Identification of Search Models with Initial Condition Problems

Gadi Barlevy
H. N. Nagaraja

Identification of Search Models
with Initial Condition Problems

Gadi Barlevy
*Federal Reserve Bank of Chicago
and IZA Bonn*

H. N. Nagaraja
Ohio State University

ABSTRACT

Identification of Search Models with Initial Condition Problems[*]

This paper extends previous work on the identification of search models in which observed worker productivity is imperfectly observed. In particular, it establishes that these models remain identified even when employment histories are left-censored (i.e. we do not get to follow workers from their initial job out of unemployment), as well as when workers set different reservation wages from one another. We further show that allowing for heterogeneity in reservation can affect the empirical estimates we obtain, specifically estimates of the rate at which workers receive job offers.

Corresponding author:

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

Introduction

This paper derives new results on the identification of job search models with imperfectly measured worker productivity. Various authors have argued that search models are helpful for understanding labor markets, especially for young workers who are still in the beginning stages of their careers. According to these models, a worker's wage depends on which random employers he happened to meet, along with his own productivity. We are interested in whether the key elements of these models can be identified when worker productivity is not perfectly observable. That is, can we infer the distribution of wages a worker of *given* productivity could earn on various jobs? Can we infer the rate at which workers receive offers from how often they choose to change jobs?

There are several applications that require recovering these objects. For example, various search models imply that the distribution of wages available to workers reflects underlying heterogeneity in productivity across producers (where less productive employers can still compete given search frictions). In this case, the distribution of wages reveals the degree of productive inefficiency across employers, which we can then use to evaluate the effects of macro shocks or policy changes on aggregate productivity. As another example, these estimates make it possible to quantify the role of mobility in wage growth and distinguish it from the role of experience which makes workers more productive. We can also examine whether workers in different job markets face different wage distributions and opportunities for job mobility.

As previous authors pointed out, identifying search models is straightforward when worker productivity is relatively homogeneous or else perfectly observable. In these cases, we can recover the distribution of wages for workers of given productivity from the empirical distribution of wages among such workers on their first job out of unemployment. To infer the rate at which workers receive offers, we can look at how the duration of a job varies with its wage. Intuitively, since lower offers are easier to beat, workers who draw low wages will leave more quickly, at a rate that depends in a precise way on the distribution of wages offered and how frequently offers arrive.

Once we acknowledge that worker productivity is hard to measure, though, this approach no longer applies. For one thing, we can no longer determine whether dispersion in wages in the first job out of unemployment reflects dispersion in worker ability or in the prices employers pay. Yet it is important that we be able to distinguish the two. For example, as noted by Flinn (2002) and Bowlus and Robin (2004), whether mobility can alleviate long-run earnings inequality ultimately depends on the source of dispersion: if wage differentials were due to price dispersion, increasing mobility would allow workers who drew low initial offers to catch up to those who drew high offers;

if wage differentials instead reflected differences in ability, increasing mobility would have no effect on inequality. Moreover, if we cannot tell whether a worker earning a low wage is less productive or just happened to meet a low paying employer, we will not be able to use the fact that workers tend to leave lower paying jobs more quickly to recover the rate at which offers arrive.

One way to cope with unobserved worker productivity is to impose parametric restrictions on the distribution of prices and unobserved abilities. However, since the conclusions we draw will be sensitive to these restrictions, we would like to know whether we can estimate these models without imposing them. At the very least, establishing that the model is identified can suggest ways of testing particular functional forms before incorporating them in parametric estimation.

Work by Nagaraja and Barlevy (2003) and Barlevy (2005) demonstrated that these models can be identified without requiring parametric assumptions. The insight in these papers is to use an approach similar to the one used for nonparametric identification of auction models, e.g. Athey and Haile (2002). Intuitively, the wages on different jobs a worker accepts represent extreme values among the offers he received, much like winning bids in auctions represent extreme values among the bids tendered. Fortunately, extreme values often characterize the distributions of variables from which these extremes are taken. Yet while the approach in the auction literature is static and concerns extremes among bids already submitted at some point in time, search models require a dynamic approach that involves following a worker over time and keeping track of how often he changes jobs and how his wages rise with mobility. Formally, these papers rely on results for record statistics rather than order statistics. One can show that the distribution of wages available to a worker of given ability can be identified from the way average wage gains of job changers vary with past mobility (i.e. the average gaps between different consecutive record values) rather than from the distribution of wages on the first job out of unemployment. Similarly, the rate at which workers receive offers can be identified from the number of jobs between unemployment spells (i.e. the number of record values) rather than by relating the duration of a job to its wage.

Unfortunately, the identification results cited above rely on various assumptions on both the model and the data used for identification. One such assumption is that when we first observe a worker, his wage represents a random draw from the distribution of wages a worker of his ability could earn. However, there are various relevant scenarios in which this assumption will be violated. For example, some datasets interview workers who are already employed at the time the survey began. In this case, the wage on the first job we observe for a worker is not a random draw from the distribution of wages a worker of his ability could earn, since some workers who drew low wages for their first job out of unemployment are likely to have moved on to a higher wage job before we first observe them. Restricting attention only to workers who leave unemployment within

2

the survey window dramatically reduces sample sizes and makes identification impractical. As another example, suppose we do get to observe workers as of their first job out of unemployment, but workers value leisure differently and therefore set different reservation wages. In that case, a worker's wage on his first job will typically be drawn not from the distribution of wages a worker of his ability could earn, but from a truncated version of this distribution.

The purpose of this paper is twofold. First, we establish that these models remain identified even in the face of such initial condition problems, i.e. situations in which the wage on the initial job we observe for a worker is not a random draw from the distribution of wages a worker of equal ability could earn. Second, using data from the National Longitudinal Survey of Youth (NLSY), we show that allowing for heterogeneity in reservation wages can affect inference. In particular, allowing for heterogeneity in reservation prices leads us to estimate a much higher rate at which employed workers receive offers in the NLSY than if we assume all workers share the same reservation price. This result is due to certain distinguishing features of the empirical distribution of the number of consecutive job-to-job transitions, patterns which previous research has overlooked. Interestingly, our estimate for the rate at which employed workers receive offers when we accommodate heterogeneity in reservation wages is on par with what previous authors have found using job duration data, although our estimate for the rate at which unemployed workers receive offers is much larger than previous authors have estimated.

The paper is organized as follows. Section 1 describes the search model we use. Section 2 derives the analytical results we require for identification. Section 3 applies our results to the case of left-censored work histories. Section 4 considers the case of unobserved differences in reservation prices across workers. In Section 5, we discuss the effects of heterogeneity in reservation prices on empirical estimates of mobility. Section 6 comments on the implications of heterogeneity in reservation prices for empirical estimates of the price offer distribution. We conclude in Section 7.

1. A Model of On-the-Job Search

There is already a vast literature on estimating search models using data from worker histories; for a recent survey of this literature, see Eckstein and van den Berg (2005). We follow previous authors in modeling search as a process in which workers periodically draw a fixed price from some distribution and choose optimally among the offers available at each point in time. However, we also allow a worker's productivity to vary over time, in a way that the econometrician may not be able to measure. This feature allows the model to be consistent with two facts that a model with fixed productivity would be unable to match. First, real wages tend to vary over the course of a

3

job, which cannot occur in our model if both the price on the job and the worker's productivity are constant. Second, workers occasionally earn less on a job they move to voluntarily than on the job they left. This could occur if the worker's productivity fell just when he happened to change jobs, but not if his productivity were constant over time. Our assumption thus plays the same role as measurement error in previous work.[1] But unlike previous work which modeled measurement error as a sequence of i.i.d. normal random variables, under our interpretation it is less obvious how to model unmeasured variation in wages, which motivates our nonparametric approach.

We begin with a description of workers. All workers supply a homogeneous labor input, but the amount they supply varies across workers and over time. Let ℓ_{it} denote the amount of labor worker i can supply per hour at date t. This amount is observable to both the employer and the worker, but is only imperfectly observable to the econometrician who collects data on this market. We follow Flinn (1986) by modelling this productivity as a log-linear function:

$$\ell_{it} = \exp\left(\beta X_{it} + \phi_i + \varepsilon_{it}\right) \tag{1.1}$$

The first term, X_{it}, represents characteristics of individual i that are observable to the econometrician, and β represents the returns to these characteristics. The next term, ϕ_i, is fixed over time, reflecting variations in innate ability that make some workers consistently more productive than others. We do not require this term to be observable to the econometrician. The last term, ε_{it}, denotes variation in the worker's productivity over time which cannot be directly measured, but can also include measurement error due to misreporting.

We impose the following assumptions on ε_{it}:

Assumption 1.1: $\{\varepsilon_{it}\}_{t=0}^{\infty}$ is independent of $\{X_{it}\}_{t=0}^{\infty}$ for any fixed i

Assumption 1.2: $E\left[E_i\left[\Delta\varepsilon_{it}\right]\right] \equiv E\left[\int_i \left(\varepsilon_{it} - \varepsilon_{i,t-1}\right) di\right] = 0$ for all dates t

Assumption 1.1 states that the observable and unobservable components of an individual's productivity are independent. The role of this assumption will become clear below. Assumption 1.2 states that if we were to average the change in ε_{it} across all workers, the unconditional expectation of this average would equal zero. This assumption is essentially a normalization; in general,

---

[1]That is, in our model a voluntary job changer will report a wage cut solely because of how wages are measured, not because he moved to a lower paying job. There are search models in which workers choose to accept lower paying jobs; one example is Postel-Vinay and Robin (2002), in which workers accept a lower wage from a more productive employer, correctly anticipating higher wage growth with such an employer. But their model also requires wages be imperfectly measured to accord with the fact that wages both rise and fall over the duration of a job. Of course, once we allow for mismeasurement, we reintroduce the possibility that most wage declines we observe are spurious.

$E\left[\int_i \Delta\varepsilon_{it} di\right]$ is a function of t, and we can always include a semiparametric function of t in X_{it} to yield a residual that satisfies Assumption 1.2. Note that we impose no restrictions on the distribution of ε_{it}, how its variance varies over time or across individuals, or its autocorrelation.

At each point in time, a worker can be classified as either employed or unemployed. He can produce $b_i\ell_{it}$ units of output per hour while unemployed, where b_i is fixed over time and represents either productivity in the home sector or the marginal value of leisure. Assuming this value is proportional to ℓ_{it} allows us to abstract from selection among employed workers. Without this assumption, workers who became less productive on the job but not at home would opt out of the labor market to enjoy leisure, and the distribution of changes in productivity $\Delta\ell_{it}$ of employed workers would not be representative of the changes among all workers. In support of this assumption, we note that recent work by Low, Meghir, and Pistaferri (2004) allowed for selection in a similar model but found it to have a negligible impact on overall fit. Nagaraja and Barlevy (2003) and Barlevy (2005) assume $b_i = b$ for all i. One of the goals of this paper is to relax this assumption, although for purposes of exposition we occasionally invoke it as well.

While unemployed, a worker receives job offers at rate λ_0 per unit time. If he accepts an offer and becomes employed, he continues to receive offers at a (possibly different) rate λ_1. In addition, he may be forced to leave his job, an event that occurs at rate δ. We assume a worker cannot recall offers he turned down, so a worker who loses his job becomes unemployed.

While working on a job, a worker supplies all of his labor to produce final goods using a linear technology specific to that job. Let z_{ij} denote the productivity of worker i on job j. Thus, he will produce $z_{ij}\ell_{it}$ units of output per hour. We assume z_{ij} on any match is fixed over time. This assumption effectively rules out match specific human capital; if a worker becomes more productive on one job, it must reflect changes in ℓ_{it} that by construction carry over to all jobs. Barlevy (2005) argues this specification is empirically plausible for young workers, since most of the wage growth of young workers on a job does appear to carry over to other jobs as well.

Let $\Gamma_i(\cdot)$ denote the cdf of productivity z_{ij} for worker i across all jobs. We assume $\Gamma_i(\cdot) = \Gamma(\cdot)$ for all i, i.e. all workers face the same production possibilities. This assumption naturally follows if we assume $z_{ij} = z_j$ for all i, i.e. productivity is job-specific rather than match-specific. We appeal to this assumption for simplicity. However, it is not essential to interpret the model this way. For example, Marimon and Zilibotti (1999), Barlevy (2002), and Gautier, Teulings, and van Vuuren (2005) develop models in which workers enjoy a comparative advantage on different jobs rather than being all best suited to the same job, but the symmetry of those models ensures the distribution of productivity across jobs is the same for all workers.

Upon meeting, a worker and employer observe match productivity z, and the latter makes an offer to the former. The exact nature of the offer will depend on how workers and employers interact. For the most part, we focus on models in which this interaction implies that all employers with the same z_j will offer in equilibrium the same constant price w_j per unit labor, independently of that worker's productivity ℓ_{it}. The hourly wage for the worker is hence

$$W_{ijt} = w_j \ell_{it} \tag{1.2}$$

Under this assumption, jobs can be ranked according to the price they pay, and workers will seek out higher paying jobs, just as in the traditional search model with fixed worker productivity. One model that is consistent with (1.2) is the Lucas and Prescott (1974) model, modified to allow for on-the-job search. In that model, employers are located on islands. Workers can instantly move between employers on an island, and all employers on the same island are equally productive. Competition on an island will drive employers to pay workers their marginal product, i.e. $w_j = z_j$. Productivity is not the same across islands, but workers only get to visit other islands at random times. In an appendix, we show that our framework also corresponds to a model in which workers cannot instantly move to other equally productive employers, and must instead bargain with their employer over the wage they earn according to a particular bargaining protocol.

Although the distinguishing feature of a job is its productivity, in the models we consider a job can also be characterized by the wage it pays. Let $F(\cdot)$ denote the distribution of wages across all jobs. From the worker's perspective, each encounter with an employer represents an independent draw from $F(\cdot)$ that does not depend on his current productivity ℓ_{it}. Given our assumptions, all workers face the same $F(\cdot)$. In practice, workers with different skills may direct their search to different labor markets. However, as long as these skills are observable, we can always carry out our analysis for each skill group separately. Although we shall focus on identifying $F(\cdot)$, in certain cases we will be able to map $F(\cdot)$ into a distribution of productivity $\Gamma(\cdot)$.[2]

What is the optimal strategy for a risk-neutral worker who discounts the future at rate ρ and whose objective is to maximize the present discounted value of his earnings? Once he is employed, it is clearly optimal for him to only accept a new offer if it pays a higher price than his current job. While unemployed, he must decide at what price it is better to work and continue searching at rate λ_1 than to remain unemployed and search at rate λ_0. This problem has a well known

---

[2]Recovering $\Gamma(\cdot)$ from $F(\cdot)$ is analogous to recovering the distribution of valuations from the distribution of bids in an auction. The Lucas and Prescott model resembles a second-price auction, since in equilibrium employers bid z_j, the amount at which they value a worker (but unlike the second-price auction, they also pay what they bid). In other models, a firm's offer will be some function of z_j. In that case, recovering $\Gamma(\cdot)$ from $F(\cdot)$ can be more involved, just as mapping bids into valuations is more involved for first-price auctions than second-price auctions.

solution. As demonstrated, among others, by Mortensen and Neumann (1988), the optimal policy for the worker is to set a reservation price w_i^* that solves

$$w_i^* = b_i + (\lambda_0 - \lambda_1) \int_{w_i^*}^{\infty} \frac{1 - F(x)}{\rho + \delta + \lambda_1 (1 - F(x))} dx \tag{1.3}$$

If $\lambda_0 = \lambda_1$, the worker should work at any price that exceeds b_i. If instead $\lambda_0 > \lambda_1$, the worker should reject offers slightly higher than b_i and retain the option to search at a higher rate λ_0.

Let us define $\underline{w}^* = \inf_i w_i^*$ as the lowest reservation price across workers. In what follows, we will find it convenient to assume that the lowest price on any job, $\underline{w} = \inf_j w_j = F^{-1}(0)$, is higher than $\underline{w}^*$. In the wage setting models we described, this condition will automatically hold if $\lambda_0 \geq \lambda_1$ and if the lowest productivity level $\underline{z} = \inf_j z_j$ exceeds the value of leisure for the worker who least enjoys it, namely $\underline{b} = \inf_i b_i$. This assumption allows us to avoid the issue of recoverability in Flinn and Heckman (1982), i.e. that it is only possible to recover the distribution of prices below $\underline{w}^*$ for particular functional forms. However, this assumption is not essential, and we can talk about a somewhat modified notion of identification without it.[3]

This completes the description of the model. We wish to recover its key parameters – λ_0, λ_1, δ, and $F(\cdot)$ – using employment history data, i.e. the hourly wage $\{W_{ijt}\}_{t=0}^{\infty} = \{w_{ij}\ell_{it}\}_{t=0}^{\infty}$ for each worker, the duration of each job, and any occurrences of unemployment. [4] As noted earlier, we may be able to map $F(\cdot)$ to the distribution $\Gamma(\cdot)$, but our discussion will focus on recovering $F(\cdot)$.

Our approach to identification builds on the fact that the jobs the worker accepts correspond to "records" among the offers he receives. More precisely, following Wolpin (1992), let us partition the data for each worker into distinct *employment cycles*, where a cycle is defined as the time between unemployment spells. Let M denote the number of offers on an employment cycle, and let $\{y_m\}_{m=1}^{M}$ denote the prices per unit labor on the respective offers he receives. As demonstrated in Barlevy (2005), M has a geometric distribution, i.e.

$$\Pr(M = m) = \left(\frac{\lambda_1}{\lambda_1 + \delta}\right)^{m-1} \left(\frac{\delta}{\lambda_1 + \delta}\right) \tag{1.4}$$

---

[3] In particular, we can still identify the truncated distribution $F(w \mid w \geq \underline{w}^*)$ and the arrival rates $\lambda_0 (1 - F(\underline{w}^*))$ and $\lambda_1 (1 - F(\underline{w}^*))$ of "viable" offers (i.e. those that some worker might accept) rather than $F(w)$, λ_0, and λ_1.

[4] More recently, economists have assembled matched employer-employee datasets that allow us to track the wages of all employees at the same workplace. If employers paid the same price per unit labor on all jobs, we could recover the distribution of prices directly using employer fixed effects as in Abowd, Kramarz, and Margolis (1999). We assume the econometrician has no access to such data. Even if such data were available, employer fixed effects would not yield identification if the price varied by job (or by match for the same job) rather than by employer.

Define $L(1) = 1$, and for any integer $n > 1$ define $L(n)$ recursively as

$$L(n) = \min\left\{m : y_m > y_{L(n-1)}\right\} \tag{1.5}$$

To simplify matters, suppose all workers share the same value b for b_i and thus the same reservation price w^*. Since we assumed $\underline{w} \geq w^*$, workers accept their first offer y_1, and then any new offer that exceeds what they currently earn. Define N as the number of *actual jobs* the worker is employed on in a given cycle, so that $N \leq M$, and let $n \in \{1, ..., N\}$ index these jobs. Since a worker always chooses the job offering the highest price, the price on the n-th job in an employment cycle, w_n, must be given by

$$w_n = y_{L(n)}$$

In extreme-value theory, $L(n)$ is called the n-th record time and w_n is called the n-th record value. That is, w_n corresponds to the n-th time in the sequence $\{y_m\}_{m=1}^M$ in which a value exceeds all elements that preceded it. Hence, the prices on accepted jobs correspond to a sequence of record values. Likewise, the number of jobs in an employment cycle N corresponds to the number of record values in the sequence $\{y_m\}_{m=1}^M$. For comprehensive surveys on record-value theory, see Arnold, Balakrishnan, and Nagaraja (1998) and Nevzorov and Balakrishnan (1998). An important distinction between our application and much of the work in these surveys is that our application involves records from a random number of observations rather than an infinite sequence $\{y_m\}_{m=1}^\infty$.

Before we turn to what is new in this paper, let us review previous work that exploits the record structure of search models for purposes of identification. Barlevy (2005) offers the following multi-step procedure for the special case where employment histories are not censored and all workers share the same reservation price w^*. The first step involves recovering the ratio $\kappa_1 = \lambda_1/\delta$. From (1.4), the number of offers M on an employment cycle has a geometric distribution that depends on κ_1. One can show that this implies the distribution of the number of records N across employment cycles is itself uniquely determined by the parameter κ_1 (and moreover must be a truncated Poisson). Hence, data on the number of jobs per employment cycle suffice for estimating the relative rate at which employed workers receive offers.

Note that we use neither wage data nor duration data to recover this ratio, in contrast to the approach outlined in the survey by Eckstein and van den Berg (2005) which identifies κ_1 from data on how the duration of the job varies with its price w_j. The fact that we cannot directly infer the price on a job w_j from the hourly wage $W_{ijt}^* = w_j \ell_{it}$ does not pose a problem. Moreover, as we reiterate below, using duration data to recover κ_1 requires that we know $F(\cdot)$, and at this stage we have yet to recover it. That said, we would need duration data to separately identify λ_1 and δ.

8

In particular, the duration of an employment cycle has an exponential distribution with hazard δ, allowing us to recover this parameter.

The next two steps describe how to recover the offer distribution $F(\cdot)$. We begin by taking the log difference of the wage in (1.2)

$$\Delta \ln W_{it} = \ln w_{n(t)} - \ln w_{n(t-1)} + \beta \Delta X_{it} + \Delta \varepsilon_{it} \tag{1.6}$$

where $n(t)$ denotes the record number of the job the worker holds at date t. For workers who remain on the same job at these two points in time, the price w per unit labor will be constant, and their log-wage growth is given by

$$\Delta \ln W_{it} = \beta \Delta X_{it} + \Delta \varepsilon_{it} \tag{1.7}$$

Given Assumption 1.1, we can recover β by ordinary least squares. Next, we turn to workers who move from their $n-1$-th job in date $t-1$ to their n-th job at date t. For these workers, we have

$$
\begin{aligned}
\Delta \ln W_{it} - \beta \Delta X_{it} &= \ln w_n - \ln w_{n-1} + \Delta \varepsilon_{it} \\
&\equiv \omega_n - \omega_{n-1} + \Delta \varepsilon_{it}
\end{aligned}
$$

where we use the estimate of β from the previous step. Note that $\omega_n = \ln w_n$ corresponds to the n-th record in the sequence of *log* price offers. Thus, net log wage growth for voluntary job changers is just an error ridden record gap. Averaging across all workers who switch from their n-th job to their $n+1$-th job, and using the fact that w_j is independent of ℓ_{it}, we obtain

$$E\left(\Delta \ln W_{it} - \beta \Delta X_{it} \mid N \geq n\right) = E\left(\omega_n - \omega_{n-1} \mid N \geq n\right) \tag{1.8}$$

This average is conditional on the there being at least n jobs in the respective employment cycle, i.e. on the event that the worker made it to his n-th job. The imputed average log wage changes thus correspond to the conditional expected record gaps $\left\{E\left(y_{L(n)} - y_{L(n-1)} \mid N \geq n\right)\right\}$ from a sequence $\{y_m\}_{m=1}^{M}$ that corresponds to log price offers over an employment cycle.[5]

These average record gaps turn out to suffice for recovering the shape of the offer distribution. As demonstrated by Nagaraja and Barlevy (2003) and Barlevy (2005), if M has a known geometric distribution, the average record spacings $\left\{E\left(y_{L(n)} - y_{L(n-1)} \mid N \geq n\right)\right\}_{n=1}^{\infty}$ from a sequence of

---

[5]While the observed log wage change for an individual $\Delta\omega + \Delta\varepsilon$ can be negative, *average* log wage changes $E(\Delta\omega)$ must be positive, an implication we can test. More generally, not every nonnegative sequence corresponds to a sequence of average record gaps from some distribution, imposing a stronger restriction on what average wage gains can be. We do not derive the admissable set here, other than to note it is isomorphic to the set of moments for which the related Hausdorff Moment Problem described in Shohat and Tamarkin (1943) is solvable.

i.i.d. observations $\{y_m\}_{m=1}^M$ identify the distribution of y_m up to a location shift. Using our estimate for κ_1 to pin down the distribution of M, we can map the average net wage gains (1.8) into a unique distribution $F(\cdot)$ up to a scaling parameter. Intuitively, average wage growth over an employment cycle is informative because the wage gains of workers leaving their first job depend more on the shape of the offer distribution near its lower support than the wage gains of workers leaving their sixth job. Tracking wage growth over an employment cycle should therefore allow us to reconstruct the offer distribution.

Several caveats are warranted regarding this identification strategy. First, it assumes all workers share the same value of leisure b_i so that prices on each employment cycle reflect records from identically distributed sequences of i.i.d. observations $\{y_m\}_{m=1}^M$. But if workers valued leisure differently, the price on the first job $y_{L(1)}$ would be drawn from a different truncated distribution $F(w \mid w \geq w_i^*)$ for each worker. Another caveat is that we need to track each employment cycle from its beginning to determine how many jobs a worker held between his last unemployment spell and his current job. If employment cycles were censored, this approach would seem doomed. It is therefore important to determine whether the model can remain identified when either of these conditions is violated.

At first, heterogeneity in reservation prices and censored employment histories might seem like distinct problems. However, one of the points of this paper is that in both cases, the wages on an employment cycle can be viewed as record values from an independent but not identically distributed (i.n.i.d.) sequence of observations $\{y_m\}_{m=1}^M$, specifically one in which the first observation y_1 is distributed differently from the remaining observations. It turns out that in this case we can continue to identify the key aspects of the model, although demonstrating this is more involved than in the i.i.d. case. In the next section we analyze records in this setup, and in subsequent sections we discuss censoring and heterogenous reservation prices, respectively.

2. Identification with a Non-Identically Distributed Initial Observation

This section contains the key mathematical results we use in subsequent sections. For a more rigorous treatment of this model, see Barlevy and Nagaraja (2005).

Let $\{y_m\}_{m=1}^M$ denote a sequence of random variables. Suppose M is a geometrically distributed random variable that is independent of the y_i, i.e. for $m \geq 1$,

$$\Pr(M = m) = (1-p)^{m-1} p \qquad (2.1)$$

10

where $0 < p < 1$. Define $L(1) = 1$ and $L(n) = \min\left\{k : y_k > y_{L(n-1)}\right\}$ as the n-th record time, and define $w_n = y_{L(n)}$ for $n \geq 1$ as the n-th record value. The total number of records that occur within the sequence $\{y_m\}_{m=1}^{M}$ is given by $N = \max\{j : L(j) \leq M\}$.

We assume $\{y_m\}_{m=1}^{M}$ are mutually independent and y_1 is distributed according to some continuous cdf $G(\cdot)$ while y_m for $m \geq 2$ is distributed according to some other continuous cdf $F(\cdot)$. We want to allow for the possibility that $G(\cdot)$ and $F(\cdot)$ are related. Thus, let us define a mapping T from the set of continuous distribution functions into itself such that $G = T(F)$. A mapping like this can always be defined. However, we add content to this formulation by imposing the following assumptions on T:

Assumption 2.1: for any distribution F, if $G = T(F)$, then the composite function $\mathcal{G}(u) \equiv G \circ F^{-1}(u)$ is absolutely continuous in $u \in (0, 1)$.

Assumption 2.2: for any distribution F, the support of $G = T(F)$ lies in the support of F.

Assumption 2.3: for any distribution F, if $G = T(F)$, then $F^{-1}(0) = G^{-1}(0)$

Assumption 2.1 is a technical assumption that allows us to assume without loss of generality that $G(\cdot)$ has a density function. Assumption 2.2 states that the first observation y_1 can only take on values in the support of all subsequent observations. In our application, this amounts to the assumption that the price on the first job we observe for a worker must lie in the support of the price offer distribution. Assumption 2.3 states that the lowest possible value for y_1 does not exceed the lowest possible value for y_2. If this were not the case, values from the lower support of $F(\cdot)$ would never show up in the record sequence $\left\{y_{L(n)}\right\}_{n=1}^{N}$, and there would be no hope of fully recovering $F(\cdot)$ from record data. Note that the identity mapping $T(F) = F$ satisfies these assumptions. Thus, the case where all observations are identically distributed that was analyzed in Nagaraja and Barlevy (2003) and Barlevy (2005) is just a special case of the model here.

Our goal is to show that the key features of this model – the parameter p in (2.1) and the distributions $F(\cdot)$ and $G(\cdot)$ – can be identified from record data. We begin by providing sufficient conditions for the expected record gaps $E\left(y_{L(n)} - y_{L(n-1)} \mid N \geq n\right)$ to exist, since our results employ these moments. The proof of this and other propositions is contained in an appendix.

Proposition 1: Consider a sequence of independent random variables $\{y_m\}_{m=1}^{M}$ where

11

(i) $\Pr(y_m \leq x) = F(x)$ for all $m \geq 2$;

(ii) $\Pr(y_1 \leq x) = G(x)$ where $G = T(F)$ satisfies Assumptions 2.1 – 2.3

(iii) M is independent of $\{y_m\}$ and $\Pr(M = m) = (1-p)^{m-1} p$ for some $p \in (0,1)$.

If $E(|y_2|) < \infty$, then $E(w_n - w_{n-1} \mid N \geq n)$ exists for all $n \geq 2$, where $w_n = y_{L(n)}$. ∎

Note that the existence of record moments for $n \geq 2$ only depends on the distribution $F(\cdot)$, not $G(\cdot)$. This may seem surprising at first, since the fact that $E(y_2)$ is finite does not imply $E(y_1)$ is finite. Indeed, if the support of $F(\cdot)$ is unbounded above, one can easily construct examples that satisfy Assumptions 2.1 – 2.3 where the mean for $F(\cdot)$ exists but the mean for $G(\cdot)$ does not. The key is that these moments are conditioned on $M \geq 2$ and $\max\{y_2, ..., y_M\} \geq y_1$. From Nagaraja and Barlevy (2003), we know $E(E(\max\{y_2, ..., y_M\} \mid M \geq 2))$ is finite whenever $E(y_2)$ is. Hence, even if the unconditional mean for y_1 does not exist, the mean conditional on y_1 being surpassed by a variable whose mean is finite does exist.

We begin with the case in which T is a known mapping, as will be true of the application in the next section. Moreover, we impose an additional assumption on T, namely that there exists a function $G_0 : [0,1] \to [0,1]$ such that for any function F and any real number x,

$$G(x) = T(F)(x) = G_0(F(x)) \tag{2.2}$$

where $T(F)(x)$ represents the value of $T(F)$ evaluated at x. In words, the probability that $y_1 < x$ under $G = T(F)$ can be expressed purely in terms of the percentile that x occupies within the distribution $F(\cdot)$. Again, this will be true for our application in the next section. We now have

Proposition 2: Consider a sequence of independent random variables $\{y_m\}_{m=1}^{M}$ where

(i) $\Pr(y_m \leq x) = F(x)$ for all $m \geq 2$;

(ii) $\Pr(y_1 \leq x) = G_0(F(x))$ where $G_0 : [0,1] \to [0,1]$ is non-decreasing and onto

(iii) M is independent of $\{y_m\}$ and $\Pr(M = m) = (1-p)^{m-1} p$ for some $p \in (0,1)$.

Then we have

a. The distribution $\{\Pr(N = n)\}_{n=1}^{\infty}$ identifies a unique $p \in [0,1]$ (which depends on G_0).

b. If $E(|y_2|) < \infty$, then $\{E(w_n - w_{n-1} \mid N \geq n)\}_{n=2}^{\infty}$ uniquely characterizes $F(\cdot)$ for a given p within the set of continuous distributions, up to a location shift. ∎

Hence, given a known mapping T that satisfies (2.2), we can use record data to recover both the parameter p and the distribution $F(\cdot)$ from which y_m for $m \geq 2$ is drawn (and since $T : F \to G$ is a known mapping, once we know F we can also deduce G). Establishing this result proves to be considerably more difficult than for the i.i.d. case. In particular, our proof hinges on a relatively obscure result in convolution theory due to Titchmarsh (1926) that would not be required if the observations $\{y_m\}_{m=1}^{M}$ were instead i.i.d.[6]

We next consider the case in which the mapping T is unknown, but is still assumed to satisfy Assumptions 2.1 – 2.3. In this case, we establish the following result:

Proposition 3: Consider a sequence of independent random variables $\{y_m\}_{m=1}^{M}$ where

(i) $\Pr(y_m \leq x) = F(x)$ for all $m \geq 2$;

(ii) $\Pr(y_1 \leq x) = G(x)$ where $G(x)$ is compatible with Assumptions 2.1 – 2.3

(iii) M is independent of $\{y_m\}$ and $\Pr(M = m) = (1-p)^{m-1} p$ for some $p \in (0,1)$.

If $E(|y_2|) < \infty$, then $\{\Pr(N = n)\}_{n=1}^{\infty}$ and $\{E(w_n - w_{n-1} \mid N \geq n)\}_{n=2}^{\infty}$ uniquely determine p and identify $F(\cdot)$ up to a location shift and its associated $G = T(F)$. ∎

In other words, given two arbitrary distributions $G(\cdot)$ and $F(\cdot)$ that comply with Assumptions 2.1 – 2.3, we can use record data to infer the shape of the two distributions as well as p. More precisely, the distribution $\{\Pr(N = n)\}_{n=1}^{\infty}$ identifies p and the distribution $G(\cdot)$ *relative to* $F(\cdot)$, i.e. the composite function $\mathcal{G}(u) \equiv G \circ F^{-1}(u)$ for all $u \in (0,1)$. Intuitively, the number of records that we will observe in a typical sequence depends on how many observations there are (hence p) and how much weight $G(\cdot)$ assigns to values near the upper support of $F(\cdot)$ that are hard to improve upon. Given p and $\mathcal{G}(u)$, the expected record gaps can then be used to determine the shape of $F(\cdot)$. By substituting $F(\cdot)$ into $\mathcal{G}(u)$, we can obtain the distribution $G(\cdot)$.

3. Censored Employment Histories

As a first application of our results, we turn to the case of censored employment histories. The problem of censoring is highly relevant for empirical applications. For example, in applying the insights of record statistics to analyze actual wage data, Barlevy (2005) was forced to turn to

---

[6] In the i.i.d. case, characterization results can be instead derived using the more well-known Müntz-Szász theorem. See Kamps (1998) for a survey of applications of the Müntz-Szász theorem to order statistics.

the National Longitudinal Survey of Youth (NLSY) dataset to obtain sufficiently complete work histories. One limitation of the NLSY is that it is not a particularly large sample, and thus yields too few observations to accurately estimate the average net wage gains beyond the third job in an employment cycle. The limited sample size also precludes implementing the approach separately for different groups of workers, making it impossible to explore such questions as whether the offer distribution and the rate at which offers arrive differ for blacks and whites. Larger panel datasets are available, such as the Census Bureau's Survey of Income and Program Participation (SIPP), but most respondents in these surveys are already employed at the time of their first interview. In other words, employment cycles are left-censored. As a result, whenever a worker we track changes jobs, we don't know how many jobs he previously worked on. We could of course wait until the worker happens to become unemployed, but this would force us to throw out much of the data. Developing a strategy for identification that can use censored employment cycles is essential for exploiting the larger sample sizes of existing datasets.

For now, let us assume as in Nagaraja and Barlevy (2003) or Barlevy (2005) that all workers share the same value of leisure b and hence the same reservation price w^*. But in contrast to these papers, suppose we only get to observe a random sample of already-employed workers. The problem with implementing the procedure outlined in Section 1 is that we no longer know what position in its respective employment cycle any job in the data represents. However, we can make use of an important feature of the labor market we consider, namely that it converges to a steady state in which both the employment rate and the distribution of employed workers across prices is constant over time. Let u denote the fraction of workers who are unemployed in steady state, and let $G(w)$ denote the fraction of employed workers who are paid no more than w per unit labor. Since both expressions are constant over time, and using the fact that unemployed workers will accept any offer they receive given our assumption that $\underline{w} = F^{-1}(0) \geq w^*$, we have

$$
\begin{aligned}
du/dt &= -\lambda_0 u + \delta (1 - u) = 0 \\
dG(w)/dt &= \lambda_0 F(w) \frac{u}{1-u} - [\delta + \lambda_1 (1 - F(w))] G(w) = 0
\end{aligned}
$$

Solving these two equations yields the following expression for the steady state distribution of prices across employed workers

$$
G(w) = \frac{F(w)}{1 + \kappa_1 (1 - F(w))} \tag{3.1}
$$

where recall $\kappa_1 = \lambda_1/\delta$. This distribution first-order stochastically dominates $F(\cdot)$, reflecting the fact that workers gravitate to higher paying jobs. Since some already-employed workers will have moved on to jobs that pay more than the first job they were offered, we will observe more workers at high price jobs than in a random sample of workers on their first job out of unemployment.

14

Suppose we chose a worker at random among all employed workers, and denote by y_1 the price on the first job we observe him on. If workers have been active in the labor market for some time, then y_1 represents an independent draw from a distribution approximately equal to the steady state distribution $G(\cdot)$ above. Let $M - 1$ denote the number of offers a worker receives from the time we first observe him until he is next laid off, and let us refer to these offers as y_2 through y_M according to the order they arrive. It is easy to show that the number of offers from now until he is laid off has the identical geometric distribution as in (1.4). Hence, M is geometric. The price offers y_m for $m \geq 2$ are drawn from $F(\cdot)$ and are independent of one another and of y_1.

It can be verified that the mapping $T : F \to G$ defined by (3.1) is consistent with Assumptions 2.1 – 2.3. Moreover, $G(w) = G_0(F(w))$, where $G_0(y) = y[1 + \kappa_1(1 - y)]^{-1}$. As long as the log price offer distribution has a finite mean, we satisfy all of the requirements of Proposition 2 above. We can therefore use the distribution of the number of jobs N from when we first observe a worker until he is eventually laid off to recover κ_1.

More precisely, as can be inferred from Appendix A, $\Pr(N = n)$ can be expressed as follows:

$$\Pr(N = n) = \int_0^1 \frac{[\ln(1 + \kappa_1(1 - u))]^{n-1}}{(n - 1)!} \frac{1 + \kappa_1}{(1 + \kappa_1(1 - u))^3} du \tag{3.2}$$

We can therefore recover κ_1 from the empirical distribution of the number of jobs per left-censored employment cycle using maximum likelihood, i.e. by choosing κ_1 to maximize the likelihood of the observed values of N across employment cycles under (3.2). Once we have an estimate for κ_1, Proposition 2 tells us we can use the net wage gains of voluntary job changers to recover the offer distribution $F(\cdot)$ up to a scaling parameter. Let n denote the number of job changes we observe for the worker since his first survey. The average net wage gains $E(\Delta \ln W_{it} - \beta \Delta X_{it} \mid N \geq n)$ then correspond to the sequence of moments $E\left(y_{L(n)} - y_{L(n-1)} \mid N \geq n\right)$ among log price offers $\{y_m\}_{m=1}^M$, and by Proposition 2 these moments characterize the distribution of y_2. While this two-step approach is similar to the one described earlier for complete cycles, the way we map the underlying data to values for κ_1 and $F(\cdot)$ will be different and will depend on the function $G_0(\cdot)$ implicit in (3.1).

At this point, we should mention earlier work by Bontemps, Robin, and van den Berg (2000) who also argue that this model can be nonparametrically identified when we only get to observe workers who are already employed. However, unlike this paper, they assumed no measurement error and no unobserved worker productivity. This allowed them to pursue a different identification strategy from the one described above. It will be useful to compare the two approaches to understand the difficulties posed by the presence of unobserved worker productivity.

When worker productivity is perfectly observable, we can deduce the price w on each job. Given this, Bontemps *et al* note that we can directly recover $G(\cdot)$ from the empirical distribution of prices among already-employed workers, assuming these workers have been active in the market for a sufficiently long time. Rearranging (3.1), we have

$$F(w) = \frac{(1 + \kappa_1) G(w)}{1 + \kappa_1 G(w)} \tag{3.3}$$

Just as with our approach, their approach requires an estimate for κ_1 to recover $F(\cdot)$. But rather than estimating κ_1 from the number of jobs a worker passes through before he is next laid off, Bontemps *et al* turn to job duration data. Since a job ends if either the worker is sent into unemployment or receives a better offer, a job offering a price w ends with hazard $\delta + \lambda_1 (1 - F(w))$. Substituting in (3.3), this becomes

$$\frac{\lambda_1 (1 + \kappa_1)}{\kappa_1 (1 + \kappa_1 G(w))}$$

As long as one can estimate the hazard for at least two different values of w, it will be possible to separately estimate λ_1 and κ_1. This provides them with the parameter κ_1 of interest, as well as separate estimates for λ_1 and $\delta = \lambda_1/\kappa_1$.

Once we allow for the possibility that worker productivity cannot be perfectly measured, though, the approach proposed by Bontemps *et al* no longer works. First and foremost, the distribution of observed wages $W = w\ell$ in a random sample of employed workers no longer corresponds to $G(\cdot)$, but to a convolution of $G(\cdot)$ and the distribution of worker ability. Without additional assumptions on unobserved worker productivity, we cannot recover $G(\cdot)$ from a cross sectional wage distribution. Second, even if $G(\cdot)$ were known, the fact that we cannot directly observe the price w makes it impossible to exploit the variation between job duration and w to identify κ_1. The virtue of recovering κ_1 from longitudinal data on the number of jobs per employment cycle is that we do not need to know either the distribution $G(\cdot)$ or the true price on each job.

Interestingly, the fact that it is possible to recover κ_1 without relying on wage data suggests our approach can be used in models where the price per unit labor is not constant over the course of the job as we have assumed. For example, consider the Postel-Vinay and Robin (2002) model, which uses the same production structure as here but assumes a different wage setting mechanism. They assume employers post wages, but can then increase the wage if their employee receives an outside offer. An implication of their model is that workers always prefer an employer who is more productive to one who is less productive. This implies the number of jobs is the number of records in the sequence $\{Z_m\}_{m=1}^{M}$ where Z_m denotes the productivity of the m-th offer and M denotes the number of offers in an employment cycle. We can therefore use the number of jobs per employment

cycle can to identify κ_1. This is true even though wages no longer correspond to record values: a worker might agree to a *lower* wage from a more productive employer, correctly anticipating that wage growth will be higher with this employer. By contrast, it will not be possible to recover κ_1 using wage and duration data as described above. However, a similar approach would work if we could observe z, since a job with productivity z will end with a hazard rate of $\delta + \lambda_1 \left(1 - \Gamma\left(z\right)\right)$.[7]

Another difference between our approach and the one proposed by Bontemps *et al* is that their strategy requires that equation (3.1) holds, whereas our approach does not. This is because the approach advocated by Bontemps *et al* proceeds by first estimating $G\left(\cdot\right)$ and then using a known mapping from $G\left(\cdot\right)$ to $F\left(\cdot\right)$ to deduce $F\left(\cdot\right)$. By contrast, our approach remains valid even if $G\left(\cdot\right)$ is unknown, since we identify $F\left(\cdot\right)$ from the evolution of log wage growth over an employment cycle, not by mapping an estimate of $G\left(\cdot\right)$ into an implied estimate of $F\left(\cdot\right)$. Formally, according to Proposition 3 we can identify $F\left(\cdot\right)$ when $T : F \rightarrow G$ is unknown.

While our approach has the advantage that it does not require knowing how $G\left(\cdot\right)$ and $F\left(\cdot\right)$ are related, its main drawback is that it is far more data intensive: we need to track workers to the end of their employment cycle, whereas Bontemps *et al* only need data on one job per worker. However, tracking each employment cycle to its very end is only necessitated by our assumption of unobservable productivity. If worker productivity were observable, we could separately identify $G\left(\cdot\right)$ and $F\left(\cdot\right)$ without as much data. Suppose we collected data on w_1 and w_2, the price per unit labor on the first and second jobs we observe for a worker, as well as the duration of the first job. As in Bontemps *et al*, we could use the empirical distribution of w_1 to recover $G\left(\cdot\right)$ and the relationship between the duration of the job and its price to recover λ_1 and δ. But rather than rely on (3.1) to infer $F\left(\cdot\right)$ from $G\left(\cdot\right)$, we can recover $F\left(\cdot\right)$ directly given estimates of κ_1 and $G\left(\cdot\right)$ by using the fact that the empirical distribution of w_2 corresponds to the distribution of the second record from $\{y_m\}_{m=1}^M$ conditional on at the occurrence of a second record. We would need to track workers beyond the first job we see them on, but not to the end of their employment cycle.

4. Heterogeneity in Reservation Prices

So far, we have maintained the assumption that all workers share the same value of leisure b. In this section we relax this assumption. Formally, let $\Upsilon\left(\cdot\right)$ denote the cdf of b_i across individuals. Given the implicit formula for w_i^* in (1.3), $\Upsilon\left(\cdot\right)$ can be mapped into a distribution $H\left(\cdot\right)$ for reservation

---

[7]Postel-Vinay and Robin (2002) appeal to precisely this strategy. Using matched employer-employee data, they estimate the productivity of each firm by the average wage of all workers at the same location. This approach is valid so long as all jobs with the same employer are equally productive.

prices w_i^* across all workers. As before, we assume that the least selective worker would be willing to accept any offer, i.e. $\underline{w}^* = H^{-1}(0) \leq F^{-1}(0) = \underline{w}$. Let us define $H_u(\cdot)$ as the distribution of w_i^* across unemployed workers. This distribution will differ from the distribution of reservation prices in the population, since workers with high reservation prices are less likely to move into employment and thus more likely to be unemployed at a given point in time. We show that it is possible to nonparametrically identify both the offer distribution $F(\cdot)$ and the distribution of reservation prices among the unemployed $H_u(\cdot)$. Under additional assumptions, it will also be possible to reconstruct $H(\cdot)$ and $\Upsilon(\cdot)$. Throughout this section, we assume the econometrician has access to complete employment histories.

Suppose we tracked an unemployed worker chosen at random and recorded the price on the first job that worker accepted. Let us call this price y_1. How is y_1 distributed? We can view y_1 as a random draw w from $F(\cdot)$ known to exceed the value of an independent draw w^* from $H_u(\cdot)$. Hence, its distribution is just the distribution of w conditional on the event that $w \geq w^*$, i.e.

$$\Pr(y_1 \leq x) = \Pr(w \leq x \mid w \geq w^*) = \frac{\Pr(w^* \leq w \leq x)}{\Pr(w^* \leq w)}$$

One can show that $\Pr(w^* \leq w \leq x) = \int_{-\infty}^{x} H_u(w)\, dF(w)$ and $\Pr(w \geq w^*) = \int_{-\infty}^{\infty} H_u(w)\, dF(w)$. Hence, if we let $G(x)$ denote the cdf for y_1, then

$$G(x) = \frac{\int_{-\infty}^{x} H_u(w)\, dF(w)}{\int_{-\infty}^{\infty} H_u(w)\, dF(w)} \tag{4.1}$$

Appealing to the change in variables $y = F(x)$, we can rewrite (4.1) as

$$G(x) = \frac{\int_{0}^{F(x)} H_u\left(F^{-1}(y)\right) dy}{\int_{0}^{1} H_u\left(F^{-1}(y)\right) dy} \tag{4.2}$$

As before, denote additional price offers over the course of an employment cycle by $\{y_2, ..., y_M\}$. The prices on the jobs the worker accepts represent records from the sequence $\{y_m\}_{m=1}^{M}$, where y_1 is distributed according to (4.1) while y_m for $m \geq 2$ is distributed according to $F(\cdot)$. It easy to verify that $G(x)$ satisfies Assumptions 2.1 and 2.2. Assumption 2.3 follows directly from our assumption that $F^{-1}(0) \geq H^{-1}(0) = \underline{w}^*$. However, as evident from (4.2), $G(x)$ cannot be expressed as a function of $F(x)$ alone, since $G(x)$ depends on $F(y)$ for all $y \leq x$. We therefore appeal to Proposition 3, which allows $G(\cdot)$ to be an unknown function.

According to Proposition 3, we can use the distribution of the number of jobs N per employment cycle to recover κ_1 and the composite function $\mathcal{G}(u) \equiv G \circ F^{-1}(u)$. Let n denote the number of jobs the worker has held since he was last unemployed. The average net wage gains

18

$E\left(\Delta \ln W_{it} - \beta \Delta X_{it} \mid N \geq n\right)$ across all workers correspond to the sequence of expected record gaps $E\left(y_{L(n)} - y_{L(n-1)} \mid N \geq n\right)$ among the log price offers the worker receives. Proposition 3 states that, given κ_1 and $\mathcal{G}(u)$, these moments uniquely determine the log price offer distribution up to a location shift, and hence the offer distribution $F(\cdot)$ up to a scaling parameter. Thus, even though the sequence of prices over an employment cycle $\{w_n\}_{n=1}^{N}$ will be distributed differently for workers with different reservation prices, we can still recover $F(\cdot)$ from the average wage gains across all job changers who are at the same position in their respective employment cycle.

Once we know $F(\cdot)$, we can substitute it into our original estimate for $\mathcal{G}(u)$ to recover $G(\cdot)$, the distribution of the price on the starting job across . Although $G(\cdot)$ itself may not be of direct interest, recovering $G(\cdot)$ is a first step towards recovering objects that are of inherent interest, such as the distribution of reservation prices $H_u(\cdot)$ across unemployed workers, the distribution of reservation prices $H(\cdot)$ across all workers, and the distribution of the value of leisure $\Upsilon(\cdot)$. For certain applications, it is important that we be able to estimate these distributions. For example, one feature of U.S. labor markets is that many unemployed workers find jobs relatively quickly. This could be because most workers are not very choosy and accept any offer that comes their way, or because search frictions for unemployed workers are small (i.e. λ_0 is large) and even choosy workers can quickly find jobs they are willing to take. The distribution of reservation prices among unemployed workers $H_u(\cdot)$ can distinguish between these two explanations.

We now describe how to recover these distributions using $G(\cdot)$. We begin with the distribution of reservation prices $H_u(\cdot)$ across unemployed workers. Intuitively, we should not expect to learn about reservation values that do not correspond to prices offered by some employers; this is because we need data on how participation increases with the price to determine how many workers have a given reservation value. Formally, we can only identify $H_u(x)$ when $x = F^{-1}(u)$ for some $u \in (0, 1)$. Setting $x = F^{-1}(u)$ in (4.2) and differentiating with respect to u yields

$$\frac{d}{du} G\left(F^{-1}(u)\right) = \frac{H_u\left(F^{-1}(u)\right)}{\int_0^1 H_u\left(F^{-1}(u)\right) du} \qquad (4.3)$$

The left-hand side of (4.3) is just $\mathcal{G}'(u)$, where recall $\mathcal{G}(u)$ is identified from data on the number of jobs N per employment cycle. Since $H_u(\cdot)$ is a cdf, it follows that $\mathcal{G}'(u)$ must be positive. This is a testable implication of the model, since not every distribution $\Pr(N = n)$ will yield a function $\mathcal{G}(u)$ that is everywhere nondecreasing.

Equation (4.3) implies that $H_u\left(F^{-1}(u)\right)$ is proportional to $\mathcal{G}'(u)$ for all $u \in (0, 1)$. To get an expression for $H_u\left(F^{-1}(u)\right)$, all we need is the constant of proportionality $\int_0^1 H_u\left(F^{-1}(u)\right) du$.

This constant is not always identified, since it depends on the fraction of workers whose reservation price exceeds the highest offered price $\overline{w} = \sup_j w_j$, which thus far we have not restricted.

If we assume $\overline{w}^* \leq \overline{w}$, an assumption that follows automatically if the offer distribution $F(\cdot)$ has unbounded support, then we can easily obtain the constant of proportionality. In particular, if all workers set their reservation price below $\overline{w} = F^{-1}(1)$, then $H_u(F^{-1}(1)) = 1$, and the constant of proportionality is given by

$$\left[\int_0^1 H_u\left(F^{-1}(u)\right) du\right]^{-1} = \mathcal{G}'(1) \tag{4.4}$$

Hence, for any $u \in (0,1)$, we have

$$H_u\left(F^{-1}(u)\right) = \frac{\mathcal{G}'(u)}{\mathcal{G}'(1)} \tag{4.5}$$

Even if we do not assume $\overline{w}^* \leq \overline{w}$, i.e. if the $F(\cdot)$ we identify has bounded support, we can still interpret (4.5) as the conditional distribution $H_u(\cdot \mid w^* \leq \overline{w})$, i.e. as the distribution of reservation prices among workers whose reservation price is no higher than $\overline{w}$.

Given an estimate for the distribution of reservation prices $H_u(\cdot)$ among unemployed workers, we next set out to construct the distribution of reservation prices $H(\cdot)$ for the population as a whole. For this, we need to impose additional assumptions on how the population of unemployed workers compares to that of all workers. Let us assume that the market we consider has been active for quite some time, so that the distribution $H_u(\cdot)$ is close to its steady-state value. In steady-state, the fraction of workers with reservation price w^* who are unemployed, which we denote u_{w^*}, remains constant over time. Using the law of motion for u_{w^*},

$$\dot{u}_{w^*} = -\lambda_0\left(1 - F\left(w^*\right)\right) u_{w^*} + \delta\left(1 - u_{w^*}\right)$$

and setting the change in u_{w^*} to zero yields the following steady-state unemployment rate:

$$u_{w^*} = \frac{1}{1 + \kappa_0\left(1 - F\left(w^*\right)\right)}$$

Hence, the steady state distribution of reservation prices across unemployed workers $H_u(\cdot)$ will be related to $H(\cdot)$ by the formula

$$H_u(w)\, u = \int_{\underline{w}^*}^w u_{w^*}\, dH\left(w^*\right) = \int_{\underline{w}^*}^w \frac{dH(x)}{1 + \kappa_0\left(1 - F(x)\right)} \tag{4.6}$$

where u denotes the steady state unemployment rate and one can show is given by

$$u = \frac{1}{1 + \kappa_0 \int_0^1 H_u\left(F^{-1}(y)\right) dy} \tag{4.7}$$

20

A similar derivation for this formula can be found in Bontemps, Robin, and van den Berg (1999). Differentiating (4.6) with respect to w yields

$$h_u(w)\,u = \frac{h(w)}{1 + \kappa_0(1 - F(w))}$$

We can therefore derive $H(\cdot)$ from $H_u(\cdot)$. To do so, we first need to estimate κ_0. One way to do so is to use the steady-state unemployment rate u from equation (4.7) and our existing estimates for $H_u(\cdot)$ to extract κ_0. Next, for any price x that lies in the support of the offer distribution, i.e. for any x that can be expressed as $F^{-1}(u)$ for some $u \in (0,1)$, we multiply the density function $h_u(x) = \frac{d}{dx}H_u(x)$ by the factor $1 + \kappa_0(1 - F(x))$. Finally, we multiply our new function by the unemployment rate to obtain the density function $h(w)$.

Once we recover the distribution $H(\cdot)$, it is fairly straightforward to obtain the distribution of leisure $\Upsilon(\cdot)$. Recall that the cutoff w^* for each worker is implicitly defined by equation (1.3). Rearranging yields

$$w^* = b + (\kappa_0 - \kappa_1)\int_{w^*}^{\infty}\frac{1 - F(x)}{\rho/\delta + 1 + \kappa_1(1 - F(x))}dx \tag{4.8}$$

where $\kappa_0 = \lambda_0/\delta$. We already described how to estimate κ_0 and κ_1. The remaining parameter we need is the ratio of the discount rate ρ to the rate of job loss δ. Recall that we can estimate δ from the duration of employment cycles. As for ρ, under certain assumptions we should be able to appeal to interest rate data to recover it, although this is beyond the scope of this paper. Assuming we can assign a value to ρ, we can appeal to (4.8) to map $H(\cdot)$ into $\Upsilon(\cdot)$. It might be possible to characterize this mapping analytically, but if not, it is always possible to use simulation methods to find a numerical approximation for $\Upsilon(\cdot)$. An example of these methods can be found in the work on identification in auctions by Guerre, Perrigne, and Vuong (2000).

5. Empirical Application: Heterogeneity and Estimates of κ_0 and κ_1

Although one of our goals is to develop an identification strategy that can be applied to new datasets where left-censoring is a concern, in our empirical analysis we limit ourselves to the same National Longitudinal Survey of Youth (NLSY) dataset already explored in previous work. This is because we wish to gauge the consequences of relaxing the assumption adopted in some of the aforementioned papers that workers share a common reservation price. In this section we consider the implications of this assumption for estimates of mobility parameters, and in the next section we consider the implications for estimates of the offer distribution $F(\cdot)$.

Our analysis suggests that allowing for heterogeneous reservation prices can dramatically affect

the estimated rate at which workers receive offers. As we discuss below, this sensitivity is driven by two features of the empirical distribution of the number of jobs N per employment cycle: (1) a majority of employment cycles end after only one job; (2) the distribution of N has a fat tail. Allowing for heterogeneous reservation values gives more weight to observation (2), which in turn leads to higher estimates for the rate at which employed workers receive offers. Interestingly, these estimates are on par with those implied by previous analysis based on duration data. However, the distribution of reservation prices required to account for observation (1) implies that the rate at which unemployed workers receive offers is much higher than previous analysis suggests.

The reason we use the NLSY dataset is that it compiles comprehensive employment histories for each worker in the survey for an extended period. This dataset tracks a cohort of workers who were between ages 14 and 22 in 1979. We focus on male workers, whom we follow up to 1993, so the oldest person in our sample is 36. Although the NLSY was continued in subsequent years, we chose not to incorporate that data, for two reasons. First, from 1994 on interviews were conducted every two years, which as Pierret (2001) shows makes it difficult to construct reliable work histories. Second, in line with our assumptions, we want to focus on young workers who have less incentive to invest in match-specific human capital, and age 36 seemed a reasonable cutoff.

According to our analysis, all we need to estimate κ_1 is the distribution of N across cycles. As in previous work, such as Flinn (2002), we use periods of non-employment to demarcate employment cycles. To avoid counting summer jobs for students as employment cycles, we restrict attention to employment cycles that end when the worker has at least one year of potential experience, i.e. he has been at least one year out of school. That is, a worker may begin an employment cycle while still in school, but we will only include the cycle in our analysis if he remains continuously employed beyond the point at which he finishes his schooling.

One concern about partitioning the data into employment cycles this way is that workers who take time off before voluntarily moving into a new job – or must wait until this job starts (e.g. teachers starting a job at a new school) – might be misclassified as starting a new employment cycle. We therefore consider an alternative classification proposed in Barlevy (2005) which combines data on non-employment spells with the reason the worker provides for ending his job. In particular, a new employment cycle is said to begin if either the worker is laid off from his job or if he spends more than 8 weeks in non-employment. The two approaches yield distributions for N that have similar qualitative features, and hence lead to similar estimates for κ_1.

After we divide the data into employment cycles, we need to count the number of consecutive jobs within each cycle. This raises the question of how to count dual job holdings, a relatively

common phenomenon in the data. We follow Barlevy (2005) in ignoring jobs that begin after and end before another job, on the grounds that these are most likely secondary jobs that supplement income and which the worker is not interested in taking on as a primary job. Indeed, in the vast majority of such cases, workers identify these jobs as secondary in response to survey questions.

Before turning to results, a final issue we must contend with is that since the NLSY spans a fixed time window, we cannot track employment cycles to their end. Formally, let t_k denote the length of the k-th cycle and T_k the time from when the k-th cycle beings until the end of the sample. Given our data, we can estimate $\Pr\left(N = n \mid t_k \leq T_k\right)$, not $\Pr\left(N = n\right)$. We will therefore oversample short employment cycles with fewer jobs. To mitigate this concern, we only consider cycles that began in the first five years of the sample. Since the first NLSY interview collected retrospective data on jobs starting from 1977, this includes all cycles that started on or before 1981. For large values of T_k, the degree of bias should be small. Indeed, among the $14,178$ employment cycles that started prior to the first week of 1982, only $1,837$ are censored (13%). Only a quarter of these are censored because they continue beyond the last year in the survey, 1993. The rest are censored because of attrition or because the worker did not provide a reason for leaving his job. Nearly half of all censored cycles (835 out of $1,837$) are censored within 2 years of when they start. Censoring is therefore unlikely to generate a large bias.

Table 1 reports the distribution of N across the $12,341$ complete employment cycles in our data that began prior to 1982. The first column partitions employment cycles by nonemployment spells. A stark feature of the data is that the vast majority (61%) of all employment cycles end after one job. The second column treats workers who quit and find a job within 8 weeks as continuing in the same employment cycle. In this case, nearly 70% of all cycles end with only one job, and the tail of the distribution looks nearly identical. Although not reported in Table 1, this feature is pervasive: even when we break down the analysis into four education groups, we systematically find that the majority of cycles end after only one job.

Armed with this data, we proceed to estimate κ_1. We first consider what happens when we assume all workers share the same reservation price. The top panel of Table 2 reports the maximum likelihood estimate for κ_1 under this assumption, and are in line with estimates reported in Barlevy (2005). The parameter κ_1 is tightly estimated around 1.9. The reason for this relatively low value is the high incidence of employment cycles with only one job. Intuitively, when workers share the same reservation price, symmetry implies that in cycles with exactly two offers, half of the time the first offer will be below the second offer. The fact that few cycles result in more than one job implies there must be few cycles in which the worker makes it to a second offer. Hence, either the rate λ_1 at which offers arrive is low, or the rate δ at which cycles end is high and workers do not

23

have enough time to accumulate multiple offers.[8] By contrast, if κ_1 were in the range of 5 to 10, in line with estimates of κ_1 based on duration data that we discuss below, the expected fraction of cycles that end after one job would be much lower, around $24 - 36\%$.

Next, we introduce heterogeneity in reservation prices. To appreciate why this can affect the estimates, suppose a significant fraction of unemployed workers were highly selective, i.e. they demanded a price near the top of the offer distribution to work. Once these workers managed to find a job within that range, they would be unlikely to improve upon that offer, even if κ_1 were high. In other words, allowing for the possibility of heterogeneous reservation prices makes it possible for a large fraction of employment cycles to end after one job even when κ_1 is high.

To integrate heterogeneity into our empirical analysis, we searched for a convenient parametric specification for $\mathcal{G}(u)$ to facilitate the estimation. Recall that the model requires $\mathcal{G}(u)$ be nondecreasing. Experimenting with different specifications confirmed that the best fit is indeed upward sloping, and moreover is convex. This prompted us to use an exponential form $\exp(u/b)$ where b is some constant. However, for this specification b plays a dual role: it determines both the shape of the distribution of reservation prices in the support of the offer distribution and the fraction of workers whose reservation price is lower than $\underline{w} = F^{-1}(0)$. We therefore turned to a two-parameter generalization, properly scaled to reflect a proper cdf:

$$\mathcal{G}(u) = au + (1-a)\frac{\exp(u/b) - 1}{\exp(1/b) - 1} \tag{5.1}$$

This generalization is natural, since when $a \to 1$ it collapses to the special case where all workers have a reservation price below $\underline{w}$ (and is thus equivalent to assuming all workers share a common reservation $w^* \leq \underline{w}$ as we previously did). From the Appendix, we know we can express $\Pr(N = n)$ as an integral involving $\mathcal{G}(u)$ and κ_1, and we choose the three parameters to maximize the likelihood of the data given these expressions. Since the results proved sensitive to outliers, we eliminated the extreme tail of the distribution and only included observations where $N \leq 7$.

The bottom panel of Table 2 reports our estimates for κ_1, a, and b. Our estimates for κ_1 by education group range between 8.4 and 16.4. Our estimate for the population as a whole is 10.8. Moreover, the point estimate for the population as a whole is quite tight, and we can safely reject the lower estimates for κ_1 we obtain when we assume workers share the same reservation price.

---

[8]More precisely, the implication is that the arrival rate of "viable" offers – that is, offers that a worker might accept – is low. This rate is equal to $\lambda_1(1 - F(w^*))$, where w^* denotes the common reservation price. Since we assumed $\underline{w} = F^{-1}(0) \geq w^*$, this is equal to λ_1. But this distinction hints at why heterogeneous reservation prices can help: even if the arrival rate of viable offers is low for most workers, implying they won't make it to a second job, the rate at which they receive offers that are deemed viable to *any* worker in the economy could still be high.

Why does heterogeneity in reservation prices lead to higher estimates for κ_1? Figure 1 sheds some insight on this. The figure plots the log of $\Pr(N = n)$ against n. The solid black line represents the data. The dashed line represents the fitted values when we assume all workers share a common reservation price (i.e. when we constrain $a = 1$). Matching the large fraction of cycles with only one job requires a low value for κ_1. However, as evident from the figure, a low value of κ_1 implies that the distribution for N has a thin tail, as reflected in the steep negative slope of the dashed line. This is intuitive: if offers arrive infrequently, few workers will make it to a second job; of those that do, few will make it to a third job; and so on. Hence, $\Pr(N = n)$ should decline sharply with n. But the empirical distribution has a much fatter tail: a non-negligible fraction make it to a third, fourth, and fifth job. This requires a high value κ_1, since workers need to receive enough offers to have enough opportunities for this much upward mobility. When we introduce heterogeneity in reservation prices, the estimation assigns a high value for κ_1 to match the fat tail of the distribution, and assigns a and b to match the large fraction of cycles with only one job. In particular, it interprets the data to imply that a large fraction of workers who are so selective that they will be unlikely to receive a more favorable offer than the first offer they are willing to take, even when the rate at which offers arrive is quite high.

Interestingly, the estimates we obtain for κ_1 when we allow for heterogeneity are compatible with the estimates previous authors have found based on duration data. For example, Flinn (2002), using the same NLSY dataset we use, estimates κ_1 between 3.3 and 7.8 across education groups, and for the group as a whole at 4.6 (see Table 4, p633). Using data from the Netherlands, van den Berg and Ridder (1996) estimate κ_1 between 6.8 to 12.3 across age groups, and for the group as a whole at 9.4 (see Table VII, p1208). Although these estimates are consistent with what we find, they are based on independent evidence. Our results are driven by the fact that a fair number of workers are observed in a large succession of jobs without a nonemployment spell, implying that they must have received offers at a fairly high rate to accumulate enough offers to move this many times. By contrast, estimates based on duration data are driven by the fact that higher wage jobs have significantly longer duration. Recall that the hazard rate for a job that pays a price w corresponds to $\delta + \lambda_1 (1 - F(w))$. The extent to which the average duration varies with the price paid on the job depends on how large λ_1 is relative to δ. Even though previous authors have imposed different assumptions on $F(\cdot)$ and differ in how they account for unobserved worker productivity, the fact that higher wage jobs last longer leads all to estimate a high value for κ_1.

While allowing for heterogeneous reservation prices leads to estimates of the arrival rate for employed workers that are consistent with what previous work has found, the nature of heterogeneity in reservation prices we estimate leads to different conclusions regarding the rate at which unem-

ployed workers encounter offers. In particular, to accord with the large fraction of employment cycles that end after only one job, our estimation requires that the distribution of reservation prices be highly skewed towards the upper support of the offer distribution, as implied by the low value we estimate for b in Table 2. Intuitively, if employed workers receive many offers, as implied by a high κ_1, it must be that a large fraction of workers begin their first job near the top of their potential earnings distribution. But if a large fraction of workers are indeed this choosy, the rate at which workers encounter offers while unemployed must be fairly high to accord with the low duration and incidence of unemployment we see in the data.[9]

The reason our estimates differ from those in previous work is that we identify the distribution of reservation prices from mobility data, whereas previous work has either abstracted from it or parameterized it in a particular way, e.g. Bontemps, Robin, and van den Berg (1999). Our estimate reveals a far more skewed distribution, and hence a far higher estimate for $\kappa_0 = \lambda_0/\delta$, the relative rate at which workers encounter offers while unemployed. To see this, substitute (4.4) into (4.7) to obtain the following expression for unemployment:

$$u = \left[1 + \kappa_0/\mathcal{G}'(1)\right]^{-1}$$

Rearranging, we get the following expression for κ_0:

$$\kappa_0 = \mathcal{G}'(1)\left(\frac{1}{u} - 1\right)$$

$\mathcal{G}'(1)$ measures how skewed the distribution is towards the upper support of the offer distribution. If all workers share a common reservation price, $\mathcal{G}'(1) = 1$. By contrast, our estimate for $\mathcal{G}'(1)$, reported in the final column in Table 2, is much larger: κ_0 is almost 16 times as large as when we assume all workers share the same reservation price. If we set u to 6%, our estimate for $\lambda_0/\lambda_1 = \kappa_0/\kappa_1$ is 23, i.e. the rate at which offers arrive while workers are unemployed is over twenty times as large as when they are employed. By contrast, since previous work has either abstracted from heterogeneity in reservation prices or considered far less skewed distributions (where the implied $\mathcal{G}'(1)$ is much smaller), it has typically estimated λ_0/λ_1 at no more than 2.

Why does our approach imply that workers receive offers at a much lower rate when they are employed than when they are unemployed? This interpretation is needed to explain why in the data most workers move fairly quickly from unemployment to employment (as evidenced by the low

---

[9]Note that one could alternatively interpret these findings as saying that a large proportion of employed workers receive offers at a lower rate than their peers, rather than that a large proportion of workers are choosier than their peers. If the ratio λ_1/λ_0 were constant across workers, this too would require that many unemployed workers receive offers at an extremely high rate to accord with the low incidence and duration of unemployment.

incidence of unemployment) but do not often move from their first job out of unemployment into a second job (as evidenced by the high incidence of cycles that end with only one job). Within our framework, the only possible explanation is that mobility slows down dramatically once workers are employed. There may be some truth to this, but the magnitude seems rather large to be plausible. More likely, our framework abstracts from some important consideration, and it is not obvious how this feature would affect our inference on the degree of labor market mobility.[10]

There are various features one could introduce into this search model that could explain these facts without implying a dramatic decline in the arrival rate once a worker becomes employed. One example is moving costs: even if employed workers receive offers at a high rate, they may not always move to a higher offer. The problem with this explanation is that it also implies a thin-tailed distribution for N, although this can be overcome if only some workers have a distaste for moving rather than all. Another possibility is to endogenize search effort, so that workers who earn a wage closer to the top of the distribution search less intensively. In this case, the number of offers M would be correlated with the realizations $\{y_m\}_{m=1}^{M}$, resulting in more spells with only one job (those where the initial draw was high). This modification poses problems for recovering arrival rates using duration data, since high wage jobs would tend to last longer even if the "true" arrival rate λ_1 were low, simply because workers on high wage jobs search less. Thus, the fact that high wage jobs have longer durations does not seem to robustly imply κ_1 is high. By contrast, if offers are independent, the fat tail in the distribution of N necessarily implies a high κ_1: records are sufficiently rare among i.i.d. draws that the number of offers workers receive over an employment cycle must be large, and hence so must κ_1.

6. Heterogeneity and Identification of the Offer Distribution

In this section, we briefly note some of the implications of heterogeneity in reservation prices for estimating the offer distribution $F(\cdot)$. Barlevy (2005) argued that in the NLSY data, average log wage gains appear to be roughly constant regardless of how many jobs the worker previously changed. In the absence of heterogeneity in reservation prices, this pattern uniquely characterizes the exponential distribution, implying that the offer distribution is Pareto (the antilog of the exponential). By contrast, if the offer distribution were lognormal, as is assumed at times, average

---

[10] However, there is some evidence to support the notion of heterogeneity in reservation wages. In particular, such heterogeneity would imply a positive relationship between the duration of an unemployment spell and the duration of the first job out of uneployment, since more selective workers will be unemployed for longer on average but are then less likely to voluntarily leave for another job. Preliminary work we carried out revealed such a pattern in the NLSY data, although a serious treatment of this analysis is beyond the scope of this paper.

wage gains would decline with n, a feature that is due to the log concave shape of the normal distribution. However, Barlevy (2005) found that the rate at which average wage gains decline with n for a lognormal distribution is not sufficiently sharp that it can be rejected statistically.

How does our interpretation of this finding change once we allow for heterogeneous reservation prices? On the one hand, to the extent that log wage growth does not vary with the number of times the worker has already changed jobs, we would still conclude that the offer distribution is Pareto. This follows directly from Proposition 3. However, distinguishing the Pareto distribution and the lognormal distribution becomes more difficult, as can be seen graphically in Figure 2. The dark line in the figure shows the implied average log wage gains under a Pareto distribution. Regardless of whether there is heterogeneity, average wage gains will be constant and assume the same value in both cases. The remaining two lines trace out $E\left(y_{L(n)} - y_{L(n-1)} \mid N \geq n\right)$ for n for a lognormal distribution that is normalized so that average wage growth across all job changers is consistent with what Barlevy (2005) estimates from the data. The dashed line is computed for the estimate of κ_1 from the top panel of Table 2 when we assume no heterogeneity, while the gray line is computed using our estimates for $\mathcal{G}(u)$ in the bottom panel of Table 2. According to these estimates, the average log wage gains decline even less rapidly with n, making it harder to distinguish these two particular shape restrictions.

To understand why heterogeneity in reservation prices results in a flatter profile of average wage gains over an employment cycle, note that when we observe a worker move multiple times, we can infer he is probably not very choosy, or else he would have started with a high wage that he would be unlikely to improve upon. But less choosy workers tend to earn lower wages on average. Since under the lognormal distribution, log wage gains are higher on average for workers in lower wage jobs, this will increase the average wage gains we would observe for workers who have already moved several times. Hence, average wage gains will not fall as much with n as when there is no heterogeneity in reservation prices. In fact, for even more skewed functions $\mathcal{G}(u)$, wage gains for workers with some mobility can increase enough to result in a profile that is not monotonic in n.

According to Proposition 3, for any $\mathcal{G}(u)$, we can in principle identify the shape of the offer distribution $F(\cdot)$. But since differences are less pronounced in the face of heterogeneity in reservation prices, we might need a much larger dataset to estimate average wage growth more precisely enough to rule out certain functional forms. Since one of the goals of this paper is to enable identification in large datasets where employment histories are left-censored, though, it might be possible to satisfy these data requirements using other datasets.

7. Conclusion

Standard approaches to estimating search models, as summarized in the recent survey article of Eckstein and van den Berg (2005), either abstract from unmeasured variation in wages or impose parametric assumptions to deal with them. This paper exploits the implicit record structure of simple search models and shows that these models remain nonparametrically identified when worker productivity is measured imperfectly, even in the presence of initial condition problems. These problems may arise because of censoring problems, or because workers set different reservation wages. Establishing this result required us to derive new results on records from observations that are independent but not identically distributed, a model which has not been analyzed so far in the statistics literature. Determining identification in even more realistic search frameworks is likely to require a more rigorous analysis of records drawn from an underlying sequence of observations that fails to satisfy the classical i.i.d. assumptions. The particular i.n.i.d structure analyzed here, and the tools we use to analyze it, hopefully represent a first step towards this goal.

In addition, this paper documented two new empirical findings: (1) the majority of employment cycles end after only one job; and (2) the distribution of the number of jobs per employment cycle has a fat tail. Viewed from the perspective of the standard search model, the first finding suggests a sluggish labor market, at least among employed searchers, since we would expect to see a fair number of workers moving on to at least a second job (namely those who drew wages slightly above their reservation price). By contrast, the second finding suggests a fluid labor market, since it implies some workers do manage to accumulate a lot of job offers. When we consider a variation of the standard search model where workers differ in their reservation values, our estimation gives more weight to the second observation. Thus, we infer that the rate at which employed workers receive offers is high, in line with estimates that use job duration data. But our approach also implies a much higher offer arrival rate for unemployed workers than previous work, perhaps implausibly so. Although we offered some directions for modifying the model, properly interpreting these facts remains as something that should be pursued more in future work.

Table 1: Distribution of N across employment cycles

Cycles than began before 1982 and end with at least one of potential experience

N	Definition 1		Definition 2	
	# of cycles	Pr(N = n)	# of cycles	Pr(N = n)
1	7555	0.612	2732	0.690
2	2853	0.231	748	0.189
3	1125	0.091	250	0.063
4	471	0.038	126	0.032
5	172	0.014	48	0.012
6	87	0.007	25	0.006
7	38	0.003	13	0.003
8	20	0.002	7	0.002
9	13	0.001	6	0.002
10	4	0.000	0	0.000
11	0	0.000	1	0.000
12	3	0.000	2	0.001

Definition 1 - cycles are partitioned according to nonemployment spells
Definition 2 - cycles are partitioned according to quit/layoff and time to next job

Table 2: Estimates for κ_1 and $G(F^{-1}(.))$

	Sample size	κ_1	a	b	Implied $\mathcal{G}'(1)$
No heterogeneity in reservation wages					
All	12,341	1.963			
		0.028			
Educ < 12	3,054	1.911			
		0.056			
Educ = 12	3,715	1.978			
		0.052			
Educ $\in$ (13,15)	3,273	2.188			
		0.060			
Educ $\geq$ 16	2,299	1.707			
		0.058			
Heterogeneity in reservation wages					
All	12,341	10.768	0.191	0.051	15.926
		1.7766	0.0350	0.0048	1.968
Educ < 12	3,054	10.082	0.1875	0.0554	14.855
		3.442	0.0752	0.0107	3.803
Educ = 12	3,715	9.489	0.224	0.054	14.647
		2.6549	0.0698	0.0079	2.942
Educ $\in$ (13,15)	3,273	16.385	0.145	0.042	20.324
		6.1542	0.0536	0.0109	6.242
Educ $\geq$ 16	2,299	8.442	0.209	0.053	15.083
		3.2509	0.0937	0.0103	4.042

Figure 1
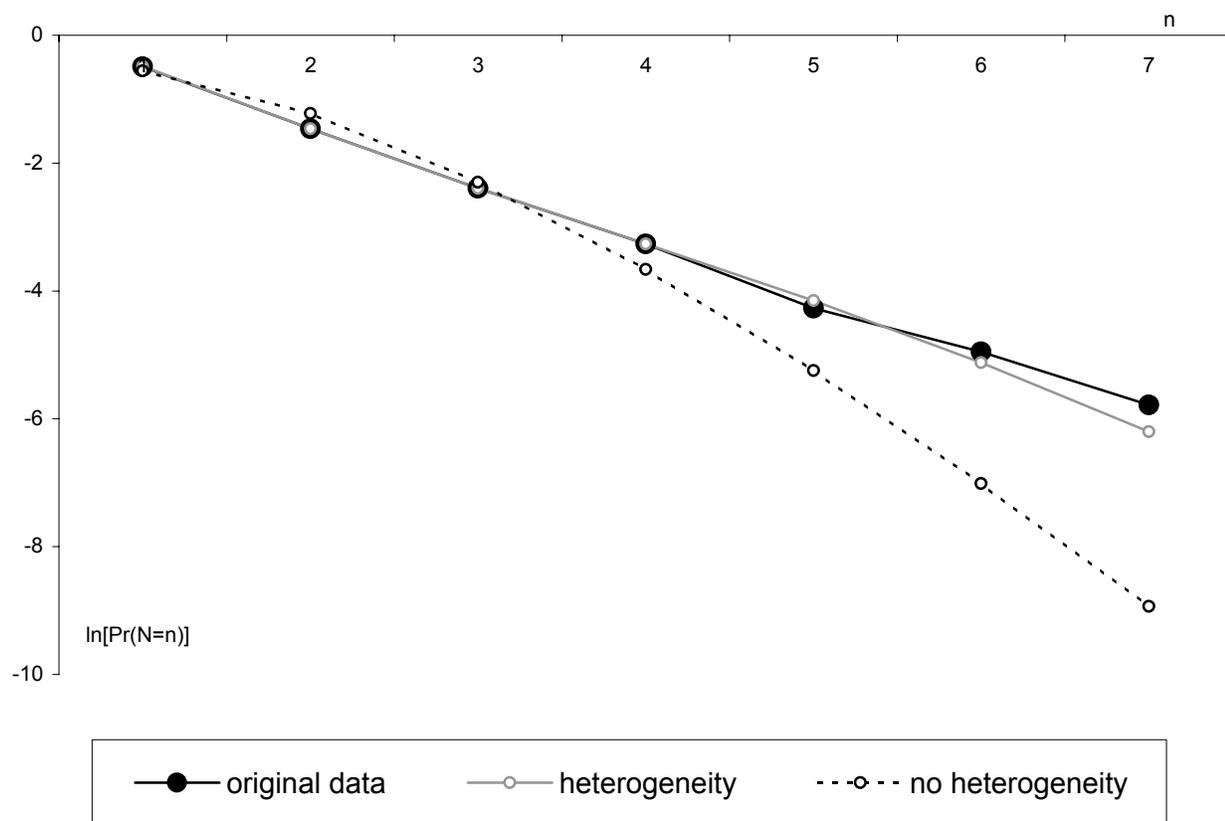
Figure 2
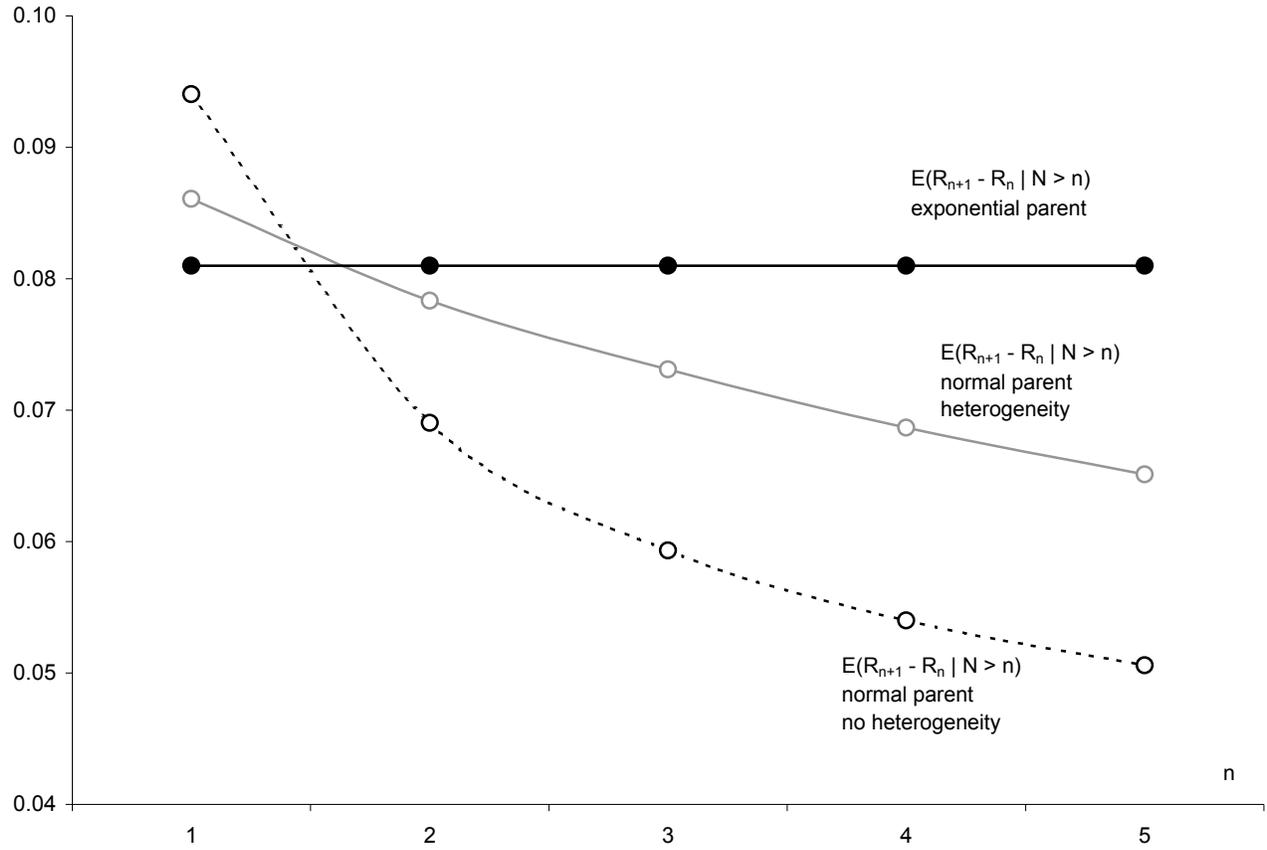
Appendix A: Proofs

Proof of Proposition 1: We begin with some necessary preliminaries. Since $G\left(F^{-1}\left(u\right)\right)$ is assumed to be absolutely continuous under Assumption 2.1, there exists a density function $\dfrac{d}{du}G\left(F^{-1}\left(u\right)\right)$ for almost all $u \in (0,1)$. Without loss of generality, we proceed as if there also exist density functions $g\left(x\right) = G'\left(x\right)$ and $f\left(x\right) = F'\left(x\right)$, even though their existence is not implied by Assumption 2.1. If either of these functions does not exist, we can always define a new set of variables $y'_m = F\left(y_m\right)$ so that y'_1 has a uniform distribution (which is absolutely continuous) and y'_m for $m \geq 2$ has an absolutely continuous distribution under Assumption 2.1. Since $F\left(\cdot\right)$ is monotonic, we can compute the likelihood of record events in the original system using record events in the analogous system where all variables have absolutely continuous distributions. The only consequence of proceeding this way is a slight abuse of notation; below, we will often write $\dfrac{g\left(F^{-1}\left(u\right)\right)}{f\left(F^{-1}\left(u\right)\right)}$ when we should write $\dfrac{d}{du}G\left(F^{-1}\left(u\right)\right)$ since the former representation may not exist. In the text we use are careful to refer to $\dfrac{d}{du}G\left(F^{-1}\left(u\right)\right)$ as $\mathcal{G}'\left(u\right)$.

Define $q = 1 - p$. From Bunge and Nagaraja (1991), we know that for $n \geq 2$, the likelihood of at least n records with values r_1 through r_n is given by

$$h\left(r_1, ..., r_n \ \cap N \geq n\right) = f\left(r_n\right) \frac{qg\left(r_1\right)}{1 - qF\left(r_1\right)} \prod_{i=2}^{n-1} \frac{qf\left(r_i\right)}{1 - qF\left(r_i\right)} \tag{7.1}$$

Integrating r_2 through r_{n-1} yields the following expression for the joint likelihood of r_1 and r_n:

$$h\left(r_1, r_n \cap N \geq n\right) = \frac{1}{(n-2)!}\left[-\ln\left(\frac{1 - qF\left(r_{n-1}\right)}{1 - qF\left(r_1\right)}\right)\right]^{n-2} \frac{qg\left(r_1\right)}{1 - qF\left(r_1\right)} f\left(r_n\right)$$

and so

$$
\begin{aligned}
E\left(|w_n| \mid N \geq n\right) &= \int_{-\infty}^{\infty}\int_{r_1}^{\infty} |r_n| \frac{\left[-\ln\left(\dfrac{1 - qF\left(r_{n-1}\right)}{1 - qF\left(r_1\right)}\right)\right]^{n-2}}{(n-2)!\Pr\left(N \geq n\right)} \frac{qg\left(r_1\right)}{1 - qF\left(r_1\right)} f\left(r_n\right) dr_n dr_1 \\
&= \int_0^1\int_0^{u_n} \left|F^{-1}\left(u_n\right)\right| \frac{\left[-\ln\left(\dfrac{1 - qu_{n-1}}{1 - qu_1}\right)\right]^{n-2}}{(n-2)!\Pr\left(N \geq n\right)} \frac{q}{1 - qu_1} \frac{g\left(F^{-1}\left(u_1\right)\right)}{f\left(F^{-1}\left(u_1\right)\right)} du_1 du_n
\end{aligned}
$$

For any pair $(u_1, u_{n-1}) \in [0,1] \times [0,1]$, it is always the case that

$$-\ln\left(\frac{1 - qu_{n-1}}{1 - qu_1}\right) \leq -\ln\left(1 - q\right) \quad \text{and} \quad \frac{q}{1 - qu_1} \leq \frac{q}{1 - q}$$

and so

$$
\begin{aligned}
E\left(|w_n| \mid N \geq n\right) &\leq \frac{\left[-\ln\left(1 - q\right)\right]^{n-2}}{(n-2)!P\left(N \geq n\right)} \frac{q}{1 - q} \int_0^1\int_0^{u_n} \left|F^{-1}\left(u_n\right)\right| \frac{g\left(F^{-1}\left(u_1\right)\right)}{f\left(F^{-1}\left(u_1\right)\right)} du_1 du_n \\
&= \frac{\left[-\ln\left(1 - q\right)\right]^{n-2}}{(n-2)!P\left(N \geq n\right)} \frac{q}{1 - q} \int_0^1 \left|F^{-1}\left(u_n\right)\right| \left[\int_0^{u_n} \frac{g\left(F^{-1}\left(u_1\right)\right)}{f\left(F^{-1}\left(u_1\right)\right)} du_1\right] du_n \\
&\leq \frac{\left[-\ln\left(1 - q\right)\right]^{n-2}}{(n-2)!P\left(N \geq n\right)} \frac{q}{1 - q} \int_0^1 \left|F^{-1}\left(u_n\right)\right| du_n < \infty
\end{aligned}
$$

where the last inequality follows from the assumption that $E\left(|y_2|\right) < \infty$. Hence, $E\left(|w_n| \mid N \geq n\right)$ is well-defined. But for any random variable Y, if $E\left(|Y|\right)$ exists then so does $E\left(Y\right)$. By a similar argument, $E\left(w_{n-1} \mid N \geq n\right)$ can also be shown to exist, and hence so does $E\left(w_n - w_{n-1} \mid N \geq n\right)$. ∎

Proof of Proposition 2: We begin by proving part (a). Let $q = 1 - p$. Using Nagaraja and Barlevy (2003), we can deduce that the likelihood of exactly n observed records with values r_1 through r_n is given by

$$h\left(r_1, ..., r_n \ \cap N = n\right) = \frac{\left(1 - q\right) g\left(r_1\right)}{1 - qF\left(r_1\right)} \prod_{i=2}^{n} \frac{qf\left(r_i\right)}{1 - qF\left(r_i\right)}.$$

Integrating out r_2 through r_n yields

$$h\left(r_1 \ \cap N = n\right) = \frac{1}{\left(n-1\right)!} \frac{1 - q}{1 - qF\left(r_1\right)} \left(\ln\left(\frac{1 - qF\left(r_1\right)}{1 - q}\right)\right)^{n-1} g\left(r_1\right)$$

Hence, $\Pr\left(N = n\right)$ is equal to

$$\Pr\left(N = n\right) = \frac{1}{\left(n-1\right)!} \int_{-\infty}^{\infty} \frac{1 - q}{1 - qF\left(r_1\right)} \left[\ln\left(\frac{1 - qF\left(r_1\right)}{1 - q}\right)\right]^{n-1} g\left(r_1\right) dr_1 \qquad (7.2)$$

Using the change of variables $u = F\left(r_1\right)$, we can rewrite (7.2) as

$$\Pr\left(N = n\right) = \frac{1}{\left(n-1\right)!} \int_0^1 \frac{1 - q}{1 - qu} \left[\ln\left(\frac{1 - qu}{1 - q}\right)\right]^{n-1} \frac{g\left(F^{-1}\left(u\right)\right)}{f\left(F^{-1}\left(u\right)\right)} du \qquad (7.3)$$

Recall that $\dfrac{g\left(F^{-1}\left(u\right)\right)}{f\left(F^{-1}\left(u\right)\right)}$ represents $\dfrac{d}{du} G\left(F^{-1}\left(u\right)\right)$. Under the additional assumption that $G\left(x\right) = G_0\left(F\left(x\right)\right)$, we have $G\left(F^{-1}\left(u\right)\right) = G_0\left(F\left(F^{-1}\left(u\right)\right)\right) = G_0\left(u\right)$. Hence, for any distribution $F\left(\cdot\right)$, $\dfrac{d}{du} G\left(F^{-1}\left(u\right)\right) = G_0'\left(u\right) \equiv g_0\left(u\right)$, i.e. the function $\dfrac{g\left(F^{-1}\left(u\right)\right)}{f\left(F^{-1}\left(u\right)\right)}$ in (7.3) is independent of the function $F\left(\cdot\right)$. To consider the most general case, we allow $g_0\left(u\right)$ to depend on the parameter q and we will write out $g_0\left(u, q\right)$. Equation (3.2) in the text offers one example of a function $g_0\left(\cdot\right)$ that depends on q, specifically $g_0\left(u\right) = \dfrac{d}{du}\left[u\left(1 + \kappa_1\left(1 - u\right)\right)^{-1}\right]$ where $\kappa_1 = \dfrac{q}{1 - q}$.

Suppose there were two values q_1 and q_2 which gave rise to the same $\Pr\left(N = n\right)$ for all n, i.e.

$$\int_0^1 \frac{1 - q_1}{1 - q_1 u} \left[\ln\left(\frac{1 - q_1 u}{1 - q_1}\right)\right]^{n-1} g_0\left(u, q_1\right) du = \int_0^1 \frac{1 - q_2}{1 - q_2 u} \left[\ln\left(\frac{1 - q_2 u}{1 - q_2}\right)\right]^{n-1} g_0\left(u, q_2\right) du$$

for $n = 1, 2, 3, ...$ We will show by contradiction that $q_1 = q_2$, which in turn implies that $\Pr\left(N = n\right)$ uniquely determines $p = 1 - q$.

Set $t = \ln\left(\left(1 - q_1 u\right)/\left(1 - q_1\right)\right)$ on the left-hand side and $t = \ln\left(\left(1 - q_2 u\right)/\left(1 - q_2\right)\right)$ on the right-hand side to rewrite the above equations as

$$\int_0^{-\ln\left(1 - q_1\right)} h_1\left(t\right) t^{n-1} dt = \int_0^{-\ln\left(1 - q_2\right)} \frac{q_1}{q_2} \frac{1 - q_2}{1 - q_1} h_2\left(t\right) t^{n-1} dt.$$

for $n = 1, 2, 3, ...$ Suppose $q_1 \neq q_2$, and assume wlog that $q_2 > q_1$, so $-\ln(1 - q_2) > -\ln(1 - q_1)$. Define

$$\widehat{h}_1(t) = \begin{cases} h_1(t) & \text{if } t \leq -\ln(1 - q_1) \\ \\ 0 & \text{if } t > -\ln(1 - q_1) \end{cases}$$

We can rewrite the above equation as

$$\int_0^{-\ln(1-q_1)} h_1(t) t^{n-1} dt = \int_0^{-\ln(1-q_2)} \widehat{h}_1(t) t^{n-1} dt = \int_0^{-\ln(1-q_2)} \frac{q_1}{q_2} \frac{1-q_2}{1-q_1} h_2(t) t^{n-1} dt$$

By the Müntz-Szász theorem (see Kamps (1998)), we can conclude that for almost all $t \in (0, -\ln(1 - q_2))$

$$\widehat{h}_1(t) = \frac{q_1}{q_2} \frac{1-q_2}{1-q_1} h_2(t)$$

Hence, $h_2(t) = 0$ for almost all $t \in (-\ln(1 - q_1), -\ln(1 - q_2))$. But this implies $g_0(u, q_2) = 0$ for almost all $u \in \left[F_2^{-1}(0), F_2^{-1}((q_2 - q_1)/[q_2(1 - q_1)]) \right]$, which violates Assumption 2.3. It follows that q_1 and q_2 must be equal.

We next establish (b). Using the likelihood function (7.1), changing variables and integrating out r_2 through r_{n-2}, we can express the average record gap $E(w_n - w_{n-1} \mid N \geq n)$ for any $n \geq 3$ as the following integral:

$$\int_0^1 \int_0^{u_{n-1}} \int_{u_{n-1}}^1 \left[F^{-1}(u_n) - F^{-1}(u_{n-1}) \right] \frac{q^2 \left[-\ln\left(\frac{1-qu_{n-1}}{1-qu_1}\right) \right]^{n-3} g_0(u_1)}{(n-3)! \Pr(N \geq n)(1 - qu_1)(1 - qu_{n-1})} du_n du_1 du_{n-1}$$

For a given value of q, the function $g_0(u, q)$ is known, as are the values of q and $\Pr(N \geq n)$. Define

$$\phi_F(u_{n-1}) = \int_{u_{n-1}}^1 \left[F^{-1}(u_n) - F^{-1}(u_{n-1}) \right] du_n$$

and introduce the change of variables

$$\begin{aligned} t &= -\ln(1 - qu_{n-1}) \\ s &= -\ln(1 - qu_1) \\ c &= -\ln(1 - q) \end{aligned}$$

so the average record gap is given by

$$E(w_n - w_{n-1} \mid N \geq n) = \frac{1}{(n-3)! \Pr(N \geq n)} \int_0^c \int_{t=s}^c g_0\left(\frac{1-e^{-s}}{q}\right) \phi_F\left(\frac{1-e^{-t}}{q}\right)(t-s)^{n-3} dt \, ds$$

Finally, we set $\omega = t - s$ and define

$$\eta_F(\omega) = \int_{t=\omega}^c g_0\left(\frac{1-e^{-(t-\omega)}}{q}\right) \phi_F\left(\frac{1-e^{-t}}{q}\right) dt.$$

to get

$$E(w_n - w_{n-1} \mid N \geq n) = \frac{1}{(n-3)! P(N \geq n)} \int_{\omega=0}^c \eta_F(\omega) \omega^{n-3} d\omega$$

Now, suppose there exist two functions F_1 and F_2 that give rise to the same sequence of expected record gaps $E(w_n - w_{n-1} \mid N \geq n)$. Then for $n = 3, 4, 5, \ldots$ it must be the case that

$$\int_{\omega=0}^{c} \eta_{F_1}(\omega) \omega^{n-3} d\omega = \int_{\omega=0}^{c} \eta_{F_2}(\omega) \omega^{n-3} d\omega$$

The Müntz-Szász theorem then implies that for almost all $\omega \in (0, c)$,

$$\eta_{F_1}(\omega) = \eta_{F_2}(\omega)$$

If we define $\phi(t) \equiv \phi_{F_2}\left(\frac{1-e^{-t}}{q}\right) - \phi_{F_1}\left(\frac{1-e^{-t}}{q}\right)$, then this implies that for almost all $\omega \in (0, c)$,

$$\int_{t=\omega}^{c} g_0\left(\frac{1 - e^{-(t-\omega)}}{q}\right) \phi(t) \, dt = 0 \tag{7.4}$$

We next argue that (7.4) requires $\phi(t) = 0$ almost surely. We first appeal to yet another change in variables, $w = c - t$ and $z = c - \omega$ to rewrite (7.4) as

$$\int_0^z a(z - w) b(w) \, dw = 0 \text{ for almost all } z \in (0, c) \tag{7.5}$$

where $a(x) = g_0\left(\frac{1 - e^{-x}}{q}\right)$ and $b(x) = \phi(c - x)$. Applying Theorem VII in Titchmarsh (1926), which is identical to Theorem 151 from the more accessible Titchmarsh (1948, p. 324-5), there exists a c^* such that $a(x) = 0$ for all $x \in (0, c^*)$ and $b(x) = 0$ for all $x \in (0, c - c^*)$. But by Assumption 2.3, there exists an $\varepsilon > 0$ such that $g_0(u) > 0$ for all $u \in (0, \varepsilon)$, and so there exists an ε' such that $a(x) > 0$ for all $x \in (0, \varepsilon')$. It follows that c^* must equal 0, and hence $b(z) = 0$ for almost all $z \in (0, c)$, which in turn implies $\phi(t) = b(c - t) = 0$ for almost all $t \in (0, c)$.

Thus far, we have shown that for any two distributions F_1 and F_2 that give rise to the same sequence of expected record gaps $E(w_n - w_{n-1} \mid N \geq n)$, it must be the case that $\phi_{F_1}(u) = \phi_{F_2}(u)$ for almost all $u \in (0, 1)$. The last step is to show that this implies F_1 and F_2 are identical almost surely up to a location shift, i.e. there exists some constant C such that for almost all $u \in (0, 1)$,

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

Our argument follows Nagaraja and Barlevy (2003). Since $\phi_{F_1}(u) = \phi_{F_2}(u)$ almost surely, then for almost all $u \in (0, 1)$, we have

$$\int_u^1 \left[F_1^{-1}(x) - F_1^{-1}(u)\right] dx = \int_u^1 \left[F_2^{-1}(x) - F_2^{-1}(u)\right] dx$$

or, rearranging,

$$\begin{aligned}
\int_u^1 \left[F_1^{-1}(x) - F_2^{-1}(x)\right] dx &= \int_u^1 \left[F_1^{-1}(u) - F_2^{-1}(u)\right] dx \\
&= (1 - u)\left[F_1^{-1}(u) - F_2^{-1}(u)\right]
\end{aligned}$$

Define $H(x) = F_2^{-1}(x) - F_1^{-1}(x)$. Then we the above equation implies that for almost all $u \in (0, 1)$,

$$\int_u^1 H(x) \, dx = (1 - u)H(u) \tag{7.6}$$

Next, we observe that

$$\frac{d}{du}[\ln \int_u^1 H(x)\,dx] = -\frac{H(u)}{\int_u^1 H(v)\,dv}$$

Hence, (7.6) implies that

$$\frac{d}{du}[\ln \int_u^1 H(x)\,dx] = -\frac{1}{1-u}$$

Integrate both sides over $u \in (s,t) \subset (0,1)$ yields

$$\ln \int_t^1 H(x)\,dx - \ln \int_s^1 H(x)\,dx = \ln(1-t) - \ln(1-s)$$

Since this is true for any s and any t, it follows that there exists a constant c such that for all $t \in (0,1)$,

$$-\log \int_t^1 H(v)dv + \log(1-t) = c$$

or

$$\int_t^1 H(x)\,dx = e^{-c}(1-t), \quad t \in (0,1).$$

Differentiating with respect to t, we obtain $H(t) = e^{-c}$ for all $t \in (0,1)$, that is $F_2^{-1}(t) - F_1^{-1}(t) = e^{-c}$ for almost all $t \in (0,1)$, as we need to show. ■

   Proof of Proposition 3: Starting with (7.2) and using the change of variable $u = F(r)$, we obtain the following expression:

$$\Pr(N = n) = \int_0^1 \frac{1-q_1}{1-q_1 u}\left[\ln\left(\frac{1-q_1 u}{1-q_1}\right)\right]^{n-1}\frac{g\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}du$$

Suppose there were two triplets $\{q_1, F_1, G_1\}$ and $\{q_2, F_2, G_2\}$ that gave rise to the same $\Pr(N = n)$ for all n, i.e.

$$\int_0^1 \frac{1-q_1}{1-q_1 u}\left[\ln\left(\frac{1-q_1 u}{1-q_1}\right)\right]^{n-1}\frac{g_1\left(F_1^{-1}(u)\right)}{f_1\left(F_1^{-1}(u)\right)}du = \int_0^1 \frac{1-q_2}{1-q_2 u}\left[\ln\left(\frac{1-q_2 u}{1-q_2}\right)\right]^{n-1}\frac{g_2\left(F_2^{-1}(u)\right)}{f_2\left(F_2^{-1}(u)\right)}du$$

for all $n = 1,2,3,...$ As in the proof of Proposition 2, set $t = \ln((1-q_1 u)/(1-q_1))$ on the left-hand side and $t = \ln((1-q_2 u)/(1-q_2))$ on the right-hand side to rewrite the above equations as

$$\int_0^{-\ln(1-q_1)} h_1(t)\,t^{n-1}dt = \int_0^{-\ln(1-q_2)} \frac{q_1}{q_2}\frac{1-q_2}{1-q_1}h_2(t)\,t^{n-1}dt.$$

for all $n = 1,2,3,...$ Just as in the proof of Proposition 2, it follows that

$$q_1 = q_2$$

   Next, since $q_1 = q_2$, the fact that $\{q, F_1, G_1\}$ and $\{q, F_2, G_2\}$ both give rise to the same distribution $\Pr(N = n)$ implies that

$$\int_0^1 \frac{1-q}{1-qu}\left[\ln\left(\frac{1-qu}{1-q}\right)\right]^{n-1}\frac{g_1\left(F_1^{-1}(u)\right)}{f_1\left(F_1^{-1}(u)\right)}du = \int_0^1 \frac{1-q}{1-qu}\left[\ln\left(\frac{1-qu}{1-q}\right)\right]^{n-1}\frac{g_2\left(F_2^{-1}(u)\right)}{f_2\left(F_2^{-1}(u)\right)}du$$

Appealing to the Müntz-Szász theorem implies that for almost all $u \in (0, 1)$,

$$\frac{g_1\left(F_1^{-1}\left(u\right)\right)}{f_1\left(F_1^{-1}\left(u\right)\right)} = \frac{g_2\left(F_2^{-1}\left(u\right)\right)}{f_2\left(F_2^{-1}\left(u\right)\right)}$$

Note that this function corresonds to $\mathcal{G}'\left(u\right) = \frac{d}{du}G\left(F^{-1}\left(u\right)\right)$, and hence $\Pr\left(N = n\right)$ uniquely identifies $\mathcal{G}\left(u\right)$ as noted in the text.

Define $g_0\left(u\right) = \mathcal{G}'\left(u\right)$. We then repeat the steps of the proof in Proposition 2 to argue that $F\left(\cdot\right)$ is uniquely determined up to an affine transformation. Since $\mathcal{G}\left(u\right) = G\left(F^{-1}\left(u\right)\right)$, it follows that for any $u \in (0, 1)$, we have $G^{-1}\left(u\right) = F^{-1}\left(\mathcal{G}^{-1}\left(u\right)\right)$. Hence, $G^{-1}\left(\cdot\right)$ can be identified up to the same constant as $F\left(\cdot\right)$. ∎

Appendix B: Bargaining

In this section, we describe a particular bargaining game whose reduced form corresponds to the model in our paper. We consider an alternating-offer bargaining game along the lines first proposed by Rubinstein (1982): the worker and the firm alternate in proposing a schedule of payments the worker should receive over the course of the job. More precisely, in line with the model we described, we require that the proposed schedule assumes the form of a function $w\left(z, \ell_{it}\right)$ that specifies the worker receive a payment that is entirely a function of his ability ℓ_{it}, although the amount can vary with the productivity z of the match. If a party proposes a schedule and the other party accepts, production takes place and the worker is paid according to this schedule. If the worker proposes a schedule that is rejected, the two parties must wait Δ_w units of time, after which the employer gets to propose his own schedule. If the employer proposes a schedule that is rejected, the two parties must wait Δ_e units of time, at which point the worker gets to propose an offer. Let

$$\beta = \frac{\Delta_w}{\Delta_e + \Delta_w}$$

We will consider taking the limit as Δ_w and Δ_e tend towards zero while holding the ratio Δ_e/Δ_w, and hence β, fixed. In the limit, it doesn't matter whether the worker or the employer makes the first offer.

As emphasized by Binmore, Rubinstein, and Wolinsky (1986), the outcome of the bargain depends crucially on what we assume occurs while the parties wait between offers. We assume the worker and the employer bargain in real time, and thus discount the future at rate ρ. While they wait, the two may be hit by a shock that causes them to separate (which recall arrives at a rate δ per unit time), and the worker may continue to encounter other employers. Hence, if Δ units of time have passed, where Δ is small, there is a probability of roughly $\delta\Delta$ the two will have separated. Moreover, if we assume that a worker will only change employers if the employer he encounters is more productive (i.e. bargaining leads to efficient mobility decisions), then there is a probability of roughly $\lambda_1\left(1 - \Gamma\left(z\right)\right)\Delta$ that the worker would leave for a more productive employer. Correspondingly, the probability the two will remain together and continue with a counteroffer is approximately $1 - \left[\delta + \lambda_1\left(1 - \Gamma\left(z\right)\right)\right]\Delta$. Finally, we assume that during the period Δ, the worker is unable to enjoy leisure. This assumption implies that all workers, regardless of their value of leisure, will negotiate to the same wage. This is a reasonable assumption: workers who value their leisure more will certainly be more choosy, but will probably not be able to use their higher value of leisure to extract higher wages. In reality this is probably because the value of leisure is hard to verify, but rather than use a model of bargaining with private information, it is simpler to assume workers cannot enjoy leisure

while bargaining. Hall and Milgrom (2005) propose a similar scheme and argue it provides a plausible description of actual wage bargaining.

Binmore, Rubinstein, and Wolinsky (1986) show that in the limit as Δ_w and Δ_e tend to 0, the schedule $w(z, \ell_{it})$ solves the Nash bargaining problem

$$\max_{w(z,\ell_{it})} [J(z, \ell_{it}) - J_0]^{1-\beta} [W(z, \ell_{it}) - W_0]^{\beta} \tag{7.7}$$

where $J(z, \ell_{it})$ denotes the expected utility of an employer who employs a worker of ability ℓ on a job with productivity z, $W(z, \ell_{it})$ denotes the expected utility of a worker of ability ℓ when working on a a job with productivity z, and J_0 and W_0 denotes the expected utility of the employer and the worker respectively if the two fail to come to agreement. Turning first to J_0, if the parties fail to agree, either the two will remain together after Δ units of time, in which case the utility of the employer will be $J(z, \ell_{it})$, or else the two will have separated by then, in which case we assume the employer has a utility of zero (as would be implied by a free entry condition). Discounting the future at the rate ρ, we have

$$J_0 \approx \frac{(1 - (\delta + \lambda_1 [1 - \Gamma(z)]) \Delta) J(z, \ell_{it+\Delta})}{1 + \rho \Delta}$$

Similarly, after Δ units of time, the worker will either be unemployed, in which case his utility is defined by the utility of an unmployed worker U, employed on a better job, which is associated with a utility of $E[W(z', \ell_{it+\Delta}) \mid z' \geq z]$, or else he will remain on the same match, which yields a utility of $W(z, \ell_{it+\Delta})$. Hence,

$$W_0 \approx \frac{(1 - (\delta + \lambda_1 [1 - \Gamma(z)]) \Delta) W(z, \ell_{it+\Delta}) + \delta \Delta U + \lambda_1 (1 - \Gamma(z)) E[W(z', \ell_{it+\Delta}) \mid z' \geq z]}{1 + \rho \Delta}$$

Substituting these expressions in, dividing by Δ^2 and taking the limit as $\Delta \to 0$ implies that $w(z, \ell_{it})$ solves

$$\max_{w(z,\ell_{it})} (z\ell - w)^{1-\beta} w^{\beta}$$

which implies

$$w(z, \ell_{it}) = \beta z \ell$$

Hence, the wage offered to a worker under this particular bargaining protocol will be proportional to his output, and all workers face the same potential offer distribution. Note that under this outcome a worker would indeed switch jobs if and only if the employer he encounteres has a higher productivity z, in line with our assumption. Hence, the wage schedule above represents a proper equilibrium. Since the wage it implies is proportional to productivity, the distribution $F(\cdot)$ is once again identically equal to $\Gamma(\cdot)$ up to a scaling factor, and so we can easily identify $\Gamma(\cdot)$ from $F(\cdot)$.

Finally, we should note that in recent work, Shimer (2005) examined bargaining in a similar model with on-the-job search. His formulation, which borrows from Binmore, Rubinstein, and Wolinsky (1986), assumes the agents do not discount between offers; rather, there is a constant rate that negotiations break down, independently of how the worker ranks his current employer relative to other employers. Under this alternative formulation, he shows that the wage schedule solves an analogous problem to (7.7), but with J_0 replaced by 0 and W_0 replaced by U. Shimer shows that when $\Gamma(\cdot)$ is a continuous distribution, there exist equilibria in which all firms with the same productivity z end up agreeing with their workers on the same wage. However, in his formulation worker productivity ℓ_{it} is assumed to be fixed over time. Once we allow for ℓ_{it} to vary over time, the solution $w(z, \ell_{it})$ that solves this alternative problem is typically not linear in ℓ_{it} or separable in z and ℓ_{it}. Thus, his proposed scheme will typically not yield the model we describe as a reduced form (except for particular processes ℓ_{it}). This underscores that while our model is consistent with some models of wage determination, it is inconsistent with others.

References

[1] Abowd, John, Francis Kramarz, and David Margolis, 1999. "High Wage Workers and High Wage Firms" *Econometrica*, March, 67(2), p251-333.

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

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

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

[5] Barlevy, Gadi, 2005. "Identification of Job Search Models using Record Statistics" Federal Reserve Bank of Chicago Working Paper.

[6] Barlevy, Gadi and H. N. Nagaraja, 2005. "Characterizations in a random record model with a non-identically distributed initial record" Federal Reserve Bank of Chicago Working Paper.

[7] Binmore, Kenneth, Ariel Rubinstein, and Asher Wolinsky, 1986. "The Nash Bargaining Solution in Economic Modelling" *Rand Journal of Economics*, Summer, 17(2), p176-88.

[8] Bontemps, Christian, Jean-Marc Robin, and Gerard van den Berg, 1999. "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms" *International Economic Review*, November, 40(4), p1039-1074.

[9] Bontemps, Christian, Jean-Marc Robin, and Gerard van den Berg, 2000. "Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation" *International Economic Review*, May, 41(2), p305-358.

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

[11] Bunge, John and H. Nagaraja, 1991. "The Distributions of Certain Record Statistics from a Random Number of Observations" *Stochastic Processes and Their Applications*, 38, p167-83.

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

[13] Eckstein, Zvi and Gerard van den Berg, 2005. "Empirical Labor Search: a Survey" Tel Aviv University Working paper and *Journal of Econometrics* (forthcoming).

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

[15] Flinn, Christopher, 2002. "Labour Market Structure and Inequality: a Comparison of Italy and the U.S." *Review of Economic Studies*, July, 69 (3), p611-45.

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

[17] Gautier, Peter, Coen Teulings, and Aico van Vuuren, 2005. "On-the-Job Search and Sorting" Tinbergen Institute Discussion Paper.

[18] Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong, 2000. "Optimal Nonparametric Estimation of First-Price Auctions" *Econometrica*, May, 68(3), p525-74.

[19] Kamps, Udo, 1998. Characterizations of distributions by recurrence relations and identities for moments of order statistics. *Handbook of Statistics, Vol. 16*, eds. N. Balakrishnan and C. R. Rao, Elsevier, Amsterdam. 291-311.

[20] Low, Hamish, Costas Meghir, and Luigi Pistaferri, 2004. "Wage Risk and Employment Risk over the Life Cycle" Stanford University Working Paper.

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

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

[23] Mortensen, Dale and George Neumann, 1988. "Esimating Structural Models of Unemployment and Job Duration" *Dynamic Econometric Modeling: Proceedings of the Third International Symposia in Economic Theory and Econometrics*, eds. W. A. Barnett, E. Berndt, and H. White. Cambridge: Cambridge University Press.

[24] Nagaraja, H. N. and Gadi Barlevy, 2003. "Characterizations Using Record Moments in a Random Record Model and Applications" *Journal of Applied Probability*, September, 40(3), p826-33.

[25] Nevzorov, V. B. and N. Balakrishnan, 1998. "A Record of Records" *Handbook of Statistics: Order Statistics, Theory and Methods*, v16, eds. N. Balakrishnan and C. R. Rao, Amsterdam: North-Holland, p515-70.

[26] Pierret, Charles, 2001. "Event History Data and Survey Recall: an Analysis of the National Longitudinal Survey of Youth 1979 Recall Experiment" *Journal of Human Resources*, Summer, 36(3), p439-66.

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

[28] Rubinstein, Ariel, 1982. "Perfect Equlibrium in a Bargaining Model" *Econometrica*, January, 50(1), p97-109.

[29] Shimer, Robert, 2005. "On-the-Job Search, Bargaining, and Wage Dispersion" Working Paper, University of Chicago.

[30] Shohat J. A. and J. D. Tamarkin, 1943. *The Problem of Moments.* New York: American Mathematical Society.

[31] Titchmarsh, E. C., 1926. "The zeros of certain integral functions" *Proceedings of the London Mathematical Society*, Series 2, **25** 283-302.

[32] Titchmarsh, E. C., 1948. *Introduction to the Theory of Fourier Integrals.* 2nd ed. Oxford: Clarendon Press.

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

[34] Wolpin, Kenneth, 1992. "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth" *Journal of Political Economy*, June, 100(3), p535-60.