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

January 18, 2005

Abstract

This paper uses record-value theory, a branch of statistics that deals with the timing and magnitude of extreme values in sequences of random variables, to non-parametrically identify the distribution of wages workers could potentially earn, even in the presence of unobserved heterogeneity across workers. Applying this approach wage data from the NLSY, I show that the data supports the hypothesis that the wage offer distribution is Pareto, but not that it is lognormal. In addition, I show how this approach to identification can be used to bound the return to employer-specific human capital, and conclude that specific human capital plays a small role for the young workers in my sample. Instead, their wage growth appears to be driven by on-the-job search and the accumulation of general human capital.

Key Words: On-the-Job Search, Wage Growth, Specific Human Capital, Hausdorff Moment Problem

*I would like to thank Joe Altonji, Susan Athey, Marco Bassetto, Jeff Campbell, Kim-Sau Chung, Chuck Manski, Francesca Molinari, Chris Taber, Nick Williams, and especially Eric French and H. N. Nagaraja for helpful discussions. I also wish to acknowledge the help of my research assistant Merritt Lyon. Finally, I have benefitted from comments at seminars in Arizona State, Michigan State, NYU, Penn State, the University of Pennsylvania, the University of British Columbia, the Federal Reserve Bank of Minneapolis, 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 used job-search models to address a variety of 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 persistence of earnings inequality when workers are free to move across jobs; papers in this vein include Flinn (2002) and Bowlus and Robin (2003). 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 (2003), or of a macroeconomic shock, as in Barlevy (2002).

Unfortunately, the results we obtain using this approach can be sensitive to functional form assumptions implicit in estimating these models, especially assumptions that govern the shape of the distribution of wages a worker could earn across potential employers. For example, we may erroneously conclude blacks and whites face the same wage offer distributions if the same distribution within the parameterized family of distributions we consider happens to provide the best overall fit for both groups. As another example, the effects of an increase in the minimum wage depends on the number of employers who choose to pay a wage just above the original minimum wage, and a functional form that provides a good general fit to the data may do poorly in matching this particular part of the distribution. Thus, it is important to ask whether 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, given the difficulty of accounting for all of the differences in worker quality that are reflected in wages, it is important that this identification be robust to the presence of unobserved heterogeneity.

This paper takes on the question of whether job search models can be identified non-parametrically. My approach is similar to the approach of Athey and Haile (2002) in their work on non-parametric identification of auction models. This is only natural given that search models similarly assume multiple agents bid on a common object, namely employers competing for the worker’s time. The key difference comes from the fact that worker surveys seldom contain information about offers workers turned down. In particular, we rarely know the number of employers a worker received offers from, and so, unlike the typical data for auctions, we do not know how many rival bids his current job has beaten. This is important, since many of Athey and Haile’s results rely on the fact that a winning bid is the maximum of a known number of bids, and as such their results cannot
be directly applied to search models.\(^1\)

Nevertheless, even without knowing how many employers already bid on a worker at a given point in time, we can use the fact that the jobs a worker accepts form a sequence of records, i.e. each such job is more attractive than all offers that preceded it. Statisticians have studied the behavior of records from random sequences, and have applied their findings to study various phenomena such as global warming, record athletic performances, road congestion, tolerance testing, and ruin probabilities.\(^2\) These results can also be applied in the study of job search.

In what follows, I describe a standard model of on-the-job search in which workers draw wages from a fixed offer distribution. I show that the shape of this distribution is identified from data on wages and job mobility, even when we allow for unobserved heterogeneity in wages. Exact identification of the distribution turns out to be impractical, since it requires computing the moments of infinitely many record values. However, even if we can only estimate a small number of moments, we can still 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 wage offer distribution is consistent with a Pareto distribution, 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 my analysis focuses on a particular approach to identifying the wage offer distribution, the model actually suggests this distribution is overidentified. Specifically, whereas I use the wage gains of voluntary job changers to recover the wage offer distribution, this same distribution uniquely determines the wage losses of involuntary job changers. I confirm that the average wage losses for involuntary job changers are consistent with the offer distribution implied by the average wage gains of voluntary job changers. Interestingly, this consistency suggests job-specific human capital cannot be a particularly important source of wage growth for the young workers in my sample, or else involuntary job changers would lose much more than what they previously accumulated through on-the-job search. This insight provides an alternative way to assess the contribution of job-specific human capital from what has been suggested by previous work on the returns to specific human capital, e.g. Altonji and Shaktoko (1987) and Topel (1991).

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

---

\(^1\)Song (2004) examines identification of auction models where the number of bidders is unknown. Her approach can be used for estimating search models, but it is not robust to the presence of unobserved heterogeneity.

\(^2\)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

\[
L_n = \min \{ m : X_m > X_{L_{n-1}} \}
\]

(1.1)

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

\[
R_n = X_{L_n}
\]

(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:

\[
\{65, 61, 68, 69, 63, 67, 64, 66, \ldots\}
\]

(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):

\[
\{65, 68, 69, \ldots\}
\]

\(^3\)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 aspects of 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, \( \max \{X_1, \ldots, X_n\} \). By contrast, the \( n \)-th record statistic \( R_n \) is the maximum of a random number \( L_n \) observations, \( \max \{X_1, \ldots, X_{L_n}\} \). 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

\[
\text{Prob} (R_n = x) = \sum_{m=n}^{\infty} \text{Prob} (L_n = m) \times \text{Prob} (X_{m:m} = x)
\]

(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 useful in 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. I will now describe a model in which workers search from a fixed offer distribution, and show how insights from record statistics can be used to identify this distribution from wage data. I treat the offer distribution as given rather than deriving it from economic fundamentals, although I show below that my model can be viewed as a reduced form of a richer model in which the equilibrium offer distribution is uniquely determined by the fundamentals of the economy. Under this interpretation, 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. I then examine whether the offer distribution in the model is identified in the benchmark case where workers are identical. Finally, I consider identification when workers differ in both observable and unobservable characteristics.

2.1. A Model of Job Search

Consider an economy populated by employers and workers. Workers supply a homogenous labor input, although they may each supply different amounts of labor. Let $\ell_{it}$ 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 the econometrician who gets to collect data on this market may only observe it imperfectly. 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 $b$.

While a worker is unemployed, he encounters potential employers at rate $\lambda_0$ 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, a fixed price per unit of effective labor will be optimal from the employer’s point of view in certain environments. If the worker accepts a job offer, his hourly wage would be

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

(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 $F_i(\cdot)$. I assume all workers face the same distribution, i.e. $F_i(\cdot) = F(\cdot)$ for all $i$. However, I could allow observably distinct groups of workers, such as blacks and whites, to search from different distributions; in this case, we would just carry out the analysis outlined below for each worker group separately.\(^4\)

Employed workers face a constant hazard $\lambda_1$, possibly different from $\lambda_0$, of encountering potential employers. Once again, they are offered a fixed price per unit of labor drawn from $F(\cdot)$. The fact that workers can continue to search while employed plays an important role in the identification results, particularly when I allow for unobserved heterogeneity.

Finally, employed workers face a constant hazard $\delta$ 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 that pay at least $w^*$ per unit of labor. The optimal cutoff will depend on the distribution $F(\cdot)$ and the parameters $b$, $\delta$, $\lambda_0$, and $\lambda_1$. Since these parameters are the same for all workers, the cutoff $w^*$ will be as well. When the worker is employed, his decision is even simpler: he should trivially accept any offer that exceeds the price on his current job and turn down any offer below it.

The key question this paper seeks to answer is whether it is possible to identify the offer distribution $F(\cdot)$ non-parametrically from hourly wage data $\{W_{it}\}$, i.e. whether we can use wage data to uncover the distribution of earning possibilities available to a worker shopping around for a job. That said, certain applications require us to identify not the offer distribution $F(\cdot)$ but the economic fundamentals that shape it. For example, to simulate the effects of changes in policy, we need to know the underlying fundamentals in order to derive the equilibrium offer distribution

---

\(^4\)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.
under the new policy. However, in various equilibrium search models, identifying $F(\cdot)$ is equivalent to identifying the deeper structural parameters of interest, since there is often a one-to-one mapping between the equilibrium offer distribution and the fundamentals that shape it.

As an illustration, consider the case where the fundamental variable 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. Several equilibrium search models predict that this distribution – together with $\lambda_0$, $\lambda_1$, and $\delta$ – determines the offer distribution $F(\cdot)$. One such model 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. Since $F(\cdot)$ is identical to the distribution of productivity $\Gamma(\cdot)$, 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 (1982) 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 from $F(\cdot)$ by applying the relevant inversion formula that previous authors already derived.

2.2. Identification with Homogeneous Labor

Let us return to the question of whether it is possible to identify the offer distribution $F(\cdot)$ from wage data $\{W_{it}\}$. I begin with the special case in which all workers are identical, i.e. where $\ell_{it} = 1$ for all $i$ and $t$. This case provides a useful benchmark, and yields important insights for the more empirically relevant case where workers are heterogeneous.
Any discussion of identification hinges on what data we observe. The most common source of wage data are surveys that follow workers over time and collect information about their various jobs: the hourly wage paid on each job, how long each job lasted, why the job ended if it did, and so on. These surveys typically ask about the jobs workers accept rather than the offers they encounter. I therefore assume that the only available data are the hourly wages workers earned on the jobs they were employed on, as well as data on work history (i.e. the number of jobs they were employed on 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 this data is only useful for identifying the arrival rates $\lambda_0, \lambda_1,$ and $\delta$, not 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 involuntary job changes, in which the worker is forced out of a job, and voluntary job changes, in which the worker meets a higher paying employer and chooses to move. 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, if the worker is laid off from the first job he accepts, $M = 1$. 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 $p = \delta / (\lambda_1 + \delta)$. 

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 currently on the \( n \)-th job in his cycle. Since \( \ell_{it} = 1 \), this wage is equal to the price per unit labor the respective employer is offering to pay, i.e. \( W_{it}^{n} = w_n \). Thus, identifying the wage offer distribution is equivalent to recovering the parent distribution from information on the record values that occur in the sequence \( \{X_m\}_{m=1}^{M} \). But unlike in the classical model, the number of observations \( M \) is itself random. The properties of records from a random number of observations have been less extensively analyzed 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 whether this truncated distribution is non-parametrically identified. For some applications, this distribution is enough. 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 \( F(\cdot \mid w \geq w^*) \).

Bontemps, Robin, and van den Berg (2000) already established that when workers are homogeneous, \( F(\cdot \mid w \geq w^*) \) can indeed be identified non-parametrically from wage data. However, their analysis does not explicitly rely on the record structure of search models. Instead, they observe that since the first job is a random draw from the truncated distribution, the empirical distribution of wages on a first job for a large enough sample should be equal to the offer distribution.\(^6\) Translated into the language of records, this implies that the distribution of the first record identifies the wage offer distribution. But as the next proposition illustrates, we do not have to restrict

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

\(^6\)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.
attention to the first record. For any integer \( n \), the empirical distribution of wages on the \( n \)-th job identifies the wage offer distribution:

**Proposition 1**: Consider a sequence of i.i.d. random variables \( \{X_m\}_{m=1}^M \) where \( \text{Prob}(M = m) = (1 - p)^{m-1}p \) 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 \( N \).

In words, the empirical distribution of wages on the \( n \)-th job in the cycle – for cycles with at least \( n \) jobs – can always be used to identify \( F(\cdot | w \geq w^*) \). However, when \( n > 1 \), we also need to use data on job mobility. To appreciate why we require 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 very 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 \( \lambda_1/\delta \), 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 \( \lambda_1 \) and \( \delta \) from job duration data. In particular, 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 can separately identify \( \lambda_1 \) and \( \delta \). But this hinges on a functional form for \( F(\cdot) \), whereas we need to know the ratio \( \lambda_1/\delta \) to determine the functional form of \( F(\cdot) \). Thus, the way previous authors have estimated these parameters is not relevant for our purposes. However, as long as \( F(\cdot) \) is continuous, the distribution of the number of records \( N \) is a function of this ratio but not of \( F(\cdot) \). Specifically, the distribution of \( N \) uniquely determines \( p \), which from Lemma 1 is a monotonic in \( \lambda_1/\delta \):

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

Thus, identification proceeds in two steps. First, we use mobility data to recover \( p \). Then, given \( p \), we can recover the offer distribution from the wages of workers on the \( n \)-th job of their employment cycle, for any value of \( n \). This is not to imply that higher values of \( n \) are somehow better than
the first job in each employment cycle. Rather, Proposition 1 reveals that the shape of the offer distribution uniquely determines the evolution of the wage distribution over an employment cycle. When labor is homogenous, this merely tells us that the wage offer distribution is overidentified, and provides a test of the internal consistency of the model: wages on different jobs in an employment cycle should consistently identify the same offer distribution. But when labor is heterogeneous, as I will next assume, the fact that the rate at which wages grow over an employment cycle is uniquely determined by the offer distribution is what allows us to cope with unobserved heterogeneity.

2.3. Identification with Unobserved Heterogeneity

Let me now consider the case where \( \ell_{it} \) varies across workers and over time. If we could observe \( \ell_{it} \), we can infer the sequence of record values \( \{w_n\} \) from hourly wages \( \{W^n_{it}\} \) and proceed as above. But productivity is typically hard to measure. When variation in \( \ell_{it} \) is unobservable, my previous approach of using the cross-sectional wage distribution no longer works. In particular, without any information on \( \ell_{it} \), it is impossible to tell if variation in the wages on the first job out of unemployment \( W^1_{it} \) is due to variation in the prices \( w_1 \) across employers or ability \( \ell_{it} \) 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 \( F(\cdot | w \geq w^*) \).

If we impose some assumptions on ability, we might still be able to use cross-sectional data to identify \( F(\cdot) \) as before. For example, suppose each worker’s productivity were constant over time, i.e. \( \ell_{it} = \ell_i \) for all dates \( t \). Intuitively, observing a worker over multiple cycles would allow us to infer his relative productivity, since if he were more productive he would earn consistently high wages. Once we know which workers are more productive, we can use the distribution of wages on the first job to recover the offer distribution. In fact, using a result due to Kotlarski (1966), we really only need to observe each worker on just two different cycles. Let \( w'_1 \) denote the price per unit labor on the first job in the first employment cycle and \( w''_1 \) denote the price per unit labor on the first job in the second employment cycle. Kotlarski’s theorem establishes that under certain regularity conditions, if \( w'_1, w''_1, \) and \( \ell_i \) are independent, the joint distribution of \( (w'_1\ell_i, w''_1\ell_i) \) uniquely 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 can prove them false; e.g. the assumption that ability is fixed over time is inconsistent with the fact that wages vary over the duration of a job). Instead, what we would like to know is whether the offer distribution can be identified under minimal assumptions on \( \ell_{it} \). This is the approach I pursue.
Consider the following specification for $\ell_{it}$, based on Flinn (1986), which allows for both observable and unobservable variation in worker ability:

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

(2.3)

The first term, $Z_{it}$, represents observable characteristics for individual $i$ that affect his productivity, and $\beta$ represents the returns to these characteristics. I will assume $Z_{it}$ is the time since the worker first entered the market, i.e. the worker’s potential experience. The next term, $\phi_i$, is fixed, 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, $\epsilon_{it}$, denotes unobserved variation in productivity, as well as multiplicative measurement error in reported wages. I impose the following assumptions on these terms:

**Assumption 1:** $\ell_{it}$ is independent of all employer-specific characteristics

**Assumption 2:** $\Delta Z_{it}$ is independent of $\Delta \epsilon_{it}$

**Assumption 3:** $E[\Delta \epsilon_{it}] = 0$

The first assumption states that the choice of employer has no effect on the worker’s productivity. This assumption insures a worker should accept any offer that pays more per unit of labor than his current job. Note that it also implies any human capital the worker accumulates must be general in nature, since it cannot be specific to any one employer. In my empirical work, I argue that several pieces of evidence confirm that employer-specific human capital is indeed negligible for the young workers 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 is essentially without loss of generality, since we $\Delta Z_{it}$ can always include intercepts to capture a non-zero mean. Intuitively, if $\epsilon_{it}$ tends to grow over time, we could infer this from workers who remain on the same job, so trend growth is effectively observable. The fact that $\epsilon_{it}$ is a martingale imposes very minimal restrictions on earnings. For example, it allows for serial correlation in earnings over the duration of a job, including the case where $\epsilon_{it}$ is non-stationary. Likewise, the variance of $\epsilon_{it}$ can vary arbitrarily over time and across individuals. In particular, productivity can follow a different martingale process for each individual.

Given such weak assumptions on productivity, it will be impossible to distinguish $w$ from $\ell_{it}$
using data on $W_{it}$. Nevertheless, I now argue we can still identify the distribution of prices $w$ that workers face. Define $\omega_n = \ln w_n$ as the log price per unit labor on the worker’s $n$-th job, so that $\omega_n$ 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 $\ell_{it}$, we obtain the following equation for the log hourly wage:

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

(2.4)

We next first-difference equation (2.4) to get rid of the fixed effect term $\phi_i$. Let $\Delta$ denote the difference in a particular variable between two distinct points in time. Then we have

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

(2.5)

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

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

(2.6)

Given that $\Delta Z_{it}$ and $\Delta \varepsilon_{it}$ are 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 $\beta$, 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

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

(2.7)

The net wage gain for a voluntary job changer who leaves his $n$-th job is therefore the sum of a noise term $\Delta \varepsilon_{it}$ and the gap between the $n$-th record and the $n+1$-th record among i.i.d. draws from the log price 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 $\Delta \omega$. However, since $\Delta \varepsilon_{it}$ has zero mean, we can recover the average record gap from wage data. That is, averaging the net wage gains across all workers who move from their $n$-th job to their $n+1$-th job, we have

$$E(\Delta \ln W_{it}^n - \beta \Delta Z_{it} \mid N > n) = E(\omega_{n+1} - \omega_n \mid N > n) + E(\Delta \varepsilon_{it} \mid N > n)$$

$$= E(\omega_{n+1} - \omega_n \mid N > n)$$

where we use the fact that $\Delta \varepsilon_{it}$ is independent of job characteristics. I shall now argue that the sequence of expected record gaps

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

(2.8)

13
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 \( E(|X_m|) < \infty \), then the conditional expectation \( E(R_{n+1} - R_n \mid N > n) \) is finite for \( n = 1, 2, 3, \ldots \). Thus, we need to assume that the offer distribution has a finite mean. If this is true, then given a value of \( p \), which recall we can identify from the distribution of \( N \), we can identify the shape of the wage offer distribution.

**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 \( E(|X_m|) < \infty \), the sequence

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

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 \( \{E(R_{n+1} - R_n)\}_{n=1}^\infty \) uniquely characterizes the parent distribution up to a location 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. When \( M \) is finite with positive probability, though, we need to condition on reaching the \( n \)-th record. This conditioning is non-trivial, and so one 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. Interestingly, 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.

In sum, the average wage gains of voluntary job changers (net of returns to experience) can be used to identify the distribution of log wage offers \( x_m = \ln X_m \) up to a location parameter, from which we can recover the distribution of wage levels up to a scale parameter. Why do record moments allow us to identify the offer distribution? Recall from my discussion of the homogenous

\[7\] I am indebted to H. N. Nagaraja for his assistance with the proof of this proposition.
labor case that this 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 all these expected gains 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 $n$. Looking at how the average record gap varies with $n$ evidently allows us to gauge how the expected gain varies with $x$, and thus to infer the shape of the distribution.

2.4. From Identification to Estimation

Proposition 2 establishes that the average net wage gains of voluntary job changers across different jobs within an employment cycle 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 moments into an underlying parent distribution. I now describe a procedure for doing this, as well as estimator based on finitely many moments that converges to the true distribution as the number of employment cycles we observe (and consequently the number of moments we can estimate) goes to infinity. I close with a discussion of how Proposition 2 can be used to test particular functional forms even when the sample size is too small to reliably estimate the distribution function non-parametrically, as will in fact be the case in my sample.

I begin with the problem of recovering the offer distribution from observations on wage gains. As demonstrated in the proof of Proposition 1, for any integer $n$, the average $n$-th record gap $E(R_{n+1} - R_n \mid N > n)$ can be expressed as an integral involving a common function $g(x)$

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

(2.9)

where $g(x)$ depends on $F(\cdot)$ and $p$. To recover $F(\cdot)$ from the sequence (2.8), I proceed in two steps. First, I determine the function $g(x)$ from the values of all expected record gaps, i.e. the function $g(x)$ which solves (2.9) for all $n \in \{1, 2, 3, \ldots\}$. Thus, from a set of estimates for $\{E(R_{n+1} - R_n \mid N > n)\}$, I can construct a function $g(x)$. Second, I invert the function $g(x)$ to obtain the distribution $F(\cdot)$.

To determine $g(x)$ from the expected record gaps, rewrite (2.9) as

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

(2.10)
Since we can identify $p$ and $E(R_{n+1} - R_n \mid N > n)$ from wage and mobility data, we can compute $\mu_{n-1}$. The task of finding a function $g(x)$ for which

$$\int_{0}^{1} g(x) x^n = \mu_n \text{ for } n = 0, 1, 2, \ldots$$

is known as the Hausdorff moment problem. That is, we want to find a function $g(x)$ over the unit interval whose moments (i.e. integrals of the function multiplied by different powers of $x$) take on particular values. Shohat and Tamarkin (1943) offer a set of sufficient conditions for this problem to have a solution, and provide an analytical representation for this solution $g(x)$ in terms of $\mu_n$. Essentially, the expressions for $\mu_n$ are used to construct coefficients for an infinite polynomial expansion. A little algebra then allows us to recover the distribution function of interest $F(\cdot)$ from the function $g(x)$ we just constructed. This procedure is summarized in the next proposition:

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

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

   \[
P_n(x) = \frac{P_n(2x - 1)}{\int_0^1 P_n^2(2x - 1) \, dx} \equiv \sum_{j=0}^{n} c_{nj}x^j \quad (2.11)
   \]

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

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

   Using the coefficients \(c_{nj}\) from (2.11), construct a new sequence \( \{\lambda_n\}_{n=1}^{\infty} \) where

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

   Define a function $g(x)$ over \((0, 1)\) as the sum of polynomials $P_n(x)$ with coefficients $\lambda_n$, i.e.

   \[
g(x) = \sum_{n=0}^{\infty} \lambda_nP_n(x) \quad (2.12)
   \]

3. The inverse parent distribution function $F^{-1}(x)$ over \((0, 1)\) can be obtained from $g(x)$ by the formula

   \[
   F^{-1}(x) = \int_{0}^{x} \frac{(1 - p) g' \left( \frac{\ln (1 - (1 - p)z)}{\ln p} \right)}{(1 - z) (1 - (1 - p)z) \ln p} \, dz + \text{constant} \quad (2.13)
   \]

   where the constant of integration denotes the unidentified location parameter.
Proposition 3 tells us how to use an infinite list of moments to recover the parent distribution. But in practice we will only be able to estimate finitely many \( \mu_n \), and these moments will be estimated imprecisely given sampling error. Talenti (1987) examines a variant of the Hausdorff moment problem with finitely many moments measured with error. He suggests replacing the infinite sum in (2.12) with the finite sum

\[
g(x) = \sum_{j=0}^{J} \lambda_j P_j(x)
\]  

where \( J \) denotes the number of moments we can observe. Talenti shows that the problem is stable, in the sense that the inaccuracy from using a finite set of moments is bounded by a function of the number of moments \( J \) and the magnitude of the sampling error, and this bound converges to zero as the number of moments goes to infinity and the error term both go to zero. Thus, the estimator for \( F(\cdot) \) in Proposition 3 is consistent, in the sense that as the number of employment cycles \( K \) goes to infinity, the expression we derive from (2.13) and (2.14) will converge to the true offer distribution. But with only a small number of moments, the approximation in (2.14) is likely to be quite imprecise. Moreover, since records tend to be rare events, the sample size required to estimate the \( n \)-th expected record gap with a given standard error can be quite large even for small values of \( n \). In other words, this estimator converges very slowly. Although we can readily construct a consistent non-parametric estimator for the wage offer distribution, it may be very noisy in practice.

While a precise estimate for \( F(\cdot) \) may be hard to come by in practice, even a small number of moments can provide us with a powerful test of whether certain functional forms are consistent with the data. As an illustration, Figure 1 displays the expected record gaps \( E(R_{n+1} - R_n \mid N > n) \) 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 what I estimate 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. Given the size of the sample I use in my empirical work, I will focus on this implication rather than try to estimate \( F(\cdot) \).
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 employer-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 a quit. However, nearly half of all workers who reported quitting did not start their next job for weeks or months. 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, either because he quit to avoid being laid off, or else 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 two months (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 for at least another two months. 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 affected 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 turn to identification, 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 \( X_{it} \), or the time since worker \( i \) entered the labor market and date \( t \):

\[
Z_{it} = \beta_1 X_{it} + \beta_2 X_{it}^2
\]

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

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

This is consistent with my assumption that \( Z_{it} \) is independent of any job-specific characteristics. However, given the prominent discussion of employer-specific human capital in labor economics, we should verify that employer-specific human capital indeed plays a small role in the wage growth of workers in my sample. Consider therefore the alternative in which the worker’s productivity
depends on the amount of time the worker has spent working with his current employer, as distinct from the time he spent working with any employer. 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 $T_{it}$:

$$Z_{it} = \beta_1 X_{it} + \beta_2 X_{it}^2 + \gamma T_{it}$$  \hspace{1cm} (3.2)

I consider higher-order terms in $T_{it}$ in my empirical application, but for notational simplicity it will be easier to proceed as though the returns to tenure are linear. Evidence that $\gamma$ is different from zero would invalidate my identification results as outlined in 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 proof). 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 $\gamma$ from the wage equation

$$\ln W_{it}^n = \omega_n + \phi_i + \beta_1 X_{it} + \beta_2 X_{it}^2 + \gamma T_{it} + \varepsilon_{it}$$  \hspace{1cm} (3.3)

Since the unobserved log price per unit labor $\omega_n$ 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 $\gamma$. Altonji and Shakotko (1987) proposed an instrumental variables approach for estimating $\gamma$, which yielded small values for $\gamma$. 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. For my sample, however, I find only modest returns to tenure, even using Topel’s approach. Since my sample consists of much younger workers than in Topel’s sample, though, this finding does not contradict his results.

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

$$\ln W_{it}^n = \omega_n + \phi_i + \beta_1 X_{0it} + \beta_2 X_{0it}^2 + (\beta_1 + \gamma) T_{it} + \varepsilon_{it}$$  \hspace{1cm} (3.4)
To estimate $\gamma$, 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 $\beta_2$ 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 $\beta_1$, adjusting the standard errors to take into account first-stage estimation error. The estimate for $\gamma$ is just the difference between the estimates for $\beta_1 + \gamma$ and $\beta_1$.

Table 2 reports the results of this two-step procedure for my dataset. The point estimates for $\beta_1 + \gamma$ and $\beta_1$ are 0.0794 and 0.0740, respectively, implying $\gamma = 0.0054$. The implied point estimate for $\gamma$ 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 $\beta_1$ 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 $\gamma$ is small, it is likely to be biased downwards, as Topel himself observed. Although the first-stage estimate for $\beta + \gamma$ is unbiased, the second stage estimate for $\beta$ is biased upwards, since initial experience $X_{0it}$ is likely to be positively correlated with $\omega_n$ given workers are more likely to find higher-paying jobs the more time they had to search. However, the bias in $\beta_1$ is probably small given the high incidence of involuntary job loss in my sample, which implies a large fraction of workers with high experience will be on the first job in their employment cycle. Moreover, I will argue in Section 5 that the absence of large returns to tenure is independently confirmed by evidence on the wage losses of involuntary job changers. Thus, the data appear consistent with my assumption that $\gamma \approx 0$. 

22
4. Empirical Results

I now turn to the question of identification. Although I could in principle try to estimate an offer distribution for distinct groups, e.g. blacks and whites or high-school and college graduates, the number of observations in my sample is already too small, so I instead group all workers together assuming they face a common distribution.

Recall that the first step in my identification is to recover the parameter \( p = \frac{\delta}{(\lambda_1 + \delta)} \). 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

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

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 \( \lambda_1/\delta \) of 1 is smaller than the value of 10 reported in some of the papers cited in the Introduction that estimate \( \delta \) and \( \lambda_1 \) 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, i.e. any instance in which the wage changed by more than a factor of 7.4 or more in the span of a year. Again, 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 $\pi_n$ are unbiased estimates of the expected moment gaps $E (R_{n+1} - R_n \mid N > n)$. Combining the two stages allows the wage growth of job changers to help in identifying the coefficient $\beta_2$, and should therefore be more efficient. Note that the variance of the residual will be different for job stayers and job changers, since for the latter the residual 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 to the $(n + 1)$-th job in the previous year for each $n$ is reported next to the corresponding dummy variable. The estimated coefficients for (4.1) are reported in the second column. The first column in the table reports the estimates for $\beta_1$ and $\beta_2$ omitting job changers, confirming that estimating $\beta_1$ and $\beta_2$ from job stayers alone would have negligible effects on my point estimates. The estimates for $\pi_n$ are all clustered around 8%, with the exception of $\pi_4$. However, this coefficient (as well as the coefficient $\pi_5$) 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 non-parametric estimator for $F (\cdot)$ based on the coefficients $\pi_n$ in Table 4 is likely to be very noisy. Clearly, we would need many more employment cycles to come up with a reasonably reliable estimator. However, as noted earlier, we can easily test particular candidate distribution functions. Recall from Figure 1 that if the wage offer is Pareto, implying the log wage offer distribution is exponential, the coefficients $\pi_n$ would be constant for all $n$. Thus, testing this particular functional form restriction amounts to testing whether the coefficients $\pi_n$ in (4.1) are all equal. Note that this is a test of whether the log offer distribution is exponential rather than a test of whether it is an exponential with particular parameter value. To the extent that we fail to reject that the $\pi_n$ are equal, we can estimate the exact exponential distribution from the estimate for $\pi_n$, which corresponds to the inverse of the hazard rate for this distribution.

The first row in the bottom panel of Table 4 reports the results for the test that all of the coefficients $\pi_n$ 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. Thus, we fail to reject the null that the wage offer distribution is Pareto at conventional significance levels. The third column of Table 4 then reports the estimate for $\pi_n$ assuming these coefficients are equal. The average net wage growth from voluntarily moving jobs is 0.0806, which is consistent with the estimated wage growth for young workers reported in Topel and Ward (1992). This estimate for the inverse hazard is smaller than the inverse hazard Flinn (2002) estimates using the same NLSY dataset of 0.2400 (Table 4,
The main source of discrepancy is that Flinn abstracts from on-the-job wage growth, and attributes all of the growth between the starting wage on the \(n\)-th job in the cycle to the starting wage on the \(n+1\)-th job in the cycle to a better price from the underlying offer distribution. It would take only two years for a worker to accumulate enough experience to account for Flinn’s estimate for how much a worker gains on average when he changes jobs.

While the data fails to reject that the wage offer distribution is Pareto, the second row in the bottom panel of Table 4 reveals that it does reject the hypothesis that the offer distribution is lognormal. Again, we can test against all lognormal distributions, and then proceed to identify the exact distribution if we fail to reject this hypothesis. In particular, if log wage offers were distributed as \(N(\mu, \sigma^2)\), the average net wage growth among workers who move from their \(n\)-th job to their \(n+1\)-th job would equal

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

(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 log-normal, the sequence \(\{\pi_n\}_{n=1}^\infty\) will be determined up to a scale parameter \(\sigma\). Given a value for \(p\), we can readily compute what this sequence should be. As can be seen from the last line of Table 4, the probability that wage gains would depart from their theoretical values under the null of a lognormal distribution as much as they do is 0.014, so we can reject this null at conventional levels of significance. This calculation does not incorporate uncertainty in our estimate for \(p\), but since the moment sequence in (4.2) is sharply declining for a variety of different \(p\), and since \(p\) is relatively tightly estimated, the rejection of the normal is likely to be robust to incorporating sampling error.\(^8\) The intuition comes from Figure 1; if the distribution were log-normal, wage gains should decline sharply with \(n\), contrary to what we see in the data. We can similarly reject other functional form that imply a rapidly diminishing wage gains. Instead, the fact that wage gains do not seem to vary much with \(n\), at least for \(n \leq 5\), thus suggests the offer distribution is Pareto.

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 are similarly informative about the offer distribution. Consider a worker who is forced out of his \(n\)-th job. The total number of jobs in the employment cycle

\(^8\)Note that under the null hypothesis that the log wage offer distribution is exponential distribution, the \(\pi_n\) would not depend on \(p\), so there is no need to adjust for sampling error in our estimate for \(p\).
he just finished is \( n \), implying the average price per unit labor on his previous job is equal to 
\[ E(R_n \mid N = n) \], the expected value of the \( n \)-th record conditional on exactly \( n \) records in the sequence \( \{X_m\}_{m=1}^{M} \). Similarly, the average price on his new job is equal to 
\[ E(R_1 \mid N \geq 1) \], the average value of the first record conditional on setting at least one record. Since every employment cycle has a first record, this is just \( E(R_1) \), the unconditional value of the first observation. Hence, the average net wage loss for this worker can be expressed as a difference of record moments
\[
E(\Delta \ln W^n_{it} - \beta \Delta Z_{it} \mid N_{t-1} = n) = E(R_n \mid N = n) - E(R_1)
\]
Adapting results in Nagaraja and Barlevy (2003), one can show that the sequence of moments 
\( \{E(R_n \mid N = n) - E(R_1)\}_{n=1}^{\infty} \) uniquely identify the parent distribution in the set of continuous 
distribution functions, up to a location parameter. To put it another way, the wage gains of 
voluntary job changers should uniquely line up with the wage losses of involuntary job changers.

As in the previous section, we can recover the respective record moments using a single wage 
regression. Let \( D^{n,1}_{it} \) denote a dummy which equals \(-1\) if worker \( i \) moved from his \( n \)-th job in date \( t - 1 \) to a job where \( n \) is reset to 1 by date \( t \). Then the coefficients \( \pi_n \) in the regression
\[
\Delta \ln W^n_{it} = \beta_1 \Delta EXP_{it} + \beta_2 \Delta EXP^2_{it} - \sum_{n=1}^{\infty} \pi_n D^{n,1}_{it} + \Delta \varepsilon_{it}
\] (5.1)
are unbiased estimators for \( E(R_n \mid N = n) - E(R_1) \). The first column in Table 5 reports the 
estimated \( \pi_n \). According to the model, this sequence should be monotonically increasing. It does 
rise with \( n \) between \( n = 1 \) and 4, although the point estimate for \( \pi_5 \) lies between \( \pi_1 \) and \( \pi_2 \).

To further check if the offer distribution is Pareto, I test whether the coefficients \( \pi_n \) are propor-
tional to \( E(R'_n \mid N = n) - E(R'_1) \), 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
\[
\{0.197, 0.762, 1.127, 1.396, 1.616, ...\}
\] (5.2)
The bottom panel of Table 5 reports the probability of observing deviations from this prediction 
that are at least as large as in the data under the null hypothesis that the log offer distribution 
is exponential. Once again, we fail to reject the null hypothesis that the log offer distribution 
is exponential. To further confirm that the offer distribution is consistent with the Pareto we 
identified from the wage gains of voluntary job changers, I replace \( \sum_{n=1}^{\infty} \pi_n D^{n,1}_{it} \) in (5.1) with
\[
\pi D_{it} = \pi \sum_{n=1}^{\infty} (E(R'_N \mid N = n) - E(R'_1)) D^{n,1}_{it}
\]
and regress
\[
\Delta \ln W_{it}^n = \beta_1 \Delta EXP_{it} + \beta_2 \Delta EXP_{it}^2 - \pi D_{it} + \Delta \varepsilon_{it}
\] (5.3)
The model implies \(\pi\) should equal the same inverse hazard we estimate from voluntary job changers. As reported in the second column of Table 5, it is equal to 0.0816, nearly identical to the value of 0.0806 implied by the wage gains of voluntary job changers. In other words, the wage losses of involuntary job changers coincide with the wage gains we observe for voluntary job changers.

We can similarly test whether the wage losses of involuntary job changers are consistent with a normal log offer distribution. If this hypothesis is true, the coefficients \(\pi_n\) would be proportional to the sequence of record moments \(E (R'_n | N = n) - E (R'_1)\) where \(R'_n\) now 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 fail to reject this hypothesis. The reason for this is illustrated in Figure 3, which shows the estimated net wage loss together with the best-fitting values for \(E (R_n | N = n) - E (R_1)\) for the normal and the exponential distribution. Although the two sequences are distinguishable, it is difficult to distinguish between the exponential and the normal distribution from the wage losses of involuntary job changers given they are both monotonically increasing and concave. By contrast, the profiles for the wage gains of voluntary job changers seen in Figure 1 are much easier to tell apart.

In sum, the average amount workers lose when they are laid off is similar to the amount they should lose given their previous job mobility; workers who are laid off from their first job lose very little, while workers who are laid off after changing jobs several times lose more, roughly in line with an exponential log offer distribution. It is tempting to infer from this that returns to specific human capital must therefore be small; after all, if workers accumulated employer specific human capital, they lose both because they have to resume searching from scratch and because they will be less productive on any new job that they start. This reasoning proves to be inaccurate. Nevertheless, we can use the wage losses of involuntary job changers to bound the returns to specific human capital \(\gamma\), and check if they are indeed small.

To appreciate why the wage losses of displaced workers need not be larger in the presence of human capital than in its absence, let us return to the specification in (3.2) in which a worker’s productivity depends on the time he spent with his current employer. Consider a worker who is laid off from his first job in an employment cycle. The wage loss for this worker is
\[
\Delta \ln W_{it} = \omega''_1 - \omega'_1 + \beta_1 + \beta_2 (2X_{it} - 1) - \gamma T_{i,t-1} + \Delta \varepsilon_{it}
\] (5.4)
where \(\omega'_1\) denotes the log price per unit labor on the first job in his first employment cycle and \(\omega''_1\)
denotes the log price per unit labor on the first job in his second employment cycle. At the same time, the wage growth of workers who remain on the same job is given by

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

Thus, the average wage change for involuntary job changers net of the estimated wage growth for job stayers is

$$E \left[ |\Delta \ln W_{it} - \hat{\beta}_1 - \hat{\beta}_2 (2X_{it} - 1)| \right| N_{t-1} = 1] = E (\omega'_1 | N_{t-1} = 1) - E (\omega''_1) + \gamma E (T_{i,t-1} + 1 | N_{t-1} = 1)$$

(5.5)

If $\gamma$ is strictly positive, $\gamma E (T_{i,t-1} + 1 | N_{t-1} = 1)$ will be as well. This is consistent with the intuition outlined above: a worker who is laid off not only has to restart at a new wage $\omega''_1$, but will be less productive on his new job than on his old one. This contributes to a larger wage loss than in the absence of specific human capital. But at the same time, the expected price on the job in the first cycle, $\omega'_1$, might be lower when $\gamma > 0$ than when $\gamma = 0$. In particular, when $\gamma = 0$, $E (\omega'_1 | N_{t-1} = 1) = E (R_1 | N = 1)$, the average value of the first record from a sequence of i.i.d. draws conditional on only exactly one record. Moreover, $E (R_1 | N = 1) > E (R_1)$, i.e. the average wage for a worker who was laid off from his first job will be higher than the average wage on a brand new job. Intuitively, the fact that the worker was only employed on one job prior to being laid off makes it less likely that he drew a low offer that could be easily beat. By contrast, when $\gamma > 0$, the worker will switch jobs only if the price on the new offer exceeds the price he had on his current job by enough to compensate him for turning less productive. Suppose returns to human capital are so large that once the worker accepts his first job, he will be unlikely to encounter another job that pays enough to compensate him for his foregone human capital. In that case, $E (\omega'_1 | N_{t-1} = 1) = E (\omega'_1)$, since the fact that the worker was employed on only one job is uninformative. Thus, a worker who is laid off from his first job may not appear to lose as much when $\gamma > 0$ than when $\gamma = 0$.

Nevertheless, we can still use the average wage loss of workers who involuntarily leave their first job to derive an upper bound for $\gamma$. In particular, we can use the fact that $E (\omega'_1 | N_{t-1} = 1) \geq E (\omega'_1)$, i.e. the average price per unit labor on a worker’s first job conditional on being laid off from a first job can be no less than the unconditional average price on a first job. Thus, the average net wage loss is given by

$$E \left( |\Delta \ln W^1_{it} - \hat{\beta} \Delta Z_{it}| \right| N_{t-1} = 1) = E (\omega'_1 | N_{t-1} = 1) - E (\omega''_1) + \gamma E (T_{i,t-1} + 1 | N_{t-1} = 1) \geq E (\omega'_1) - E (\omega''_1) + \gamma E (T_{i,t-1} + 1 | N_{t-1} = 1)$$

$$= \gamma E (T_{i,t-1} + 1 | N_{t-1} = 1)$$
Rearranging yields the following upper bound on the returns to tenure $\gamma$:

$$
\gamma \leq \frac{E \left( |\Delta \ln W_{l_1}^1 - \hat{\beta} \Delta Z_{l_1}| \mid N_{t-1} = 1 \right)}{E \left( T_{i,t-1} + 1 \mid N_{t-1} = 1 \right)}
$$

(5.6)

To estimate this bound, note that the coefficient $\pi_1$ from (5.1) is an unbiased estimator for the numerator in (5.6). Let $\bar{T}_1$ denote the average tenure in the sample for workers who were laid off from their first job. Then $\bar{T}_1 + 1$ forms an unbiased estimator for the denominator in (5.6). A natural estimator for the upper bound in (5.6) is the ratio of the two, i.e. $\hat{\gamma}_1 = \frac{\pi_1}{\bar{T}_1 + 1}$.

The first row of Table 6 shows how this estimator is constructed. Column (1) repeats the value of $\pi_1$ from Table 5. Column (3) reports the average tenure of workers who are laid off from their first job, which is 1.28 years. The implied value of $\hat{\gamma}_1$ is therefore 0.001, reported in column (4) of the table, even smaller than I obtain using Topel’s approach. To construct a confidence interval for $\hat{\gamma}_1$, I apply the delta method to derive an asymptotic standard error for $\hat{\gamma}_1$. That is, $\hat{\gamma}_1$ must asymptotically converge to a normal random variable with variance

$$
\sigma^2_{\hat{\gamma}_1} = \frac{\text{Var}(\pi_1) + 2\hat{\gamma}_1 \text{Cov}(\pi_1, \bar{T}_1) + \hat{\gamma}_1^2 \text{Var}(\bar{T}_1)}{\bar{T}_1 + 1}^2
$$

I can then estimate $\text{Var}(\pi_1)$ using the standard error for $\pi_1$. For my estimate of $\text{Var}(\bar{T}_1)$, I use the sample variance for tenure across all workers who are laid off from their first job, divided by the number of such workers. The more difficult part is how to estimate $\text{Cov}(\pi_1, \bar{T}_1)$. Using a Monte Carlo simulation, I verified that, at least under the null of $\gamma = 0$, this covariance term is positive but negligible, as is $\text{Var}(\bar{T}_1)$. Essentially, $\bar{T}_1$ is estimated sufficiently precisely that sampling error has only a negligible effect on our estimate for $\hat{\gamma}_1$, so we can essentially treat $\bar{T}_1$ as a known constant. Using a one-tailed t-test, we can reject that $\gamma$ 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 the difference between this approach to bounding the returns to tenure and the approaches of Altonji and Shakotko (1987) and Topel (1991) for directly measuring $\gamma$.

What about workers who are laid off from jobs later on in their respective employment cycles? We could use their wage losses to construct bounds on $\gamma$ as well. However, unlike the case of workers who are displaced from their first job, the bounds depend on the exact shape of the wage offer distribution. Since the previous section suggested the data is consistent with a Pareto offer distribution, we could proceed with this as an assumption. This is not quite legitimate, though, since this identification relies on the assumption that $\gamma = 0$. But suppose we had independent evidence that the wage offer distribution was Pareto which did not require any assumptions on $\gamma$.  

29
In this case, we can show that for any \( n \),
\[
E(\omega_n \mid N_{t-1} = n) - E(\omega_1) \geq E(R_n \mid N \geq n) - E(R_1)
\]
where \( R_n \) denotes the \( n \)-th record as conventionally defined. Using this, we can similarly construct an upper bound \( \hat{\gamma}_n \) based on the wage losses of workers who lose their \( n \)-th job
\[
\hat{\gamma}_n = \frac{\pi_n - (E(R_n \mid N \geq n) - E(R_1))}{T_n + 1}
\]
where \( \pi_n \) is defined in (5.1), \( T_n \) 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 for an exponential distribution with mean 0.0816 and \( p = 0.42 \). Table 6 reports the point estimates for \( \hat{\gamma}_n \). For \( n = 2, 3, \) and \( 4 \), we can assign a 95% probability that \( \gamma \leq 0.023 \), half as large as Topel’s estimate. This is suggestive, but given that it \( \hat{\gamma}_n \) for \( n > 1 \) uses a distribution that we identified by ruling out the case where \( \gamma > 0 \), these bounds are less compelling than the bound based on workers who are laid off from their first job and is independent of the exact shape of the offer distribution.

In sum, the wage growth of the young workers in my sample appear to be driven by the accumulation of general human capital and on-the-job search, and the wage losses of involuntary job changers can be entirely explained as the consequence of having to resume search from scratch. At first, this might seem at odds with the work of Neal (1999), which uses NLSY data to argue that young workers tend to search in stages, i.e. workers first settle on a career and then search for across employer where they can work in this career. We would therefore expect that laid-off workers should not resume searching from scratch, since they can focus their search on jobs in the career they learned they are best suited for. However, Neal also finds that workers tend to settle on a career fairly quickly; for example, he reports that 67% of workers who start a new career change careers on their first job change, but only 48% of those with two consecutive jobs in the same career will change careers on their next job change.\(^9\) Since I only assign \( n \) after the worker reported his first involuntary job change, many of the workers in my sample may have already settled on a career before their first layoff, and what I estimate corresponds to the distribution of potential earnings within a career rather than across careers.

6. Alternative Models of Search

The approach to identification described in this paper relies heavily on the fact that observed wages over an employment cycle \( \{W_n^i\} \) correspond to a sequence of records \( \{w_n\}_{n=1}^N \) from a set of price

\(^9\)These estimates are based on the corrected estimates of Neal (1999) reported in a 2004 erratum.
offers \( \{X_m\}_{m=1}^M \). This will not always be the case for richer search models. Nevertheless, I now argue that even when the underlying sequence \( \{w_n\}_{n=1}^N \) does not correspond to a list of record values, it will often be the case that a model with on-the-job search has some implicit record structure which we might be able to exploit for purposes of identification. The case where returns to tenure \( \gamma \) are positive provides one example. There, \( \{w_n\}_{n=1}^N \) do 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), but where the observations \( \{X_m\}_{m=1}^M \) are not identically distributed. Even with employer specific human capital, search models give rise to a record structure that could potentially be used for identification.

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, 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 a record from i.i.d. draws from the implied distribution for income across all offers, and we can adapt my approach to identify the distribution of income levels across job offers from observations on income \( I_{it} = (w_n h_n) \times \ell_{it} \).

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 precise estimates of this distribution, the analysis is still informative: it rejects the lognormal distribution as a candidate for the offer
distribution, and suggests the offer distribution may be Pareto. This result is distinct from the oft-noted fact that the cross-sectional distribution of wages exhibits a Pareto tail.\textsuperscript{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. Interestingly, appealing to this implicit records structure also proves useful for constructing bounds on the returns to tenure, offering an alternative approach to what has been emphasized in previous literature. For my sample of young workers, I conclude that these returns are not economically meaningful.

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.

\textsuperscript{10}On the presence of a Pareto tail in cross-sectional earnings distributions, see Neal and Rosen (2000).
### Table 1: Summary Statistics for Entire Sample

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

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 $\gamma$

#### Linear returns to tenure

<table>
<thead>
<tr>
<th>within-job wage growth</th>
<th>experience effect</th>
<th>tenure effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 + \gamma$</td>
<td>$\beta_1$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>0.0794</td>
<td>0.0740</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.0065</td>
<td>0.0061</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1 year</th>
<th>2 years</th>
<th>5 years</th>
<th>7 years</th>
<th>10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>implied returns to tenure</td>
<td>0.0054</td>
<td>0.0108</td>
<td>0.0271</td>
<td>0.0380</td>
<td>0.0542</td>
</tr>
<tr>
<td></td>
<td>0.0024</td>
<td>0.0049</td>
<td>0.0122</td>
<td>0.0171</td>
<td>0.0245</td>
</tr>
</tbody>
</table>

|                       | 0.0723 | 0.1411 | 0.3270 | 0.4337 | 0.5680 |
|                       | 0.0058 | 0.0109 | 0.0226 | 0.0274 | 0.0300 |

#### Quadratic returns to tenure

<table>
<thead>
<tr>
<th>within-job wage growth</th>
<th>experience effect</th>
<th>tenure effect</th>
<th>tenure squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 + \gamma_1$</td>
<td>$\beta_1$</td>
<td>$\gamma_1$</td>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>0.0826</td>
<td>0.0661</td>
<td>0.0165</td>
<td>-0.0016</td>
</tr>
<tr>
<td>0.0065</td>
<td>0.0067</td>
<td>0.0024</td>
<td>0.00048</td>
</tr>
</tbody>
</table>

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 $\Delta X (= \text{constant})$ and $\Delta X^2$. This is the same regression in column (1) of Table 4, where $\beta_1 + \gamma$ corresponds to the coefficient on $\Delta X$. The second stage regresses the log real wage net of the estimated $(\beta_1 + \gamma)T + \beta_2X^2$ on initial experience and individual fixed-effects. The coefficient on initial experience corresponds to the estimate of $\beta_1$, and the difference corresponds to the estimate of $\gamma$ above. Standard errors for $\beta_1$ and $\gamma$ 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 $\gamma$, $\beta_1$, and $\beta_2$. In the bottom panel, the first stage regression is amended to allow for a $\Delta T^2$ term, which is then subtracted from the log real wage at the second stage.
Table 3: Estimates for p

<table>
<thead>
<tr>
<th>Sample size</th>
<th>p</th>
<th>Standard error</th>
<th>Implied $\lambda_{i}/\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>22,135</td>
<td>0.4823</td>
<td>0.0031</td>
</tr>
<tr>
<td>Educ &lt; 12</td>
<td>6,515</td>
<td>0.5008</td>
<td>0.0055</td>
</tr>
<tr>
<td>Educ = 12</td>
<td>6,648</td>
<td>0.4797</td>
<td>0.0058</td>
</tr>
<tr>
<td>Educ ∈ (13,15)</td>
<td>5,436</td>
<td>0.4504</td>
<td>0.0062</td>
</tr>
<tr>
<td>Educ ≥ 16</td>
<td>3,536</td>
<td>0.5049</td>
<td>0.0082</td>
</tr>
</tbody>
</table>

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 $p = (1 + \lambda_{i}/\delta)^{-1}$. 
Table 4: The Wage Gains of Voluntary Job Changers, by $n$

<table>
<thead>
<tr>
<th></th>
<th>sample size</th>
<th>(1)</th>
<th>(2)</th>
<th>(3) exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>∆X</td>
<td>∆X$^2$</td>
<td>∆X$^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0767</td>
<td>0.0809</td>
<td>0.0816</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0046</td>
<td>0.0050</td>
<td>0.0050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.0016</td>
<td>-0.0018</td>
<td>-0.0018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0090</td>
<td>0.0090</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0711</td>
<td>0.0711</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0137</td>
<td>0.0137</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0799</td>
<td>0.0799</td>
<td>0.0806</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0200</td>
<td>0.0200</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0331</td>
<td>0.0331</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0799</td>
<td>0.0799</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0520</td>
<td>0.0520</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th># obs</th>
<th>stayers</th>
<th>changers</th>
<th>stayers</th>
<th>changers</th>
<th>stayers</th>
<th>changers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27,712</td>
<td>27,712</td>
<td>27,712</td>
<td>31,868</td>
<td>27,712</td>
<td>27,712</td>
<td>27,712</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4,156</td>
<td>4,156</td>
<td>31,868</td>
<td>27,712</td>
<td>27,712</td>
<td>27,712</td>
</tr>
</tbody>
</table>

Test of particular functional forms:

<table>
<thead>
<tr>
<th></th>
<th>Exponential</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F(4, 31861) = 1.31$</td>
<td>$F(4, 31861) = 3.12$</td>
</tr>
<tr>
<td></td>
<td>Prob &gt; $F = 0.2639$</td>
<td>Prob &gt; $F = 0.0140$</td>
</tr>
</tbody>
</table>

The dependent variable is the annual growth rate of real wages. The independent variables are the growth $\Delta$EXP, which is identically equal to 1, $\Delta X^2$, which is equal to $2X - 1$, and a set of dummy variables $D^{n\rightarrow n+1}$ 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 $\Delta X$ and $\Delta X^2$ using job stayers only. Column (2) adds job changers and estimates the coefficients on the dummy variables as well. Column (3) estimates the same regression as in column (2) assuming the coefficients on all the dummy variables are equal, which from the text is true if and only if the log wage offer distribution is exponential. The 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

<table>
<thead>
<tr>
<th>sample size</th>
<th>(1)</th>
<th>(2) exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔX</td>
<td>--</td>
<td>0.0837</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0062</td>
</tr>
<tr>
<td>ΔX^2</td>
<td>--</td>
<td>-0.0020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0002</td>
</tr>
<tr>
<td>D11</td>
<td>2,767</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0094</td>
</tr>
<tr>
<td>D21</td>
<td>873</td>
<td>0.0843</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0153</td>
</tr>
<tr>
<td>D31</td>
<td>305</td>
<td>0.0904</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0278</td>
</tr>
<tr>
<td>D41</td>
<td>137</td>
<td>0.0942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0432</td>
</tr>
<tr>
<td>D51</td>
<td>50</td>
<td>0.0754</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0726</td>
</tr>
<tr>
<td># obs</td>
<td></td>
<td>31,844</td>
</tr>
<tr>
<td>stayers</td>
<td></td>
<td>27,712</td>
</tr>
<tr>
<td>changers</td>
<td></td>
<td>4,132</td>
</tr>
</tbody>
</table>

Test of particular functional forms:

<table>
<thead>
<tr>
<th></th>
<th>Exponential</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>F (4, 31837) = 1.24</td>
<td>Prob &gt; F = 0.2895</td>
<td></td>
</tr>
<tr>
<td>F (4, 31837) = 1.08</td>
<td>Prob &gt; F = 0.3622</td>
<td></td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>n</th>
<th>( \pi_n )</th>
<th>( \frac{E(R_n \mid N \geq n) - E(R_1)}{\bar{T}_n} )</th>
<th>( \frac{\gamma}{\hat{\gamma}_n} )</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0029</td>
<td>0.0000</td>
<td>1.28</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.0843</td>
<td>0.0475</td>
<td>1.66</td>
<td>0.014</td>
</tr>
<tr>
<td>3</td>
<td>0.0904</td>
<td>0.0797</td>
<td>1.52</td>
<td>0.004</td>
</tr>
<tr>
<td>4</td>
<td>0.0942</td>
<td>0.1036</td>
<td>1.70</td>
<td>-0.003</td>
</tr>
<tr>
<td>5</td>
<td>0.0754</td>
<td>0.1224</td>
<td>1.39</td>
<td>-0.020</td>
</tr>
</tbody>
</table>

Assuming log offer distribution is exponential with mean 0.0816

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 coefficients 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


