identifies 8,234 as secondary jobs. As a check, 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 by the worker as his primary job, and only 9% are ranked as the primary job each year the job is reported.

Figure 2 displays 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 could not be assigned. Since we can only assign n following the first involuntary job change, this fraction is small in the first few years of the sample when workers experienced a limited amount of mobility. By 1993, though, I could assign a value of n to 87% of all the jobs reported. 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 cross-sectional 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 make use of this data, a few issues need to be settled. First, I need to decide the horizon at which to compute the differences in equation (2.7). Since the NLSY only asks for one wage per job per interview, I 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. My constructed sample consists of 40,370 observations in which the worker reports a wage 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 X_{it}, i.e. the time from when worker i entered the labor market up to date t:

$$Z_{it} = \beta_1 X_{it} + \beta_2 X_{it}^2 \tag{3.1}$$

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

$$\Delta Z_{it} \equiv Z_{it} - Z_{i,t-1} = \beta_1 + \beta_2 (2X_{it} - 1)$$

Assuming that potential experience is the only observable worker characteristic is faithful to my assumption that Z_{it} is independent of any job-specific characteristics. To assess the plausibility of this assumption, I also consider the possibility that the worker's ability depends on certain job-specific characteristics, specifically the time the worker has spent working for his current employer. This measure can be viewed as a proxy for the amount of job-specific human capital the worker could have accumulated. Let T_{it} denote the tenure of worker i on the job he holds at date t, and let us amend (3.1) to include T_{it}:

$$Z_{it} = \beta_1 X_{it} + \beta_2 X_{it}^2 + \gamma T_{it} \tag{3.2}$$

I will consider higher-order terms in T_{it} in my empirical implementation below, but for notational simplicity it will be easier to proceed as if returns to tenure are linear. Evidence that γ is different from zero would invalidate my identification results from the previous section. In particular, under (3.2) one can show that optimal search will imply that the sequence of prices $\{w_n\}_{n=1}^N$ correspond not to simple records as defined in Section 1 but to records in which an observation is counted as a record if it beats the previous record by some (random) threshold, enough to compensate the worker for the returns to tenure he loses when changing jobs. Although wages still correspond to records in an appropriately defined sense, the proofs of the various propositions in the previous section no longer apply (although this doesn't rule out that analogous identification results could be obtained by appealing to a different argument). For the analysis above to be relevant, we need to establish returns to tenure are in fact small in my sample.

Previous authors have already tackled the question of how to estimate the returns to tenure from wage data, i.e. to uncover γ from the wage equation

$$\ln W_{it}^n = \omega_n + \phi_i + \beta_1 X_{it} + \beta_2 X_{it}^2 + \gamma T_{it} + \varepsilon_{it} \tag{3.3}$$

Since the unobserved log price per unit labor ω_n is likely to be correlated with T_{it} – for example, workers are more likely to remain on a job that pays a relatively high price – ordinary least squares will yield a biased estimate for γ. Altonji and Shakotko (1987) proposed an instrumental variables approach for estimating γ, which yielded small values for γ. Topel (1991) proposed a two-step estimator that yielded fairly large returns to tenure. Altonji and Williams (1997) critique Topel's implementation, but even after they take their critiques into account, they find that his approach yields somewhat larger estimates for the returns to tenure than the original Altonji and Shakotko estimates. To bias against finding small returns to tenure, I focus on Topel's approach. However, since my sample consists of much younger workers than in Topel's sample, my results may not be comparable to his.

Topel's approach uses the fact that $X_{it} = X_{0it} + T_{it}$, where X_{0it} is the worker's experience when he started working on the job he holds at date t. Substituting this into (3.2), we have

$$\ln W_{it}^n = \omega_n + \phi_i + \beta_1 X_{0it} + \beta_2 X_{it}^2 + (\beta_1 + \gamma) T_{it} + \varepsilon_{it} \tag{3.4}$$

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

$$\Delta \ln W_{it}^n = (\beta_1 + \gamma) + \beta_2 (2X_{it} - 1) + \Delta\varepsilon_{it}$$

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

$$\ln W_{it}^n - (\beta_1 + \gamma) T_{it} - \beta_2 X_{it}^2 = \omega_n + \phi_i + \beta_1 X_{0it} + \varepsilon_{it}$$

We regress the difference on the left-hand side on X_{0it} and individual fixed effects to arrive at an estimate for β_1, adjusting the standard errors to take into account first-stage estimation error. The estimate for γ is just the difference between the estimates for $\beta_1 + \gamma$ and β_1.

Table 2 reports the results of this two-step procedure for my dataset. The point estimates 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 quite small. 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 quadratic returns to tenure, i.e. $\gamma_1 T_{it} + \gamma_2 T_{it}^2$. The estimated returns remain small; although not reported, returns to tenure attain a maximum of only 0.0433 log points at 5 years with the same employer and decline from that point on.

Interestingly, the returns to potential experience in Table 2 are consistent with those found in Altonji and Shakotko (1987), Topel (1991), and Altonji and Williams (1997, 1998). Topel's point estimate for β_1 of 0.0713 in particular is close to mine. The reason I find such small returns to tenure is that wage growth on the job in my sample is smaller than in Topel's sample; whereas he estimated $\beta_1 + \gamma$ at 0.1258, my estimate is only 0.0794. That is, on-the-job wage growth among young workers is not much larger than the consensus estimates for the returns to experience that are found in the literature, leaving little room for wage to grow with tenure.

While my point estimate for γ is small, Topel himself observed that it is likely to be biased downwards given that the estimate for β_1 is biased upwards. The source of the bias in estimating β_1 is that workers with more experience have had more time to search for better matches, so initial experience X_{0it} will be positively correlated with n and thus ω_n. However, this bias is likely to be small given the high incidence of involuntary job loss in my sample, which weakens the correlation between X_{0it} and n. Moreover, when I revisit the question of how large returns to tenure are in Section 5, I find additional evidence that γ is small.

4. Empirical Results

Having described the variables I use in my analysis, I can now estimate the average record gaps implied by the wage growth of voluntary job changers. While I would have liked to estimate the offer distribution separately for distinct worker groups, e.g. blacks and whites or high-school and college graduates, the number of observations in my sample is sufficiently small that I am forced to group all workers together and assume they face a common offer distribution.

Recall that the first step in identification is to recover the parameter $p = \delta / (\lambda_1 + \delta)$. Let K_n denote the number of employment cycles with exactly n records. From Lemma 2, the maximum likelihood estimator for p is given by

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

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 the rate of involuntary job loss is high relative to the rate at which workers encounter offers, i.e. p should be relatively high. The maximum likelihood estimates for p are reported in Table 3. I estimate p at 0.48, implying $\lambda_1/\delta \approx 1$. To check whether grouping workers together overlooks important differences across 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 value of 10 reported in some of the papers cited in the Introduction that estimate δ and λ_1 from duration data as opposed to mobility data, including some that use the same NLSY dataset. However, it agrees with Bowlus, Keifer, and Neumann (2001), who also use duration data and estimate $\lambda_1/\delta \approx 1$.

Next, I estimate the average wage gains in (2.7). Once again, to mitigate the effect of outliers, I eliminated the extreme 1% of my sample for which $|\Delta \ln W_{it}|$ was largest. Most of these deletions appear to be due to coding errors, since nearly all were followed by equally large wage changes in the opposite direction in the subsequent year. 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 that 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, I combine job stayers and job changers into a single regression

$$\Delta \ln W_{it}^n = \beta_1 \Delta X_{it} + \beta_2 \Delta X_{it}^2 + \sum_{n=1}^{\infty} \pi_n D_{it}^{n,n+1} + \Delta \varepsilon_{it} \tag{4.1}$$

The coefficients π_n are unbiased estimates of the expected moment gaps $E\left(R_{n+1} - R_n \mid N > n\right)$. Combining the two stages allows the wage growth of job changers to help in identifying the coefficient β_2, and should therefore be more efficient. Note that the variance of the residual will be different for job stayers and job changers, since the residual for the latter also contains deviations of $\omega_{n+1} - \omega_n$ from its average. I therefore report only robust (White) standard errors.

The results of this regression are reported in Table 4. The number of workers who are observed to change from the n-th job in the previous year to the $n+1$-th job this year is reported for each n next to the corresponding dummy variable. The estimated coefficients in (4.1) are reported in the second column. The first column in the table reports the estimates for 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 estimates for π_n are all clustered around 8%, with the exception of π_4. However, this coefficient (as well as the coefficient π_5) is rather imprecisely estimated given the small number of job changers for this value of n. The large standard errors in Table 4 illustrate the difficulty of further dividing this sample by education or race.

Given that we can only estimate a small number of moments very precisely, a fully non-parametric estimator for $F(\cdot)$ based on π_n is likely to be very noisy. Clearly, we would need many more employment cycles to come up with a reliable estimator. However, as noted earlier, we can still test particular candidate distribution functions. Recall that the offer distribution is Pareto – and consequently the log offer distribution is exponential – if and only if the coefficients π_n are constant for all n. Thus, testing whether the coefficients π_n in (4.1) are equal for all n is equivalent to testing whether the offer distribution is Pareto. Note that this is a test of a general shape restriction, i.e. it tests whether the offer distribution is Pareto rather than whether it is a Pareto with a particular parameter value. To the extent that we fail to reject that the π_n are equal, we can estimate the exact Pareto distribution from (4.1) restricting all π_n to be equal.

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 this degree of variation in wage gains under the null that they are all the same equals 0.264. We thus fail to reject the null that the wage offer distribution is Pareto at conventional significance levels. The third column of Table 4 estimates (4.1) imposing that π_n are all equal. The average net wage growth from voluntarily moving jobs is 0.0806, in line with the average wage growth for young workers reported in Topel and Ward (1992). Under the null of a Pareto offer distribution, this value represents the inverse hazard rate of the implied exponential log offer distribution. Flinn (2002) also estimated a Pareto offer distribution using the same NLSY data, but he finds an inverse hazard of 0.2400 (Table 4, p633). The discrepancy arises because Flinn abstracts from on-the-job wage changes and attributes any growth between the starting wage on the n-th job in the cycle and the starting wage on the $n+1$-th job in the cycle to a better price from the underlying offer distribution. Using my estimates for the returns to experience, workers would have to spend about two years on a job to reconcile this discrepancy. This is a little larger than the average tenure of workers on their first job in the NLSY (which Flinn uses in his estimation), but it is certainly within reason.

While we fail to reject a Pareto offer distribution, the second row in the bottom panel of Table 4 reveals that we can reject the hypothesis that the offer distribution is lognormal. In particular, if log wage offers were distributed as $N\left(\mu, \sigma^2\right)$, the average net wage growth among workers who move from their n-th job to their $n+1$-th job would equal

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

where R'_n denotes the n-th record from the sequence $\{X'_m\}_{m=1}^M$ where X'_m are i.i.d. standard normals. Thus, if the wage offer distribution is lognormal, the sequence $\{\pi_n\}_{n=1}^\infty$ will be proportional to $\left\{E\left(R'_{n+1} - R'_n \mid N > n\right)\right\}_{n=1}^\infty$. Using my estimate for p from Table 3, we can readily compute the latter sequence. Table 4 shows we can reject the null of a lognormal offer distribution at almost a 1% significance level. This calculation does not incorporate uncertainty in our estimate for p, but since the moment sequence in (4.2) is sharply declining for a various p, and since p is tightly estimated, the rejection of the normal is likely to be robust to incorporating sampling error.[8] The intuition comes from Figure 1; if the distribution were lognormal, wage gains would decline sharply with n. We can likewise reject other functional forms that imply similarly sharp declines.

5. Involuntary Job Changers and Specific Human Capital

So far, I have focused exclusively on the wage gains of voluntary job changers. Yet the wage losses of involuntary job changers also contain useful information. Consider a worker who is forced out of his n-th job. The total number of jobs in his last employment cycle is n, implying the price per unit labor on his previous job will on average equal $E(R_n \mid N = n)$, the expected value of the n-th record conditional on exactly n records in the sequence $\{X_m\}_{m=1}^{M}$. Similarly, the price per unit labor on his new job will, on average, equal $E(R_1 \mid N \geq 1)$, the expected value of the first record conditional on at least one record. Since every employment cycle has a first record, this is just $E(R_1)$. Hence, the average net wage loss for this worker is given by

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

Adapting results in Nagaraja and Barlevy (2003), one can show that the sequence of moments $\{E(R_n \mid N = n) - E(R_1)\}_{n=1}^{\infty}$ identifies the parent distribution among continuous distribution functions up to a location parameter. That is, the offer distribution uniquely determines not just the wage gains of voluntary job changers but also the wage losses of involuntary job changers.

As in the previous section, we can recover the relevant differences between record moments with a single wage regression. 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 first job in date t. Then the coefficients π_n in the regression

$$\Delta \ln W_{it}^{n} = \beta_1 \Delta X_{it} + \beta_2 \Delta X_{it}^{2} - \sum_{n=1}^{\infty} \pi_n D_{it}^{n,1} + \Delta \varepsilon_{it} \tag{5.1}$$

are unbiased estimators for $E(R_n \mid N = n) - E(R_1)$. The first column in Table 5 reports my estimates of π_n. According to the model, π_n should be monotonically increasing in n. It indeed rises with n between $n = 1$ and 4, although the point estimate for π_5 falls below that of π_4.

[8] Note that under the null hypothesis that the log wage offer distribution is exponential distribution, the π_n would not depend on p, so there is no need to adjust for sampling error in our estimate for p.

Once again, I can test whether the offer distribution has a Pareto shape. Specifically, the offer distribution is Pareto if and only if $\{\pi_n\}_{n=1}^{\infty}$ is proportional to $\{E(R'_n \mid N=n) - E(R'_1)\}_{n=1}^{\infty}$, where R'_n denotes the n-th record from a sequence of standard exponentials (with mean 1). Setting $p = 0.48$ in line with Table 3, I numerically compute $E(R'_n \mid N=n) - E(R'_1)$ to be

$$\{0.197,\ 0.762,\ 1.127,\ 1.396,\ 1.616, ...\} \tag{5.2}$$

The bottom panel of Table 5 reports the probability of observing deviations from this proportionality condition at least as large as those in the data under the null of a Pareto offer distribution. Once again, we fail to reject the null hypothesis. In the second column of Table 5, I estimate (5.1) under the constraint that π_n is proportional to (5.2); the constant of proportionality corresponds to the inverse hazard of the implied exponential log offer distribution. I estimate this inverse hazard to equal 0.0816. By comparison, the wage gains of voluntary job changers imply an inverse hazard of 0.0806. The wage losses of involuntary job changers and the wage gains of voluntary job changers thus consistently identify the same offer distribution as required by the model.

Although data on voluntary job changers reject the lognormal specification, I did check if the wage losses of involuntary job changers provide additional evidence against this functional form. The offer distribution is lognormal if and only if $\{\pi_n\}_{n=1}^{\infty}$ is proportional to the sequence $\{E(R'_n \mid N=n) - E(R'_1)\}_{n=1}^{\infty}$, where R'_n denotes the n-th record from a sequence of standard normals. As the bottom row of Table 5 reveals, in contrast to the data for voluntary job changers, we cannot reject this hypothesis using involuntary job changers. The reason for this is illustrated in Figure 3, which shows the estimated net wage loss together with the best-fitting values for $E(R_n \mid N=n) - E(R_1)$ assuming a normal and exponential distribution respectively. Although the two sequences are distinct, it is difficult to distinguish them empirically given both are increasing and concave. By contrast, the implied average wage gains of voluntary job changers are sufficiently different for these two distributions that they can be easily distinguished.

To recap, the wage losses of involuntary job changers do not help to narrow down the set of functional forms for the offer distribution beyond what we learn from voluntary job changers. However, they do allow us to test a particular overidentifying restriction implied by the model. Specifically, the model predicts that the average wage losses of involuntary job changers should equal record moments from i.i.d. observations whose parent distribution is the offer distribution we identify from voluntary job changers, and the data are consistent with this prediction.

There are different ways to interpret this consistency between the wage growth of voluntary job changers and the wage losses of involuntary job changers. On the one hand, if we take as given that returns to specific human capital are negligible, comparing gains and losses allows us to test whether workers really search from a fixed offer distribution without recall. Conversely, if we instead take as given that workers do in fact search from a fixed offer distribution, we can use wage losses to test whether returns to specific human capital are indeed as small as suggested by the evidence in Section 3. The latter interpretation is particularly intriguing, since it offers a new way to tackle an old question in labor economics, namely whether wages rise with seniority. The remainder of this section develops this idea.

The intuition behind my approach to identifying returns to seniority is as follows. When a worker loses his job, he loses both the human capital that was specific to his last job and the returns to previous on-the-job search. We can use the moments of record statistics to directly account for the latter. Any remaining losses must then be due to specific human capital, from which we can infer a value for the returns to seniority γ.

Formally, suppose the worker's productivity ℓ_{it} is linear in the time spent with his current employer as in (3.2). The implied wage change associated with an involuntary job change is

$$\Delta \ln W_{it} = \omega_n - \omega_1 + \beta_1 + \beta_2 \Delta X_{it}^2 - \gamma T_{i,t-1} + \Delta\varepsilon_{it} \tag{5.3}$$

At the same time, the wage change for workers who remain on the same job is given by

$$\Delta \ln W_{it} = (\beta_1 + \gamma) + \beta_2 \Delta X_{it}^2 + \Delta\varepsilon_{it} \equiv \beta \Delta Z_{it} + \Delta\varepsilon_{it}$$

Averaging the wage losses of involuntary job changers net of the expected wage growth of job stayers $\beta\Delta Z_{it}$ yields

$$\begin{aligned} E\left[\left|\Delta \ln W_{it} - \beta\Delta Z_{it}\right| \mid N_{t-1} = 1\right] = E\left(\omega_n \mid N = n\right) - E\left(\omega_1\right) \\ + \gamma E\left(T_{i,t-1} + 1 \mid N_{t-1} = n\right) \end{aligned} \tag{5.4}$$

Knowing that workers search from a fixed offer distribution, we could compute $E\left(\omega_n \mid N = n\right) - E\left(\omega_1\right)$ and proceed to estimate γ. However, recall that in the presence of specific human capital, ω_n no longer represent simple record statistics; a worker will only switch jobs if the price on his new job exceeds the price on his previous job by enough to compensate him for the returns to tenure on his old job. To compute $E\left(\omega_n \mid N = n\right) - E\left(\omega_1\right)$ would therefore require us to know γ. But this is precisely the parameter we wish to estimate. Rather than compute $E\left(\omega_n \mid N = n\right) - E\left(\omega_1\right)$, then, I derive a bound for this term using moments of ordinary record statistics. While this does not allow me to estimate γ, it does provide an upper bound on its value.

I begin with workers who are laid off from their very first job in an employment cycle. Let ω_1' denote the log price per unit labor on the first (and only) job in his first employment cycle, and let ω_1'' denote the log price per unit labor on the first job in his second employment cycle. Since ω_1'' is just a random draw from the truncated offer distribution, it is identical to the first record statistic from a sequence of i.i.d. with the offer distribution as the parent distribution. Hence, $E(\omega_1'') = E(R_1)$. As for $E(\omega_1' \mid N_{t-1} = 1)$, I show in the Appendix that

$$E\left(\omega_1' \mid N = 1\right) \geq E\left(\omega_1'\right) = E(R_1) \tag{5.5}$$

The average net wage loss for a worker laid off from his first job thus satisfies

$$\begin{aligned} E\left(\left|\Delta \ln W_{it} - \beta \Delta Z_{it}\right| \mid N_{t-1} = 1\right) &= E\left(\omega_1' \mid N = 1\right) - E\left(\omega_1''\right) + \gamma E\left(T_{i,t-1} + 1 \mid N_{t-1} = 1\right) \\ &\geq E(R_1) - E(R_1) + \gamma E\left(T_{i,t-1} + 1 \mid N_{t-1} = 1\right) \\ &= \gamma E\left(T_{i,t-1} + 1 \mid N_{t-1} = 1\right) \end{aligned}$$

Rearranging yields the following upper bound on the returns to tenure γ:

$$\gamma \leq \frac{E\left(\left|\Delta \ln W_{it} - \widehat{\beta} \Delta Z_{it}\right| \,\Big|\, N_{t-1} = 1\right)}{E\left(T_{i,t-1} + 1 \mid N_{t-1} = 1\right)} \tag{5.6}$$

In other words, given that $E(\omega_1' \mid N = 1) - E(\omega_1'')$ is nonnegative, the average wage loss of workers who are laid off from their first job provides an upper bound on the average returns to tenure for these workers. When returns to tenure are linear, this allows us to estimate an upper bound on the returns to tenure γ. Note that this bound will be true for any fixed offer distribution, so we do not need to identify the offer distribution to derive it.[9]

To estimate the upper bound in (5.6), note that the coefficient π_1 in (5.1) is an unbiased estimator for the numerator in (5.6). Let $\overline{T}_1$ denote the average tenure in the sample for workers who were laid off from their first job. Then $\overline{T}_1 + 1$ forms an unbiased estimator for the denominator in (5.6). Thus, a natural estimator for the upper bound in (5.6) is the ratio of the two, $\dfrac{\pi_1}{\overline{T}_1 + 1} \equiv \widehat{\gamma}_1$.

The first row of Table 6 constructs $\widehat{\gamma}_1$. Column (1) reports the value of π_1 in Table 5. Column (3) reports the average tenure of workers who are laid off from their first job, 1.28 years. The implied value of $\widehat{\gamma}_1$ is 0.001, reported in column (4) of the table. To construct a confidence interval for $\widehat{\gamma}_1$, I apply the delta method to derive an asymptotic standard error for $\widehat{\gamma}_1$. In particular, $\widehat{\gamma}_1$

[9] The bound I derive might seem trivial at first; if workers gravitate to higher paying jobs, isn't the price on the job a worker lost always at least as large on average as the price on a brand new job? Surprisingly, this is not true for any arbitrary offer distribution. Intuitively, workers who find a job they prefer to their first job even when they already have tenure on that first job probably drew a very low initial offer. Observing a worker hold two or more jobs in an employment cycle could lower our assessment of $E(\omega_n \mid N = n)$ by enough to fall below $E(\omega_1)$.

asymptotically converges to a normal random variable with variance

$$\sigma^2_{\widehat{\gamma}_1} = \frac{\mathrm{Var}\left(\pi_1\right) + 2\widehat{\gamma}\mathrm{Cov}\left(\pi_1, \overline{T}_1\right) + \widehat{\gamma}^2\mathrm{Var}\left(\overline{T}_1\right)}{\left(\overline{T}_1 + 1\right)^2}$$

I estimate $\mathrm{Var}(\pi_1)$ using the standard error for π_1 in Table 5. To estimate $\mathrm{Var}\left(\overline{T}_1\right)$, I use the sample variance for tenure across all workers laid off from their first job, divided by the number of such workers. This leaves $\mathrm{Cov}\left(\pi_1, \overline{T}_1\right)$. Using a Monte Carlo simulation, I verified that this covariance is small (but positive) when $\gamma = 0$. Thus, for small values of γ, this covariance should not have a noticeable effect on my estimate of $\sigma^2_{\widehat{\gamma}_1}$. Intuitively, since we have a reasonably large sample of workers who are laid off from their first job, we can estimate $\overline{T}_1$ quite precisely, and the main source of variation in estimating $\widehat{\gamma}_1$ comes from variance in π_1. Using a one-tailed t-test, we can reject that γ exceeds 0.008 at the 5% level. Thus, the wage losses of workers who are laid off from their first job suggest only modest returns to tenure. Note that this result is distinct from the evidence of small returns to tenure using Topel's two-step estimator. In particular, the results in Table 6 are based on the wage losses of workers laid off from the first job in their employment cycle, while Topel's estimator compares the wage growth of job stayers and the way wages grow with initial experience for all workers. My approach does require imposing additional assumptions on how workers search for jobs that are not required for Topel's approach to be valid, but it still allows for fairly arbitrary unobserved differences in ability across workers and over time.

What about workers who are laid off from jobs later on in their respective employment cycles? Once again, we can try to use record statistics to bound $E\left(\omega_n \mid N_{t-1} = n\right) - E\left(\omega_1\right)$ and thereby bound γ. However, unlike in the case where $n = 1$, establishing these bounds requires us to know the exact shape of the wage offer distribution. While I argued above that the offer distribution is consistent with a Pareto shape, recall that this relied on assuming $\gamma = 0$. What can we infer about the shape of the offer distribution if we don't impose that $\gamma = 0$? Suppose returns to tenure were linear. Since this implies wages always grow at the same rate on all jobs, the worker should change jobs if and only if his new job pays more than he earns on his current job after accounting for his tenure on his current job, i.e. if $\omega_n \geq \omega_{n-1} + \gamma\left(T_{i,t-1} + 1\right)$. In this case, the average wage growth for a voluntary job changer net of the expected wage growth of job stayers, $E\left(\Delta \ln W_{it} - \beta\Delta Z_{it} \mid N \geq n\right)$, must equal

$$E\left(\omega_n - \omega_{n-1} - \gamma\left(T_{i,t-1} + 1\right) \mid \omega_n \geq \omega_{n-1} + \gamma\left(T_{i,t-1} + 1\right), N \geq n\right) \tag{5.7}$$

If the offer distribution is Pareto, implying the log offer distribution is exponential, the above expectation will be constant for all n, and corresponds to the inverse hazard of the implied log offer distribution. If returns to tenure are linear, then, we can still take the results in Table 4 to mean that we fail to reject a Pareto offer distribution even when we allow $\gamma > 0$.

If the offer distribution is indeed Pareto, it is possible to show that

$$E(\omega_n \mid N_{t-1} = n) \geq E(R_n \mid N \geq n)$$

i.e. the average wage on the worker's n-th job is at least as large as the n-th record from i.i.d. draws from this distribution. Since $E(\omega_1) = E(R_1)$, it follows that

$$\gamma \leq \frac{\pi_n - (E(R_n \mid N \geq n) - E(R_1))}{\overline{T}_n + 1} \equiv \widehat{\gamma}_n$$

where π_n is defined in (5.1), $\overline{T}_n$ is the average tenure for workers who lose their n-th job, and $E(R_n \mid N \geq n) - E(R_1)$ is computed using the implied exponential distribution for the log offer distribution. Relying on Tables 3 and 4, I compute record moments for an exponential distribution with mean 0.0806 and $p = 0.42$. Table 6 reports the point estimates for $\widehat{\gamma}_n$. For $n = 2$, 3, and 4, we can assign a 95% probability that $\gamma \leq 0.023$, half of Topel's point estimate using older workers from the PSID. This last estimate should be interpreted with some caution, since my standard error ignores estimation error in either the mean of the exponential or in p. But the point estimates for $\widehat{\gamma}_n$ for $n \geq 2$ are consistent with those of $\widehat{\gamma}_1$ in suggesting very modest returns to seniority.

To summarize, a model in which workers search from a fixed offer distribution and accumulate only general human capital can account quite well for the wage dynamics of young workers early in their careers. Returns to experience are the main force for wage growth in my sample. According to Table 2, ten years of experience on average add 0.57 log points to wages, an increase of over 75% over their starting wage. Search also plays a significant role; workers raise their log wages by 0.08 on average each time they change jobs voluntarily. However, the benefit to job search is inherently limited, since once a worker finds a high wage job it will be harder for him to find an even higher paying job. Workers only benefit from mobility if they earn low wages. Thus, for example, the model implies that workers who change jobs five times only increase their wages by 0.13 log points on average; this is because their first several job changes must not have involved high wage increases if they went on to find even better jobs. By contrast, job-specific human capital seems no to contribute at all to wage growth in the first several years of a worker's career.

6. Alternative Models of Search

My identification results above rely heavily on the fact that observed wages over an employment cycle $\{W_{it}^n\}$ are associated with a sequence of prices $\{w_n\}_{n=1}^N$ that represent records from the set of offers $\{X_m\}_{m=1}^M$. However, it will not be true for more general search models that the prices on the jobs workers accept correspond to ordinary record statistics. Nevertheless, I now argue that even when prices $\{w_n\}_{n=1}^N$ do not correspond to record values, it will often be the case that there is still some implicit record structure implied by the model, and we might still be able to exploit this structure. The case where returns to tenure γ are positive provides one example. Recall that in this case $\{w_n\}_{n=1}^N$ does not correspond to a list of records in the conventional sense; rather, w_n must exceed w_{n-1} by some threshold amount. However, by normalizing wages in a particular way, we can transform the data so that the normalized wages are a sequence of records in the conventional sense (i.e. with no threshold). The key difference is that $\{X_m\}_{m=1}^M$ are no longer identically distributed as in the case where $\gamma = 0$.

As another example, suppose a job offer specifies both a price w and a number of hours h that the worker must 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 $W_{it} = w\ell_{it}$, and an income $I_{it} = w\ell_{it}h$. Once again, we can define an employment cycle as the time between forced layoffs, 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, and will typically not 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*. Formally, the sequence $\{U(w_n, h_n)\}_{n=1}^\infty$ represents the 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 knew the function $U(\cdot, \cdot)$, e.g. by estimating it from observed choices, we might be able to use data on wages and hours to identify the distribution of utility across job offers. As an illustration, suppose agents do not care about leisure, and 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 the n-th record from i.i.d. draws in which the parent distribution is the offer distribution for incomes, and we can adapt my approach to identify this distribution from observations on income $I_{it}^n = (w_n h_n) \times \ell_{it}$.

7. Conclusion

This paper proposes a way to estimate the wage offer distribution non-parametrically by exploiting the underlying record structure implicit in standard search models. While the number of observations in the NLSY dataset I use is too small to provide a fully non-parametric estimator for this distribution, we can reject the lognormal distribution as a candidate for the offer distribution in favor of the Pareto distribution. This result is distinct from the oft-noted fact that the cross-sectional distribution of wages exhibits a Pareto tail.[10] 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.

The implicit record structure of the standard search model also proves useful for constructing bounds on the returns to tenure, offering an alternative approach to estimating these returns to the one advanced in previous work. For my sample of young workers, I conclude that these returns are not economically meaningful, and that it is instead general human capital and on-the-job search that account for wage growth of these workers.

Finally, while this paper only examines search applications, record theory is potentially applicable in a variety of contexts. Record statistics arise whenever we get to observe the extremes from an unknown number of observations, a feature that characterizes various 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 we rarely get to observe when a worker receives an outside offer. A related example is the problem of optimal contracting with one-sided commitment in Beaudry and DiNardo (1991), where the optimal contract stipulates that the wage is a monotonic function of the record economic conditions since the employment relationship began. 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 be useful in both empirical and theoretical economic applications.

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

8. Appendix

Proof of Lemma 1: To derive the expression for 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

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

This establishes the claim. ■

Proof of Lemma 2: Theorem 4.1 in Bunge and Nagaraja (1991) shows that for an i.i.d. sequence $\{X_m\}_{m=1}^M$ where X_m has a continuous distribution and where Prob($M = m$) $= (1-p)^{m-1} p$ establishes that $\Pr(N = n) = \dfrac{p}{1-p} \dfrac{(-\ln p)^n}{n!}$. For $n = 1$, differentiating this expression with respect to p yields

$$\frac{-\ln p - (1 - p)}{(1 - p)^2} \tag{8.1}$$

Using the inequality $\ln p \leq p - 1$, we can establish (8.1) is positive for all $p > 0$. Thus, the expression for $\Pr(N = 1)$ is an invertible function of p. It trivially follows that we can identify p from the distribution $\{\Pr\,(N = n)\}_{n=1}^\infty$ since it includes this value. ■

Proof of Proposition 1: From lemma 2, we know that $\{\Pr\,(N = n)\}_{n=1}^\infty$ identifies p. We therefore need to show that given p and the distribution of R_n, we can identify the parent distribution of X_m, which I denote by $F(\cdot)$. Define $q \equiv 1 - p$. Using Theorem 4.1 in Bunge and Nagaraja (1991), the probability density for the first n records given at least n records in the sequence is given by

$$h\,(r_1, r_2, ..., r_n \cap N \geq n) = f\,(r_n) \prod_{i=1}^{n-1} \frac{qf\,(r_i)}{1 - qF\,(r_i)} \tag{8.2}$$

where $f(\cdot) = dF(\cdot)$. Integrating out r_1 through r_{n-1} in (8.2) 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

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

Define the inverse cdf $F^{-1}(x)$ for $x \in (0,1)$ as $\sup\{y : F(y) \leq x\}$. Using the change of variables $u = F(r_{n+1})$ and $du = f(r_{n+1}) dr_{n+1}$, we have

$$\begin{aligned}
\Pr(R_n \leq x \mid N \geq n) &= \frac{\int_0^x [-\ln(1-qF(r_n))]^{n-1}}{(n-1)! \Pr(N \geq n)} f(r_n) dr_n \\
&= \frac{\int_0^{F(x)} [-\ln(1-qu)]^{n-1} du}{(n-1)! \Pr(N \geq n)}
\end{aligned}$$

Since the right-hand side above is monotonic in $F(x)$, we can indeed recover $F(\cdot)$ from the conditional distribution of the n-th record as claimed. ■

Proof of Lemma 3: As in the proof of Proposition 1, we Theorem 4.1 in Bunge and Nagaraja (1991) to derive the density function for the n-th record as

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

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

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

Since $E\left(|R_n| \;\middle|\; N > n \right) < E\left(|R_{n+1}| \;\middle|\; N > n \right)$, the former is also finite. The lemma follows from the fact that $E(|R_n|) < \infty$ implies $E(R_n) < \infty$. ■

Proof of Proposition 2: Integrating out (8.2) yields the following densities:

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

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

$$\begin{aligned} E(\Delta \mid N > n) &= E(\mathcal{F}(r_n) \mid N > n) \\ &= \int_{-\infty}^{\infty} \mathcal{F}(r_n) \frac{h(r_n \cap N > n)}{\Pr(N > n)} dr_n \\ &= \int_{-\infty}^{\infty} \mathcal{F}(r_n) \frac{q - qF(r_n)}{1 - qF(r_n)} \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)! \Pr(N > n)} f(r_n) \, dr_n \\ &= \int_{-\infty}^{\infty} \left[\int_0^\infty [1 - F(r_n + \Delta)] \, d\Delta\right] \frac{q}{1 - qF(r_n)} \frac{[-\ln(1 - qF(r_n))]^{n-1}}{(n-1)! \Pr(N > n)} f(r_n) \, dr_n \end{aligned} \tag{8.3}$$

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 $n = 1, 2, 3, ...$ Then we have

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

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

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

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 $n = 1, 2, 3, ...$ 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

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

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

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

Since $F_1\left(F_1^{-1}(u)\right) = F_2\left(F_2^{-1}(u)\right) = u$, 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. ■

Deriving inequality (5.5) in text: Consider a sequence of i.i.d. random variables $\{X_m\}_{m=1}^M$, and any sequence of nonnegative numbers $\{\Delta_m\}_{m=2}^M$. Define

$$Z = \begin{cases} \max\{X_2 + \Delta_2, ..., X_M + \Delta_M\} & \text{if } M \geq 2 \\ -\infty & \text{if } M = 1 \end{cases}$$

We now use the fact that $E(\omega_1 \mid N = 1) \equiv E(X_1 \mid X_1 \geq Z)$. However,

$$\begin{aligned} E(X_1 \mid X_1 \geq Z) &= E_z[E(X_1 \mid X_1 \geq z)] \\ &\geq E_z[E(X_1)] \\ &= E(X_1) \end{aligned}$$

Since $E(R_1) = E(X_1)$, the claim follows. ■

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 full 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		0.0740		0.0054
	0.0065		0.0061		0.0024

	1 year	2 years	5 years	7 years	10 years
implied returns to tenure	0.0054	0.0108	0.0271	0.0380	0.0542
	0.0024	0.0049	0.0122	0.0171	0.0245
implied returns to experience	0.0723	0.1411	0.3270	0.4337	0.5680
	0.0058	0.0109	0.0226	0.0274	0.0300

quadratic returns to tenure

within-job wage growth $\beta_1 + \gamma_1$	experience effect β_1	tenure effect γ_1	tenure squared γ_2
0.0826	0.0661	0.0165	-0.0016
0.0065	0.0067	0.0024	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 ΔX (= constant) and ΔX^2. This is the same regression in column (1) of Table 4, where β_1+γ corresponds to the coefficient on ΔX. The second stage regresses the log real wage net of the estimated $(\beta_1 + \gamma)T + \beta_2 X^2$ 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 ΔT^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
ΔX	--	0.0767	0.0809	0.0816
		0.0046	0.0050	0.0050
ΔX^2	--	-0.0016	-0.0018	-0.0018
		0.0002	0.0002	0.0002
D^{12}	2,443		0.0900	
			0.0094	
D^{23}	982		0.0711	
			0.0137	
D^{34}	452		0.0799	0.0806
			0.0200	0.0072
D^{45}	204		0.0168	
			0.0331	
D^{56}	75		0.0799	
			0.0520	
# obs		27,712	31,868	31,868
stayers		27,712	27,712	27,712
changers		0	4,156	4,156

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, ΔX^2, which is equal to 2 X - 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 ΔX and ΔX 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 coefficient reported in column (3) corresponds to the inverse hazard of this exponential distribution. 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
ΔX	--	0.0837	0.0849
		0.0062	0.0050
ΔX^2	--	-0.0020	-0.0020
		0.0002	0.0002
D^{11}	2,767	0.0029	
		0.0094	
D^{21}	873	0.0843	
		0.0153	
D^{31}	305	0.0904	0.0816
		0.0278	0.0130
D^{41}	137	0.0942	
		0.0432	
D^{51}	50	0.0754	
		0.0726	
# obs		31,844	31,844
stayers		27,712	27,712
changers		4,132	4,132

Test of particular functional forms:

Exponential	$F(4, 31837) = 1.24$	Prob > F = 0.2895
Normal	$F(4, 31837) = 1.08$	Prob > F = 0.3622

The dependent variable is the annual growth rate of real wages. The independent variables are ΔX and ΔX^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 coefficient reported in column (2) corresponds to the inverse hazard of this exponential distribution. The numbers below the coefficient denote robust standard errors. The F -statistic 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.

Table 6: Estimates for Upper Bounds on Returns to Tenure

	(1)	(2)	(3)	(4)	(5)
n	π_n	$E(R_n \mid N \geq n) - E(R_1)$	$\overline{T}_n$	$\widehat{\gamma}_n$	standard error
1	0.0029	0.0000	1.28	0.001	0.004
Assuming log offer distribution is exponential with mean 0.0816					
2	0.0843	0.0475	1.66	0.014	0.006
3	0.0904	0.0797	1.52	0.004	0.011
4	0.0942	0.1036	1.70	-0.003	0.016
5	0.0754	0.1224	1.39	-0.020	0.030

Column (1) reports the average net wage loss for workers who are laid off from their n-th job in an employment cycle. These correspond to the cofficients reported in column (1) of Table 5. Column (2) reports the average value of the n-th record conditional on there being at least n records net of the average value of the first record as computed from an exponential distribution with mean 0.0816 and where the number of observations is geometric with success probability 0.48. Column (3) reports the average tenure on the n-th job for workers who left that job involuntarily. Column (4) constructs the bound on returns to tenure based on workers who were laid off from their n-th job. It is equal to the difference between column (1) and column (2), divided by one plus the value in column (3). The derivation of this formula is described in the text. Column (5) reports the asymptotic standard error for the estimator in column (4). For $n = 1$, the bound holds for any distribution. For $n \geq 2$, the bound applies only if the offer distribution is exponential with mean 0.0816.

Figure 1: Expected Record Gaps
for Different Parent Distributions

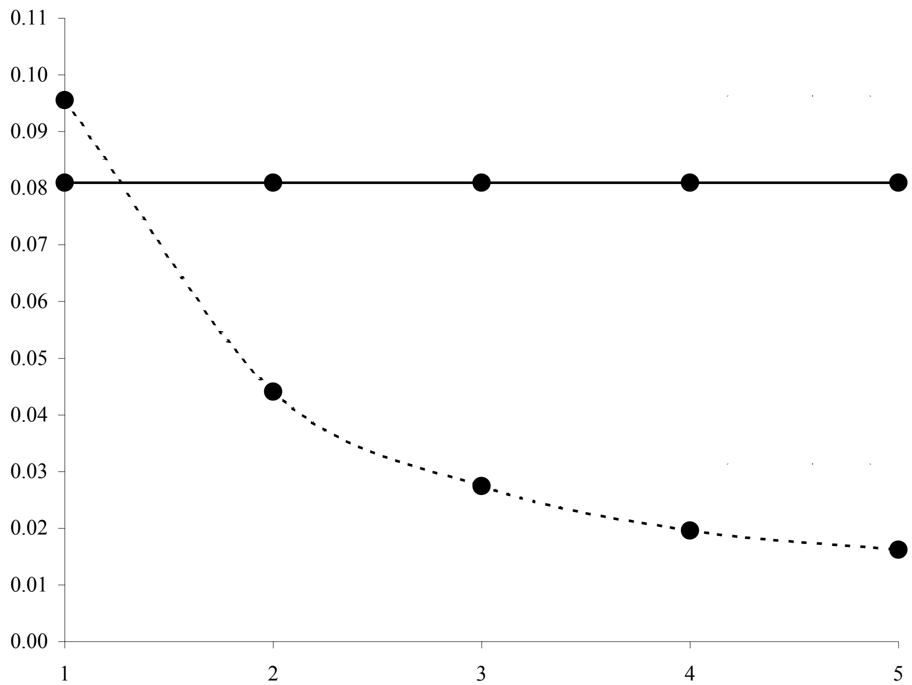


Figure 2: Summary Statistics for *n*

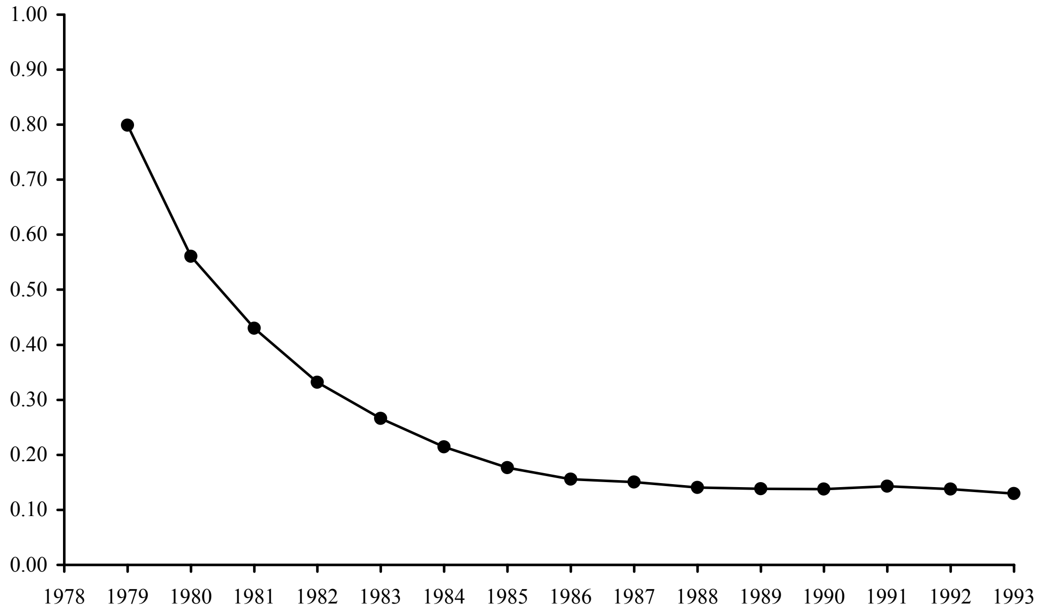


Figure 1a: Proportion of observations where no value for n was assigned

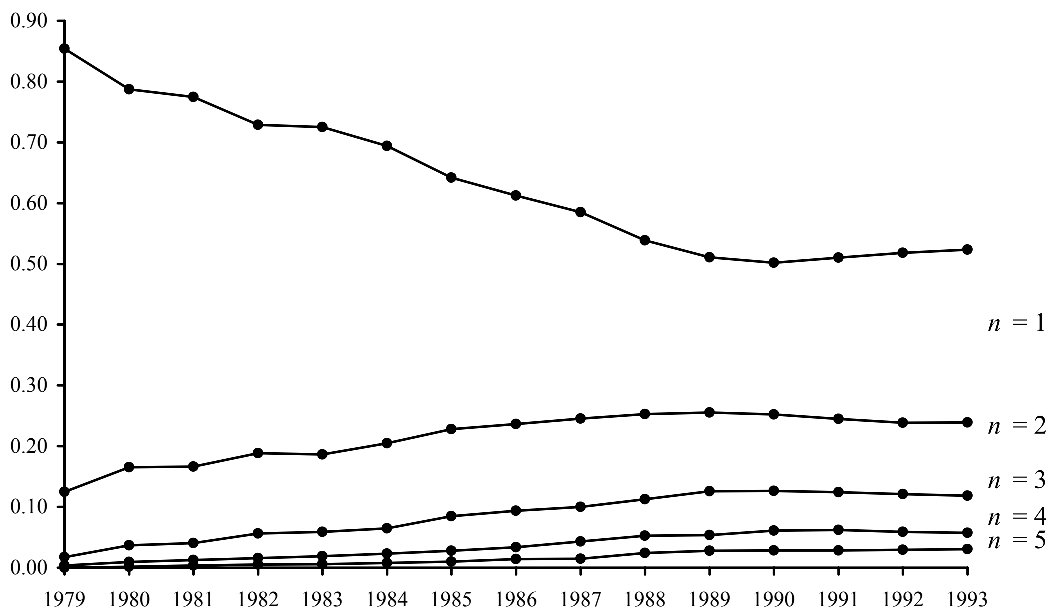
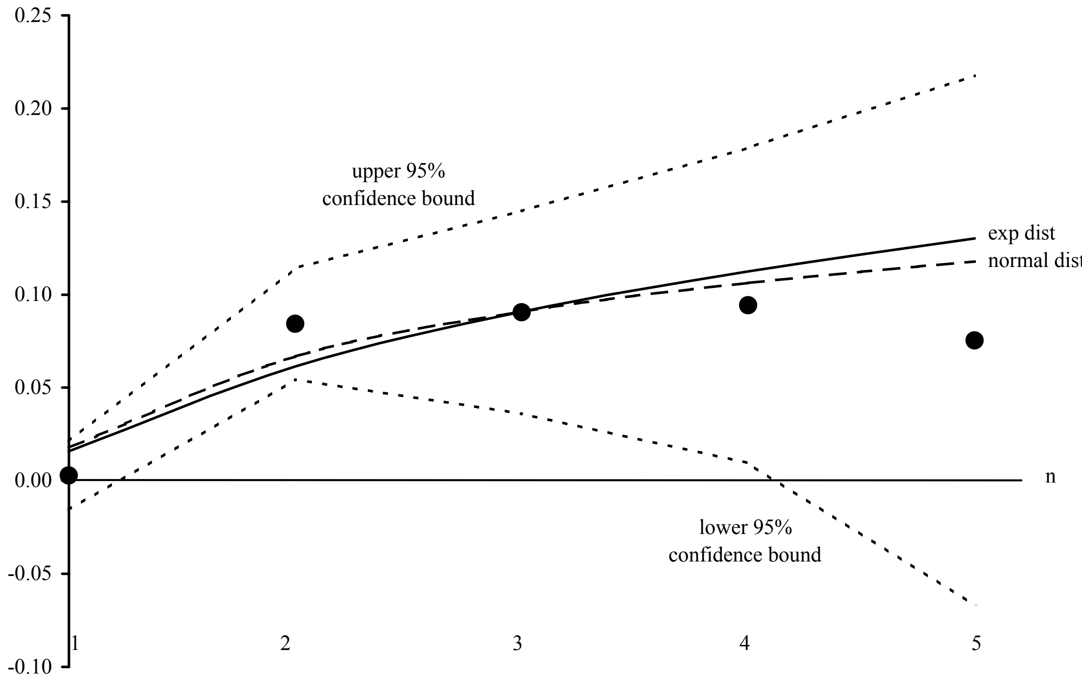


Figure 1b: Share of all observations with $n \geq 1$ for each level of n

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



References

[1] Altonji, Joseph and Robert Shakotko, 1987. "Do Wages Rise with Job Seniority?" *Review of Economic Studies*, July, 54 (3), p437-59.

[2] Altonji, Joseph and Nicolas Williams, 1997. "Do Wages Rise with Job Seniority? A Reassessment" Mimeo, Northwestern University.

[3] 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.

[4] Arnold, Barry, N. Balakrishnan and H. Nagaraja, 1992. *A First Course in Order Statistics*. New York: John Wiley and Sons.

[5] Arnold, Barry, N. Balakrishnan and H. Nagaraja, 1998. *Records*. New York: John Wiley and Sons.

[6] Athey, Susan and Philip Haile, 2002. "Identification in Standard Auction Models" *Econometrica*, November, 70(6), p2107-40.

[7] Barlevy, Gadi, 2002. "The Sullying Effect of Recessions" *Review of Economic Studies*, January, 69 (1), p65-96.

[8] 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.

[9] 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.

[10] 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.

[11] Bowlus, Audra and Jean-Marc Robin, 2004. "Twenty Years of Rising Inequality in US Lifetime Labor Income Values" *Review of Economic Studies*, July, 71(3), p709-742.

[12] 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.

[13] Burdett, Kenneth and Dale Mortensen, 1998. "Wage Differentials, Employer Size, and Unemployment" *International Economic Review*, 39, p257-273.

[14] Chandler, K. N., 1952. "The Distribution and Frequency of Record Values" *Journal of the Royal Statistical Society*, Series B, 14, p220-8.

[15] Flinn, Christopher, 1986. "Wages and Job Mobility of Young Workers" *Journal of Political Economy* 94(3, Part 2), pS88-S110.

[16] 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.

[17] Flinn, Christopher, 2004. "Minimum Wage Effects on Labor Market Outcomes Under Search, Bargaining, and Endogenous Contact Rates" Mimeo, New York University.

[18] Flinn, Christopher and James Heckman, 1982. "New Methods for Analyzing Structural Models of Labor Force Dynamics" *Journal of Econometrics*, January, 18(1), p115-68.

[19] Glick, Ned, 1978. "Breaking Records and Breaking Boards" *American Mathematical Monthly*, 85(1), p2-26.

[20] Gupta, Ramesh, 1984. "Relationships Between Order Statistics and Record Values and Some Characterization Results" *Journal of Applied Probability*, June, 21(2), p425-430.

[21] Kirmani, S. N. U. A. and Beg, M. I., 1984. "On characterization of distributions by expected records" *Sankhyā A*, 46(3), p463-465.

[22] Kortum, Samuel, 1997. "Research, Patenting, and Technological Change" *Econometrica*, November, 65(6), p1389-1419.

[23] Kotlarski, I., 1967. "On Characterizing the Gamma and the Normal Distributions" *Pacific Journal of Mathematics*, 20, p69-76.

[24] Lin, G. D., 1987. "On Characterizations of Distributions via Moments on Record Values" *Probability Theory and Related Fields*, 74, p479-83.

[25] Lucas, Robert and Edward Prescott, 1974. "Equilibrium Search and Unemployment" *Journal of Economic Theory* February, 7(2), p188-209.

[26] Marimon, Ramon and Fabrizio Zilibotti, 1999. "Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits" *Economic Journal*, April, 109(127), p266-91.

[27] Mortensen, Dale, 1986. "Job Search and Labor Market Analysis" *Handbook of Labor Economics*, v2, eds. O. Ashenfelter and J Layard, Amsterdam: North Holland, p849-920.

[28] 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.

[29] 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.

[30] Neal, Derek and Sherwin Rosen, 2000. "Theories of the Distribution of Earnings" *Handbook of Income Distribution*, 1, Amsterdam: Elsevier Science, North-Holland, p379-427.

[31] 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.

[32] Paxson, Christina and Nachum Sicherman, 1996. "The Dynamics of Dual-Job Holding and Job Mobility" *Journal of Labor Economics*, July, 14(3), p357-93.

[33] Postel-Vinay, Fabien and Jean-Marc Robin, 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity" *Econometrica*, November, 70(6), p2295-2350.

[34] Shimer, Robert, 2004. "On-the-Job Search, Bargaining, and Wage Dispersion" Working Paper, University of Chicago.

[35] Song, Unjy, 2004. "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders" Mimeo, University of British Colubmia.

[36] Topel, Robert, 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority" *Journal of Political Economy*, February, 99(1), p145-76.

[37] Topel, Robert and Michael Ward, 1992. "Job Mobility and the Careers of Young Men" *Quarterly Journal of Economics*, May, 107(2), p439-79.

[38] van den Berg, Gerard and Geert Ridder 1998. "An Empirical Equilibrium Search Model of the Labor Market" *Econometrica*, September, 66(5), p1183-1221.

[39] 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.