Decomposing Duration Dependence
in a Stopping Time Model

Fernando Alvarez

University of Chicago

Katarína Borovičková

New York University

Robert Shimer

University of Chicago

November 10, 2015

Abstract

We develop a simple dynamic model of a worker's transitions between employment and non-employment. Our model implies that a worker finds a job at an optimal stopping time, when a Brownian motion with drift hits a barrier. The model has structural duration dependence in the job finding rate, in the sense that the hazard rate of finding a job changes during a non-employment spell for a given worker. In addition, we allow for arbitrary parameter heterogeneity across workers, so dynamic selection also affects the average job finding rate at different durations. We show that our model has testable implications if we observe at least two completed non-employment spells for each worker. Moreover, we can identify the distribution of a subset of our model's parameters using data on the duration of repeated non-employment spells and use the estimated parameters to understand the determinants of duration dependence. We use a large panel of social security data for Austrian workers to test and estimate the model. Our model is not rejected by the data. Our parameter estimates indicate that dynamic selection is critical for understanding the evolution of the aggregate job finding rate.

1 Introduction

The hazard rate of finding a job is higher for workers who have just exited employment than for workers who have been out of work for a long time. Economists and statisticians have long understood that this reflects a combination of two factors: structural duration dependence in the job finding probability for each individual worker, and changes in the composition of workers at different non-employment durations (Cox, 1972). The goal of this paper is develop a flexible but testable model of the job finding rate for any individual worker and use it to provide nonparametric decompositions of these two factors. We find that most of the observed decline in the job finding probability is due to changes in the composition of searching workers over time.

Our analysis is built around a structural model which views finding a job as an optimal stopping problem. One interpretation of our structural model is a classical theory of employment. All individuals always have two options, working at some wage $w(t)$ or not working and receiving some income and utility from leisure $b(t)$. The difference between these values is persistent but changes over time. If there were no cost of switching employment status, an individual would work if and only if the wage is sufficiently high relative to the value of not working. We add a switching cost to this simple model, so a worker starts working when the difference between the wage and the value of leisure is sufficiently large and stops working when the difference is sufficiently small. Given a specification of the individual's preferences, a level of the switching cost, and the stochastic process for the wage and non-employment income, this theory generates a structural model of duration dependence for any individual worker. For instance, the model allows parameters where, as in Ljungqvist and Sargent (1998), "workers accumulate skills on the job and lose skills during unemployment".

An alternative interpretation of our structural model is a classical theory of unemployment. According to this interpretation, a worker's productivity $p(t)$ and her wage $w(t)$ follow a stochastic process. Again, the difference is persistent but changes over time. If the worker is unemployed, a monopsonist has the option of employing the worker, earning flow profits $p(t) - w(t)$, by paying a fixed cost. There may similarly be a fixed cost of firing the worker. Given a specification of the hiring cost and the stochastic process for productivity and the wage, the theory generates the same structural duration dependence for any individual worker.

We also allow for arbitrary individual heterogeneity in the parameters describing for preferences, fixed costs, and stochastic processes. For example, some individuals may expect the residual duration of their non-employment spell to increase the longer they stay out of work while others may expect it to fall. We maintain two key restrictions: for each

1

individual, the evolution of a latent variable, the net benefit from employment, follows a geometric Brownian motion with drift during a non-employment spell; and each individual starts working when the net benefit exceeds some fixed threshold and stops working when it falls below some (weakly) lower threshold. In the first interpretation of our structural model, this threshold is determined by the worker while in the second interpretation it is determined by the firm. These assumptions imply that the duration of a non-employment spell is given by the first passage time of a Brownian motion with drift, a random variable with an inverse Gaussian distribution. The parameters of the inverse Gaussian distribution are fixed over time for each individual but may vary arbitrarily across individuals.

In this environment, we ask four key questions. First, we ask whether the distribution of unobserved heterogeneity is identified. We prove that an economist armed with data on the joint distribution of the duration of two non-employment spells can identify the population distribution of the parameters of the inverse Gaussian distribution, except for the sign of the drift in the underlying Brownian motion. We discuss this important limitation to identification and analyze how it affects the interpretation of our result.

Second we ask whether the model has testable implications. We show that an economist armed with the same data on the joint distribution of the duration of two spells can potentially reject the model. Moreover, the test has power against competing models. We prove that if the true data generating process is one in which each individual has a constant hazard of finding a job, the economist will always reject our model. Similarly, we prove that if the true data generating process is one in which each individual has a log-Normal distribution for duration, the economist will always reject our model. The same result holds if the data generating process is a finite mixture of such models.

Third, we ask whether we can use the partial identification of the model parameters to decompose the observed evolution of the hazard of exiting non-employment into the portion attributable to structural duration dependence and the portion attributable to unobserved heterogeneity. We propose a simple decomposition of both the hazard rate and of the residual duration of a non-employment spell.

Finally, we show that we can use duration data as well as information about wage dynamics to infer the size of the fixed cost of switching employment status. Even small fixed costs give rise to a large region of inaction, which in turn affects the duration of job search spells. We show how to invert this relationship to recover the fixed costs.

We then use data from the Austrian social security registry from 1986 to 2007 to test our model, estimate the distribution of unobserved parameters, and evaluate the decomposition. Using data on nearly one million individuals who experience at least two non-employment spells, we find that we cannot reject our model and we uncover substantial heterogeneity

across individuals. Although the raw hazard rate is hump-shaped with a peak at around 13 weeks, the hazard rate for the average individual increases until about 33 weeks and then scarcely declines. Similarly, the bulk of the increase in the residual duration of an in-progress job-search spell is a consequence of the changing composition of the searching population, with little of it explained by conditions getting worse for an individual as his duration increases. We also estimate tiny fixed costs. For most individuals, the total cost of taking a job and later leaving it are approximately equal to one hour of leisure.

There are a few other papers that use the first passage time of a Brownian motion to model duration dependence. Lancaster (1972) examines whether such a model does a good job of describing the duration of strikes in the United Kingdom. He creates 8 industry groups and observes between 54 and 225 strikes per industry group. He then estimates the parameters of the first passage time under the assumption that they are fixed within industry group but allowed to vary arbitrarily across groups. He concludes that the model does a good job of describing the duration of strikes, although subsequent research armed with better data reached a different conclusion (Newby and Winterton, 1983). In contrast, our testing and identification results require only two observations per individual and allow for arbitrary heterogeneity across individuals.

Shimer (2008) assumes that the duration of an unemployment spell is given by the first passage time of a Brownian motion but does not allow for any heterogeneity across individuals. The first passage time model has also been adopted in medical statistics, where the latent variable is a patient's health and the outcome of interest is mortality (Aalen and Gjessing, 2001; Lee and Whitmore, 2006, 2010). For obvious reasons, such data do not allow for multiple observations per individual, and so bio-statistical researchers have so far not introduced unobserved individual heterogeneity into the model. These papers have also not been particularly concerned with either testing or identification of the model.

Abbring (2012) considers a more general model than ours, allowing that the latent net benefit from employment is spectrally negative Lévy process, e.g. the sum of a Brownian motion with drift and a Poisson process with negative increments. On the other hand, he assumes that individuals differ only along a single dimension, the distance between the barrier for stopping and starting an employment spell. In contrast, we allow for two dimensions of heterogeneity, and so our approach to identification is completely different.

Within economics, the mixed proportional hazard model (Lancaster, 1979) has received far more attention than the first passage time model. This model assumes that the probability of finding a job at duration t is the product of three terms: a baseline hazard rate that varies depending on the duration of non-employment, a function of observable characteristics of individuals, and an unobservable characteristic. Our model neither nests the

mixed proportional hazard model nor is it nested by that model. A large literature, starting with Elbers and Ridder (1982) and Heckman and Singer (1984a), show that such a model is nonparametrically identified using a single spell of non-employment and appropriate variation in the observable characteristics of individuals. Heckman and Singer (1984b) illustrates the perils of parametric identification strategies in this context. Closer to the spirit of our paper, Honoré (1993) shows that the mixed proportional hazard model is also nonparametrically identified with data on the duration of at least two non-employment spells for each individual.

The remainder of the paper proceeds as follows. In Section 2, we describe our structural model and show how to use the model to address the questions of interest. We prove that a subset of the parameters is nonparametrically identified if we observe at least two non-employment spells for each individual, that the model has testable implications under the same conditions, and that we can use the model to decompose changes in the hazard of exiting non-employment and the residual duration of a non-employment spell into the portion that is structural and the portion that is attributable to changes in the composition of the non-employment pool. Section 4 summarizes the Austrian social security registry data. Section 5 presents our results, including tests and estimates of the model, decomposition of hazard rates and residual duration, and inference of the distribution of fixed costs.

2   Theory

2.1   Structural Model

We consider the problem of a risk-neutral, infinitely-lived worker with discount rate r, who can either be employed, $s(t) = e$, or non-employed, $s(t) = n$, at each instant in continuous time t. We describe and solve the worker's problem in Appendix A and here focus on the key results.

We assume an employed worker earns a wage $e^{w(t)}$ and a non-employed worker gets flow utility $b_0 \, e^{b(t)}$, where b_0 is a positive constant. Both $w(t)$ and $b(t)$ follow correlated Brownian motions with drift, both when the worker is employed and when the worker is non-employed. The drift and standard deviation of each may depend on the worker's employment status. In order for the problem to be well-behaved, we impose in Appendix A restrictions on the drift and volatility of $w(t)$ and $b(t)$ both while employed and non-employed to ensure that the worker's value is finite.

A non-employed worker can become employed at t by paying a fixed cost $\psi_e e^{b(t)}$ for a constant $\psi_e \geq 0$. Likewise, an employed worker can become non-employed by paying a

cost $\psi_n e^{b(t)}$ for a constant $\psi_n \geq 0$. The worker must decide optimally when to change her employment status $s(t)$.

It will be convenient to define $\omega(t) \equiv w(t) - b(t)$. This inherits the properties of w and b, following a random walk with state-dependent drift and volatility given by:

$$d\omega(t) = \mu_{s(t)}dt + \sigma_{s(t)}dB(t), \tag{1}$$

where $B(t)$ is a standard Brownian motion.

The worker's log net benefit from employment is $\omega(t) - \log b_0$. With a slight abuse of terminology, we will refer to $\omega(t)$ alone as the net benefit from employment. We prove in Appendix A that the worker's employment decision depends only on her employment status $s(t)$ and her net benefit from employment. In particular, the worker's optimal policy involves a pair of thresholds. If $s(t) = e$ and $\omega(t) \geq \underline{\omega}$, the worker remains employed, while she stops working the first time $\omega(t) < \underline{\omega}$. If $s(t) = n$ and $\omega(t) \leq \bar{\omega}$, the worker remains non-employed, while she takes a job the first time $\omega(t) > \bar{\omega}$. Assuming the sum of the fixed costs $\psi_e + \psi_n$ is strictly positive, the thresholds satisfy $\bar{\omega} > \underline{\omega}$, while the thresholds are equal if both fixed costs are zero.

Proposition 4 in Appendix A provides an approximate characterization of the distance between the thresholds, $\bar{\omega} - \underline{\omega}$, as a function of the fixed costs when the fixed costs are small for arbitrary parameter values. Here we consider a special case, where the utility from unemployment is constant, $b(t) = 0$ for all t. We still allow the stochastic process for wages to depend on a worker's employment status. Then

$$(\bar{\omega} - \underline{\omega})^3 \approx \frac{12r\sigma_e^2\sigma_n^2}{(\mu_e + \sqrt{\mu_e^2 + 2r\sigma_e^2})(-\mu_n + \sqrt{\mu_n^2 + 2r\sigma_n^2})} \frac{\psi_e + \psi_n}{b_0} \tag{2}$$

An increase in the fixed costs relative to the utility of unemployment increases the distance between the thresholds $\bar{\omega} - \underline{\omega}$, as one would expect. An increase in the volatility of the net benefit from employment, σ_n or σ_e, has the same effect because it raises the option value of delay. An increase in the drift in the net benefit from employment while out of work, μ_n, or a decrease in the drift in the net benefit from employment while employed, μ_e, also increases the distance between the thresholds. Intuitively, an increase in μ_n or a reduction in μ_e reduces the amount of time it takes to go between any fixed thresholds. The worker optimally responds by increasing the distance between the thresholds.

We have so far described a model of voluntary non-employment, in the sense that a worker optimally chooses when to work. But a simple reinterpretation of the objects in the model turns it into a model of involuntary unemployment. In this interpretation, the wage

is $b_0\, e^{b(t)}$, while a worker's productivity is $e^{w(t)}$. If the worker is employed by a monopsonist, it earns flow profits $e^{w(t)} - b_0 e^{b(t)}$. If the worker is unemployed, a firm may hire her by paying a fixed cost $\psi_e e^{b(t)}$, and similarly the firm must pay $\psi_n e^{b(t)}$ to fire the worker. In this case, the firm's optimal policy involves the same pair of thresholds. If $s(t) = e$ and $\omega(t) \geq \underline{\omega}$, the firm retains the worker, while she is fired the first time $\omega(t) < \underline{\omega}$. If $s(t) = n$ and $\omega(t) \leq \bar{\omega}$, the worker remains unemployed, while a firm hires her the first time $\omega(t) > \bar{\omega}$.

This structural model is similar to the one in Alvarez and Shimer (2011) and Shimer (2008). In particular, setting the switching cost to zero ($\psi_e = \psi_n = 0$) gives a decision rule with $\bar{\omega} = \underline{\omega}$, as in the version of Alvarez and Shimer (2011) with only rest unemployment, and with the same implication for non-employment duration as Shimer (2008). Another difference is that here we allow the process for wages to depend on a worker's employment status, $(\mu_e, \sigma_e) \neq (\mu_n, \sigma_n)$. The difference in the drift μ_e and μ_n allows us to capture structural features such as those emphasized by Ljungqvist and Sargent (1998), who explain the high duration of European unemployment by using "...a search model where workers accumulate skills on the job and lose skills during unemployment."

The most important difference is that this paper allows for arbitrary time-invariant worker heterogeneity. An individual worker is described by a large number of structural parameters, including her discount rate r, her fixed costs ψ_e, and ψ_n, and all the parameters governing the joint stochastic processes for her potential wage and benefit, both while the worker is employed and while she is non-employed. Our analysis allows for arbitrary distributions of these structural parameters in the population, subject only to the constraints that the utilities are finite.

2.2   Duration Distribution

We turn next to the determination of non-employment duration. All non-employment spells start when an employed worker's wage hits the lower threshold $\underline{\omega}$. The log net benefit from employment then follows the stochastic process $d\omega(t) = \mu_n dt + \sigma_n dB(t)$ and the non-employment spell ends when the worker's log net benefit from employment hits the upper threshold $\bar{\omega}$. Therefore the length of a non-employment spell is given by the first passage time of a Brownian motion with drift. This random variable has an inverse Gaussian distribution with density function

$$f(t; \alpha, \beta) = \frac{\beta}{\sqrt{2\,\pi}\, t^{3/2}} e^{-\frac{(\alpha\, t - \beta)^2}{2\, t}} \tag{3}$$

where $\alpha \equiv \mu_n / \sigma_n$ and $\beta \equiv (\bar{\omega} - \underline{\omega})/\sigma_n$. Note $\beta \geq 0$ by assumption, while α may be positive or negative. If $\alpha \geq 0$, $\int_0^\infty f(t; \alpha, \beta) dt = 1$, so a worker almost surely returns to work. But if $\alpha < 0$, the probability of eventually returning to work is $e^{2\alpha\beta} < 1$, so there is a probability

the worker never finds a job. Thus a non-employed worker with $\alpha < 0$ is characterized by a severe form of long term non-employment, since with probability $1 - e^{2\alpha\beta}$ it stays forever non-employed.

The inverse Gaussian is a flexible distribution but the model still imposes some restrictions on behavior. Assuming $\beta > 0$, the hazard rate of exiting non-employment always starts at 0 when $t = 0$, achieves a maximum value at some finite time t which depends on both α and β, and then declines to a long run limit of $\alpha^2/2$. If $\beta = 0$, the hazard rate is initially infinite and declines monotonically towards its long-run limit. At the start of a non-employment spell, the expected duration is β/α with variance β/α^3. Asymptotically the residual duration of an in-progress non-employment spell converges to $2/\alpha^2$, which may be bigger or smaller. The model is therefore consistent with both positive and negative duration dependence in the structural exit rate from non-employment.

In our model, this structural duration dependence may be exacerbated by dynamic selection. For example, take two types of workers characterized by reduced-form parameters (α_1, β_1) and (α_2, β_2). Suppose $\alpha_1 \leq \alpha_2$ and $\beta_1 \geq \beta_2$, with at least one inequality strict. Then type 2 workers have a higher hazard rate of finding a job at all durations t and so the population of long-term non-employed workers is increasingly populated by type 1 workers, those with a lower hazard of exiting non-employment. .

2.3   Magnitude of the Switching Costs

This section shows how we can use an individual's estimated type (α, β) to infer the magnitude of his switching costs. We focus on the special case highlighted in equation (2), where the utility from unemployment is constant at $b_0 > 0$, so that $b(t) = 0$ for all t. Suppose we observe a non-employed worker's type (α, β), as well as the parameters of the wage process when working (μ_e, σ_e), the drift of the wage when not working μ_n and the discount rate r. We find that

$$\frac{\psi_e + \psi_n}{b_0} \approx \frac{(\mu_e + \sqrt{\mu_e^2 + 2r\sigma_e^2})(-\alpha + \sqrt{\alpha^2 + 2r})\beta^3\mu_n^2}{12\, r\, \alpha^2\, \sigma_e^2} \sim \begin{cases} \frac{\mu_e\,\mu_n^2}{6\,\sigma_e^2}\frac{\beta^3}{|\alpha|^3} & \text{if } \alpha > 0 \\ \frac{\mu_e\,\mu_n^2}{6\,\sigma_e^2}\frac{\beta^3}{|\alpha|^3}\frac{\alpha^2}{r} & \text{if } \alpha < 0 \end{cases} \tag{4}$$

Equation (4) expresses the fixed costs as a function of four parameters, $\mu_e, \sigma_e, \mu_n, r$ and α, β.[1] Since the interest rate r is typically small, in (4) we derive two expressions for the limit as $r \to 0$, one for positive and one for negative α.[2]

---

[1]The sense in which we use the approximation $\approx$ in expression (4), as well as its derivation for the general model, is in Proposition 4 in Appendix A.

[2]In 4 we use $\sim$ to mean that as $r \downarrow 0$ the ratio of the two functions converge to one.

As shown above, the magnitude of fixed costs depends also on parameters which we do not estimate, $\mu_n, \mu_e, \sigma_e, r$. To infer the magnitude of fixed costs, we will have to choose some value for them. Since we expect the estimated fixed costs to be small, our strategy is going to be to choose their values to make the fixed costs as large as possible, while staying within reasonable range for the values of $\mu_n, \mu_e, \sigma_e, r$. In Section 5.5, we then use estimated distribution of α and β to calculate distribution of the fixed costs at positive values of r.

Equation (4) implies that for given value of α, β, higher $\mu_e, |\mu_n|$ and lower σ_e, r increase the fixed costs. We choose $\mu_e = 0.01, \sigma_e = 0.05$ at the annual frequency. Wages of employed workers grow at 1% per year. Estimates of the average wage growth of employed workers are often higher than 1%, but these are for workers who choose to stay employed, thus it applies to a selected sample. In our case we think of the parameter μ_e as governing the wage growth for all workers without selection, and thus we view the choice of $\mu_e = 0.01$ as a large number. The standard deviation of log wages of 5% is rather low, typical estimates in the literature are around 10%. The drift of latent wages when non-employed μ_n is not observable, but we can infer its value relative to μ_e from observed completed employment and unemployment spells. The model implies that the expected duration of completed employment and non-employment spells are given by $(\bar{\omega} - \underline{\omega})/\mu_e$ and $(\bar{\omega} - \underline{\omega})/|\mu_n|$, respectively, and thus $|\mu_n|/\mu_e$ determines the relative expected duration. In our sample, the average duration of non-employment spells is 29.6 weeks, while the average duration between two non-employment spells is 96.4 weeks, implying that $|\mu_n|/\mu_e = 3.25$. Finally, we choose a low value for r. Since agents in the model are infinitely lived, we can think of r as being a sum of two values, $r = \hat{r} + \delta$ where $\hat{r}$ is worker's discount rate and δ is the rate at which workers drop out of the labor force. We choose $\hat{r} = 0$ and $\delta = 0.02$, so workers do not discount and have an expected working lifetime of 50 years.

To illustrate that the fixed costs are small, take $\alpha = 0.96, \beta = 0.54$ at the annual frequency. This choice of parameters implies that the mean and standard deviation of completed non-employment spells is 29.6 and 40.7 weeks, respectively, as measured in the data. The implied values of fixed cost for $\alpha < 0$ is 3.4% of the annual flow value of non-employment. For $\alpha > 0$, the costs are two orders of magnitude smaller.

We can also use a simple calculation to deduce whether switching costs are necessarily positive. If switching costs were zero, the distance between the barriers would be zero as well, i.e. $\beta = 0$. In that limit, the duration density (17) is ill-behaved. Nevertheless, we can compute the density condition on durations lying in some interval $T = [\underline{t}, \bar{t}]$:

$$ f(t; \alpha, \beta | t \in [\underline{t}, \bar{t}]) = \frac{t^{-3/2} e^{-\frac{\alpha^2 t}{2}}}{\int_{\underline{t}}^{\bar{t}} \tau^{-3/2} e^{-\frac{\alpha^2 \tau}{2}} d\tau}. $$

8

The expected value of a random draw from this distribution is

$$\frac{\big(\Phi(\alpha\bar{t}^{\frac{1}{2}}) - \Phi(\alpha\underline{t}^{\frac{1}{2}})\big)/\alpha}{\Phi'(\alpha\underline{t}^{\frac{1}{2}})/\underline{t}^{\frac{1}{2}} - \Phi'(\alpha\bar{t}^{\frac{1}{2}})/\bar{t}^{\frac{1}{2}} - \alpha\big(\Phi(\alpha\bar{t}^{\frac{1}{2}}) - \Phi(\alpha\underline{t}^{\frac{1}{2}})\big)} \leq \big(\underline{t}\,\bar{t}\big)^{\frac{1}{2}},$$

with the inequality binding when $\alpha = 0$. Thus if we observe that the expected duration conditional on duration lying in some interval $[\underline{t}, \bar{t}]$ exceeds the geometric mean of $\underline{t}$ and $\bar{t}$, we can conclude that switching costs must be positive for at least some individuals in the data set.

3 Duration Analysis

Suppose we have a lot of duration data, possibly generated from this model. This section examines how we can use that data to say something about the model. We have three goals. First, assume the data is generated by the model. We examine whether the joint distribution of α and β in the population, $G(\alpha, \beta)$, is nonparametrically identified using non-employment duration data. The distribution of these reduced-form parameters reflects the underlying joint distribution of the structural parameters, but only these two reduced form parameters affect duration and so only the distribution of these two parameters can possibly be identified using duration data alone. The second is understand whether our model is testable. If we allow for an arbitrary distribution of the structural parameters in the model, are there non-employment duration data that are inconsistent with our theory? The third is to examine how the joint distribution of α and β can be used to decompose the overall evolution of the hazard of exiting non-employment into two components: the portion attributable to changes in the hazard for each individual worker as non-employment duration changes, and the portion attributable to changes in the population of non-employed workers at different durations.

In performing this analysis, we assume that the reduced-form parameters α and β are fixed over time for each worker, consistent with our model. In principle, variation in these parameters across workers may reflect some time-invariant observable characteristics of the workers or it may reflect time-invariant unobserved heterogeneity. We do not attempt to distinguish between these two possibilities. Our analysis precludes the possibility of time-varying heterogeneity. For example, a worker's experience cannot affect the stochastic process for the net benefit from employment, (μ_s, σ_s), nor can it affect the switching costs ψ_s, $s \in \{e, n\}$. Note, however, that our model does allow for learning-by-doing, since a worker's wage may increase faster on average when employed than when non-employed, $\mu_e > \mu_n$.

3.1 Nonparametric Identification

We start by examining whether our model is nonparametrically identified. With a single non-employment spell, our model is in general not identified.[3] To see this, suppose that the true model is one in which there is a single type of worker (α, β), which gives rise to non-employment duration density $f(t; \alpha, \beta)$, as in equation (3). This could alternatively have been generated by an economy with many types of workers. A worker who takes d periods to find a job has $\sigma_n = 0$ and $\mu_n = (\bar{\omega} - \underline{\omega})/d$, which implies that both α and β converge to infinity with $\beta/\alpha = d$. Moreover, the distribution of this ratio differs across workers so as to recover the empirical non-employment duration density $f(t; \alpha, \beta)$. More generally, this and many other type distributions can fit any non-employment duration distribution, so long as the density is strictly positive at all durations $t > 0$.[4]

Our approach to identification uses the duration of two non-employment spells. To see how the duration of two spells can help with identification, go back to the example in the previous paragraph. If there is a single type of worker, the correlation of the duration of two spells is zero, while the alternative model would imply that the duration of the two spells are equal. Thus repeated spells opens up the possibility of identification.

Now consider a population of individuals, each of whom has completed two spells. Let $G(\alpha, \beta)$ denote the distribution of (α, β) in the population. For some of those individuals, both spells have duration $(t_1, t_2) \in T^2$, where $T \subseteq \mathbb{R}_+$ is a set with non-empty interior. Let $\phi : T^2 \to \mathbb{R}_+$ denote the joint distribution of the durations for this population:

$$\phi(t_1, t_2) = \frac{\iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\iint_{T^2} \iint f(t'_1; \alpha, \beta) f(t'_2; \alpha, \beta) dG(\alpha, \beta) \, dt'_1 dt'_2}. \tag{5}$$

We allow for the possibility that T is a subset of the positive reals to prove that our model is identified even if we do not observe spells of certain durations.

Our main identification result, Theorem 1 below, is that the joint density of spell lengths ϕ identifies the joint distribution of characteristics (α, β) if we know the sign of α. We prove this result through a series of Propositions. The first shows that the partial derivatives of ϕ exist at all points where $t_1 \neq t_2$:

---

[3]Restricted versions of our model are identified with one spell. We describe two cases in Appendix C: in one example, every individual has the same expected duration β/α but there is a non-degenerate distribution of β. In the second example, there are no switching costs, $\psi_e = \psi_n = 0$, so $\beta = 0$. Both of these examples reduce the unknown joint distribution of (α, β) to one dimension.

[4]If the density of the non-employment duration distribution is ever 0, it must be the case that α and β are infinite for all workers. Of course, in any empirical application with a finite sample of data, the realized density may be zero at some durations even if α and β are finite, and this approach to identification does not seem robust.

Proposition 1 *Take any $(t_1, t_2) \in T^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$. The density ϕ is infinitely many times differentiable at (t_1, t_2).*

We prove this proposition in Appendix B. The proof verifies the conditions under which the Leibniz formula for differentiation under the integral is valid. This requires us to bound the derivatives in appropriate ways, which we accomplish by characterizing the structure of the partial derivatives of the product of two inverse Gaussian densities. Our bound uses that $t_1 \neq t_2$. Indeed an example shows that this condition is indispensable:

Example 1 *Consider the case where G is such that among all individuals with a particular mean duration, $\bar{\mu} = \beta/\alpha$, the parameter β has density $g(\beta|\bar{\mu}) = \theta\beta^{-1-\theta}$ on $[1, \infty)$, a Pareto distribution. In this case the joint density of two spells for these individuals is:*

$$\phi(t_1, t_2|\bar{\mu}) = \frac{\theta\Delta^{\frac{\theta}{2}-1}}{4\pi t_1^{3/2} t_2^{3/2}} \Gamma\left(1 - \tfrac{\theta}{2}, \Delta\right) \tag{6}$$

where $\Delta \equiv \frac{1}{2}\left(\frac{(t_1/\bar{\mu}-1)^2}{t_1} + \frac{(t_2/\bar{\mu}-1)^2}{t_2}\right)$ and $\Gamma(s, x) = \int_x^\infty z^{s-1} e^{-z} dz$ is the incomplete Gamma function.

When either $t_1 \neq \bar{\mu}$ or $t_2 \neq \bar{\mu}$ or both, Δ is strictly positive and hence $\phi(t_1, t_2|\bar{\mu})$ is infinitely differentiable. But when $t_1 = t_2 = \bar{\mu}$, $\Delta = 0$ and so both the Gamma function and $\Delta^{\frac{\theta}{2}-1}$ can either diverge or be non-differentiable. In particular, for $\theta \in (0, 2)$, the right tail of g is so thick that $\lim_{t \to \bar{\mu}} \phi(t, t|\bar{\mu}) = \infty$. For $\theta \in [2, 4)$ the right tail of β is thick enough so that the density is non-differentiable at $t_1 = t_2 = \bar{\mu}$. For higher values of θ, the density can only be differentiated a finite number of times at this critical point.

The source of the non-differentiability is that for very large β, the volatility of the Brownian motion vanishes, and thus the spells end with certainty at duration $\bar{\mu}$. Equivalently the corresponding distribution tends to a Dirac measure concentrated at $t_1 = t_2 = \bar{\mu}$. For a distribution with a sufficiently thick right tail of β, the same phenomenon happens, but only at points with $t_1 = t_2$, since individuals with vanishingly small volatility in their Brownian motion almost never have durations $t_1 \neq t_2$. Instead, for values of $t_1 \neq t_2$, the density ϕ is well-behaved because randomness from the Brownian motion smooths out the duration distribution, regardless of the underlying type distribution.

Finally, if for different values of $\bar{\mu} = \beta/\alpha$, we have a Pareto distribution of β, ϕ may be non-differentiable at multiple points, each of which has $t_1 = t_2$.

For the next step, we look at the conditional distribution of (α, β) among individuals

whose two spells last exactly (t_1, t_2) periods:

$$\tilde{G}(\alpha, \beta | t_1, t_2) = \frac{f(t_1, \alpha, \beta) \, f(t_2, \alpha, \beta) \, dG(\alpha, \beta)}{\iint f(t_1, \alpha', \beta') \, f(t_2, \alpha', \beta') \, dG(\alpha', \beta')}, \tag{7}$$

We prove that the partial derivatives of ϕ uniquely identify all the *even* moments of $\tilde{G}$ for any $t_1 \neq t_2$:

Proposition 2 *Take any $(t_1, t_2) \in T^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$, and any strictly positive integer m. The set of partial derivatives $\partial^{i+j}\phi(t_1, t_2)/\partial t_1^i \partial t_2^j$ for all $i \in \{0, 1, \ldots, m\}$ and $j \in \{0, 1, \ldots, m-i\}$ uniquely identifies the set of moments*

$$\mathbb{E}(\alpha^{2i}\beta^{2j}|t_1, t_2) \equiv \iint \alpha^{2i}\beta^{2j} d\tilde{G}(\alpha, \beta | t_1, t_2) \tag{8}$$

for all $i \in \{0, 1, \ldots, m\}$ and $j \in \{0, 1, \ldots, m-i\}$.

Note that the statement of the proposition suggests a recursive structure, which we follow in our proof in appendix B. In the first step, set $m = 1$. The two first partial derivatives $\partial\phi(t_1, t_2)/\partial t_1$ and $\partial\phi(t_1, t_2)/\partial t_2$ determine the two first even moments, $\mathbb{E}(\alpha^2|t_1, t_2)$ and $\mathbb{E}(\beta^2|t_1, t_2)$. In the second step, set $m = 2$. The three second partial derivatives and the results from first step then determine the three second even moments, $\mathbb{E}(\alpha^4|t_1, t_2)$, $\mathbb{E}(\alpha^2\beta^2|t_1, t_2)$, and $\mathbb{E}(\beta^4|t_1, t_2)$. In the mth step, the $m+1$ m^{th} partial derivatives and the results from the previous steps determine the $m+1$ m^{th} even moments of $\tilde{G}$. The proof, which is primarily algebraic, shows how this works.

In the third step of the proof, we recover the joint distribution $\tilde{G}(\alpha, \beta | t_1, t_2)$ from the moments of (α^2, β^2) among individuals who find jobs at durations (t_1, t_2). There are two pieces to this. First, we need to know the sign of α; we assume this is either always positive or always negative since, as we previously noted, this is not identified. Second, we need to ensure that the moments uniquely determine the distribution function. A sufficient condition is that the moments not grow too fast; our proof verifies that this is the case.

Proposition 3 *Assume that $\alpha \geq 0$ with G-probability 1 or that $\alpha \leq 0$ with G-probability 1. Take any $(t_1, t_2) \in T^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$. The set of conditional moments $\mathbb{E}(\alpha^{2i}\beta^{2j}|t_1, t_2)$ for $i = 0, 1, \ldots$ and $j = 0, 1, \ldots$, defined in equation (8), uniquely identifies the conditional distribution $\tilde{G}(\alpha, \beta | t_1, t_2)$.*

The proof of this proposition in Appendix B.

Our main identification result follows immediately from these three propositions:

Theorem 1 *Assume that $\alpha \geq 0$ with G-probability 1 or that $\alpha \leq 0$ with G-probability 1. Take any function $\phi : T^2 \to \mathbb{R}_+$. There is at most one distribution function G such that equation (5) holds.*

Proof. Proposition 1 shows that ϕ is infinitely many times differentiable. Proposition 2 shows that for any $(t_1, t_2) \in T^2$, $t_1 \neq t_2$, $t_1 > 0$, and $t_2 > 0$, there is one solution for the moments of (α^2, β^2) conditional on durations (t_1, t_2), given all the partial derivatives of ϕ at (t_1, t_2). Proposition 3 shows that these moments uniquely determine the distribution function $\tilde{G}(\alpha, \beta | t_1, t_2)$ with the additional assumption that $\alpha \geq 0$ with G-probability 1 or $\alpha \leq 0$ with G-probability 1. Finally, given the conditional distribution $\tilde{G}(\cdot, \cdot | t_1, t_2)$, we can recover $G(\cdot, \cdot)$ using equation (7) and the known functional form of the inverse Gaussian density f:

$$\frac{dG(\alpha, \beta)}{dG(\alpha', \beta')} = \frac{d\tilde{G}(\alpha, \beta | t_1, t_2)}{d\tilde{G}(\alpha', \beta' | t_1, t_2)} \frac{f(t_1; \alpha', \beta') f(t_2; \alpha', \beta')}{f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)} \tag{9}$$

■

Our theorem states that the density ϕ is sufficient to recover the joint distribution G if we know the sign of α. Our proof uses all the derivatives of ϕ evaluated at a point (t_1, t_2) to recover all the moments of the conditional distribution $\tilde{G}(\cdot, \cdot | t_1, t_2)$ to recover the necessary moments. Intuitively, if one thinks of a Taylor expansion around (t_1, t_2), we are indeed using the entire empirical density ϕ for $(t_1, t_2) \in T^2$ to recover the distribution function G.

We stress that the sign of α is not identified. The basic problem is that $f(t; \alpha, \beta) = e^{2\alpha\beta} f(t; -\alpha, \beta)$ for all (α, β). Take any two distributions G_1 and G_2 and for simplicity assume they have density functions, g_1 and g_2. Then if

$$g_1(\alpha, \beta) + e^{-4\alpha\beta} g_1(-\alpha, \beta) = g_2(\alpha, \beta) + e^{-4\alpha\beta} g_2(-\alpha, \beta) \tag{10}$$

for all (α, β), G_1 and G_2 generate the same distribution of completed spells $\phi(t_1, t_2)$ for all $(t_1, t_2) \in T^2$. It is possible to use incomplete spells to derive bounds on the fraction of the population with the positive or negative α, as we discuss in more detail in Section 3.3.

We comment briefly on an alternative but ultimately unsuccessful proof strategy. Proposition 2 establishes that we can measure $\mathbb{E}(\beta^m | t_1, t_2)$ at almost all (t_1, t_2). It might seem we could therefore integrate the conditional moments using the density $\phi(t_1, t_2)$ to compute the unconditional m^{th} moment of β. This strategy might fail, however, because the integral need not converge. Indeed, this is the case whenever the appropriate moment of G does not exist. We continue Example 1 to illustrate this possibility:

Example 2 *Assume that β is distributed Pareto with parameter θ while $\mu_n = \alpha/\beta > 0$ is degenerate. The distribution G thus does not have all its moments. We turn to the*

conditional distribution $\tilde{G}(\beta|t_1, t_2)$ for some $t_1 \neq t_2$. The mth moment of the conditional distribution is

$$\mathbb{E}(\beta^m|t_1, t_2) = \Delta^{-\frac{m}{2}} \frac{\Gamma(1 + \frac{m-\theta}{2}, \Delta)}{\Gamma(1 - \frac{\theta}{2}, \Delta)},$$

where again $\Delta = \frac{1}{2}\left(\frac{(\mu_n t_1 - 1)^2}{t_1} + \frac{(\mu_n t_2 - 1)^2}{t_2}\right)$ and $\Gamma(s, x) = \int_x^\infty z^{s-1}e^{-z}dz$ is the incomplete Gamma function; this follows from equation (7).

If $\Delta > 0$, all moments $M_m \equiv \mathbb{E}(\beta^m|t_1, t_2)$ exist and are finite. To prove that the moments uniquely describe a distribution, we use the D'Alembert criterium (see for example Theorem A.5 in Coelho, Alberto, and Grilo (2005)). It suffices to show that $\lim_{m\to\infty} \frac{1}{m-1}\frac{M_m}{M_{m-1}}$ is finite:

$$\frac{1}{m}\frac{\mathbb{E}(\beta^{m+1}|t_1, t_2)}{\mathbb{E}(\beta^m|t_1, t_2)} = \frac{1}{m}\frac{\Delta^{-1/2}\Gamma(1 + \frac{m+1-\theta}{2}, \Delta)}{\Gamma(1 + \frac{m-\theta}{2}, \Delta)},$$

which converges to 0 as $m \to \infty$. Therefore, conditional moments uniquely determine the conditional distribution $\tilde{G}(\beta|t_1, t_2)$, even though some moments of G do not exist.

The proof of our identification theorem instead establishes only that the conditional moments exist and are well behaved. We then use the conditional distribution to infer the unconditional distribution without reference to its moments.

3.2 Comments on Identification with Two Spells

Identification with a finite number of types. $2 - 4$ types are enough to fit the distribution of single spell data but here the identification can be coming from the functional form. The advantage of the multi-spell data is that its correlation structure is informative about heterogeneity.

3.3 Long-Term Non-employment: Identifying the Sign of α

We argued that data on two completed spells cannot identify the sign of α, which in turn determines the chances that a non-employed worker never finds a job. In this section we argue that the fraction of incomplete spells provides some information on the sign of α.

First we note that for any distribution $G^+(\alpha, \beta)$ with $\alpha \geq 0$ with G-probability 1, there is another distribution $G^-(\alpha, \beta)$ with $\alpha \leq 0$ with G-probability 1 such that both G^+ and G^- imply the same joint distribution of durations $\phi : T^2 \to \mathbb{R}_+$. In particular, since $f(t; -\alpha, \beta) = e^{-2\alpha\beta}f(t; \alpha, \beta)$, equation (5) implies these distributions are related via

$$dG^-(-\alpha, \beta) = \frac{e^{4\alpha\beta}dG^+(\alpha, \beta)}{\int\int e^{4\alpha'\beta'}dG^+(\alpha', \beta')}. \tag{11}$$

The expression in the denominator of equation (11) ensures that the dG^- integrates to 1. More generally, any convex combination of G^+ and G^- yields the same joint distribution of durations ϕ.

While the sign of α does not affect the distribution of durations of completed spells, it does affect the fraction of the population with two completed spells with durations $(t_1, t_2) \in T^2$, a statistic that we refer to as c. For any distribution ϕ, G^+ provides an upper bound on the fraction of the population whose first two completed spells each have duration in any set $T \subset \mathbb{R}$. We denote the upper bound for the fraction c by $\bar{c}$ which is given by:

$$c \leq \bar{c} = \iint_{T^2} \iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG^+(\alpha, \beta) \, dt_1 dt_2. \tag{12}$$

The lower bound $\underline{c}$ for c is obtained using the distribution G^-,

$$\underline{c} = \iint_{T^2} \iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG^-(\alpha, \beta) \, dt_1 dt_2 = \frac{\bar{c}}{\int \int e^{4\alpha\beta} dG^+(\alpha, \beta)}. \tag{13}$$

Mixtures between these two distributions yield intermediate values of the fraction of the population with two completed spells.

We address this lack of identification as follows. First, we use our model to estimate the distribution function G^+. We then compute $\bar{c}$ and $\underline{c}$ using equations (12) and (13). We compare these values with the empirical fraction of individuals who have at least two completed spells, the first two with durations $(t_1, t_2) \in T^2$, say $\hat{c}$. If it were the case that $\bar{c} = \hat{c}$, then we would conclude that the whole population has positive α. Similarly, if we find that $\underline{c} = \hat{c}$, then we conclude that the whole population has a negative α. For any intermediate case, $\underline{c} < \hat{c} < \bar{c}$, we conclude that there are some fraction of people, call it μ^- with negative α, and it holds that $0 < \mu^- < 1$ and moreover, the distribution with μ^- has to be such that the share workers with first two spells each in T has to be $\hat{c}$.

There are many distributions of (α, β) which are consistent with ϕ and imply that $c = \hat{c}$, and each of them will imply a different value of μ^-. Since the data does not contain enough information to identify one of these distributions exactly, we find a distribution which maximizes μ^- and one which minimizes μ^-. This would give us an upper and a lower bound on the share of workers with negative α in the population. We will then use these distributions as well as G^+ to conduct the decomposition.

Let $\underline{G}$ be the distribution of (α, β) which minimizes μ^-. How can this distribution be constructed? Consider G^+. This distribution has no types with $\alpha < 0$ and so its $\mu^- = 0$ but at the same time it generates c which is too high compared to $\hat{c}$. The idea is then to add a small number of types with $\alpha < 0$ which have a very low probability to experience

two spells shorter in T as this will decrease c. The lowest probability of experiencing two completed spells is zero, and in fact, any type with $\alpha < 0$ and $\beta \to \infty$ has a zero probability of completing any non-employment spell. To understand why this is the case, recall that these restrictions on α, β imply that $\mu_n < 0$ and $\sigma_n \to 0$. The process for the net benefit from working is deterministic with a negative drift. Starting at the lower threshold $\underline{\omega}$, the probability of reaching the upper threshold $\bar{\omega}$ (and thus completing the spell) is zero. We need to add a fraction x of these workers where x solves that $\hat{c} = (1-x)\bar{c} + x \cdot 0$, or $x = 1 - \hat{c}/\bar{c}$.

Finally, let $\bar{G}$ be the distribution for which μ^- is maximized. We know that the upper bound μ^- is 1 and that this upper bound is achieved with G^-, but G^- implies c which is too low compared to $\hat{c}$ measured in the data. To increase c and keep μ^- high, we thus need to revert some of the types back to α positive. The most effective way of doing this is to keep types with low $|\alpha\beta|$ as negative types, as we show formally in Appendix D, have types with high $\alpha\beta$ as positive. To understand this result, let's first think about the trade-off between keeping a type α, β positive or turning it negative. Our constraint is that the same number of workers have to be observed as having two completed spells (remember the constraint on the distribution of completed spells). If the type with negative α has a low probability of experiencing two completed spells, it has to be compensated by a large share of these types in the population so that a given number of workers is observed with two completed spells. But this is going to be difficult to reconcile with the constraint that c cannot be lower than $\hat{c}$. Thus, we want to keep this types as positive. Which types tend to have a low probability of experiencing two completed spells? As discussed in the previous paragraph, if σ_n is close to zero and μ_n is negative, then the probability of having two completed spells are very low. For other values of σ_n and $\mu_n < 0$, the probability of observing two completed spells is lower for higher $|\mu_n|/\sigma_n$. This two examples justify why we want to keep types with high $\alpha\beta$ as positive and turn those with low $\alpha\beta$ into negative.

3.4 Testable Implications

The stopping time model and inverse Gaussian densities are flexible, and so a natural question is whether they can explain any data. If we observe only a single non-employment spell for each individual, the model indeed has no testable implications. Any single-spell duration data can be explained perfectly though an assumption that an individual who takes d periods to find a job has $\sigma_n = 0$ and $\mu_n = (\bar{\omega} - \underline{\omega})/d$. We focus instead on a data set that includes two completed non-employment spells for each individual.

Our approach to identification yields the model's overidentifying restrictions. First, Proposition 1 tells us that the joint density ϕ is infinitely differentiable at any $(t_1, t_2) \in T^2$

with $t_1 > 0$, $t_2 > 0$, and $t_1 \neq t_2$. We can reject the model if this is not the case. This test is not useful in practice, however, since ϕ is never differentiable in any finite data set.

Second, Proposition 2 tells us how to construct the even-powered moments of the joint distribution function $\tilde{G}(\alpha, \beta | t_1, t_2)$. Even-powered moments must all be positive, and so this prediction yields additional tests of the model.

Third, Proposition 3 tells us that we can use the moments to reconstruct the distribution function $\tilde{G}$. These moments must satisfy certain restrictions in order for them to be generated from a valid CDF. For example, Jensen's inequality implies that

$$\mathbb{E}(\alpha^{2i} | t_1, t_2)^{1/i} \leq \mathbb{E}(\alpha^{2j} | t_1, t_2)^{1/j}$$

for all integers $0 < i < j$.

In practice, measuring higher moments can be difficult and so we focus on the simplest restriction that comes from the model, $\mathbb{E}(\alpha^2 | t_1, t_2) \geq 0$ and $\mathbb{E}(\beta^2 | t_1, t_2) \geq 0$ for all $t_1 \neq t_2$. Following the proof of Proposition 2, our model implies that these moments satisfy

$$\mathbb{E}(\alpha^2 | t_1, t_2) = \frac{2\left(t_2^2 \frac{\partial \phi(t_1, t_2)}{\partial t_2} - t_1^2 \frac{\partial \phi(t_1, t_2)}{\partial t_1}\right)}{\phi(t_1, t_2)(t_1^2 - t_2^2)} - \frac{3}{t_1 + t_2} \geq 0 \tag{14}$$

$$\text{and } \mathbb{E}(\beta^2 | t_1, t_2) = t_1 t_2 \left(\frac{2 t_1 t_2 \left(\frac{\partial \phi(t_1, t_2)}{\partial t_2} - \frac{\partial \phi(t_1, t_2)}{\partial t_1} \right)}{\phi(t_1, t_2)(t_1^2 - t_2^2)} + \frac{3}{t_1 + t_2} \right) \geq 0. \tag{15}$$

These inequality tests have considerable power against alternative theories, as some simple examples illustrate.

Example 3 *Consider the canonical search model where the hazard of finding a job is a constant θ and so the density of completed spells is $\phi(t_1, t_2) = \theta^2 e^{-h(t_1 + t_2)}$. Then applying conditions (14) and (15) gives*

$$\mathbb{E}(\alpha^2 | t_1, t_2) = 2\theta - \frac{3}{t_1 + t_2} \geq 0 \text{ and } \mathbb{E}(\beta^2 | t_1, t_2) = \frac{3 t_1 t_2}{t_1 + t_2} \geq 0.$$

In particular, $\mathbb{E}(\alpha^2 | t_1, t_2) < 0$ whenever $t_1 + t_2 < 3/2\theta$, where $1/\theta$ represents the mean duration of a non-employment spell. We conclude that our model cannot generate this density of completed spells for any joint distribution of parameters.

More generally, suppose the constant hazard θ has a population distribution $\tilde{G}$, with some abuse of notation. The density of completed spells is $\phi(t_1, t_2) = \int \theta^2 e^{-\theta(t_1 + t_2)} d\tilde{G}(\theta)$. Then

$$\mathbb{E}(\alpha^2 | t_1, t_2) = 2 \frac{\int \theta^3 e^{-\theta(t_1 + t_2)} d\tilde{G}(\theta)}{\int \theta^2 e^{-\theta(t_1 + t_2)} d\tilde{G}(\theta)} - \frac{3}{t_1 + t_2} \geq 0,$$

17

while $\mathbb{E}(\beta^2|t_1, t_2)$ *is unchanged. If the ratio of the third moment of* θ *to the second moment is finite—for example, if the support of the distribution* $\tilde{G}$ *is bounded—this is always negative for sufficiently small* $t_1 + t_2$ *and hence the more general model is rejected.*

One might think that the constant hazard model is rejected because the implied density ϕ is decreasing, while the density of a random variable with an inverse Gaussian distribution is hump-shaped. This is not the case. The next two examples illustrate this. The first looks at a log-normal distribution

Example 4 *Suppose that the density of durations is log-normally distributed with mean* μ *and standard deviation* σ. *For each individual, we observe two draws from this distribution and test the model using conditions (14) and (15). Then our approach implies*

$$\mathbb{E}(\alpha^2|t_1, t_2) = \frac{2}{\sigma^2(t_1 + t_2)} \left(\frac{t_1 \log t_1 - t_2 \log t_2}{t_1 - t_2} - \left(\mu + \tfrac{1}{2}\sigma^2 \right) \right) \geq 0$$

$$\text{and } \mathbb{E}(\beta^2|t_1, t_2) = \frac{2t_1 t_2}{\sigma^2(t_1 + t_2)} \left(\frac{t_2 \log t_1 - t_1 \log t_2}{t_1 - t_2} + \left(\mu + \tfrac{1}{2}\sigma^2 \right) \right) \geq 0.$$

One can prove that $\frac{t_1 \log t_1 - t_2 \log t_2}{t_1 - t_2}$ *is increasing in* (t_1, t_2), *converging to minus infinity when* t_1 *and* t_2 *are sufficiently close to zero. Therefore, for any* μ *and* $\sigma > 0$, *the first condition is violated at small values of* (t_1, t_2). *Similarly,* $\frac{t_2 \log t_1 - t_1 \log t_2}{t_1 - t_2}$ *is decreasing in* t_1 *and* t_2, *converging to minus infinity when* t_1 *and* t_2 *are sufficiently large. Therefore, for any* μ *and* $\sigma > 0$, *the second condition is violated at large values of* (t_1, t_2).

The same logic implies that any mixture of log-normally distributed random variables generates a joint density ϕ *that is inconsistent with our model, as long as the support of the mixing distribution is compact. Thus even though the log normal distribution generates hump-shaped densities, the test implied by conditions (14) and (15) would never confuse a mixture of log normal distributions with a mixture of inverse Gaussian distributions.*

The final example relates our results to data generated from the proportional hazard model, a common statistical model in duration analysis

Example 5 *Each individual has a hazard rate equal to* $\theta\, h(t)$ *at times* $t \geq 0$, *and where* $h(\cdot)$ *is common function with unrestricted shape and* θ *is an individual characteristic with distribution function again denoted by* $\tilde{G}$. *If* $h(t)$ *and* $|h'(t)/h(t)|$ *are both bounded as* t *conveges to 0, the test will imply* $\mathbb{E}(\alpha^2|t_1, 0) < 0$ *for* t_1 *small enough. Appendix E gives a more detailed description and proves this result.*

3.5 Decomposition: Structure versus Heterogeneity

We turn now to the relative importance of structural duration dependence and dynamic selection of heterogeneous individuals for the evolution of the hazard rate of exiting non-employment. For notational convenience alone, assume that the type distribution G has a density g.

3.5.1 Decomposition of the Hazard Rate

We decompose the aggregate hazard rate conditional on two completed spells using a Divisia index. To start, assume $\alpha \geq 0$ with G-probability 1. Let $h(t; \alpha, \beta)$ denote the hazard rate for type (α, β) at duration t,

$$h(t; \alpha, \beta) = \frac{f(t; \alpha, \beta)}{1 - F(t; \alpha, \beta)}. \tag{16}$$

Also let $g(t; \alpha, \beta)$ denote the density of the type distribution among individuals who complete two spells and whose duration exceeds t periods:

$$g(t; \alpha, \beta) = \frac{(1 - F(t; \alpha, \beta))g(\alpha, \beta)}{\iint (1 - F(t; \alpha', \beta'))g(\alpha', \beta') \, d\alpha' \, d\beta'}. \tag{17}$$

The aggregate hazard rate $H(t)$ is an average of individual hazard rates weighted by their share among workers with duration t,

$$H(t) = \frac{\iint f(t; \alpha, \beta)g(\alpha, \beta) \, d\alpha \, d\beta}{\iint (1 - F(t; \alpha', \beta'))g(\alpha', \beta') \, d\alpha' \, d\beta'} = \iint h(t; \alpha, \beta)g(t; \alpha, \beta) \, d\alpha d\beta,$$

as can be confirmed directly from the definitions of $h(t; \alpha, \beta)$ and $g(t; \alpha, \beta)$. Taking a derivative with respect to t, $\dot{H}(t) = \dot{H}^s(t) + \dot{H}^h(t)$, where

$$\dot{H}^s(t) = \iint \dot{h}(t; \alpha, \beta)g(t; \alpha, \beta) \, d\alpha d\beta \tag{18}$$

$$\dot{H}^h(t) = \iint h(t; \alpha, \beta)\dot{g}(t; \alpha, \beta) \, d\alpha d\beta. \tag{19}$$

We interpret the term $\dot{H}^s(t)$ as the instantaneous contribution of structural duration dependence since it is based on the change in the hazard rates of individual worker types. Observe that if the hazard rate were constant (and thus there were no structural duration dependence), this term would be zero. The second term $\dot{H}^h(t)$ captures the instantaneous role of heterogeneity because it captures how the distribution of worker types changes with

unemployment duration.

The sign of $\dot{H}^s(t)$ can be either positive or negative, but the contribution of heterogeneity $\dot{H}^h(t)$ is equal to the minus the cross-sectional variance of the hazard rates:[5]

$$\dot{H}^h(t) = -\iint (h(t;\alpha,\beta) - H(t))^2 g(t;\alpha,\beta)\, d\alpha\, d\beta < 0. \tag{20}$$

This result is a version of the *fundamental theorem of natural selection* (Fisher, 1930), which states that "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time."[6] Intuitively, types with a higher than average hazard rate are always declining as a share of the population.

Now suppose $\alpha \leq 0$ with G-probability 1, but we measure and decompose hazard rates as described here. This has two effects. First, when we fail to account for negative values of α, we mismeasure the G distribution, effectively conditioning on a particular individual with type (α, β) having two completed spells, an event that in truth only has probability $e^{4\alpha\beta}$. That is, the measured distribution $G(\alpha, \beta)$ is the distribution of $(|\alpha|, \beta)$ conditional on two completed spells. Second, the hazard rate itself conditions on a completed spell. That is $h(t; \alpha, \beta)$ is the hazard at duration t conditional on the spell ending in finite time. This leads to general interpretation of our decomposition: it is the hazard rate of each spell conditional on an individual completing two spells. This interpretation remains valid if α is positive for some individuals and negative for others.

3.5.2 Hazard Rate Conditional on Realized Duration

An alternative way of showing the relative importance of heterogeneity versus structural duration dependence is to look at the difference in the hazard rates of workers who found jobs at different durations.

Each worker i in the dataset is characterized by the realized duration of his two nonemployment spells, (t_1^i, t_2^i). We use the estimated distribution $G(\alpha, \beta)$ to form a posterior belief

---

[5]To prove this, first take logs and differentiate $g(t;\alpha,\beta)$:

$$\frac{\dot{g}(t;\alpha,\beta)}{g(t;\alpha,\beta)} = -\frac{f(t;\alpha,\beta)}{1 - F(t;\alpha,\beta)} + \frac{\iint f(t;\alpha',\beta')dG(\alpha',\beta')}{\iint (1 - F(t;\alpha',\beta'))dG(\alpha',\beta')} = -h(t;\alpha,\beta) + H(t).$$

Substituting this result into the expression for $\dot{H}^h(t)$ gives

$$\dot{H}^h(t) = -\iint h(t;\alpha,\beta)(h(t;\alpha,\beta) - H(t))g(\alpha,\beta)\, d\alpha\, d\beta.$$

Since $\iint (h(t;\alpha,\beta) - H(t))g(\alpha,\beta)\, d\alpha\, d\beta = 0$, we can add $H(t)$ times this to the previous expression to get the formula in equation (20).

[6]We are grateful to Jörgen Weibull for pointing out this connection to us.

about his type, $dG^i(\alpha, \beta) \equiv dG^{post}(\alpha, \beta | t_1^i, t_2^i)$, where we the Bayes formula

$$dG^{post}(\alpha, \beta | t_1, t_2) = \frac{f(t_1, \alpha, \beta) f(t_1, \alpha, \beta) dG(\alpha, \beta)}{\int f(t_2, \alpha', \beta') f(t_2, \alpha', \beta') dG(\alpha', \beta')}. \tag{21}$$

We then split workers into groups conditional on the realized duration of their first spell. For example, we consider workers whose first spell lasted between 0 and 13 weeks in one group, 14 to 26 weeks in another group, and so on. Defining $I(\underline{t}, \bar{t})$ as the set of workers whose first spell lasted between $\underline{t}$ and $\bar{t}$, $I(\underline{t}, \bar{t}) \equiv i : t_1^i \in [T_1, T_2]$, the distribution of types in $I(\underline{t}, \bar{t})$ is given by

$$dG^{I(\underline{t}, \bar{t})}(\alpha, \beta) = \frac{1}{N_{I(\underline{t}, \bar{t})}} \sum_{i \in I(\underline{t}, \bar{t})} g^i(\alpha, \beta), \tag{22}$$

where $N_{I(\underline{t}, \bar{t})}$ is the number of workers in this group.

Finally, we compute the hazard rate of each group as $h^{[T_1, T_2](t)}$ as

$$h^{I(\underline{t}, \bar{t})}(t) = h(t, \alpha, \beta) dG^{I(\underline{t}, \bar{t})}(\alpha, \beta | t) \tag{23}$$

4   Austrian Data

We test our theory, estimate our model, and evaluate the role of structural duration dependence using data from the Austrian social security registry. The data set covers the universe of private sector workers over the years 1986–2007 (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf, and Buchi, 2009). It contains information on individual's employment, registered unemployment, maternity leave, and retirement, with the exact begin and end date of each spell.[7]

The use of the Austrian data is compelling for two reasons. First, the data set contains the complete labor market histories of the majority of workers over a 21 year period, which allows us to construct multiple non-employment spells per individual. Second, the labor market in Austria remains flexible despite institutional regulations, and responds only very mildly to the business cycle. Therefore, we can treat the Austrian labor market as a stationary environment and use the pooled data for our analysis. Figure 11 in the Appendix shows the mean duration of in progress unemployment spells shorter than 260 weeks. One can clearly see seasonal variation but movements driven by a business cycle is very mild.

We discuss the key regulations of the Austrian labor market. Almost all private sector jobs are covered by collective agreements between unions and employer associations at the region and industry level. The agreements typically determine the minimum wage and wage

---

[7]We have data available back to 1972, but can only measure registered unemployment after 1986.

increases on the job, and do not directly restrict the hiring or firing decisions of employers. The main firing restriction is the severance payment, with size and eligibility determined by law. A worker becomes eligible for the severance pay after three years of tenure if he does not quit voluntarily. The pay starts at two month salary and increases gradually with tenure.

The unemployment insurance system in Austria is similar to the one in the U.S. The duration of the unemployment benefits depends on the previous work history and age. If a worker has been employed for more than a year during two years before the layoff, she is eligible for 20 weeks of the unemployment benefits. The duration of benefits increases to 30 weeks and 39 weeks for workers with longer work history.

Temporary separations and recalls are prevalent in Austria. Around 40 percent of non-employment spells end with an individual returning to the previous employer. Our structural model naturally allows for this possibility.

We work with complete and incomplete non-employment spells. We define complete non-employment spells as the time from the end of one full-time job to the start of the following full-time job. We further impose that a worker has to be registered as unemployed for at least one day during the non-employment spell. We drop spells involving a maternity leave. Although in principle we could measure non-employment duration in days, disproportionately many jobs start on Mondays and end on Fridays, and so we focus on weekly data. We measure spells in calendar weeks. A calendar week starts on Monday and ends on Sunday. If a worker starts and ends a spell in the same calendar week, we code it as duration of 0 weeks. The duration of 1 week means that the spell ended in the calendar week following the calendar week it has started, and so on.

A non-employment spell is not complete if it does not end by a worker taking another job. Instead, one of the following can happen: 1) the non-employment spell is still in progress when the dataset ends, 2) a worker retires, 3) a worker goes on a maternity leave, 4) a worker disappears from the sample. We consider any of these as incomplete spells.

We consider only individuals who were no older than 45 in 1986 and no younger than 40 in 2007, and have at least one non-employment spell which started after the age of 25.[8] Imposing the age criteria guarantees that each individual has at least 15 years when he could potentially be at work. To estimate the model, we will use information on two complete spells shorter than 260 weeks, which means that we are choosing $T = [0, 260]$. We keep incomplete spells only if they are longer than 260 weeks.

---

[8] We do this because older individuals in 1986 or younger individuals in 2007 are less likely to experience two such spells in the data set we have available. Moreover, Theorem 1 tells us that we can identify the type distribution G using the duration density $\phi(t_1, t_2)$ on any subset of durations $(t_1, t_2) \in T^2$. Here we set $T = [0, 260]$.

starting sample	1,012,342
only one spell longer than 260	122,316
two spells, the second spell incomplete	37,456
two complete spells, but either first or second longer than 260	56,760
final sample	795,810

Table 1: Construction of the data

We further restrict our sample to workers who either have only one spell longer than 260 weeks (including incomplete), or have at least two spells (including incomplete). This is our starting sample. There are 1,012,342 such workers, with 122,316 workers having only one spell. There are 890,026 workers with at least two spells, but only 795,810 have first two spells shorter than 260 weeks. Table 1 shows further details of the data construction.

Our final sample which we use for estimation consists of 795,810 workers. These workers have 6.3 such non-employment spells on average, but we only use information on the first two spells. We use 78.6% of the starting sample, and thus our measured $\hat{c} = 0.786$.

In this sample, the average duration of a completed non-employment spell is 29.6 weeks, and the average employment duration between these two spells is 96.4 weeks. Figure 1 depicts the marginal distribution of non-employment spells during each of the first two non-employment spells for all workers who experience at least two spells. The two distributions are very similar. They rise sharply during the first five weeks, hover near four percent for the next ten weeks, and then gradually start to decline. The density during the first spell is slightly lower than the density during the second spell at short durations and slightly higher at long durations, a difference we suppress in our analysis.

Figure 2 depicts the joint density $\phi(t_1, t_2)$ for $(t_1, t_2) \in \{0, \ldots, 80\}^2$. Several features of the joint density are notable. First, it has a noticeable ridge at values of $t_1 \approx t_2$. Many workers experience two spells of similar durations. Second, the joint density is noisy, even with 800,000 observations. This does not appear to be primarily due to sampling variation, but rather reflects the fact that many jobs start during the first week of the month and end during the last one. There are notable spikes in the marginal distribution of nonemployment spells every fourth or fifth week and, as Figure 1 shows, these spikes persist even at long durations.
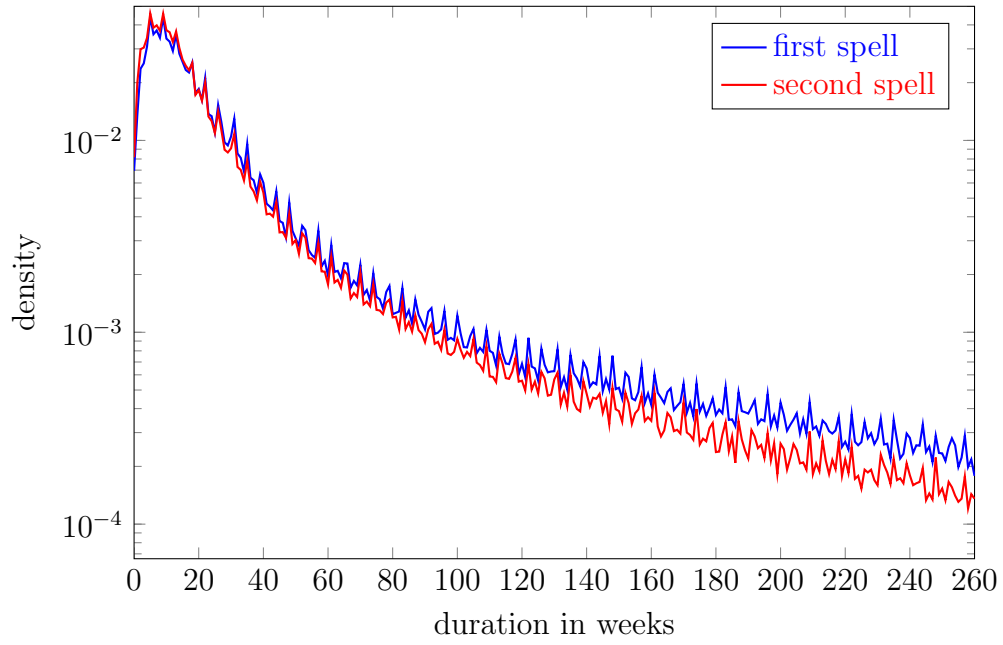
Figure 1: Marginal distribution of the first two non-employment spells, conditional on duration less than or equal to 260 weeks.
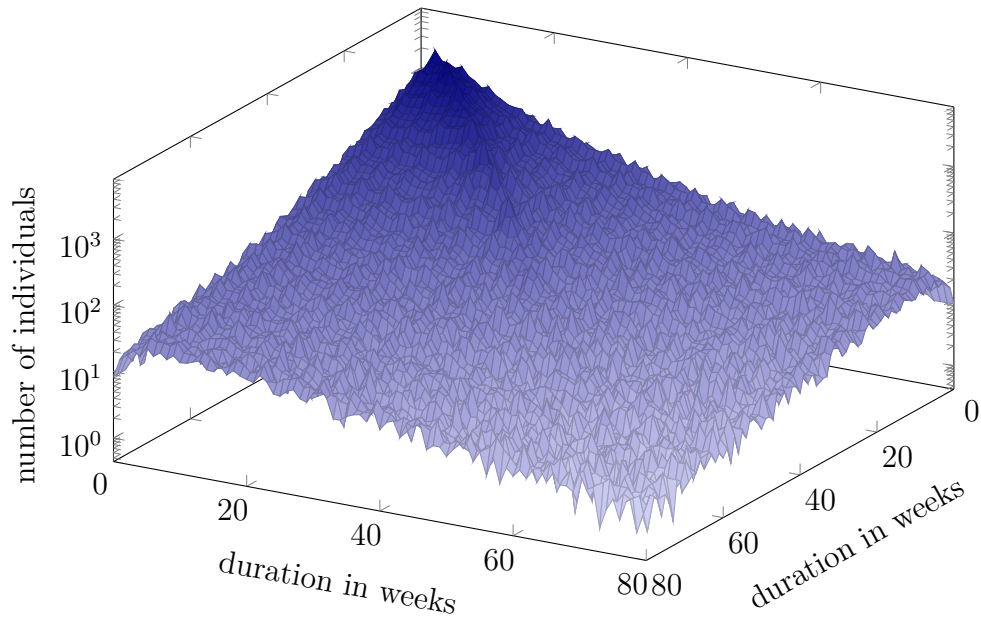

Figure 2: Nonemployment exit joint density during the first two non-employment spells, conditional on duration less than or equal to 60 weeks.

5   Results

5.1   Test of the Model

We propose a test of the model inspired by Section 3.4. We make three changes to accommodate the reality of our data. The first is that the data are only available with weekly durations, and so we cannot measure the partial derivatives of the reemployment density ϕ. Instead, we propose a discrete time analog of equations (14)–(15):

$$\mathbb{E}(\alpha^2|t_1, t_2) = \frac{t_2^2 \log\left(\frac{\phi(t_1,t_2+1)}{\phi(t_1,t_2-1)}\right) - t_1^2 \log\left(\frac{\phi(t_1+1,t_2)}{\phi(t_1-1,t_2)}\right)}{t_1^2 - t_2^2} - \frac{3}{t_1 + t_2} \geq 0$$

$$\text{and } \mathbb{E}(\beta^2|t_1, t_2) = t_1 t_2 \left(\frac{t_1 t_2 \log\left(\frac{\phi(t_1,t_2+1)}{\phi(t_1,t_2-1)} \frac{\phi(t_1-1,t_2)}{\phi(t_1+1,t_2)}\right)}{t_1^2 - t_2^2} + \frac{3}{t_1 + t_2} \right) \geq 0,$$

where we have approximated partial derivatives using

$$\frac{\partial\phi(t_1,t_2)/\partial t_1}{\phi(t_1,t_2)} \approx \frac{1}{2}\log\left(\frac{\phi(t_1+1,t_2)}{\phi(t_1-1,t_2)}\right)$$

$$\text{and } \frac{\partial\phi(t_1,t_2)/\partial t_2}{\phi(t_1,t_2)} \approx \frac{1}{2}\log\left(\frac{\phi(t_1,t_2+1)}{\phi(t_1,t_2-1)}\right).$$

The second is that the density ϕ is not exactly symmetric in real world data, as seen in Figure 1. We instead estimate ϕ as $\frac{1}{2}(\phi(t_1,t_2)+\phi(t_2,t_1))$. The third is that the raw measure of ϕ is noisy, as we discussed in the previous section. This noise is amplified when we estimate the slope $\log\left(\frac{\phi(t_1+1,t_2)}{\phi(t_1-1,t_2)}\right)$ and $\log\left(\frac{\phi(t_1,t_2+1)}{\phi(t_1,t_2-1)}\right)$. In principle, we could address this by explicitly modeling calendar dependence in the net benefit from employment, but we believe this issue is secondary to our main analysis. Instead, we smooth the symmetric empirical density ϕ using a multidimensional Hodrick-Prescott filter and run the test on the trend $\bar{\phi}$.[9] Since Proposition 1 establishes that ϕ should be differentiable at all points except possibly along the diagonal, we also do not impose that $\bar{\phi}$ is differentiable on the diagonal. See Appendix G for more details on the filter we use.

Figure 3 displays our test results. Without any smoothing, we reject the model for over 40 percent of pairs (t_1, t_2) with $0 \leq t_1 < t_2 \leq 60$. Setting the smoothing parameter to at least 5 reduces the rejection rate below five percent. Setting it to at least 20 reduces the rejection rate below two percent. When we look at higher values of (t_1, t_2), we reject the model more often, even in smoothed data. This may be due to a reduction in the signal-to-noise ratio in

---

[9]In practice we smooth the function $\log(1+\phi(t_1, t_2))$, rather than ϕ, where ϕ is the number of individuals whose two spells have durations (t_1, t_2).
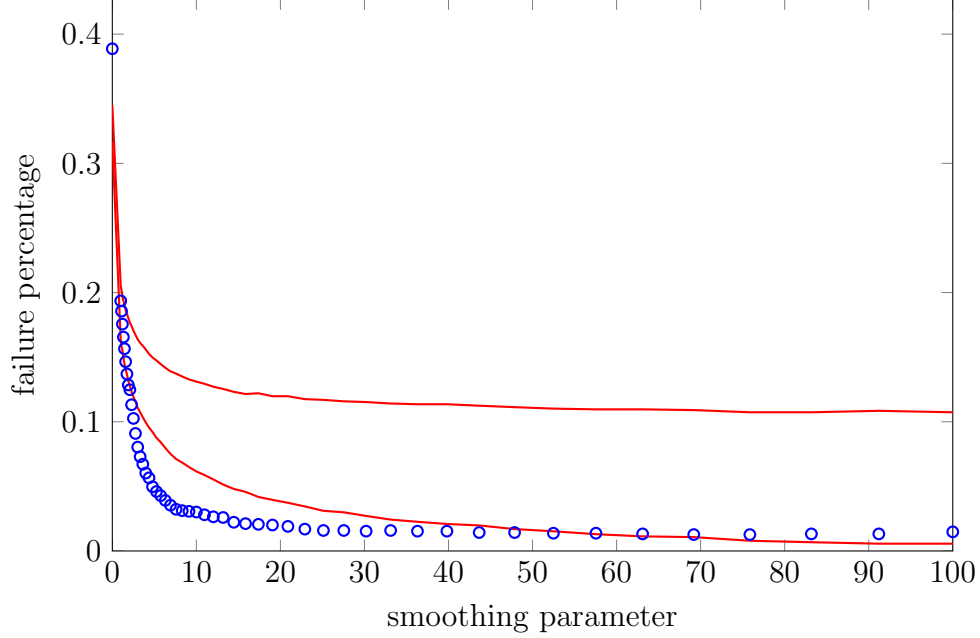
Figure 3: Nonparametric test of model. The blue circles show the percent of observations in the data with $a(t_1, t_2) < 0$ or $b(t_1, t_2) < 0$ with $0 \leq t_1 < t_2 \leq 60$ for different values of the filtering parameter. The red lines show a bootstrapped 95% confidence interval.

our data set.

To interpret the magnitude of the rejection rates, we show a bootstrapped 95% confidence interval. To compute the standard errors of the test statistic via bootstrapping, we need to sample under the null hypothesis, which is that the data are generated from the inverse Gaussian model, yet still stay as close as possible to the original data. Since the best description of our data through the lens of the model is the captured by the distribution G estimated in the later sections, we use it to draw samples. In particular, we draw 500 samples of two nonemployment spells for 800,000 individuals, and keep individuals with two completed spells between 0 and 260 weeks. We then proceed the same way as with the data. We construct the empirical distribution $\phi(t_1, t_2)$, and smooth it with our 2-dimensional HP filter for different values of the smoothing parameter λ. For each sample and each value of λ, we apply our test and evaluate the fraction of pairs (t_1, t_2) with $0 \leq t_1 < t_2 \leq 60$ with $a(t_1, t_2) < 0$ or $b(t_1, t_2) < 0$. The rejection rate in the 95% of our samples lies within the bands shown in red in Figure 3.

For low values of λ, our test statistic lies above the 95% confidence interval. This is because there are two sources of noise in the data but only one in the bootstrapped samples. The measured distribution of spells in the data is not smooth both because we have a finite number of people, and because there is seasonality. We observe spikes in the distribution

	minimum distance estimate				EM estimate			
	mean	median	st.dev.	min	mean	median	st.dev	min
α	0.36	0.20	0.51	0.007	391	0.12	2776	0.0803
β	7.48	5.03	5.94	1.466	2510	6.01	15623	1.4306
$\frac{\mu_n}{\bar{\omega}-\underline{\omega}}$	0.04	0.04	0.03	0.005	0.04	0.04	0.04	0.0177
$\frac{\sigma_n}{\bar{\omega}-\underline{\omega}}$	0.21	0.20	0.12	0.005	0.22	0.17	0.14	0.00001

Table 2: Summary statistics from estimation.

at one month, two months, three months etc., but this type of noise is not present in the bootstrapped samples. With low values of λ, there is very little smoothing, the real data are more noisy than the simulated one, and thus the rejection rate is higher. Higher values of λ remove both sources of noise and the measured rejection rate lies within or even below the 95% confidence interval. We thus conclude that our data could be generated by the proposed model and the non-zero rejection rate is due to small sample properties.

5.2 Estimation

We estimate our model in two steps. For a given type distribution $G(\alpha, \beta)$, the probability that any individual has duration $(t_1, t_2) \in T^2$ is

$$\frac{\iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\iint_{T^2} \iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \, dG(\alpha, \beta) \, dt_1 \, dt_2}.$$

We can therefore compute the likelihood function by taking the product of this object across all the individuals in the economy. Combining individuals with the same realized duration into a single term, we obtain that the log-likelihood of the data $\phi(t_1, t_2)$ is equal to

$$\sum_{(t_1, t_2) \in T^2} \phi(t_1, t_2) \log \left(\frac{\iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\iint_{T^2} \iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \, dG(\alpha, \beta) \, dt_1 \, dt_2} \right).$$

Our basic approach to estimation chooses a distribution function G to maximize this objective. More precisely, we follow a two-step procedure. In the first step we use a minimum distance estimator to obtain an initial estimate of G, constraining α and β to lie on a discrete grid. We then use the EM algorithm to perform maximum likelihood. In the second step, we allow α and β to take on values off of the grid. See appendix H for more details.

Our parameter estimates place a positive weight on 50 different types (α, β). Table 2 summarizes our estimates. We report mean, median, minimum and standard deviation of α, β and also μ and σ. The latter two are the drift and standard deviation of log-wages during
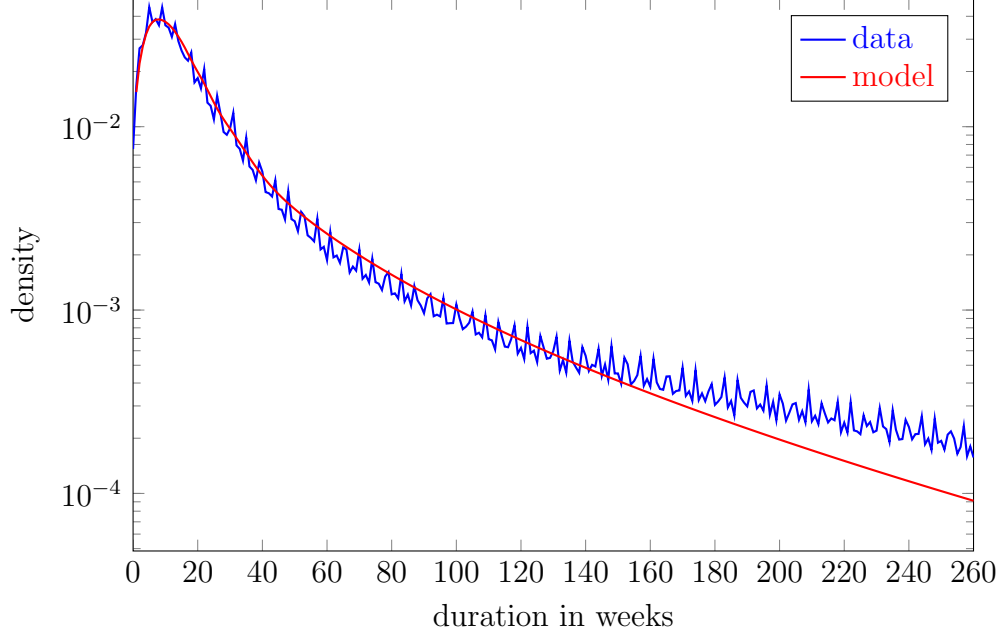
Figure 4: Marginal distribution of nonemployment spells in the data and in the model.

non-employment relative to the width of the inaction region, $\mu \equiv \mu_n/(\bar{\omega} - \underline{\omega}) = \alpha/\beta$ and $\sigma \equiv \sigma_n/(\bar{\omega} - \underline{\omega}) = 1/\beta$. Columns 2–5 summarize the estimates from the first estimation step, the last 4 columns show results after refining the initial estimates using the EM algorithm.

The mean and standard deviation of μ and σ are similar in both cases, but statistics for α and β differ substantially. The difference is due to the fact that there are now several types with a small value of σ which is then reflected in a very large value of α and β. Indeed, the EM step of our estimates reduces the smallest value of σ by two orders of magnitude, which has a big impact on the mean of α and β. The median values of α and β remain similar.

We find that there is a considerable amount of heterogeneity. For example the cross-sectional standard deviation of α is seven times its mean, while the cross-sectional standard deviation of β is around six times its mean. Moreover, α and β are positively correlated in the cross-section, with correlation 0.77 in the initial stage and 0.85 in the EM stage.

Figure 4 shows the fitted marginal distribution of spells up to 260 weeks of a non-employment. To make the marginal distribution from the model exactly comparable to the data, we normalize the distribution of spells between 0 and 260 weeks so that it adds up to one. The model matches the initial increase in distribution during the first thirteen weeks, as well as the gradual decline the subsequent five years. We miss the distribution at the very long durations because the mass of workers at these durations is very low.

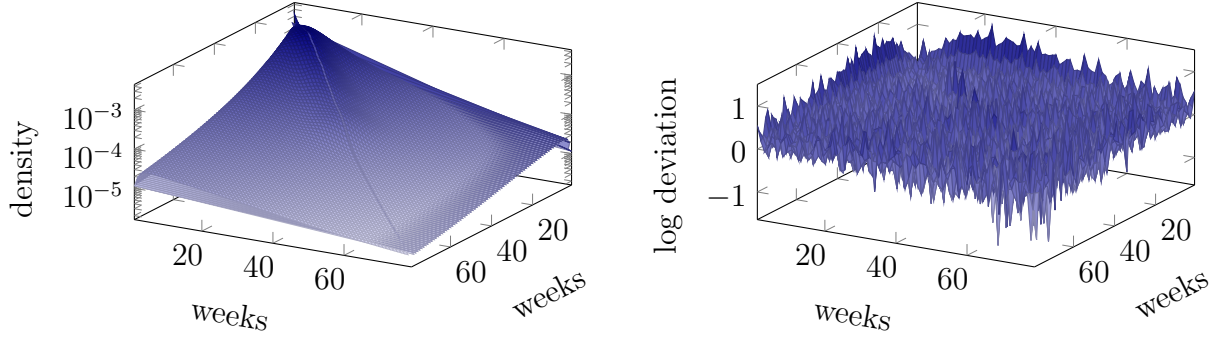Of course, it is not surprising that we can match the univariate hazard rate, since it

Figure 5: Nonemployment exit density: model (left) and log ratio of model to data (right)

is theoretically possible to match any univariate hazard rate with a mixture of (possibly degenerate) inverse Gaussian distributions. More interesting is that we can also match the joint density of the duration of the first two spells. The first panel in figure 5 shows the theoretical analog of the joint density in Figure 2. The second panel shows the log of the ratio of the empirical density to the theoretical density. The root mean squared error is about 0.17 times the average value of the density ϕ, with the model able to match the major features of the empirical joint density, leaving primarily the high frequency fluctuations that we previously indicated we would not attempt to match.

Finally, we use our analysis in 3.3 to infer bounds on the fraction of the population with a negative α. Using the estimated distribution and formulas (12) and (13), we find that for $\bar{t} = 260$, $\bar{c}(\bar{t}) = 0.9845$ and $\underline{c}(\bar{t}) \approx 0$. In the data we find that 78.6 percent of the starting sample has first two completed spells shorted than 260 weeks, we can soundly reject that the whole population has a negative α and conclude that majority has a positive value of α.

5.3   Robustness of estimates

Theorem 1 establishes that our model is identified using repeated spells, but this does not necessarily imply that our maximum likelihood estimates are consistent. For example, in Appendix H, we identify several biases that complicate our estimation procedure.[10] We use the following procedure to check if our estimator can recover the parameters if we were to have an arbitrarily large sample. We start with the joint distribution of types that we estimate in the previous subsection. We generate the population distribution for two consecutive non-employment spells both with duration between 0 and 260 weeks, i.e. we use equation (5) for

---

[10]One source of bias is that duration in measured in weeks, as opposed to be measured continuously. The other is that even the estimation a single type inverse gaussian with α close to zero has standard errors that diverge to infinity.

	minimum distance estimate				EM estimate			
	mean	median	st.dev.	min	mean	median	st.dev	min
α	0.49	0.03	1.77	0.007	728	0.14	4532	0.10
β	9.04	5.18	39.5	1.299	3640	6.74	18712	1.30
$\frac{\mu_n}{\bar{\omega}-\underline{\omega}}$	0.04	0.01	0.04	0.005	0.05	0.04	0.04	0.019
$\frac{\sigma_n}{\bar{\omega}-\underline{\omega}}$	0.22	0.20	0.14	0.001	0.21	0.15	0.14	0.00001

Table 3: Summary statistics from estimation using artificial data. The data are simulated from our model using the estimated distribution of types, with summary statistics shown in Table 2.

the estimated parameters to generate ϕ. We use our two step procedure on the distribution of duration ϕ to re-estimate the type distribution G. We then compare the results with the ones from the data generating process.

Table 3 summarizes our results. Our procedure recovers 45 types, even though the data were generated by 50 types. Nevertheless, the moments of α, β, $\mu = \alpha/\beta$, and $\sigma = 1/\beta$ are largely unchanged. More importantly, the results from the decomposition exercise and our estimates of the size of switching costs are also unchanged. We are therefore confident that we have recovered an accurate characterization of the joint distribution $G(\alpha, \beta)$.

5.4 Decomposition

We now use our estimated type distribution to decompose the evolution of the hazard rate and residual duration. We start with the hazard rate. Define

$$H^s(t) = \int_0^t \dot{H}^s(s)ds \text{ and } H^h(t) = \int_0^t \dot{H}^h(s)ds,$$

so $H(t) = H^s(t) + H^h(t)$ for all t.[11] The structural hazard rate represents the evolution of the hazard rate for the average individual at each point in time, without regards to dynamic selection. Plugging equations (16) and (17) into equation (18) gives

$$\dot{H}^s(t) = \frac{\iint \dot{f}(t; \alpha, \beta)dG(\alpha, \beta)}{\iint (1 - F(t; \alpha', \beta')).dG(\alpha', \beta')}$$

The model tells us $f(t; \alpha, \beta)$ and $F(t; \alpha, \beta)$ for all (α, β); these are the density and cumulative distribution of the inverse Gaussian. We use our estimate of the type distribution G to recover the structural contribution to the hazard rate.

We consider three cases for G: G^+ with all types positive, and $\bar{G}$ and $\underline{G}$ with a non-zero

---

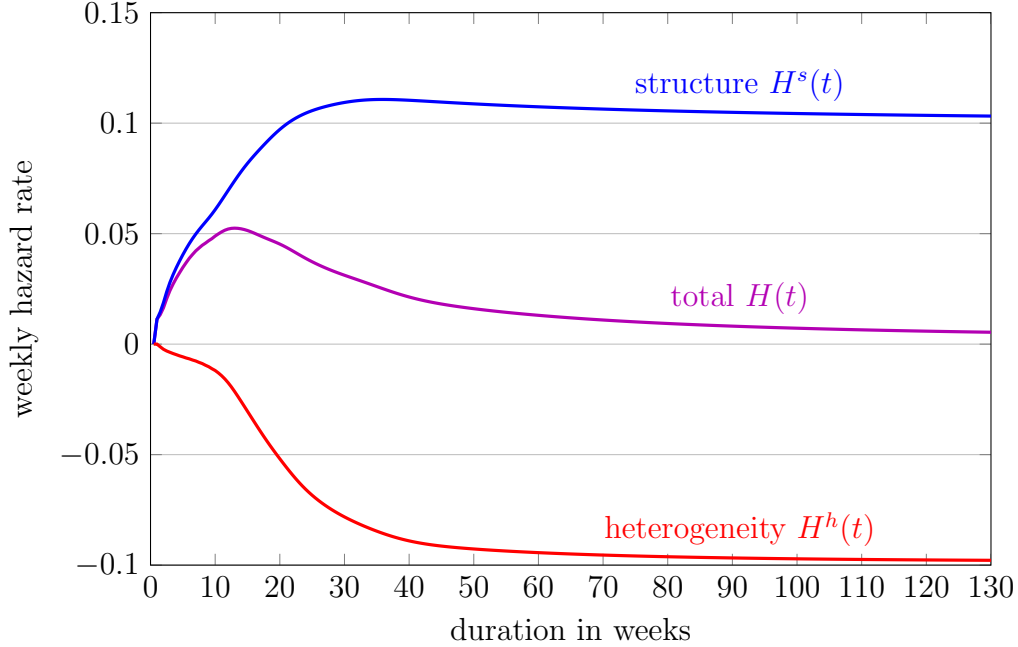[11]With an inverse Gaussian distribution, $H(0) = 0$, avoiding an additional term in this expression.

Figure 6: Decomposition of the hazard rate for distribution G^+. The blue line shows the structural hazard rate $H^s(t)$. The red line shows the contribution of heterogeneity, $H^h(t)$, which dynamically selects survivors to have a lower hazard rate. The sum of the two is the raw hazard rate $H(t)$, shown as a purple line. Note that these hazard rates do not conditional on the spell ending within 260 weeks and so may continue to decline indefinitely.

fraction of types with negative α. Even though we know that G^+ is not consistent with measured $\hat{c}$, we consider this case to be a useful benchmark. The hazard rate decomposition will be affected by the choice of distribution, mainly because the hazard rate for a type with α, β depends on the sign of α. While in both cases the hazard rate equals 0 at $t = 0$ and is hump-shaped, the hazard rate converges to an asymptote $\alpha^2/2$ for $\alpha > 0$ and to zero for $\alpha < 0$. With $\alpha < 0$, there is a positive probability of never completing a non-employment spell, so the hazard rate has to eventually go to zero.

Figure 6 shows that for G^+ structural hazard rate increases substantially, peaking at 11.1 percent after 33 weeks before falling to 10.3 percent after two years. In contrast, dynamic selection necessarily pushes down the empirical hazard rate, as showed in (20). After two years, this has cumulatively reduced the hazard rate by 9.8 percent. The difference, 0.5 percent, is the level of the raw hazard rate after two years. In other words, heterogeneity pulls the peak in the hazard rate forward by about 20 weeks and wipes out more than 90 percent of the long-run increase in the hazard rate. The contribution of heterogeneity to a declining hazard rate is substantial.

The decomposition with distributions $\underline{G}, \bar{G}$ is depicted in Figure 7. The level of aggregate
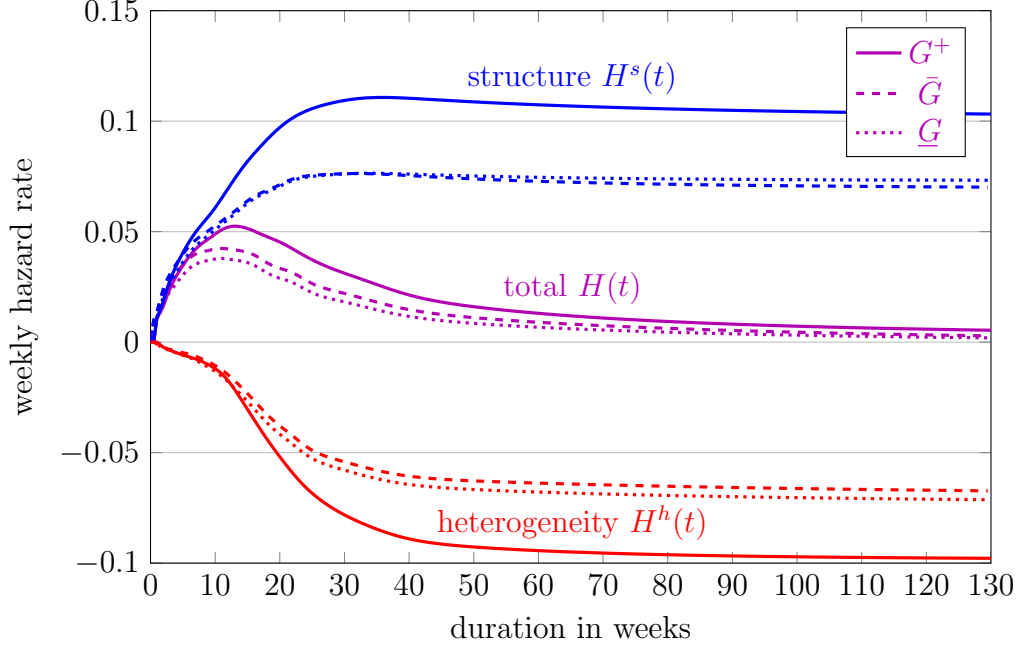
Figure 7: Decomposition of the hazard rate for distribution $G^+, \underline{G}, \bar{G}$. The blue lines show the structural hazard rate $H^s(t)$. The red lines show the contribution of heterogeneity, $H^h(t)$, which dynamically selects survivors to have a lower hazard rate. The sum of the two is the raw hazard rate $H(t)$, shown as purple lines. Solid lines correspond to distribution G^+, dashed lines to $\bar{G}$ and dotted lines to $\underline{G}$.

and structural hazard rate is lower than with G^+, which is a consequence of having types with negative α whose hazard rate converges to zero. However, qualitatively the decomposition result remains unchanged, leaving us again with the conclusion that the contribution of heterogeneity to the decreasing hazard rate is substantial during the first week of non-employment. The differences in the decomposition using $\underline{G}$ and $\bar{G}$ are very small.

We turn next to expected residual duration. Aggregate residual duration $R(t)$ is the expected number of additional weeks of non-employment anticipated by the average worker with duration t. In our estimated model, this number is huge, starting at over 236 weeks for a worker at the start of a jobless spell and reaching nearly 1200 weeks after one year out of work. To understand why, recall that expected duration for a newly displaced worker with type (α, β) is $1/\alpha$, while asymptotic residual duration is $2/\alpha^2$. Even a few workers with a value of α close to zero imply a high and rapidly rising residual duration in the whole population. Even worse, if α is less than or equal to zero, residual duration is infinite for that worker, an issue that our analysis has assumed away.

We focus instead on discounted residual duration. For any value of $\delta > 0$, discounted residual duration is capped at $1/\delta$, even if α is negative. This downplays the unreasonable
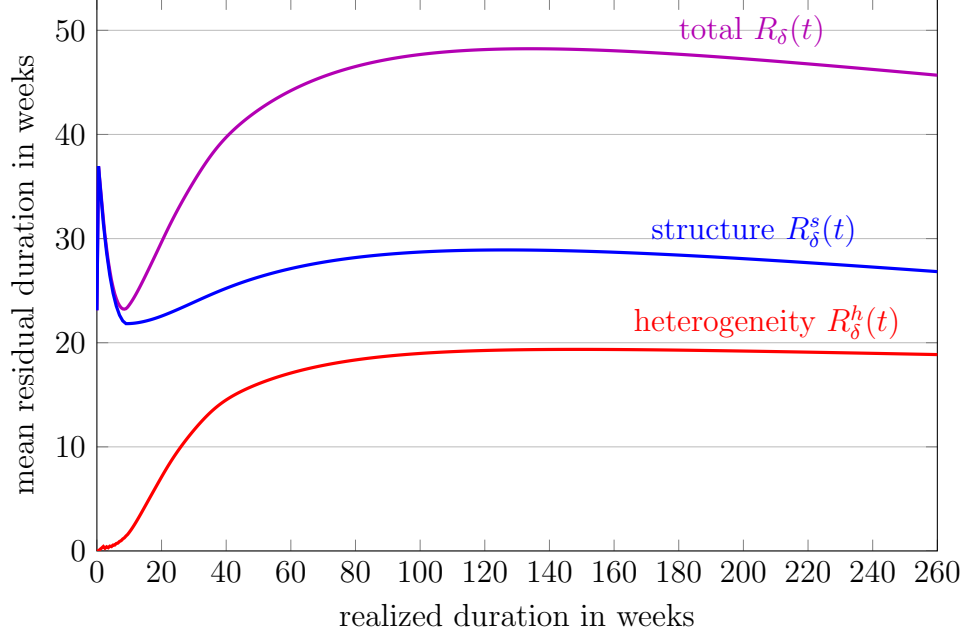
Figure 8: Decomposition of discounted residual duration. The blue line shows structural discounted residual duration $R_\delta^s(t)$. The red line shows the contribution of heterogeneity, $R_\delta^h(t)$, which generally slects survivors with higher residual duration. The sum of the two is the discounted residual duration $R_\delta(t)$, shown as a purple line.

influence of a small number of individuals with extremely large expected residual durations. We set the discount rate to $\delta = 0.0006$, which implies a worker's expected lifetime is just over 32 years. Discounting therefore has its main effect in truncating the very longest durations.

Figure 8 shows the decomposition of discounted residual duration, $R_\delta(t)$, into the components $R_\delta^s(t)$ and $R_\delta^h(t)$, defined as

$$R_\delta^s(t) = R_\delta(0) + \int_0^t \dot{R}_\delta^s(s)ds \text{ and } R_\delta^h(t) = \int_0^t \dot{R}_\delta^h(s)ds.$$

Note that we normalize all of the initial level of residual duration to be structural but focus on the changes in the outcomes. Expected residual duration initially falls because few workers find a job. Heterogeneity plays little role during this period. After about ten weeks, it starts to increase rapidly, eventually more than doubling in value to a peak of almost a year. The bulk of this increase is a consequence of heterogeneity, with structure playing a comparatively small role. Finally, expected residual duration declines at durations above 2.5 years. This primarily reflects the structure of the model, with little change in the composition of job searchers.

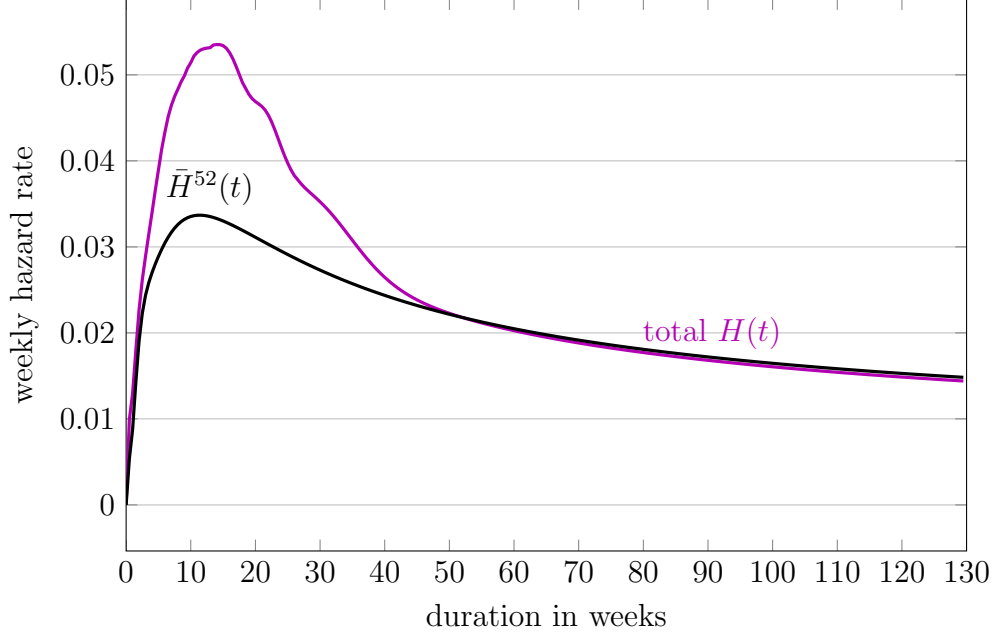To reiterate the importance of heterogeneity we have recovered, we consider another

Figure 9: Hazard rate of workers still out of work at duration of 52 weeks. The black line shows the hazard rate of an average worker who is non-employed at 52 weeks, the purple line is the aggregate hazard rate.

experiment. We take the distribution of types at duration of 52 weeks, $dG(52; \alpha, \beta)$, and calculate an average hazard rate at different durations for this group, denote it $\bar{H}^{52}(t)$. We have

$$\bar{H}^{52}(t) = \int \int h(t; \alpha, \beta;) dG(52; \alpha, \beta). \tag{24}$$

Figure 9 shows $\bar{H}^{52}(t)$ together with $H(t)$. We observe that $\bar{H}^{52}(t)$ is much lower than $H(t)$ at short durations, reaching the highest value of 3.4 percent at duration of 12 weeks and declining to 2.1 at 52 weeks. This indicates that prospects of an average worker out of work for one year have not looked much better when she became nonemployed than they look now.

5.5 Estimated Switching Costs

Section 2.3 shows that the knowledge of α, β together with other four parameters of the model can be used to estimate the magnitude of the fixed costs. Here we use the estimated distribution of α, β to find an upper bound on the distribution of the fixed costs in the population. We assume that there are no costs of switching from employment to non-employment, $\psi_n = 0$, and we focus on costs of switching from non-employment to employment, relative to the value of leisure, ψ_e/b_0.

34

	switching costs $\psi_e/b_0 \times 100$					standard deviation σ_n	
	mean	std	10th per.	50th per.	90th per.	mean	std
$\mu_n^* = +0.0325$	0.06	0.07	0.0021	0.026	0.001	0.028	0.015
$\mu_n^* = -0.0325$	2.92	2.98	0.64	0.95	10	0.028	0.015

Table 4: Summary statistics for the estimated switching cost.

We choose $r = 0.02, \mu_e = 0.01, \sigma_e = 0.05, |\mu_n| = 3.25\mu_e$. We consider both positive and negative μ_n. The fixed costs are larger for $\mu_n < 0$.

For both values of μ_n, the estimated width of the range of inaction as well as the corresponding values of ψ_e/b_0 are rather small. The costs are larger for the negative values of μ_n so we focus on this case in our discussion.

The median value of the switching costs is only 2.92 percent of the annual non-employment flow value. The costs vary a lot across types: the highest cost is 16.7, the lowest 0.18 percent of the annual non-employment flow value. Recall that we choose values of other parameters to make the fixed costs large, yet the median value is 3 percent, and the highest 16.7 percent.

It has been argued in the literature that even small fixed costs can generate large regions of inaction. In our model, however, not only the fixed costs are small, but so is the region of inaction.[12] The mean width of the inaction region is 0.0194 and the median is 0.0187. To understand how small these values are, consider a worker who has just started working and thus her current log wage equals $\bar{\omega} + b_0$. The wage that induces an average worker to quit to non-employment is merely 2 percent lower than the starting wage. By construction, these values are the same for the positive or negative μ_n^*.

Previous work by Mankiw (1985), Dixit (1991), Abel and Eberly (1994), and others has shown that even small costs can have a large economic impact. We are unaware of other papers which study the cost of switching between employment and non-employment at the level of an individual worker. In other areas, empirical results on the size of fixed costs are more mixed. Cooper and Haltiwanger (2006) find a large fixed cost of capital adjustment, around 4 percent of the average plant-level capital stock. Nakamura and Steinsson (2010) estimate a multisector model of menu costs and find that the annual cost of adjusting prices is less than 1 percent of firms' revenue. In a model of house selling, Merlo, Ortalo-Magne, and Rust (2013) find a very small fixed cost of changing the listing price of a house, around 0.01 percent of the house value.

---

[12]On the other hand, a non-degenerate region of inaction is important for our results. If the region of inaction were degenerate, we would be unable to match the mean duration of a spell, for the reasons we discussed in Section 2.3.

6 Conclusion

We develop a dynamic model of a worker's transitions in and out of employment. Our model features a structural duration dependence in the job finding rate, in the sense that the hazard rate of finding a job changes during a non-employment spell for a given worker. Moreover, the job finding rate as a function of duration varies across workers. We use the model to answer two questions. First, what are the costs of switching between employment and non-employment, and second, what is the relative importance of heterogeneity versus structural duration dependence for explaining the decreasing aggregate job finding rate. We find that the costs of switching between employment and non-employment are very small, with only 18 percent of workers having this costs larger than 1 percent of annual non-employment flow benefit. Even though switching costs are small in magnitude, they are economically relevant for non-employment duration, because we can soundly reject any version of the model without switching costs. We further find that the decline in the job-finding rate is mostly driven by changes in the composition of the pool of non-employed workers, rather than by declines in the job-finding rate for the typical worker.

References

Aalen, Odd O., and Håkon K. Gjessing, 2001. "Understanding the shape of the hazard rate: A process point of view (with comments and a rejoinder by the authors)." *Statistical Science.* 16 (1): 1–22.

Abbring, Jaap H., 2012. "Mixed Hitting-Time Models." *Econometrica.* 80 (2): 783–819.

Abel, Andrew B., and Janice C. Eberly, 1994. "A Unified Model of Investment Under Uncertainty." *American Economic Review.* 84 (5): 1369–1384.

Alvarez, Fernando, and Robert Shimer, 2011. "Search and Rest Unemployment." *Econometrica.* 79 (1): 75–122.

Coelho, Carlos A., Rui P. Alberto, and Luis P. Grilo, 2005. "When Do the Moments Uniquely Identify a Distribution." Preprint 17/2005, Mathematics Department, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

Cooper, Russell W, and John C Haltiwanger, 2006. "On the nature of capital adjustment costs." *The Review of Economic Studies.* 73 (3): 611–633.

Cox, David R., 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society. Series B (Methodological).* 34 (2): 187–220.

Dixit, Avinash, 1991. "Irreversible Investment with Price Ceilings." *Journal of Political Economy.* 99 (3): 541–557.

Elbers, Chris, and Geert Ridder, 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies.* 49 (3): 403–409.

Feller, William, 1966. *An introduction to probability theory and its applications. Vol. II*, John Wiley & Sons Inc., New York, second edn.

Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*, Oxford University Press.

Heckman, James J., and Burton Singer, 1984a. "The Identifiability of the Proportional Hazard Model." *Review of Economic Studies.* 51 (2): 231–241.

Heckman, James J., and Burton Singer, 1984b. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica.* 52 (2): 271–320.

Hodrick, Robert James, and Edward C Prescott, 1997. "Postwar US Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking.* 29 (1): 1–16.

Honoré, Bo E., 1993. "Identification Results for Duration Models with Multiple Spells." *Review of Economic Studies.* 60 (1): 241–246.

Lancaster, Tony, 1972. "A Stochastic Model for the Duration of a Strike." *Journal of the Royal Statistical Society. Series A (General).* 135 (2): 257–271.

Lancaster, Tony, 1979. "Econometric Methods for the Duration of Unemployment." *Econometrica.* 47 (4): 939–956.

Lee, Mei-Ling Ting, and G. A. Whitmore, 2006. "Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary." *Statistical Science.* 21 (4): 501–513.

Lee, Mei-Ling Ting, and G. A. Whitmore, 2010. "Proportional Hazards and Threshold Regression: Their Theoretical and Practical Connections." *Lifetime data analysis.* 16 (2): 196–214.

Lemeshko, Boris Yu, Stanislav B. Lemeshko, Kseniya A. Akushkina, Mikhail S. Nikulin, and Noureddine Saaidia, 2010. "Inverse Gaussian Model and Its Applications in Reliability and Survival Analysis." in *Mathematical and Statistical Models and Methods in Reliability*Statistics for Industry and Technology, pp. 433–453. Birkhäuser Boston.

Ljungqvist, Lars, and Thomas J Sargent, 1998. "The European Unemployment Dilemma." *Journal of Political Economy.* 106 (3): 514–550.

Loeve, Michel, 1977. *Probability Theory I* . , vol. 45, Graduate texts in mathematics.

Mankiw, N Gregory, 1985. "Small menu costs and large business cycles: A macroeconomic model of monopoly." *The Quarterly Journal of Economics.* 100 (2): 529–537.

Merlo, Antonio, Francois Ortalo-Magne, and John Rust, 2013. "The home selling problem: Theory and evidence." PIER Working Paper.

Nakamura, Emi, and Jón Steinsson, 2010. "Monetary Non-neutrality in a Multisector Menu Cost Model." *The Quarterly Journal of Economics.* 125 (3): 961–1013.

Newby, Martin, and Jonathan Winterton, 1983. "The Duration of Industrial Stoppages." *Journal of the Royal Statistical Society, Series A (General).* 146 (1): 62–70.

Shimer, Robert, 2008. "The Probability of Finding a Job." *American Economic Review.* 98 (2): 268–273.

Zweimuller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philipe Wuellrich, Oliver Ruf, and Simon Buchi, 2009. "Austrian Social Security Database." Mimeo.

Appendix

A Structural Model

A.1 Characterization of Thresholds

This section describes the structural model in Section 2.1 precisely and characterizes the solution to it. We assume that $b(t)$ and $w(t)$ follow a state-contingent Brownian motions:

$$db(t) = \begin{cases} \mu_{b,e}\, dt + \sigma_{b,e}\, dB_b(t) & \text{if worker is employed}, s = e \\ \mu_{b,n}\, dt + \sigma_{b,n}\, dB_b(t) & \text{if worker is non-employed}, s = n, \end{cases}$$

$$dw(t) = \begin{cases} \mu_{w,e}\, dt + \sigma_{w,e}\, dB_w(t) & \text{if worker is employed}, s = e \\ \mu_{w,n}\, dt + \sigma_{w,n}\, dB_w(t) & \text{if worker is non-employed}, s = n. \end{cases}$$

$B_b(t)$ and $B_w(t)$ are correlated Brownian motions. We let $\rho_s \in [-1,1]$ be the instantaneous correlation between dw and db in state $s \in \{e,n\}$:

$$\mathbb{E}\left[dw(t)\, db(t)\right] = \begin{cases} \sigma_{w,e}\sigma_{b,e}\, \rho_e\, dt & \text{if worker is employed}, s = e \\ \sigma_{w,n}\sigma_{b,n}\, \rho_n\, dt & \text{if worker is non-employed}, s = n. \end{cases}$$

The state of worker's problem is triplet (s,w,b) where $s \in \{e,n\}$ denotes whether the worker is employed or non-employed. Denote value function of an employed worker with state (w,b) as $\tilde{E}(w,b)$ and the value for a non-employed worker with state (w,b) by $\tilde{N}(w,b)$. These satisfy:

$$\tilde{E}(w,b) = \max_{\tau_e} \mathbb{E}\left[\int_0^{\tau_e} e^{-rt}e^{w(t)}dt + e^{-r\tau_e}\left(\tilde{N}(w(\tau_e),b(\tau_e)) - \psi_n e^{b(\tau_e)}\right)|w(0) = w, b(0) = b\right] \tag{25}$$

$$\tilde{N}(w,b) = \max_{\tau_n} \mathbb{E}\left[\int_0^{\tau_n} e^{-rt}b_0 e^{b(t)}dt + e^{-r\tau_n}\left(\tilde{E}(w(\tau_n),b(\tau_n)) - \psi_e e^{b(\tau_n)}\right)|w(0) = w, b(0) = b\right] \tag{26}$$

In equation (25), the employed worker chooses the stopping time τ_e at which to switch to non-employment. Similarly in equation (26), the non-employed worker chooses the first time τ_n at which to change his status to employment. The expectation in equation (25) and Equation (26) is taken with respect of the law of motion for $w(t)$ and $b(t)$ between $0 \le t \le \tau_e$ when $s = e$, or $0 \le t \le \tau_n$ when $s = n$.

For the problem to be well-defined, we require that

$$r > \mu_{w,s} + \tfrac{1}{2}\sigma_{w,s}^2 \quad \text{for } s \in \{e, n\} \tag{27}$$

$$r > \mu_{b,s} + \tfrac{1}{2}\sigma_{b,s}^2 \quad \text{for } s \in \{e, n\} \tag{28}$$

The conditions in (27) guarantee that the value of being employed (non-employed) forever is finite. Moreover, if the conditions in (28) hold, then being non-employed (employed) for T periods and then switching to employment (non-employment) forever is also finite in the limit as T converges to infinity.

From equation (25) and equation (26) it is immediate to show that we can restrict our attention to functions that satisfy the following homogeneity property. For any pair (w, b) and any constant a:

$$\tilde{E}(w + a, b + a) = e^a \, \tilde{E}(w, b),$$
$$\tilde{N}(w + a, b + a) = e^a \, \tilde{N}(w, b).$$

By choosing $a = -b$, we get

$$\tilde{E}(w, b) = e^b \, \tilde{E}(w - b, 0) \equiv e^b \, E(w - b)$$
$$\tilde{N}(w, b) = e^b \, \tilde{N}(w - b, 0) \equiv e^b \, N(w - b)$$

which implicitly defines E and N only as a function the scalar $w - b$. We define $\omega(t)$, the log net benefit to work, as $\omega(t) \equiv w(t) - b(t)$, so

$$d\omega(t) = \mu_s dt + \sigma_s dB(t)$$

where $\{B\}$ is a standard Brownian motion defined in terms of Brownian motions $\{B_b, B_w\}$. The process for $\omega(t)$ has a drift and a diffusion coefficient for $s \in \{e, n\}$ given by:

$$\mu_s = \mu_{w,s} - \mu_{b,s} \text{ and } \sigma_s^2 = \sigma_{w,s}^2 - 2\sigma_{w,s}\sigma_{b,s}\rho_s + \sigma_{b,s}^2.$$

The optimal decision of switching from employment to non-employment and vice versa is described by thresholds $\underline{\omega}$ and $\bar{\omega}$ such that a non-employed worker chooses to become employed if the net benefit from working is sufficiently high, $\omega(t) > \bar{\omega}$, and an employed worker switches to non-employment if the benefit is sufficiently low, $\omega(t) < \underline{\omega}$. To see this,

note that since switching after paying the fixed cost is feasible it must be the case that:

$$\tilde{E}(w, b) \geq \tilde{N}(w, b) - e^b \psi_n \text{ for all } (w, b), \text{ or } E(\omega) \geq N(\omega) - \psi_n \text{ for all } \omega \text{ and}$$

$$\tilde{N}(w, b) \geq \tilde{E}(w, b) - e^b \psi_e \text{ for all } (w, b), \text{ or } N(\omega) \geq E(\omega) - \psi_e \text{ for all } \omega$$

with equality at the states where switching is optimal.

To solve for the thresholds, we formulate the Hamilton-Jacobi-Bellman (HJB) equation for the worker's problem. Start with HJB for $\tilde{E}(w, b)$ and $\tilde{N}(w, b)$:

$$r\tilde{E}(w, b) = e^w + \tilde{E}_1(w, b)\mu_{w,e} + \tilde{E}_2(w, b)\mu_{b,e} + \tilde{E}_{11}(w, b)\frac{\sigma_{w,e}^2}{2} + \tilde{E}_{22}(w, b)\frac{\sigma_{b,e}^2}{2} + \tilde{E}_{12}(w, b)\sigma_{w,e}\sigma_{b,e}\rho_e$$

for all w and b with $w - b \geq \underline{\omega}$. Similarly, if the worker is non-employed,

$$r\tilde{N}(w, b) = b_0 e^b + N_1(w, b)\mu_{w,n} + \tilde{N}_2(w, b)\mu_{b,n} + \tilde{N}_{11}(w, b)\frac{\sigma_{w,n}^2}{2} + \tilde{N}_{22}(w, b)\frac{\sigma_{b,n}^2}{2} + \tilde{N}_{12}(w, b)\sigma_{w,n}\sigma_{b,n}\rho_n$$

for all w and b with $w - b \leq \bar{\omega}$. The boundary conditions for the problem are given by

$$\tilde{E}(\underline{\omega}, 0) = \tilde{N}(\underline{\omega}, 0) - \psi_n, \ \tilde{E}_1(\underline{\omega}, 0) = \tilde{N}_1(\underline{\omega}, 0), \ \tilde{E}_2(\underline{\omega}, 0) = \tilde{N}_2(\underline{\omega}, 0)$$

$$\tilde{N}(\bar{\omega}, 0) = \tilde{E}(\bar{\omega}, 0) - \psi_e, \ \tilde{E}_1(\bar{\omega}, 0) = \tilde{N}_1(\bar{\omega}, 0), \ \tilde{E}_2(\bar{\omega}, 0) = \tilde{N}_2(\bar{\omega}, 0)$$

Thus, worker's problem leads to two partial differential equations. These are difficult to solve in general, and therefore we use the homogeneity property and rewrite them as a system of second-order ordinary differential equations for $E(\omega)$ and $N(\omega)$.

We write the the derivatives of E and N in terms of $\tilde{E}$ and $\tilde{N}$:

$$\tilde{E}_1(w, b) = e^b E'(w - b) \text{ and } \tilde{E}_2(w, b) = e^b E(w - b) - e^b E'(w - b)$$

Differentiate again to obtain the second derivatives. The expressions for the derivatives of N are analogous. Use these to get a second-order ODE for $E(\omega)$ and $N(\omega)$:

$$r_e E(\omega) = e^\omega + \mu_e E'(\omega) + \tfrac{1}{2}\sigma_e^2 E''(\omega) \tag{29}$$

$$r_n N(\omega) = b_0 + \mu_n N'(\omega) + \tfrac{1}{2}\sigma_n^2 N''(\omega) \tag{30}$$

where the parameters are

$$r_s \equiv r - \mu_{b,s} - \tfrac{1}{2}\sigma_{b,s}^2$$

$$\mu_s \equiv \mu_{w,s} - \mu_{b,s} - \sigma_{b,s}^2 + \sigma_{w,s}\sigma_{b,s}\rho_s$$

$$\sigma_s^2 \equiv \sigma_{w,s}^2 + \sigma_{b,s}^2 - 2\sigma_{w,s}\sigma_{b,s}\rho_s$$

for $s \in \{e, n\}$. Conditions (27) and (28) reduce to

$$r_s > \mu_s + \tfrac{1}{2}\sigma_s^2 \text{ and } r_s > 0 \text{ for } s \in \{e, n\}. \tag{31}$$

We can also rewrite the boundary conditions as

$$E(\underline{\omega}) = N(\underline{\omega}) - \psi_n \text{ and } E'(\underline{\omega}) = N'(\underline{\omega}) \tag{32}$$

$$N(\bar{\omega}) = E(\bar{\omega}) - \psi_e \text{ and } N'(\bar{\omega}) = E'(\bar{\omega}). \tag{33}$$

The solution to equation (29) and equation (30) with boundary conditions equation (32) and equation (33) is of a form

$$E(\omega) = \frac{e^\omega}{r_e - \mu_e - \sigma_e^2/2} + c_{e,1}e^{\lambda_{e,1}\omega} + c_{e,2}e^{\lambda_{e_2}\omega} \tag{34}$$

$$N(\omega) = \frac{b_0}{r_n} + c_{n,1}e^{\lambda_{n,1}\omega} + c_{n,2}e^{\lambda_{n,2}\omega} \tag{35}$$

where

$$\lambda_{e,1} < 0 < \lambda_{e,2} \text{ and } \lambda_{n,1} < 0 < \lambda_{n,2}$$

are the roots of the equations $r_e = \lambda_e\left(\mu_e + \lambda_e\sigma_e^2/2\right)$ and $r_n = \lambda_n\left(\mu_n + \lambda_n\sigma_n^2/2\right)$. Hence we have six equations, (32)–(35), in six unknowns, $(c_{e,1}, c_{e,2}, c_{n,1}, c_{n,2}, \underline{\omega}, \bar{\omega})$. We turn to their solution.

Two non-bubble conditions require that

$$\lim_{\omega \to -\infty} N(\omega) = \frac{b_0}{r_n} \text{ and} \tag{36}$$

$$\lim_{\omega \to +\infty} \frac{E(\omega)}{e^\omega} = \frac{1}{r_e - \mu_e - \sigma_e^2/2} \tag{37}$$

Equation (36) requires that the value function for arbitrarily small ω converges to the value of being non-employed forever. Likewise equation (37) requires that for an arbitrarily large ω the value function converges to the value of being employed forever. These no-bubble conditions imply that $c_{e,2} = c_{n,1} = 0$. Simplifying the notation, we let $c_e = c_{e,1} > 0$,

$\lambda_e = \lambda_{e,1} < 0$, $c_n = c_{n,2} > 0$, and $\lambda_n = \lambda_{n,2} > 0$. Using this, we rewrite the value functions (34) and (35) as:

$$E(\omega) = \frac{e^\omega}{r_e - \mu_e - \sigma_e^2/2} + c_e e^{\lambda_e \omega} \text{ for all } \omega \geq \underline{\omega} \tag{38}$$

$$N(\omega) = \frac{b_0}{r_n} + c_n e^{\lambda_n \omega} \text{ for all } \omega \leq \bar{\omega} \tag{39}$$

with

$$\lambda_e = \frac{-\mu_e - \sqrt{\mu_e^2 + 2r_e \sigma_e^2}}{\sigma_e^2} < -1 \text{ and } \lambda_n = \frac{-\mu_n + \sqrt{\mu_n^2 + 2r_n \sigma_n^2}}{\sigma_n^2} > 1. \tag{40}$$

Condition (31) ensures that the roots are real and satisfy the specified inequalities.

We now have four equations, two value matching and two smooth pasting, in four unknowns $(c_e, c_n, \underline{\omega}, \bar{\omega})$. Rewrite these as

$$\psi_n + \frac{e^{\underline{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e e^{\lambda_e \underline{\omega}} = \frac{b_0}{r_n} + c_n e^{\lambda_n \underline{\omega}} \tag{41}$$

$$-\psi_e + \frac{e^{\bar{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e e^{\lambda_e \bar{\omega}} = \frac{b_0}{r_n} + c_n e^{\lambda_n \bar{\omega}} \tag{42}$$

$$\frac{e^{\underline{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e \lambda_e e^{\lambda_e \underline{\omega}} = c_n \lambda_n e^{\lambda_n \underline{\omega}} \tag{43}$$

$$\frac{e^{\bar{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e \lambda_e e^{\lambda_e \bar{\omega}} = c_n \lambda_n e^{\lambda_n \bar{\omega}} \tag{44}$$

Note that the values of c_e and c_n have to be positive, since it is feasible to choose to be either employed forever or non-employed forever, and since the value of being employed forever and non-employed forever are the obtained in equations (38) and equations (39) by setting $c_e = 0$ and $c_n = 0$ respectively.

Figure 10 displays an example of the value functions $E(\cdot)$ and $N(\cdot)$ for the case $\psi_n = 0$. We plot the net benefit from employment on the horizontal axis and indicate the thresholds $\underline{\omega} < \bar{\omega}$. The domain of the employment value function E is $[\underline{\omega}, \infty)$ and the domain of the non-employment value function is $(-\infty, \bar{\omega}]$. We also plot the value of non-employment forever, i.e. b_0/r_n, and the value of employment forever, i.e. $e^\omega/(r_e - \mu_e - \sigma_e^2/2)$. It is readily seen that as $\omega \to -\infty$, the value function $N(\omega)$ converges to the value of non-employment forever, and that as $\omega \to \infty$, the value function $E(\omega)$ converges to the value of employment forever. Additionally the level and slope of E and N coincide at $\underline{\omega}$, while at $\bar{\omega}$ the slopes coincide, but the value of E is ψ_e higher than N, since a non-employed worker must pay the fixed cost to become employed.
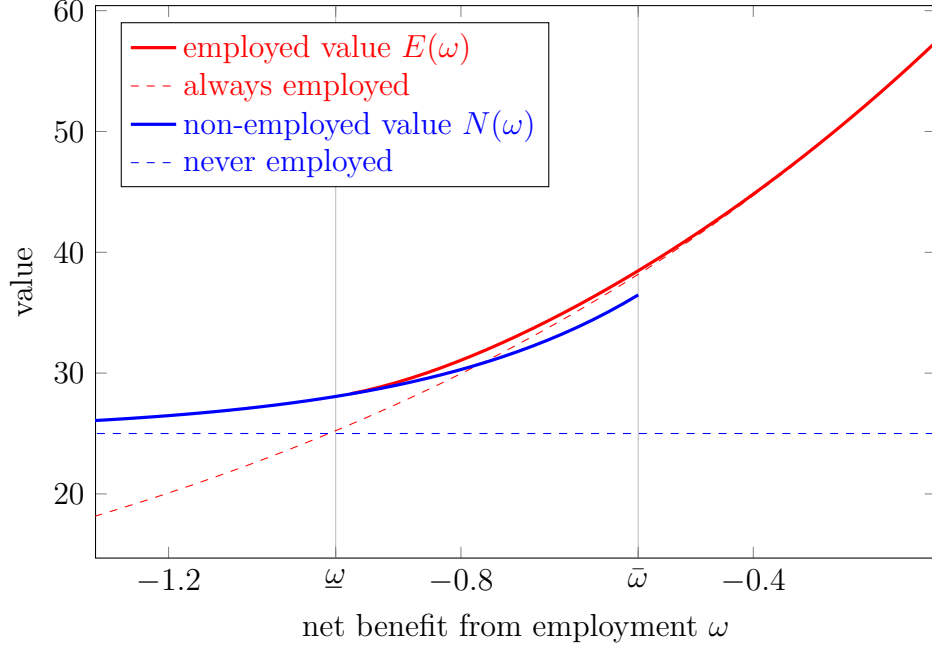
Figure 10: Example of Value Functions. The parameters values are $r = 0.04$, $\mu_e = 0.02$, $\sigma_e = 0.1$, $\mu_n = 0.01$, $\sigma_n = 0.04$, $b_0 = 1$, $\mu_{b,s} = \sigma_{b,s} = 0$, $\psi_e = 2$, and $\psi_n = 0$.

A.2 Determinants of Barriers

Our goal in this section is to understand how duration data alone can be used to infer information about the size of switching costs. We start with a result on the units of switching costs.

Lemma 1 *Fix $\lambda_n > 1$, $\lambda_e < -1$ and $r_e - \mu_e - \sigma_e^2/2 > 0$. Suppose that $(c_e, c_n, \underline{\omega}, \bar{\omega})$ solve the value functions for fixed cost and flow utility of non-employment (ψ_e, ψ_n, b_0). Then for any $k > 0$, $(e', n', \underline{\omega}', \bar{\omega}')$ solve the value function for flow utility of non-employment $b_0' = kb_0$ and fixed cost $\psi_e' = k\psi_e, \psi_n' = k\psi_n$ with $\bar{\omega}' = \bar{\omega} + \log k$, $\underline{\omega}' = \underline{\omega} + \log k$, $c_e' = c_e k^{1-\lambda_e}$, and $n' = nk^{1-\lambda_n}$.*

To prove Lemma 1, multiply the appropriate objects in equations (41) and (42) by k and then simplifying those equations as well as equations (43) and (44) using the expressions in the statement of the proof. We omit the algebraic details. The lemma implies that the size of the region of inaction, $\bar{\omega} - \underline{\omega}$, depends only on $(\psi_e + \psi_n)/b_0$.

We can invert this logic to express the implied size of the fixed costs $(\psi_e + \psi_n)/b_0$ as a function of the width of the region of inaction and other model parameters. In the first step, solve equations (43) and (44) for c_e and c_n. Because $\lambda_e < -1$, $\lambda_n > 1$, and $\bar{\omega} > \underline{\omega}$, the equations are linearly independent and so there is a unique solution. Moreover, these

same parameter restrictions ensure that $c_e > 0 > c_n$. Next take the difference between equations (41) and (42) to find the sum of the fixed costs, $\psi \equiv \psi_e + \psi_n$, again a positive number. Finally, let $\gamma \equiv \psi_e/\psi$ denote the share of the fixed costs paid during employment. Then solve equation (42) for b_0 as a function of model parameters, including γ. Once again this is positive. Finally, we take the ratio of these last two expressions to obtain ψ/b_0, the relative size of the fixed costs, as a function of model parameters (including γ) and the barriers $\bar\omega$ and $\underline\omega$. Lemma 1 implies that this depends on $\bar\omega - \underline\omega$ alone.

The resulting expression is messy, but we obtain a simple approximation when the barriers are close together:

Proposition 4 *The distance between the barriers is approximately proportional to the cube root of the size of fixed costs. More precisely,*

$$\frac{\psi}{b_0} = -\frac{\lambda_e \lambda_n (\bar\omega - \underline\omega)^3}{12 r_n} + o\big((\bar\omega - \underline\omega)^3\big),$$

where λ_e and λ_n are given in equation (40).

We use this equation in the text to infer the size of the fixed costs from the distance between the barriers and known values of the other parameters. Numerical simulations indicate that this approximation is very accurate at empirically plausible values of $\bar\omega - \underline\omega$.

B   Proof of Identification

We start by proving a preliminary lemma that describes the structure of the partial derivatives of the product of two inverse Gaussian distributions.

Lemma 2 *Let m be a nonnegative integer and $i = 0, \ldots, m$. The partial derivative of the product of two inverse Gaussian distributions at (t_1, t_2) is:*

$$\frac{\partial^m \big(f(t_1; \alpha, \beta)\, f(t_2; \alpha, \beta)\big)}{\partial t_1^i\, \partial t_2^{m-i}} = f(t_1; \alpha, \beta)\, f(t_2; \alpha, \beta) \left(\sum_{r,s=0}^{r+s \le m} \kappa_{r,s}(t_1, t_2; i, m-i) \alpha^{2r} \beta^{2s} \right) \quad (45)$$

where $\kappa_{r,s}(t_1, t_2; i, m-i)$ are polynomials functions of (t_1, t_2),

$$\kappa_{r,s}(t_1, t_2; i, m-i) = \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \theta_{k,\ell,r,s}(i, m-i) t_1^{-k}\, t_2^{-\ell}, \quad (46)$$

and the coefficients $\theta_{k,\ell,r,s}(i, m-i)$ are independent of t_1, t_2, α, and β.

Proof of Lemma 2. The lemma holds trivially when $m = i = 0$, with $\kappa_{0,0}(t_1, t_2, 0, 0) = 1$. We now proceed by induction. Assume equation (45) holds for some $m \geq 0$ and all $i \in \{0, \ldots, m\}$. We first prove that it holds for $m+1$ and all $i+1 \in \{1, \ldots, m+1\}$, then verify that it also holds for $i = 0$. We start by differentiating the key equation:

$$\frac{\partial^{m+1}\big(f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\big)}{\partial t_1^{i+1} \, \partial t_2^{m-i}}$$

$$= \frac{\partial}{\partial t_1}\left(\frac{\partial^m \big(f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\big)}{\partial t_1^i \, \partial t_2^{m-i}}\right)$$

$$= f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \left(\frac{\beta^2}{2t_1^2} - \frac{3}{2t_1} - \frac{\alpha^2}{2}\right)\left(\sum_{r,s=0}^{r+s\leq m} \kappa_{r,s}(t_1, t_2; i, m-i)\alpha^{2r}\beta^{2s}\right)$$

$$+ f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \left(\sum_{r,s=0}^{r+s\leq m} \frac{\partial \kappa_{r,s}(t_1, t_2; i, m-i)}{\partial t_1}\alpha^{2r}\beta^{2s}\right)$$

or

$$\frac{1}{f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)} \frac{\partial^{m+1}\big(f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\big)}{\partial t_1^{i+1} \, \partial t_2^{m-i}}$$

$$= -\frac{1}{2} \sum_{r,s=0}^{r+s\leq m} \kappa_{r,s}(t_1, t_2; i, m-i)\alpha^{2(r+1)}\beta^{2s}$$

$$+ \frac{1}{2t_1^2} \sum_{r,s=0}^{r+s\leq m} \kappa_{r,s}(t_1, t_2; i, m-i)\alpha^{2r}\beta^{2(s+1)}$$

$$+ \sum_{r,s=0}^{r+s\leq m} \left(-\frac{3}{2t_1}\kappa_{r,s}(t_1, t_2; i, m-i) + \frac{\partial \kappa_{r,s}(t_1, t_2; i, m-i)}{\partial t_1}\right)\alpha^{2r}\beta^{2s}.$$

This expression defines the new functions $\kappa_{r,s}(t_1, t_2; i, m+1-i)$, and it can be verified that they are polynomial functions by induction. Finally, an analogous expression obtained by differentiating with respect to t_2 gives the result for $m+1$ and $i = 0$. ∎

Proof of Proposition 1. We seek conditions under which we can apply Leibniz's rule and differentiate equation (5) under the integral sign:

$$\frac{\partial^m \phi(t_1, t_2)}{\partial t_1^i \partial t_2^{m-i}} = \int \frac{\partial^m \big(f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\big)}{\partial t_1^i \, \partial t_2^{m-i}} G(d\alpha, d\beta)$$

for $m > 0$ and $i \in \{0, \ldots, m\}$. Let B represent a bounded, non-empty open neighborhood of (t_1, t_2) and let $\bar{B}$ denote its closure. Assume that there are no points of the form (t, t),

$(t_1, 0)$, or $(0, t_2)$ in $\bar{B}$. In order to apply Leibniz's rule, we must check two conditions:

1. The partial derivative $\partial^m \big(f(t_1; \alpha, \beta) \, f(t_2; \alpha, \beta)\big)/\partial t_1^i \, \partial t_2^{m-i}$ exists and is a continuous function of (t_1', t_2') for every $(t_1', t_2') \in B$ and G-almost every (α, β); and

2. There is a $G-$integrable function $h_{i,m-i} : \mathbb{R}_+^2 \to \mathbb{R}_+$, i.e. a function satisfying

$$\int h_{i,m-i}(\alpha, \beta) \, G(d\alpha, d\beta) < \infty$$

   such that for every $(t_1', t_2') \in B$ and G-almost every (α, β)

$$\left| \frac{\partial^m \big(f(t_1; \alpha, \beta) \, f(t_2; \alpha, \beta)\big)}{\partial t_1^i \, \partial t_2^{m-i}} \right| \leq h_{i,m-i}(\alpha, \beta) \ .$$

Existence of the partial derivatives follows from Lemma 2. The bulk of our proof establishes that the constant

$$h_{i,m-i} \equiv \max_{(t_1, t_2) \in \bar{B}} \sum_{r,s=0}^{r+s \leq m} \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \frac{\big|\theta_{k,\ell,r,s}(i, m-i)\big|}{2\pi} t_1^{-k-\frac{3}{2}} t_2^{-\ell-\frac{3}{2}} \left(\frac{r+s+1}{\tau(t_1, t_2)} \right)^{r+s+1} e^{-(r+s+1)}, \tag{47}$$

where

$$\tau(t_1, t_2) = \frac{(t_1 - t_2)^2}{2\big(t_1(1+t_2)^2 + t_2(1+t_1)^2\big)}, \tag{48}$$

is a suitable bound. Note that $h_{i,m-i}$ is well-defined and finite since it is the maximum of a continuous function on a compact set; the exclusion of points of the form (t, t), $(t_1, 0)$, or $(0, t_2)$ is important for this continuity. This bound on the $(i, m-i)$ partial derivatives ensures that the lower order partial derivatives are continuous.

We now prove that $h_{i,m-i}$ is an upper bound on the magnitude of the partial derivative. Using Lemma 2, the partial derivative is the product of a polynomial function and an exponential function:

$$\frac{\partial^m \big(f(t_1; \alpha, \beta) \, f(t_2; \alpha, \beta)\big)}{\partial t_1^i \, \partial t_2^{m-i}} = \left(\sum_{r,s=0}^{r+s \leq m} \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \frac{\theta_{k,\ell,r,s}(i, m-i)}{2\pi} t_1^{-k-\frac{3}{2}} t_2^{-\ell-\frac{3}{2}} \alpha^{2r} \beta^{2(s+1)} \right)$$

$$\times \exp\left(-\frac{(\alpha t_1 - \beta)^2}{2 t_1} - \frac{(\alpha t_2 - \beta)^2}{2 t_2} \right).$$

Only the constant terms θ may be negative.

To bound the partial derivative, first note that for any nonnegative numbers α and β, r, and s,

$$(\alpha + \beta)^{2(r+s+1)} \geq \alpha^{2r} \beta^{2(s+1)}. \tag{49}$$

To prove this, observe that the inequality holds when $r = s = 0$, and the difference between the right hand side and left hand side is increasing in r and s whenever the two sides are equal; therefore it holds at all nonnegative r and s. Next note that

$$\exp\left(-(\alpha + \beta)^2 \tau(t_1, t_2)\right) \geq \exp\left(-\frac{(\alpha t_1 - \beta)^2}{2t_1} - \frac{(\alpha t_2 - \beta)^2}{2t_2}\right). \tag{50}$$

This can be verified by finding a maximum of the right hand side of (50) with respect to α, β subject to the constraint that $\alpha + \beta = K$ for some $K > 0$. Next, consider the function $a^x \exp(-ay)$ for a and x nonnegative and y strictly positive. This is a single-peaked function of a for fixed x and y, achieving its maximum value at $a = x/y$. Letting $(\alpha + \beta)^2$ play the role of a, this implies in particular that

$$\left(\frac{r + s + 1}{\tau(t_1, t_2)}\right)^{r+s+1} e^{-(r+s+1)} \geq (\alpha + \beta)^{2(r+s+1)} \exp\left(-(\alpha + \beta)^2 \tau(t_1, t_2)\right) \tag{51}$$

for all nonnegative r, s, α, and β, as long as $\tau(t_1, t_2) \neq 0$, i.e. $t_1 \neq t_2$. Finally, combine inequalities (49)–(51) to verify the bound on the partial derivative,

$$h_{i,m-i} \geq \left|\frac{\partial^m \left(f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\right)}{\partial t_1^i \, \partial t_2^{m-i}}\right|,$$

where $h_{i,m-i}$ is defined in equation (47). $\blacksquare$

Proof of Proposition 2. Start with $m = 1$. Using the functional form of $f(t; \alpha, \beta)$ in equation (3), the partial derivatives satisfy

$$\frac{\partial \phi(t_1, t_2)}{\partial t_i} = \frac{\iint \left(\frac{\beta^2}{2t_i^2} - \frac{3}{2t_i} - \frac{\alpha^2}{2}\right) f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\iint_{T^2} \iint f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta) \, dt_1 dt_2}$$

or

$$\frac{2t_i^2}{\phi(t_1, t_2)} \frac{\partial \phi(t_1, t_2)}{\partial t_i} = \mathbb{E}(\beta^2 | t_1, t_2) - 3t_i - t_i^2 \mathbb{E}(\alpha^2 | t_1, t_2),$$

where

$$\mathbb{E}(\alpha^2 | t_1, t_2) \equiv \iint \alpha^2 d\tilde{G}(\alpha, \beta | t_1, t_2) \text{ and } \mathbb{E}(\beta^2 | t_1, t_2) \equiv \iint \beta^2 d\tilde{G}(\alpha, \beta | t_1, t_2).$$

For any $t_1 \neq t_2$, we can solve these equations for these two expected values as functions of $\phi(t_1, t_2)$ and its first partial derivatives.

For higher moments, the approach is conceptually unchanged. First express the $(i, j)^{th}$ partial derivatives of $\phi(t_1, t_2)$ as

$$\frac{2^{i+j} t_1^{2i} t_2^{2j}}{\phi(t_1, t_2)} \frac{\partial^{i+j} \phi(t_1, t_2)}{\partial t_1^i \partial t_2^j} = \mathbb{E}\left((\beta^2 - \alpha^2 t_1^2)^i (\beta^2 - \alpha^2 t_2^2)^j | t_1, t_2\right) + v_{ij}(t_1, t_2)$$

$$= \sum_{x=0}^{i+j} \sum_{y=\max\{0, x-j\}}^{\min\{x,i\}} \frac{i! j! (-t_1)^y (-t_2)^{x-y} \mathbb{E}(\alpha^{2x} \beta^{2(i+j-x)} | t_1, t_2)}{y!(x-y)!(i-y)!(j-x+y)!} + v_{ij}(t_1, t_2), \quad (52)$$

where v_{ij} depends only on lower moments of the conditional expectation. The first line can be established by induction. Express $\frac{\partial^{i+j} \phi(t_1, t_2)}{\partial t_1^i \partial t_2^j}$ from the first line and differentiate with respect to t_1. One can realize that all terms except one contain conditional expected moments of order lower than $i + j$ and thus could be grouped into the term $v_{i+1,j}$. The only term of order $m + 1$ has a form $\mathbb{E}\left((\beta^2 - \alpha^2 t_1^2)^{i+1} (\beta^2 - \alpha^2 t_2^2)^j | t_1, t_2\right)$ which follows directly from the derivative of $f(t_1, \alpha, \beta)$ with respect to t_1. The second line of (52) follows from the first by expanding the power functions.

Now let $i = \{0, \ldots, m\}$ and $j = m - i$. As we vary i, equation (52) gives a system of $m + 1$ equations in the $m + 1$ m^{th} moments of the joint distribution of α^2 and β^2 among workers who find jobs at durations (t_1, t_2), as well as lower moments of the joint distribution. These functions are linearly independent, which we show by expressing them using an LU decomposition:

$$\begin{pmatrix} \frac{2^m t_1^{2m}}{\phi(t_1,t_2)} \frac{\partial^m \phi(t_1,t_2)}{\partial t_1^m} \\ \frac{2^m t_1^{2(m-1)} t_2^2}{\phi(t_1,t_2)} \frac{\partial^m \phi(t_1,t_2)}{\partial t_1^{m-1} \partial t_2} \\ \frac{2^m t_1^{2(m-2)} t_2^4}{\phi(t_1,t_2)} \frac{\partial^m \phi(t_1,t_2)}{\partial t_1^{m-2} \partial t_2^2} \\ \vdots \\ \frac{2^m t_2^{2m}}{\phi(t_1,t_2)} \frac{\partial^m \phi(t_1,t_2)}{\partial t_2^m} \end{pmatrix} = L(t_1, t_2) \cdot U(t_1, t_2) \cdot \begin{pmatrix} \mathbb{E}(\alpha^{2m} | t_1, t_2) \\ \mathbb{E}(\alpha^{2(m-1)} \beta^2 | t_1, t_2) \\ \mathbb{E}(\alpha^{2(m-2)} \beta^4 | t_1, t_2) \\ \vdots \\ \mathbb{E}(\beta^{2m} | t_1, t_2) \end{pmatrix} + v_m(t_1, t_2), \quad (53)$$

where $L(t_1, t_2)$ is a $(m + 1) \times (m + 1)$ lower triangular matrix with element $(i + 1, j + 1)$ equal to

$$L_{ij}(t_1, t_2) = \frac{(m-j)!}{(m-i)!(i-j)!} (-t_2)^{2(i-j)} (t_2^2 - t_1^2)^{j/2}$$

for $0 \leq j \leq i \leq m$ and $L_{ij}(t_1, t_2) = 0$ for $0 \leq i < j \leq m$; $U(t_1, t_2)$ is a $(m+1) \times (m+1)$ upper triangular matrix with element $(i+1, j+1)$ equal to

$$U_{ij}(t_1, t_2) = \frac{j!}{i!(j-i)!}(t_2^2 - t_1^2)^{i/2}$$

for $0 \leq i \leq j \leq m$ and $U_{ij}(t_1, t_2) = 0$ for $0 \leq j < i \leq m$; and $v_m(t_1, t_2)$ is a vector that depends only on $(m-1)^{st}$ and lower moments of the joint distribution, each of which we have found in previous steps.[13] It is easy to verify that the diagonal elements of L and U are nonzero if and only if $t_1 \neq t_2$. This proves that the mth moments of the joint distribution are uniquely determined by the mth and lower partial derivatives. The result follows by induction. ∎

Before proving Proposition 3, we state a preliminary lemma, which establishes sufficient conditions for the moments of a function of two variables to uniquely identify the function. Our proof of Proposition 3 shows that these conditions hold in our environment.

Lemma 3 *Let $\hat{G}(\alpha, \beta)$ denote the cumulative distribution of a pair of nonnegative random variables and let $\mathbb{E}(\alpha^{2i}\beta^{2j}) \equiv \iint \alpha^{2i}\beta^{2j} d\hat{G}(\alpha, \beta)$ denote its $(i,j)^{th}$ even moment. For any $m \in \{1, 2, \ldots\}$, define*

$$M_m = \max_{i=0,\ldots,m} \mathbb{E}(\alpha^{2i}\beta^{2(m-i)}). \tag{54}$$

Assume that

$$\lim_{m\to\infty} \frac{[M_m]^{\frac{1}{2m}}}{2m} = \lambda < \infty. \tag{55}$$

Then all the moments of the form $\mathbb{E}(\alpha^{2i}\beta^{2j})$, $(i,j) \in \{0, 1, \ldots\}^2$ uniquely determine $\hat{G}$.

Proof of Lemma 3. First recall the sufficient condition for uniqueness in the Hamburger moment problem. For a random variable $u \in \mathbb{R}$, its distribution is uniquely determined by its moments $\{\mathbb{E}[|u|^m]\}_{m=1}^{\infty}$ if the following condition holds:

$$\limsup_{m\to\infty} \frac{(\mathbb{E}[|u|^m])^{\frac{1}{m}}}{m} \equiv \lambda' < \infty, \tag{56}$$

---

[13]If $t_2 > t_1$, the elements of L and U are real, while if $t_1 > t_2$, some elements are imaginary. Nevertheless, $L.U$ is always a real matrix. Moreover, we can write a similar real-valued LU decomposition for the case where $t_1 > t_2$. Alternatively, we can observe that $\tilde{G}(\alpha, \beta | t_1, t_2) = \tilde{G}(\alpha, \beta | t_2, t_1)$ for all (t_1, t_2), and so we may without loss of generality assume $t_2 \geq t_1$ throughout this proof.

as shown in the Appendix of Feller (1966) chapter XV.4. We will, however, use an analogous condition but for even moments only,

$$\lim_{m \to \infty} \frac{\left(\mathbb{E}\left[u^{2m} \right] \right)^{\frac{1}{2m}}}{2m} \equiv \lambda < \infty. \tag{57}$$

Note that if condition (57) holds, then condition (56) holds as well. To prove this, assume that $\lambda' = \infty$ and $\lambda < \infty$. Then there must be an odd integer m which very large, and in particular

$$\frac{\left(\mathbb{E}\left[|u|^m \right] \right)^{\frac{1}{m}}}{m} > (1 + \varepsilon) \frac{m+1}{m} \lambda, \tag{58}$$

where $\varepsilon > 0$ is any number. For any positive number m, as shown in Loève (1977) Section 9.3.e', it holds that $\left(\mathbb{E}\left[|u|^m \right] \right)^{\frac{1}{m}} < \left(\mathbb{E}\left[|u|^{m+1} \right] \right)^{\frac{1}{m+1}}$, and thus

$$\frac{\left(\mathbb{E}\left[|u|^m \right] \right)^{\frac{1}{m}}}{m} \leq \frac{m+1}{m} \frac{\left(\mathbb{E}\left[|u|^{m+1} \right] \right)^{\frac{1}{m+1}}}{m+1} \leq \frac{m+1}{m} \lambda(1 + \varepsilon), \tag{59}$$

which is a contradiction with (58).

We combine this result with the Cramér-Wold theorem, stating that the distribution of a random vector, say (α, β), is determined by all its one-dimensional projections. In particular, the distribution of the sequence of random vectors (α_m, β_m) converges to the distribution of the random vector (α_*, β_*) if and only if the distribution of the scalar $x_1\alpha_m + x_2\beta_m$ converges to the distribution of the scalar $x_1\alpha_* + x_2\beta_*$ for all vectors $(x_1, x_2) \in \mathbb{R}^2$.

Thus we want to ensure that for any (x_1, x_2) the distribution of $(x_1\alpha + x_2\beta)$ is determined by its moments. For this we want to check the condition in Equation (57) for $u(x) \equiv (x_1\alpha + x_2\beta)$. We note that:

$$\mathbb{E}\left[u(x)^{2m} \right] = \mathbb{E}\left((x_1\alpha + x_2\beta)^{2m} \right) = \sum_{i=0}^{2m} \frac{2m!}{i!(2m-i)!} x_1^i x_2^{2m-i} \mathbb{E}\left(\alpha^{2i}\beta^{2(m-i)} \right)$$

$$\leq \sum_{i=0}^{2m} \frac{2m!}{i!(2m-i)!} |x_1|^i |x_2|^{2m-i} \mathbb{E}\left(\alpha^{2i}\beta^{2(m-i)} \right) \leq M_m \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} |x_1|^i |x_2|^{m-i}$$

$$= M_m \left(|x_1| + |x_2| \right)^{2m}$$

where we use the (α, β) are non-negative random variables, and where M_m is defined in equation (54) in the statement of lemma 3.

Now we check that the limit in equation (57) is satisfied given our the assumptions in equation (54) and equation (55), i.e.:

$$\frac{\left(\mathbb{E}\left[u(x)^{2m}\right]\right)^{\frac{1}{2m}}}{2m} \leq \frac{\left[M_m\right]^{\frac{1}{2m}}}{2m}$$

Hence, since the distribution of each linear combination is determined, the joint distribution is determined. ∎

Proof of Proposition 3. Write the conditional moments as

$$\mathbb{E}(\alpha^{2i}\beta^{2(m-i)}|t_1, t_2) = \frac{\iint q(\alpha, \beta, i, m; t_1, t_2)dG(\alpha, \beta)}{\iint f(t_1; \alpha, \beta)f(t_2; \alpha, \beta)dG(\alpha, \beta)},$$

where

$$q(\alpha, \beta, i, m; t_1, t_2) \equiv \alpha^{2i}\beta^{2(m-i)}f(t_1; \alpha, \beta)f(t_2; \alpha, \beta).$$

Using the definition of f, we have

$$q(\alpha, \beta, i, m; t_1, t_2) = \frac{\alpha^{2i}\beta^{2(m+1-i)}}{2\pi t_1^{3/2}t_2^{3/2}}\exp\left(-\frac{(\alpha t_1 - \beta)^2}{2t_1} - \frac{(\alpha t_2 - \beta)^2}{2t_2}\right)$$

$$\leq \frac{1}{2\pi t_1^{3/2}t_2^{3/2}}\left(\frac{m+1}{\tau(t_1, t_2)}\right)^{m+1}\exp(-(m+1)),$$

where $\tau(t_1, t_2)$ is defined in equation (48) and the inequality uses the same steps as Proposition 1 to bound the function. In the language of Lemma 3, this implies

$$M_m = \frac{((m+1)/\tau(t_1, t_2))^{m+1}\exp(-(m+1))}{2\pi t_1^{3/2}t_2^{3/2}\iint f(t_1; \alpha, \beta)f(t_2; \alpha, \beta)dG(\alpha, \beta)}. \tag{60}$$

We use this to verify condition (55) in Lemma 3.

Taking the log transformation of $(1/2m)(M_m)^{1/2m}$ and using the expression (60) we get:

$$\log\left(\frac{[M_m]^{\frac{1}{2m}}}{2m}\right) = \frac{1}{2m}\varphi(t_1, t_2) - \frac{1+m}{2m}\log(\tau(t_1, t_2))$$

$$+ \frac{1+m}{2m}\log(m+1) - \frac{1+m}{2m} - \log(m)$$

where φ is independent of m. We argue that the limit of this expression as $m \to \infty$ diverges to $-\infty$, or that $(1/2m)(M_m)^{1/2m} \to 0$ as $m \to \infty$. To see this

$$\log\left(\frac{[M_m]^{\frac{1}{2m}}}{2m}\right) = \frac{1}{2m}\varphi(t_1, t_2) - \frac{1+m}{2m}\log\left(\tau(t_1, t_2)\right)$$

$$+ \frac{1}{2m}\log(m+1) + \frac{1}{2}\left[\log(m+1) - \log(m)\right] - \frac{1+m}{2m} - \frac{1}{2}\log(m)$$

Note that $\log(1+m) \le \log(m) + 1/m$ and $\log(1+m) \le m$ thus taking limits we obtain the desired result. $\blacksquare$

C Identification with One Spell

Special cases of our model are identified with one spell. We discuss two of them. First, we consider an economy where every worker has the same expected duration of unemployment $1/\mu$. Second, we consider the case of no switching costs $\psi_e = \psi_n = 0$. These special cases reduce the dimensionality of the unknown parameters. In the first case, the distribution of α is just a scaled version of the distribution of β. In the second case, the distribution of $\beta = 0$ and we are after recovering the distribution of α.

C.1 Identifying the Distribution of β with a Fixed μ

Consider the case where every individual has the same expected unemployment duration and thus the same value of μ_n, $\mu_n^i = \mu_n$ for all i, and σ_n is distributed according to some non-degenerate distribution. In our notation, we have that $\alpha = \mu_n \beta$ for some fixed μ_n and β is distributed according to $g(\beta)$. We argue that we can identify μ_n and all moments of the distribution g from data on one spell. The distribution of spells in the population is then given by

$$\phi(t) = \int f(t; \mu_n\beta, \beta)\, g(\beta)\, d\beta. \tag{61}$$

Since the expected duration is $1/\mu$,

$$\frac{1}{\mu_n} = \int_0^\infty t f\left(t; \mu_n\beta, \beta\right) g\left(\beta\right) d\beta dt = \int_0^\infty t\phi(t)dt,$$

which we can use to identify μ_n.

Let's now identify the moments of g. Our approach is based on relating the kth moment of the distribution $\phi(t)$ to the expected values of β^{2k}. Let $M(k)$ and $m(k, \mu_n\beta, \beta)$ be the kth

moment of the distribution $\phi(t)$ and $f(t; \mu_n\beta, \beta)$, respectively,

$$m(k, \mu_n\beta, \beta) \equiv \int_0^\infty t^k f(t; \mu_n\beta, \beta)dt \tag{62}$$

$$M(k) \equiv \int_0^\infty t^k \phi(t)dt = \int \left[\int_0^\infty t^k f(t; \mu_n\beta, \beta)\, dt \right] g(\beta)\, d\beta \tag{63}$$

$$= \int m(k, \mu_n\beta, \beta)g(\beta)\, d\beta \tag{64}$$

Lemeshko, Lemeshko, Akushkina, Nikulin, and Saaidia (2010) show that the kth moment of the inverse Gaussian distribution $m(k, \alpha, \beta)$ can be written as

$$m(k, \alpha, \beta) = \left(\frac{\beta}{\alpha} \right)^k \sum_{i=0}^{k-1} \frac{(k-1+i)!}{i!\,(k-1-i)!} (2\alpha\beta)^{-i}.$$

Specialize it to our case with $\alpha = \mu_n\beta$ to get

$$m(k, \mu_n\beta, \beta) = \sum_{i=0}^{k-1} a(k, i, \mu_n)\, \beta^{-2i}$$

$$a(k, i, \mu_n) \equiv 2^{-i} \frac{(k-1+i)!}{i!\,(k-1-i)!} \left(\frac{1}{\mu_n} \right)^{k+i}$$

Then the kth moment of the distribution ϕ is

$$M(k) = \int \sum_{i=0}^{k-1} a(k, i, \mu_n)\, \beta^{-2i} g(\beta)\, d\beta$$

$$= \sum_{i=0}^{k-1} a(k, i, \mu_n)\, \mathbb{E}\left[\beta^{-2i} \right] \tag{65}$$

Note that since μ_n is known, the values of $a(k, i, \mu_n)$ are known for all $k, i \geq 0$. For $k = 2$, equation (65) can be solved to find $\mathbb{E}\left[\beta^{-2}\right]$. By induction, if $\mathbb{E}\left[\beta^{-2i}\right]$ are known for $i = 1, \ldots, k-1$, then equation (65) for $M(k)$ can be used to find $\mathbb{E}\left[\beta^{-2k}\right]$.

C.2 The Case of Zero Switching Costs

Consider now the special case of no switching costs, $\psi_e = \psi_n = 0$. The region of inaction is degenerate, $\bar{\omega} = \underline{\omega}$ and hence $\beta = 0$. The distribution of spells for any given type is described by a single parameter α distributed according to density $g(\alpha)$. For any given α,

the distribution of spells is again given by the inverse Gaussian distribution

$$f(t; \alpha, 0) = \frac{1}{\sigma_n \sqrt{2\pi} t^{3/2}} \exp\left(-\frac{1}{2}\alpha^2 t\right). \tag{66}$$

and thus the distribution of spells in the population is

$$\phi(t) = \int f(t; \alpha, 0)\, g(\alpha) d\alpha. \tag{67}$$

We argue that the derivatives of ϕ can be used to identify even moments of the distribution of g.

Let start by deriving the kth derivative of $f(t; \alpha, 0)$. Use the Leibniz formula for the derivative of a product to get

$$\frac{\partial^m f(t; \alpha, 0)}{\partial t^m} = \frac{1}{\sqrt{2\pi}} \sum_{s=0}^{m} \binom{m}{s} \frac{\partial^s}{\partial t^s}\left(t^{-3/2}\right) \frac{\partial^{m-s}}{\partial t^{m-s}} \exp\left(-\frac{1}{2}\alpha^2 t\right).$$

Observe that

$$\frac{\partial^s}{\partial t^s}\left(t^{-3/2}\right) = t^{-3/2} \prod_{i=0}^{s}\left(-\frac{3}{2} - i\right)$$

$$\frac{\partial^{m-s}}{\partial t^{m-s}} \exp\left(-\frac{1}{2}\alpha^2 t\right) = \exp\left(-\frac{1}{2}\alpha^2 t\right)\left(-\frac{1}{2}\alpha^2\right)^{r-s},$$

and thus we can write an equation for the mth derivative of ϕ,

$$\frac{\partial^m \phi(t)}{\partial t^m} = \int \frac{\partial^m f(t; \alpha, 0)}{\partial t^m} g(\alpha)\, d\alpha$$

$$= \int f(t; \alpha, 0) \sum_{s=0}^{m} \binom{m}{s} t^{-s} \prod_{i=0}^{s}\left(-\frac{3}{2} - i\right)\left(-\frac{1}{2}\alpha^2\right)^{r-s} g(\alpha)\, d\alpha.$$

Finally, rearrange the terms

$$\frac{\partial^m \phi(t)}{\partial t^m} = \sum_{s=0}^{m} \binom{m}{s} t^{-s} \prod_{i=0}^{s}\left(-\frac{3}{2} - i\right)\left(-\frac{1}{2}\right)^{r-s} \mathbb{E}\left[(\alpha^2)^{r-s} | t\right],$$

to find the mth derivative of ϕ as a sum of mth and lower moments of α^2.

D   Deriving Distribution $\bar{G}$

We want to find a distribution of (α, β) which maximizes the fraction of workers with negative α subject to a constraint that the distribution is consistent with the distribution of completed spells ϕ and implies that $c = \hat{c}$. We start from the distribution G^+ estimated under assumption that $\alpha > 0$ for all types. For each type (α, β), we will turn a fraction $x(\alpha, \beta) \in [0, 1]$ of the type to $(-\alpha, \beta)$, and leave the fraction $(1 - x(\alpha, \beta))$ with a positive α.

As we already argued before, in order for a distribution with negative types to be consistent with ϕ, the share of $(-\alpha, \beta)$ types has to be proportional to $e^{4\alpha\beta}g^+(\alpha, \beta)$. Thus, our new distribution is of a form

$$g(-\alpha, \beta) = \frac{(1 - x(\alpha, \beta))\, e^{4\alpha\beta} dG^+(\alpha, \beta)}{Z}, \quad g(\alpha, \beta) = \frac{x(\alpha, \beta)\, dG^+(\alpha, \beta)}{Z},$$

where Z is the defined as

$$Z \equiv \int \left(1 - x(\alpha, \beta) + x(\alpha, \beta)\, e^{4\alpha\beta}\right) dG^+(\alpha, \beta),$$

and guarantees that the density integrates to one.

We wish to solve the following maximization problem

$$\max_{\{x(\alpha,\beta)\}} \int \frac{(1 - x(\alpha, \beta))\, e^{4\alpha\beta} dG^+(\alpha, \beta)}{Z}$$

$$s.t. \quad \hat{c} \leq \frac{1}{Z} \int \left[g(\alpha, \beta) F^2(T, \alpha, \beta) + g(-\alpha, \beta) F^2(T, -\alpha, \beta)\right] d(\alpha, \beta)$$

$$x(\alpha, \beta) \geq 0 \quad \forall (\alpha, \beta)$$

$$x(\alpha, \beta) \leq 1 \quad \forall (\alpha, \beta)$$

After some algebra, it can be shown that the constraint simplifies to

$$\int \left[g(\alpha, \beta) F^2(T, \alpha, \beta) + g(-\alpha, \beta) F^2(T, -\alpha, \beta)\right] d(\alpha, \beta) = \frac{1}{Z} \int \left[g^+(\alpha, \beta) F^2(T, \alpha, \beta)\right] d(\alpha, \beta) = \frac{\bar{c}}{Z}$$

where $\bar{c}$ is defined as in the main text, the upper bound.

So we have a maximization problem

$$\max_{\{x(\alpha,\beta)\}} \int \frac{(1 - x(\alpha,\beta)) e^{4\alpha\beta} g^+ (\alpha,\beta) \, d(\alpha,\beta)}{Z}$$

$$s.t. \quad Z \leq \frac{\bar{c}}{\hat{c}}$$

$$x(\alpha,\beta) \geq 0 \quad \forall (\alpha,\beta)$$

$$x(\alpha,\beta) \leq 1 \quad \forall (\alpha,\beta)$$

Recall that Z also depends on the choice of $x(\alpha,\beta)$. Let λ be the lagrange multiplier on the first constraint. Also, let $\lambda_0(\alpha,\beta)$ and $\lambda_1(\alpha,\beta)$ be the multipliers on the other two constraints, which now depend on (α,β). The first order condition is

$$\frac{\bar{c}}{\hat{c}} \cdot \frac{e^{4\alpha\beta}}{e^{4\alpha\beta} - 1} = \frac{1}{(\bar{c}/\hat{c})^2} Z - \lambda - (\lambda_0(\alpha,\beta) - \lambda_1(\alpha,\beta)),$$

which implies that we have a cutoff rule. In particular, there will be only one type for which $\lambda_0(\alpha,\beta) = \lambda_1(\alpha,\beta) = 0$ and the $0 < x(\alpha,\beta) < 1$. The FOC in this case becomes

$$\frac{\bar{c}}{\hat{c}} \cdot \frac{e^{4\alpha\beta}}{e^{4\alpha\beta} - 1} = \frac{1}{(\bar{c}/\hat{c})^2} Z - \lambda,$$

and since the right-hand side of this equation does not depend on α, β but the left-hand side does, there can be only one such a type. For all other types will have a corner solution, with either $x(\alpha,\beta) = 0$ or $x(\alpha,\beta) = 1$. Since the left-hand side of the first order condition is decreasing in $\alpha\beta$, it will be the case that for $\alpha\beta$ larger than the cutoff, $\lambda_0(\alpha,\beta) > 0, \lambda_1(\alpha,\beta) = 0$ and $x(\alpha,\beta) = 0$ while for $\alpha\beta$ smaller than the cutoff, $\lambda_1(\alpha,\beta) > 0, \lambda_0(\alpha,\beta) = 0$ and $x(\alpha,\beta) = 1$.

E   Power of the First Moment Test

We consider two interesting specification of the data generating mechanism which fail our test for the first moments of (α^2, β^2). Both cases are elaborations around examples introduced in Section 3.4. In both cases we obtain that if the data is generated according to these models, the test for $\mathbb{E}[\alpha^2|t1, t_2]$ fails for $t_2 = 0$ and t_1 sufficiently small. We also note that the first example has the property that ϕ is not differentiable at points where $t_1 = t_2$.

First, consider an extension of a canonical search model where an unemployed individual starts actively searching for a job only after τ periods, after which she finds a job at the rate θ. Thus, the hazard rate of exiting unemployment is zero for $t \leq \tau$, and θ for $t \geq \tau$. Each

worker is thus described by a pair (θ, τ), which are distributed in the population according to cumulative distribution function G. The joint density of the two spells is given by:

$$\phi(t_1, t_2) = \int_0^\infty \int_0^{\min\{t_1, t_2\}} \theta^2 \, e^{-\theta \, (t_1 + t_2 - 2\tau)} \, dG(\theta, \tau)$$

Suppose we apply our test to this model. If $t_1 > t_2$, then

$$\mathbb{E}(\alpha^2 | t_1, t_2) = \frac{2 \, t_2^2}{t_1^2 - t_2^2} \frac{\int \theta^2 e^{-\theta(t_1 - t_2)} dG(\theta|t_2)}{\int \theta^2 e^{-\theta(t_1 + t_2 - 2\tau)} dG(\theta, \tau)} + 2 \frac{\int \theta^3 e^{-\theta(t_1 + t_2 - 2\tau)} dG(\theta, \tau)}{\int \theta^2 e^{-\theta(t_1 + t_2 - 2\tau)} dG(\theta, \tau)} - \frac{3}{t_1 + t_2},$$

$$\mathbb{E}(\beta^2 | t_1, t_2) = t_1 t_2 \left(\frac{2 t_1 t_2}{(t_1^2 - t_2^2)} \frac{\int \theta^2 e^{-\theta(t_1 - t_2)} dG(\theta|t_2)}{\int \theta^2 e^{-\theta(t_1 + t_2 - 2\tau)} dG(\theta, \tau)} + \frac{3}{t_1 + t_2} \right) \geq 0.$$

Assume that the following regularity conditions hold:

$$\frac{\int \theta^3 e^{-\theta(t_1 - 2\tau)} d\tilde{G}(\theta|t_2)}{\int \theta^2 e^{-\theta(t_1 - 2\tau)} d\tilde{G}(\theta, \tau)} < \infty \quad \text{and} \quad \frac{\int \theta^2 e^{-\theta t_1} d\tilde{G}(\theta|0)}{\int \theta^2 e^{-\theta(t_1 - 2\tau)} d\tilde{G}(\theta, \tau)} < \infty$$

Setting $t_2 = 0$, the term $\mathbb{E}(\alpha^2 | t_1, t_2)$ becomes

$$\mathbb{E}(\alpha^2 | t_1, t_2) = 2 \frac{\int \theta^3 e^{-\theta(t_1 - 2\tau)} dG(\theta, \tau)}{\int \theta^2 e^{-\theta(t_1 - 2\tau)} dG(\theta, \tau)} - \frac{3}{t_1}.$$

For t_1 small enough, the negative term $\frac{3}{t_1}$ will dominate and the test fails, i.e. $\mathbb{E}(\alpha^2 | t_1, 0) < 0$.

The second example is a version of the multiplicative hazard rate model, with a baseline hazard rate $h(t)$ and a multiplicative constant θ distributed according to G. The distribution of two spells t_1, t_2 is given by

$$\phi(t_1, t_2) = \int_0^\infty \theta^2 \, h(t_1) \, h(t_2) \, e^{-\theta \left(\int_0^{t_1} h(s) ds + \int_0^{t_2} h(s) ds \right)} \, dG(\theta) \tag{68}$$

Differentiate with respect to i^{th} spell

$$\phi_i(t_1, t_2) = \int_0^\infty \left[\frac{h'(t_i)}{h(t_i)} - \theta \, h(t_i) \right] \theta^2 \, h(t_1) \, h(t_2) \, e^{-\theta \left(\int_0^{t_1} h(s) ds + \int_0^{t_2} h(s) ds \right)} \, dG(\theta)$$

and thus:

$$\frac{\phi_i(t_1, t_2)}{\phi(t_1, t_2)} = \frac{h'(t_i)}{h(t_i)} - h(t_i) \, \mathbb{E}\left[\theta \mid t_1, t_2 \right] \text{ where}$$

$$\mathbb{E}\left[\theta \mid t_1, t_2 \right] \equiv \frac{\int_0^\infty \theta^3 \, h(t_1) \, h(t_2) \, e^{-\theta \left(\int_0^{t_1} h(s) ds + \int_0^{t_2} h(s) ds \right)} \, dG(\theta)}{\int_0^\infty \theta^2 \, h(t_1) \, h(t_2) \, e^{-\theta \left(\int_0^{t_1} h(s) ds + \int_0^{t_2} h(s) ds \right)} \, dG(\theta)}$$

We thus have:

$$\mathbb{E}(\alpha^2|t_1,t_2) = \frac{2}{t_1^2 - t_2^2}\left[t_2^2\frac{h'(t_2)}{h(t_2)} - t_1^2\frac{h'(t_1)}{h(t_1)} + \mathbb{E}\left[\theta\,|\,t_1,t_2\right]\left[t_1^2\,h(t_1) - t_2^2\,h(t_2)\right]\right] - \frac{3}{t_1 + t_2},$$

$$\mathbb{E}(\beta^2|t_1,t_2) = t_1 t_2\left[\frac{2\,t_1 t_2}{t_1^2 - t_2^2}\left(\frac{h'(t_2)}{h(t_2)} - \frac{h'(t_1)}{h(t_1)} + \mathbb{E}\left[\theta\,|\,t_1,t_2\right]\left[h(t_1) - h(t_2)\right]\right) + \frac{3}{t_1 + t_2}\right] \geq 0.$$

Assume that the baseline hazard rate h is bounded and has bounded derivative around $t = 0$, so that $|h'(t_1)/h(t_1)| \leq b$ and $|h(t_1)| < B$ for two constants B, b. Set $t_2 = 0$ in which case we have:

$$\mathbb{E}(\alpha^2|t_1,0) = 2\left[-\frac{h'(t_1)}{h(t_1)} + \mathbb{E}\left[\theta\,|\,t_1,0\right]h(t_1)\right] - \frac{3}{t_1} \geq 0$$

Then the test fails, i.e. $\mathbb{E}(\alpha^2|t_1,0) < 0$, for t_1 small enough because the negative term $-3/t_1$ will dominate.

F   Austrian Data

During the analyzed period, the business cycle in Austria is very mild and the labor market does not respond to it. To document this, we plot the mean of the in progress unemployment spells, conditional on them being shorter than 5 years. We make this restriction for two reasons. First, it is the same restriction as we impose on the data which we use for estimation. Second, looking at the unconditional mean of in progress unemployment spell would show a hump-shaped profile, simply due to sampling scheme. Our data start in 1986 hence we start the measurement in year 1991.

G   Multidimensional Smoothing

We start with a data set that defines the density on a subset of the nonnegative integers, say $\psi : \{0, 1, \ldots, T\}^2 \mapsto \mathbb{R}$. We treat this data set as the sum of two terms, $\psi(t_1, t_2) \equiv \bar\psi(t_1, t_2) + \tilde\psi(t_1, t_2)$, where $\bar\psi$ is a smooth "trend" and $\tilde\psi$ is the residual. According to our model, the trend is smooth except possibly at points with $t_1 = t_2$ (Proposition 1). We therefore define a separate trend on each side of this "diagonal."

The spirit of our definition of the trend follows Hodrick and Prescott (1997), but extended to a two dimensional space. For any value of the smoothing parameter λ, we find $\bar\psi(t_1, t_2)$
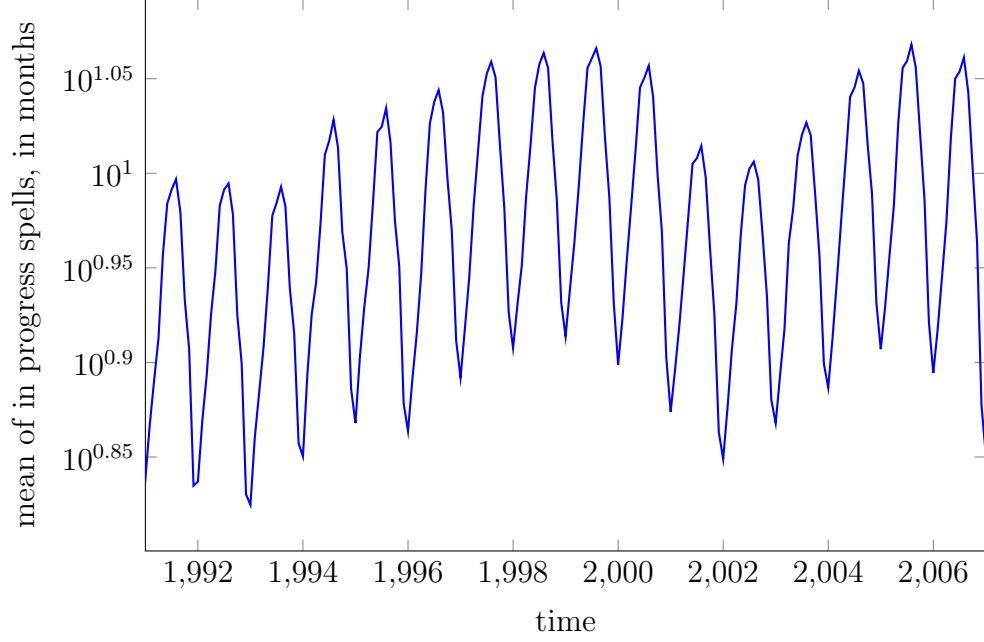
Figure 11: Mean in-progress unemployment spells.

at $t_2 \geq t_1$ to solve

$$\min_{\{\bar{\psi}(t_1,t_2)\}} \left(\sum_{t_1=1}^{T} \sum_{t_2=t_1}^{T} (\psi(t_1,t_2) - \bar{\psi}(t_1,t_2))^2 + \right.$$

$$\lambda \sum_{t_2=3}^{T} \sum_{t_1=2}^{t_2-1} (\bar{\psi}(t_1+1,t_2) - 2\bar{\psi}(t_1,t_2) + \bar{\psi}(t_1-1,t_2))^2 +$$

$$\left. \lambda \sum_{t_1=1}^{T-2} \sum_{t_2=t_1+1}^{T-1} (\bar{\psi}(t_1,t_2+1) - 2\bar{\psi}(t_1,t_2) + \bar{\psi}(t_1,t_2-1))^2 \right).$$

The first line penalizes the deviation between ψ and its trend at all points with $t_2 \geq t_1$. The remaining lines penalize changes in the trend along both dimensions, with weight λ attached to the penalty. If $\lambda = 0$, the trend is equal to the original series, while as λ converges to infinity, the trend is a plane in (t_1,t_2) space. More generally, the first order conditions to this problem define $\bar{\psi}$ as a linear function of ψ and so can be readily solved.

The optimization problem for (t_1,t_2) with $t_1 \leq t_2$ is analogous. If ψ is symmetric, $\psi(t_1,t_2) = \psi(t_2,t_1)$ for all (t_1,t_2), the trend will be symmetric as well.

H   Estimation

The link between the model and the data is given by equation (17). Our goal is to find the distribution $G(\alpha, \beta)$ using the data on the distribution of two non-employment spells. We do so in two steps. In the first step, we discretize equation (17) and solve it by minimizing the sum of squared errors between the data and the model-implied distribution of spells. In the second step, we refine these estimates by applying the expectation-maximization (EM) algorithm.

Each method has advantages and disadvantages. The advantage of the first step is that it is a global optimizer. The disadvantage is that we optimize only on a fixed grid for (α, β). The EM method does not require to specify bounds on the parameter space, but needs a good initial guess because it is a local method.

It also turns out that the maximum likelihood method suffers from two potential biases: one inherited from inverse Gaussian distribution, and one from working with discrete rather than continuous durations. We elaborate on these issues in our detailed discussion of the EM algorithm below.

H.1   Step 1: Minimum Distance Estimator

To discretize equation (17), we view $\phi(t_1, t_2)$ and $g(\alpha, \beta)$ as vectors in finite dimensional spaces. We consider a set $\mathbb{T} \subset \mathbb{R}_+^2$ of duration pairs (t_1, t_2), and refer to its typical elements as $(t_1(i), t_2(i)) \in \mathbb{T}$ with $i = 1, \ldots, I$. Guided by our data selection and the fact that the model is symmetric, we choose $\mathbb{T}$ to be the set of all integer pairs (t_1, t_2) satisfying $0 \leq t_1 \leq t_2 \leq 260$. We also replace $\phi(t_1, t_2)$ with the average of $\phi(t_1, t_2)$ and $\phi(t_2, t_1)$ to take advantage of the fact that our model is symmetric.

For the pairs of (α, β) we choose a set $\Theta \subset \mathbb{R}_+^2$ and again refer to its typical element $(\alpha(k), \beta(k)) \in \Theta$ with $k = 1, \ldots, K$.[14] The distribution of types is then represented by $g(k), k = 1, \ldots K$ such that $g(k) \geq 0$ and $\sum_{k=1}^{K} g(k) = 1$. Naturally, $\beta(k) > 0$ for all k. Given the limitation of our identification, we choose $\alpha(k) > 0$ for all k.

Equation (17) in the discretized form is

$$\phi = \frac{\mathbf{F}\,\mathbf{g}}{\mathbf{H}'\mathbf{g}},$$

where ϕ is a vector $\phi(t_1, t_2), (t_1, t_2) \in \mathbb{T}$, $\mathbf{g}$ is $K \times 1$ vector of $g(k)$, $\mathbf{F}$ is a $T \times K$ matrix with

---

[14]We experiment with different square grids on Θ both on equally spaced values on levels and in logs. We also set the grid in terms of $\sigma_n/(\bar{\omega} - \underline{\omega}) = 1/\beta$ and $\mu_n/(\bar{\omega} - \underline{\omega}) = \alpha/\beta$.

elements $F_{i,j}$, and $\mathbf{H}$ is a $K \times 1$ vector with element H_j defined below

$$F_{i,j} = f(t_1(i), \alpha(j), \beta(j)) f(t_2(i), \alpha(j), \beta(j))$$

$$H_j = \sum_{(t_1,t_2) \in \mathbb{T}^2} f(t_1, \alpha(j), \beta(j)) f(t_2, \alpha(j), \beta(j)).$$

In the first stage, we solve the minimization problem

$$\min_{\mathbf{g}} \left((\mathbf{F} - \phi\mathbf{H}') \, \mathbf{g} \right)' \left((\mathbf{F} - \phi\mathbf{H}') \, \mathbf{g} \right)$$

$$\text{s.t.} \quad \mathbf{g} \geq 0, \quad \sum_{k=1}^{K} g(k) = 1.$$

In practice, this problem is ill-posed. The kernel $f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)$ which maps the distribution $g(\alpha, \beta)$ into the joint distribution of spells $\phi(t_1, t_2)$ is very smooth, and dampens any high-frequency components of g. Thus, when solving the inverted problem of going from the data ϕ to the distribution g, high-frequency components of ϕ get amplified. This is particularly problematic when data are noisy, as is our case, since standard numerical methods lead to an extremely noisy estimate of g. Moreover, the solution of ill-posed problems is very sensitive to small perturbation in ϕ. In order to stabilize the solution and eliminate the noise, we do two things: first, we use smoothed rather than raw data as a vector ϕ, and second, we stabilize the solution by replacing $\tilde{\mathbf{F}} \equiv \mathbf{F} - \phi\mathbf{H}'$ with $\tilde{\mathbf{F}} + \lambda\mathbf{I}$ where λ is a parameter of choice. This effectively adds a penalty λ on the norm of g, and one minimizes $||\tilde{\mathbf{F}}\mathbf{g}||^2 + \lambda||\mathbf{g}||^2$ subject to the same constraints as above. We use the so-called L-curve to determine the optimal choice of λ.[15]

We apply the EM method in the second stage. This is an iterative method for finding maximum likelihood estimates of parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$

$$\log \ell(\mathbf{t}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}) = \sum_{(t_1,t_2) \in \mathbb{T}} \phi(t_1, t_2) \log \left(\sum_{k=1}^{K} \frac{f(t_1; \alpha(k), \beta(k)) f(t_2; \alpha(k), \beta(k))}{(1 - F(\bar{t}; \alpha(k), \beta(k)))^2} g(k) \right),$$

where $F(t; \alpha, \beta$ is the cumulative distribution function of the inverse Gaussian distribution with parameters α, β, and $\bar{t} = 260$ is the maximum measured duration. The mth iteration step of the EM has two parts. In the first part, the E step, we use estimates from $(m-1)^{st}$ iteration to calculate probabilities that i^{th} pair of spells $t_1(i), t_2(i)$ comes from each of the

---

[15]The L-curve is a graphical representation of the tradeoff between $|| (\mathbf{Fg} - \phi) ||^2$ and $|g||^2$. When plotted in the log-log scale, it has the L shape, hence its name. We choose value of λ which corresponds to the "corner" of the L-curve because it is a compromise between fitting the data and smoothing the solution.

type k. In the second part, the M step, we use these probabilities to find new values of $\alpha(k), \beta(k), g(k)$ from the first order conditions of the maximum likelihood problem.

H.2 Step 2: Maximum Likelihood using the EM algorithm

The EM algorithm is an iterative procedure to solve a maximum likelihood problem. To simplify the notation, denote data as $x_i = (t_{1i}, t_{2i})$, $i = 1, \ldots N$ and parameter $\theta_k = (\alpha_k, \beta_k)$ and g_k for $k = 1, \ldots K$. Also, let $\mathbf{x} = \{x_i\}_{i=1}^N$, $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$, $\mathbf{g} = \{g_k\}_{k=1}^K$. The likelihood is

$$l(\mathbf{x}; \boldsymbol{\theta}, \mathbf{g}) = \prod_{i=1}^N \left[\sum_{k=1}^K h(x_i, \theta_k) \, g_k \right]$$

where we use the following notation

$$h(x_i, \theta_k) \quad = \quad \frac{f(t_{1i}, \alpha_k, \beta_k) f(t_{2i}, \alpha_k, \beta_k)}{(F(\bar{t}, \alpha_k, \beta_k) - F(\underline{t}, \alpha_k, \beta_k))^2}.$$

Here, $\underline{t}$ and $\bar{t}$ are the bounds on t. In our case, $\underline{t} = 0$ and $\bar{t} = 260$. The log-likelihood is then given by

$$\log \ell(\mathbf{x}; \boldsymbol{\theta}, \mathbf{g}) = \sum_{i=1}^N \log \left(\sum_{k=1}^K h(x_i, \theta_k) \, g_k \right), \tag{69}$$

which we want to maximize by choosing $\boldsymbol{\theta}, \mathbf{g}$.

This problem has first order conditions:

$$0 = \frac{\partial \log \ell(x; \theta, g)}{\partial \theta_k} = \sum_{i=1}^N \frac{h(x_i, \theta_k) \, g_k}{\sum_{k'=1}^K h(x_i, \theta_{k'}) \, g_{k'}} \frac{\partial \log h(x_i, \theta_k)}{\partial \theta_k}$$

$$0 = \frac{\partial \log \ell(x; \theta, g)}{\partial g_k} = \sum_{i=1}^N \frac{h(x_i, \theta_k)}{\sum_{k'=1}^K h(x_i, \theta_{k'}) \, g_{k'}}$$

Define $z_{k,i}$ as the probability that the ith pair of spells comes from the type k, for all $i = 1, \ldots, N$ and $k = 1, \ldots, K$, as

$$z_{k,i}(x_i; \theta, g) \equiv \frac{h(x_i, \theta_k) \, g_k}{\sum_{k'=1}^K h(x_i, \theta_{k'}) \, g_k'}. \tag{70}$$

Notice that for all $i = 1, \ldots, N$, we have $\sum_{k=1}^K z_{k,i} = 1$. We can write the first order conditions

64

using z as follows:

$$0 = \sum_{i=1}^{N} z_{k,i}\left(x_i; \theta, g\right)\ \frac{\partial \log h(x_i, \theta_k)}{\partial \theta_k} \tag{71}$$

$$g_k = \frac{\sum_{i=1}^{N} z_{k,i}\left(x_i; \theta, g\right)}{\sum_{k'=1}^{K} \sum_{i=1}^{N} z_{k',i}\left(x_i; \theta, g\right)} \tag{72}$$

This is a system of $(3 + N)\, K$ equations in $(3 + N)\, K$ unknowns, namely $\{\alpha_k, \beta_k, g_k\}$ and $\{z_{k,i}\}$. These equations are not recursive; for instance $\mathbf{g}$ enters in all of them.

The EM algorithm is a way of computing the solution to the above system iteratively. It can be shown that this procedure converges to a local maximum of the log-likelihood function. Given $\{\boldsymbol{\theta}^m, \mathbf{g}^m\}$ we obtain new values $\{\boldsymbol{\theta}^{m+1}, \mathbf{g}^{m+1}\}$ as follows:

1. (E-step) For each $i = 1, .., N$ compute the weights $z_{k,i}^m$ as :

$$z_{k,i}^m = \frac{h(x_i, \theta_k^m)\, g_k^m}{\sum_{k'=1}^{K} h(x_i, \theta_{k'}^m)\, g_{k'}^m} \quad \text{for all } k = 1, ..., K\,. \tag{73}$$

2. (M-step) For each $k = 1, ..., K$ define θ_k^{m+1} as the solution to:

$$0 = \sum_{i=1}^{N} z_{k,i}^m\ \frac{\partial \log h(x_i, \theta_k^{m+1})}{\partial \theta_k}\,, \tag{74}$$

for all $k = 1, ..., K$.

3. (M-step) For each $k = 1, ..., K$ let g_k^{m+1} as :

$$g_k^{m+1} = \frac{\sum_{i=1}^{N} z_{k,i}^m}{\sum_{k'=1}^{K} \sum_{i=1}^{N} z_{k',i}^m}\,. \tag{75}$$

H.3 Potential Biases in ML Estimation

There are two biases in the maximum likelihood estimation, one related to estimation of μ, and one related to estimation of σ. These then lead to biases in estimation of α and β.

It is instructive to derive the maximum likelihood estimators for μ and σ in a simple case, where data on (single spell) duration $t(i), i = 1, \dots N$ come from an inverse Gaussian distribution. Straightforward algebra leads to

$$\hat{\mu}_{MLE} = \frac{1}{N} \sum_{i=1}^{N} t(i) = E[t], \quad \hat{\sigma}^2{}_{MLE} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{t(i)} - \frac{1}{N} \sum_{i=1}^{N} t(i) = E[\frac{1}{t}] - E[t]. \tag{76}$$
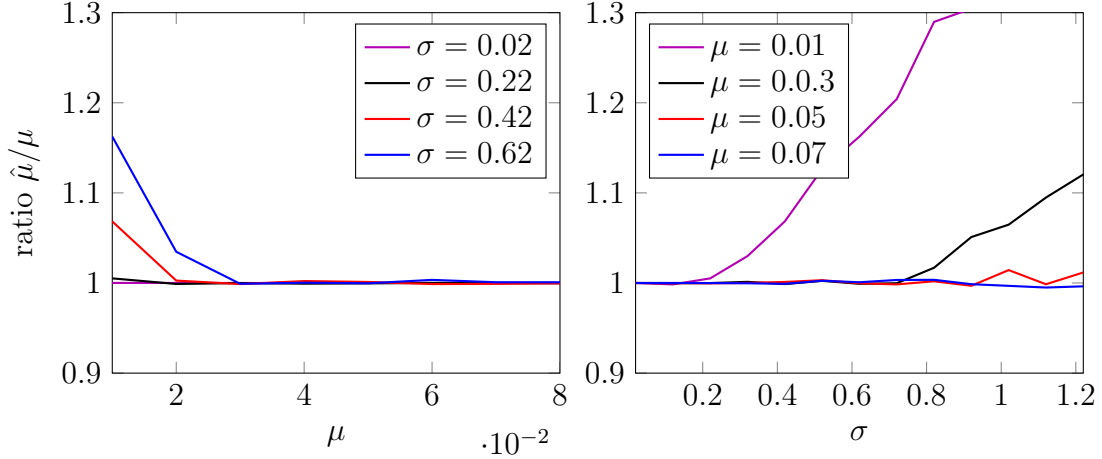
Figure 12: Maximum likelihood estimates of $\hat{\mu}$, relative to the true value of μ.

Notice that $E[t]$ and $E[1/t]$ are sufficient statistics.

The bias in μ is inherited from the inverse Gaussian distribution. In particular, it is very difficult to estimate μ precisely if μ is close to zero, which can be seen from the Fisher information matrix. This is given by, see for example Lemeshko, Lemeshko, Akushkina, Nikulin, and Saaidia (2010),

$$ I(\mu, \sigma) = \begin{pmatrix} \mu^3/\sigma^2 & 0 \\ 0 & \frac{1}{2}\sigma^4 \end{pmatrix} $$

and thus the lower bound on any unbiased estimate of μ is proportional to $1/\mu^3$. This diverges to infinity as μ approaches zero. Therefore, any estimate of μ, and thus also any estimate of α, will have a high variance for small μ (α). To illustrate this point, we generate 1,000,000 unemployment spells from a single inverse Gaussian distribution with parameters μ, σ, assuming that $\mu \in [0.01, 0.08]$ and $\sigma \in [0.02, 1.2]$. For different combinations of μ and σ, we find the maximum likelihood estimates $\hat{\mu}$ and $\hat{\sigma}$, and plot $\hat{\mu}$ relative to true value of μ in Figure 12. The left panel shows this ratio as a function of μ, the right panel as a function of σ. The estimate of μ has a high variance for small μ and thus is likely to be further away from the true value. This bias is somewhat worse for a larger value of σ, in line with the lower bound on the variance σ^2/μ^3, which is higher for smaller μ and larger σ.

To illustrate the performance of the ML estimator, we worked with continuous data. The real-world data differ from simulated in terms of measurement, as these can be measured only in discrete times. In particular, anybody with duration between, say 12 and 13 weeks, will be used in our estimation as having duration of 12.5 weeks. We study what bias this
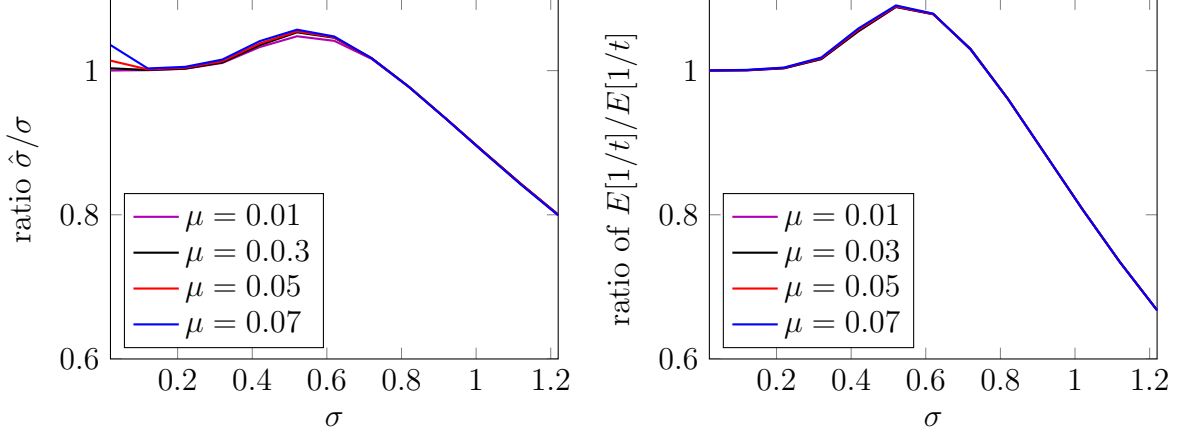
Figure 13: Maximum likelihood estimates of σ (left panel) and of the mean of $1/t$ (right panel) using discretized data, relative to their true values. The ratios are plotted as a function of σ, each line corresponds to one value of μ.

measurement introduces by by treating the simulated data as if they were measured in discrete times too. We find that this measurement affects estimates of σ, see the left panel of Figure 13, and the bias comes through the bias in estimating $E[1/t]$, see the right panel of the same figure. The bias in estimation of μ is small for values of $\sigma < 0.6$, which the range we estimate in the Austrian data. The magnitude of the bias for estimation of σ does not depend on the value of μ, but is also larger for larger values of σ, see the right panel of Figure 13. Discretization affects the mean of t only very mildly, and thus it does not affect estimation of μ. However, the mean of $1/t$ is sensitive to discretization. Since the mean of t is very similar for discretized and real values of t, this suggests that the distribution of spells between t and $t+1$ is not very different from symmetric. If this distribution was uniform, the bias in $E[1/t]$ can be mitigated by using a different estimator for $E[1/t]$. For example, noticing that $\log(t+1) - \log(t) = \int_t^{t+1} 1/t dt$, one can use the sample average of $\log(t+1) - \log(t)$ to measure $E[1/t]$. In practice, we find that this estimator reduces the bias in $E[1/t]$ if spells are measured at some starting duration $\underline{t}$ larger than 0, say 2 weeks. However, if spells are measured starting at zero, the bias is worse.