NBER WORKING PAPER SERIES

DECOMPOSING DURATION DEPENDENCE IN A STOPPING TIME MODEL

Fernando E. Alvarez
Katarína Borovičková
Robert Shimer

Working Paper 22188
http://www.nber.org/papers/w22188

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2016

We are grateful for comments from Jaap Abbring, Bo Honoré, Stephane Bonhomme, Jan Eberly, Rasmus Lentz, Andreas Mueller, Emi Nakamura, Pedro Portugal, Andrea Rotnitzky, Jon Steinsson, Quang Vuong, and Josef Zweimüller, as well as participants in numerous seminars. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w22188.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Fernando E. Alvarez, Katarína Borovičková, and Robert Shimer. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Decomposing Duration Dependence in a Stopping Time Model
Fernando E. Alvarez, Katarína Borovičková, and Robert
Shimer NBER Working Paper No. 22188
April 2016
JEL No. E24,J64

ABSTRACT

We develop a dynamic model of transitions in and out of employment. A worker finds a job at an optimal stopping time, when a Brownian motion with drift hits a barrier. This implies that the duration of each worker's jobless spells has an inverse Gaussian distribution. We allow for arbitrary heterogeneity across workers in the parameters of this distribution and prove that the distribution of these parameters is identified from the duration of two spells. We use social security data for Austrian workers to estimate the model. We conclude that dynamic selection is a critical source of duration dependence.

Fernando E. Alvarez
University of Chicago
Department of Economics
1126 East 59th Street
Chicago, IL 60637
and NBER
f-alvarez1@uchicago.edu

Robert Shimer
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
shimer@uchicago.edu

Katarína Borovičková
New York University
Department of Economics
19 West 4th Street
New York, NY 10003
kb103@nyu.edu

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 rate for each individual worker, and changes in the composition of workers at different non-employment durations (Cox, 1972). The goal of this paper is to explore a flexible but testable model of the job finding rate for any individual worker, allow for arbitrary heterogeneity across workers, and decompose these two factors. To do this, we develop a structural economic model which views finding a job as an optimal stopping problem for each worker. On top of this, we layer an arbitrary degree of individual heterogeneity. Thus, while the structure of our economic model is pinned down by economic theory, we treat heterogeneity flexibly, since here economic theory offers us no guidance. In the end, we attribute about twice as much of the observed decline in the job finding rate to changes in the composition of jobless workers at different durations, compared to what we would have obtained from a more standard statistical approach to the decomposition.

One interpretation of our structural model is a classical theory of employment. All individuals always have two options, working at some time-varying wage or not working and receiving some time-varying income and utility from leisure. These values are persistent but change over time. If there were no cost of switching employment status, an individual would work if and only if the wage were 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 non-employment income 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 that workers gradually accumulate skills while employed and lose them while out of work, as in Ljungqvist and Sargent (1998).

An alternative interpretation of our structural model is a classical theory of unemployment. According to this interpretation, a worker's productivity and her wage follow a stochastic process. Again, the difference is persistent but changes over time. If the worker is unemployed, a monopsonist has the option of paying a fixed cost to employ the worker, then earning flow profits equal to the difference between productivity and the wage. 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 preferences, fixed costs, and stochastic processes. For example, some individuals may expect their labor market productivity to increase the longer they stay out of work while others may expect it to fall. We maintain two key restrictions: for each 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, these thresholds are determined by the worker while in the second interpretation they are 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 we allow that they vary arbitrarily across individuals.

Given 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 completed 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 show how information on incomplete spells can help overcome this.

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 completed 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 multiplicative decomposition. With a mixed proportional hazard model (Lancaster, 1979), this decomposition would recover the baseline hazard, so this generalizes that well-known approach.

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, evaluate the decomposition, and bound the size of fixed costs. Using data on over 750,000 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 10 weeks, the hazard rate for the average individual increases until about 20 weeks and then declines by much less. Overall, changes in the composition of jobless workers accounts for about a 74 percent reduction in the hazard rate during the first two years of non-employment. We also estimate tiny fixed costs. For the median individual, the total cost of taking a job and later leaving it are approximately equal to six minutes of leisure time. As a result, the median newly employed worker leaves her job if she experiences a 1.6 percent drop in the wage. Previous work has shown that small fixed costs can generate large regions of inaction (Dixit, 1991; Abel and Eberly, 1994). We find that not only are the fixed costs small, but so is the region of inaction.

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-statistics researchers have so far not introduced unobserved individual heterogeneity into the model. These papers have also not explored 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. We also go beyond Abbring (2012) by confronting the model with real-world data.

Within economics, the mixed proportional hazard model 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. Relaxing the mixed proportional hazard assumption, which is feasible because of our large data set, is important for our finding that heterogeneity plays a critical role in the evolution of the hazard rate. We show that if we would impose the mixed proportional hazard assumption on our data, we would find that heterogeneity accounts for much less of the observed duration dependence.

Despite the difference in our conclusions, our work harkens back to an older literature on identification of the mixed proportional hazard model. Elbers and Ridder (1982) and Heckman and Singer (1984a) show that such a model is 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. Even closer to the spirit of our paper, Honoré (1993) shows that the mixed proportional hazard model is also identified with data on the duration of at least two non-employment spells for each individual.

Finally, some recent papers analyze duration dependence using models that are identified through assumptions on the extent of unobserved heterogeneity. For example, Krueger, Cramer, and Cho (2014) argue that observed heterogeneity is not important in accounting for duration dependence and so conclude that unobserved heterogeneity must also be unimportant. Hornstein (2012) and Ahn and Hamilton (2015) both assume there are two types of workers with different job finding hazards at all durations. We show that in our model and with our data set, identification through assumptions on the number of unobserved types leads us to understate the role of unobserved heterogeneity.

The remainder of the paper proceeds as follows. In Section 2, we describe our structural model, show that the model generates an inverse Gaussian distribution of duration for each worker, and discuss how we can use duration data to infer the magnitude of switching costs. Section 3 contains our main theoretical results on duration analysis. We prove that a subset of the parameters is identified if we observe at least two non-employment spells for each individual, discuss how information on incomplete spells can provide additional information that helps identify the model, and mention the limitations of any analysis that relies on single-spell data. In Section 4, we propose a multiplicative decomposition of the aggregate hazard rate into the portion attributable to structural duration dependence and the proportion attributable to heterogeneity. Section 5 summarizes the Austrian social security registry data. Section 6 presents our empirical results, including tests and estimates of the model, decomposition of hazard rates, comparison to the mixed proportional hazard model, and inference of the distribution of fixed costs. Finally, Section 7 briefly concludes.

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. The worker earns a wage $e^{w(t)}$ when employed and gets flow utility $e^{b(t)}$ when non-employed. Both $w(t)$ and $b(t)$ follow correlated Brownian motions with drift, but the drift and standard deviation of each may depend on the worker's employment status. If the worker is non-employed at t, she can become employed by paying a fixed cost $\psi_e e^{b(t)}$ for a constant $\psi_e \geq 0$. Likewise, the worker can switch from employment to non-employment 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 is convenient to define $\omega(t) \equiv w(t) - b(t)$, the worker's (log) net benefit from employment. 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 and $\mu_{s(t)}$ and $\sigma_{s(t)}$ are the drift and instantaneous standard deviation when the worker is in state $s(t)$.

Appendix A describes and solves the worker's problem fully. There we impose 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. We then prove that the worker's employment decision depends only on her employment status $s(t)$ and her net benefit from employment $\omega(t)$. 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.

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 $e^{b(t)}$, while a worker's productivity is $e^{w(t)}$. If the worker is employed by a monopsonist, it earns flow profits $e^{w(t)} - e^{b(t)}$. If the worker is unemployed, the 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 $\omega(t) > \bar{\omega}$.

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, $b(t) = \bar{b}$ for all t. We still allow the stochastic process for wages to depend on a worker's employment status. Then

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

An increase in the fixed costs 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.

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 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 wage and benefit, both while the worker is employed and while she is non-employed. We allow for arbitrary distributions of these structural parameters in the population, subject only to the constraint that the utility is finite.

2.2 Duration Distribution

We use the structural model to determine the distribution of non-employment duration for any single individual. All non-employment spells start when an employed worker's log net benefit from employment 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 it hits the upper threshold $\bar{\omega}$. Therefore the duration of a completed 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 at duration t

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

where $\alpha \equiv \mu_n/\sigma_n$ and $\beta \equiv (\bar{\omega} - \underline{\omega})/\sigma_n$. Hence, even though each worker is described by a large number of structural parameters, only two reduced-form parameters α and β determine how long a worker stays without a job. Note β is nonnegative by assumption, while α may be positive or negative. If α is nonnegative, $\int_0^\infty f(t;\alpha,\beta)dt = 1$, so a worker almost surely returns to work. But if α is negative, the probability of eventually returning to work is $e^{2\alpha\beta} < 1$, so there is a probability the worker has a *defective* spell and never finds a job. Thus a non-employed worker with α negative faces a risk of a severe form of long term non-employment, since with probability $1 - e^{2\alpha\beta}$ she stays forever non-employed.

The inverse Gaussian is flexible, but the model still imposes some restrictions on behavior. Figure 1 shows hazard rates for different values of α and β. It reveals that for the most part, β controls the shape of the hazard rate and α controls its level. Assuming β is strictly positive, 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 α is positive and 0 otherwise. If $\beta = 0$, the hazard rate is initially infinite and declines monotonically towards its long-run limit.

If α is positive, the expected duration of a completed non-employment spell is β/α and the variance of duration is β/α^3. As a spell progresses, the expected residual duration converges to $2/\alpha^2$, which may be bigger or smaller than the initial expected duration. The model is therefore consistent with both positive and negative duration dependence in the structural exit rate from non-employment.

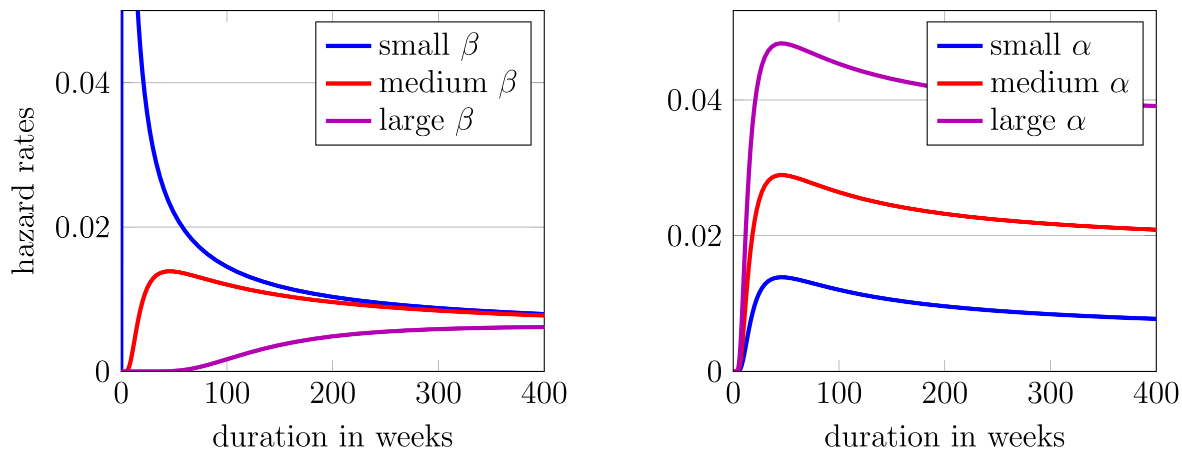

Figure 1: Hazard rates implied by the inverse Gaussian distribution for different values of $\alpha > 0$ and β. The left panel shows hazard rates for $\alpha = 0.1$ and three different values of β, 1, 10, and 30. The right panel shows hazard rates for three different values of α, 0.10, 0.18, and 0.27. We also adjust the value of β to keep the peak of the hazard rate at the same duration, which gives $\beta = 10$, 9.5, and 9.2, respectively.

Our analysis explicitly assumes that there is no time-varying heterogeneity beyond that captured by the model. 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, $s \in \{e, n\}$. However, our model does allow for learning-by-doing, since a worker's wage may increase faster on average when employed than when non-employed, $\mu_e > \mu_n$. Our maintained assumption is that the parameters α and β are constant for each worker throughout her lifetime.

On the other hand, we allow for an arbitrary population distribution of these parameters, which we label $G(\alpha, \beta)$. A major goal of this paper is to recover G from data. This heterogeneity exacerbates the structural duration dependence. Take two types of workers characterized by reduced-form parameters (α_1, β_1) and (α_2, β_2) and 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

Non-employment duration is determined by two reduced-form parameters, α and β, the distribution of which we will estimate. Although we will not recover the underlying structural parameters, we show in this section that we can use this distribution and a small amount of additional information to bound the magnitude of workers' switching costs.

We focus on the special case highlighted in equation (2), where the utility from non-employment is constant, $b(t) = \bar{b}$ for all t. Suppose we observe 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. Assuming that $\mu_e > 0$, we find that

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

Equation (4) expresses the fixed costs as a function of four parameters, μ_e, σ_e, μ_n, and r, as well as α and β.[1] Since the discount 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]

To back out the magnitude of switching costs, we need to choose values for the parameters μ_n, μ_e, σ_e, and r. Since we expect the estimated fixed costs to be small, our strategy will be to choose their values to make the fixed costs as large as possible while still staying within a range that can be supported empirically. We use the second part of equation (4) to guide our choice, since it tells us whether a given structural parameter increases or decreases fixed costs. In Section 6.6, we use estimated distribution of α and β to calculate the distribution of the fixed costs.

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 (13) is ill-behaved. Nevertheless, we can compute the density conditional on durations lying in some interval $[\underline{t}, \bar{t}]$:

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

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 viewed through the lens of our model, if we observe that the average duration of all spells with duration in the 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

[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 the ratio of the two functions converges to one as r converges to 0.

some individuals. We show below that this is the case in our data set.

3 Duration Analysis

This section examines how we can use duration data to evaluate this model. We start by showing that the joint distribution of the reduced-form parameters α and β, $G(\alpha, \beta)$, is identified using data on the completed duration of two spells for each individual and a sign restriction on the drift in the net benefit from employment while non-employed. We then show that incorporating information on the frequency of defective spells allows us to relax the sign restriction. We also show that the model is in fact overidentified and develop testable implications using data on the completed duration of two spells. Finally, we show that the model is identified with a single spell only under strong auxiliary assumptions, such as that there is a known, finite number of types in the population.

Our analysis builds on the structure of our economic model. We assume that each individual is characterized by a pair of parameters (α, β) and that density of each completed spell length is an inverse Gaussian $f(t; \alpha, \beta)$ for that individual. In particular, we impose that α and β are fixed over time for each individual, although the parameters may vary arbitrarily across individuals, reflecting some time-invariant observed or unobserved heterogeneity.

3.1 Intuition for Identification

Consider the following two data generating processes. In the first, there is a single type of worker (α, β), giving rise to the duration density $f(t; \alpha, \beta)$ in equation (3). In the second, there are 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 generate the same population duration density as the first data generating process. There is no way to distinguish these two data generating processes using a single non-employment spell.

With two completed spells for each individual, however, distinguishing these two models is trivial. In the first model without any heterogeneity, the duration of an individual's first spell tells us nothing about the duration of her second spell. In particular, the correlation between the durations of the two spells is zero. In the second model without any uncertainty in the duration of a spell for each individual, the duration of an individual's two spells is identical. In particular, the correlation between the durations of the two spells is one.

This simple example suggests that the distribution of the duration of the second spell conditional on the duration of the first spell should provide some information on the underlying type distribution. Our main result is that this information, together with some prior information about the sign of drift in the net benefit from employment while non-employed, identifies the type distribution.

3.2 Proof of Identification

Let $G(\alpha, \beta)$ denote the distribution of (α, β) in some population. Assume that all members of this population completed their first spell and started their second spell. The second spell may be completed or defective. The assumption that we observe (at least) two spells for each individual lies at the heart of our identification, and hence our results apply only to this population.

For some individuals in our population, the first two spells have duration $(t_1, t_2) \in T^2$, where $T \subseteq \mathbb{R}_+$ is a set with non-empty interior.[3] Let $\phi : T^2 \to \mathbb{R}_+$ denote the joint distribution of the durations for this subpopulation. According to the model,

$$\phi(t_1, t_2) = \frac{\int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\int_{T^2} \int f(t_1'; \alpha, \beta) f(t_2'; \alpha, \beta) dG(\alpha, \beta)\, d(t_1', t_2')}. \tag{5}$$

Our main identification result, Theorem 1 below, is that the joint density of spell lengths ϕ identifies the joint distribution of characteristics G if we know the sign of α, i.e. the drift in the net benefit from employment while non-employed.

We prove this result through a series of Propositions. The first shows that the partial derivatives of ϕ exist at all points where $t_1 \neq t_2$:

Proposition 1 *Take any $(t_1, t_2) \in T^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$. For any G, 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 $t_1 \neq t_2$, and indeed an example shows that this condition is indispensable:

Example 1 *Assume that β is distributed Pareto with parameter θ while $\alpha = d\beta$ for some constant d, equal to the common mean duration of all individuals' spells. Solving equation* (5)

[3] 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.

implies that the joint density of two spells is

$$\phi(t_1,t_2) = \frac{\theta\Delta^{\theta/2-1}}{4\pi t_1^{3/2}t_2^{3/2}}\Gamma\big(1-\theta/2,\Delta\big) \tag{6}$$

where $\Delta \equiv \frac{1}{2}\big(\frac{(t_1/d-1)^2}{t_1} + \frac{(t_2/d-1)^2}{t_2}\big)$ *and* $\Gamma(1-\theta/2,\Delta) \equiv \int_\Delta^\infty z^{-\theta/2}e^{-z}dz$ *is the incomplete Gamma function.*

When either $t_1 \neq d$ *or* $t_2 \neq d$ *or both,* Δ *is strictly positive and hence* $\phi(t_1,t_2)$ *is infinitely differentiable. But when* $t_1 = t_2 = d$, $\Delta = 0$ *and so both the Gamma function and* $\Delta^{\frac{\theta}{2}-1}$ *can either diverge or be non-differentiable. In particular, for* $\theta \in (0,2)$, $\lim_{t\to d}\phi(t,t) = \infty$*. For* $\theta \in [2,4)$*, the density is finite but non-differentiable at* $t_1 = t_2 = d$*. 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 when β *converges to infinity, the volatility of the Brownian motion vanishes, and thus the spells end with certainty at duration* d*. Equivalently the corresponding distribution tends to a Dirac measure concentrated at* $t_1 = t_2 = d$*. 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.*

For the next Proposition, we look at the conditional distribution of (α,β) among individuals whose two spells last exactly (t_1,t_2) periods:

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

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*

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

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 m^{th} step, the $m+1$ m^{th} partial derivatives and the results from the previous steps determine the $m+1$ m^{th} even moments of $\tilde{G}$.

In the third Proposition, 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; here we assume this is either always positive or always negative, although other assumptions would work. We show in Section 3.3 that the sign of α is not identified using completed spell data alone. 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 $(i,j) \in \{0,1,\dots\}^2$, 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 density 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 for any G, 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 $\tilde{G}(\cdot,\cdot|t_1,t_2)$. Intuitively, if one thinks of a Taylor expansion around (t_1,t_2), we are using the entire empirical density ϕ for $(t_1,t_2)\in T^2$ to recover the distribution function G.

We comment briefly on an alternative but ultimately unsuccessful proof strategy. Proposition 2 establishes that we can measure $\mathbb{E}(\alpha^{2i}\beta^{2j}|t_1,t_2)$ at almost all (t_1,t_2) and all i and j. It might seem we could therefore integrate the conditional moments using the density $\phi(t_1,t_2)$ to compute the unconditional $(2i,2j)^{\text{th}}$ moment of G. 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 $\alpha=d\beta$ for some constant d. The distribution G thus does not have all its moments. Nevertheless, we find that the conditional moments are well-defined:*

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

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

If $\Delta>0$, all conditional moments $M_m\equiv\mathbb{E}(\beta^m|t_1,t_2)$ exist and are finite. Moreover, the moments do not grow too fast, so we can use the D'Alembert criterium (see for example Theorem A.5 in Coelho, Alberto, and Grilo (2005)) to prove that the conditional moments uniquely determine the conditional distribution $\tilde{G}(\alpha,\beta|t_1,t_2)$. Finally, we can use Bayes rule to recover the unconditional distribution G, even though that distribution only has finitely many moments.

3.3 Share of Population with Negative Drift

Our theoretical model makes no predictions about the sign of the drift in the net benefit from employment while an individual is non-employed, i.e. about the sign of α. Completed spell data alone also cannot identify the sign of this reduced-form parameter. This is a consequence of the functional form of the inverse Gaussian distribution, which implies

$$f(t;\alpha,\beta)=e^{2\alpha\beta}f(t;-\alpha,\beta)$$

for all α, β, and t; see equation (3). Proportionality of $f(t;\alpha,\beta)$ and $f(t;-\alpha,\beta)$ implies that, once we condition on an individual having $n\geq 1$ completed spells, the probability distribution over completed durations $(t_1,\ldots,t_n)$ is the same if the individual is described by reduced-form parameters (α, β) or $(-\alpha, \beta)$.

On the other hand, the possibility that individuals have a negative drift in the net benefit from employment is economically important because it affects the hazard rate of exiting non-employment, particularly at long durations. This insight motivates our approach to identifying the distribution of the sign of α using data on incomplete spells. Recall that our population of interest consists of individuals with at least two spells, the first of which is completed within an interval T. We proceed in two steps. In the first step, we use ϕ, the distribution of the duration of two completed spells, together with the auxiliary assumption that $\alpha \geq 0$ with G-probability 1, in order to identify a candidate type distribution, which we call $G^+(\alpha, \beta)$. Theorem 1 tells us that this is feasible.

In the second step, we let c denote the fraction of the population whose second spell also has duration $t_2 \in T$. The candidate type distribution G^+ provides an upper bound on c:

$$\bar{c} \equiv \int_T \int f(t_2; \alpha, \beta) dG^+(\alpha, \beta)\, dt_2.$$

The model implies that $\bar{c} \geq c$, with equality if and only if the true type distribution is given by the candidate, $G = G^+$. Any other type distribution which gives rise to the same completed spell distribution ϕ must have some negative values of α and so must generate a smaller fraction of completed second spells.

We aim to construct a distribution of types which is consistent with both ϕ and c. To do this, take any individual of type (α, β), $\alpha > 0$, and replace her with $e^{4\alpha\beta}$ individuals of type $(-\alpha, \beta)$. These individuals have the same duration distribution conditional on a completed spell, but they have a lower probability of completing a spell. The choice of $e^{4\alpha\beta}$ ensures that the same number of individuals complete two spells. By flipping the sign of α for enough individuals, we generate the observed value of c from the model.

There are many ways of selecting individual types from G^+ for flipping the sign of α. We focus on two versions of this exercise, choosing alternatively the largest and smallest possible fraction of individuals. Assume $T = [0, \bar{t}]$ and define

$$p(\alpha, \beta) \equiv c(1 - e^{-4\alpha\beta}) - (e^{-2\alpha\beta} - e^{-4\alpha\beta})F(\bar{t}; \alpha, \beta).$$

We first choose individuals from G^+ with the smallest value of $p(\alpha, \beta)$, flip the sign of α for these individuals, and augment their share by $e^{4|\alpha|\beta}$, so as to keep the completed spell distribution unchanged. We do this until we achieve the desired value of the fraction of completed spells c.[4] We prove in Online Appendix OA.B that the resulting type distribution,

[4] In theory, we might flip the sign of everyone's drift without achieving the desired completed spell share c. In that case, we would reject the model.

which we label $\bar{G}$, has the largest fraction of individuals with negative drift consistent with our data. We then choose individuals with the largest value of $p(\alpha, \beta)$ to construct the distribution $\underline{G}$, which has the smallest fraction of individuals with negative drift. We view our model as partially identified, with $\bar{G}$ and $\underline{G}$ providing bounds on the numbers of individuals with negative drift.

3.4 Overidentifying Restrictions

This model has many overidentifying restrictions. First, Proposition 1 tells us that the joint density of two completed spells 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 $\mathbb{E}(\alpha^{2i}\beta^{2j}|t_1, t_2)$ for all $\{i, j\} \in \{0, 1, \ldots\}^2$. Even-powered moments must all be nonnegative, 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 $0 < i < j$. Any completed spell distribution ϕ that satisfies these three types of restrictions could have been generated by some type distribution G.

In practice, measuring higher moments can be difficult and so we focus on the simplest overidentifying test 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

$$\mathbb{E}(\alpha^2|t_1, t_2) = \frac{2\left(t_2^2 \frac{\partial\phi(t_1,t_2)}{\partial t_2} - t_1^2 \frac{\partial\phi(t_1,t_2)}{\partial t_1}\right)}{\phi(t_1, t_2)(t_1^2 - t_2^2)} - \frac{3}{t_1 + t_2} \geq 0 \tag{10}$$

$$\text{and } \mathbb{E}(\beta^2|t_1, t_2) = t_1 t_2 \left(\frac{2t_1 t_2 \left(\frac{\partial\phi(t_1,t_2)}{\partial t_2} - \frac{\partial\phi(t_1,t_2)}{\partial t_1}\right)}{\phi(t_1, t_2)(t_1^2 - t_2^2)} + \frac{3}{t_1 + t_2} \right) \geq 0. \tag{11}$$

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 $\phi(t_1, t_2) = \theta^2 e^{-\theta(t_1+t_2)}$. Then conditions* (10) *and* (11) *impose*

$$\mathbb{E}(\alpha^2|t_1,t_2) = 2\theta - \frac{3}{t_1+t_2} \geq 0 \text{ and } \mathbb{E}(\beta^2|t_1,t_2) = \frac{3t_1t_2}{t_1+t_2} \geq 0.$$

The first inequality is violated whenever $t_1 + t_2 < \frac{3}{2\theta}$. *Weighting this by the density* ϕ, *we find that* $1 - \frac{5}{2}e^{-3/2} \approx 44\%$ *of individuals experience these short durations. 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* G, *with some abuse of notation. The density of completed spells is* $\phi(t_1,t_2) = \int \theta^2 e^{-\theta(t_1+t_2)} dG(\theta)$. *Then*

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

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* 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* (10) *and* (11). *Then our approach implies*

$$\mathbb{E}(\alpha^2|t_1,t_2) = \frac{2}{\sigma^2(t_1+t_2)}\left(\frac{t_1\log t_1 - t_2\log t_2}{t_1-t_2} - \left(\mu + \tfrac{1}{2}\sigma^2\right)\right) \geq 0$$
$$\text{and } \mathbb{E}(\beta^2|t_1,t_2) = \frac{2t_1t_2}{\sigma^2(t_1+t_2)}\left(\frac{t_2\log t_1 - t_1\log t_2}{t_1-t_2} + \left(\mu + \tfrac{1}{2}\sigma^2\right)\right) \geq 0.$$

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 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,t_2), *converging to minus infinity when* t_1 *and* t_2 *are sufficiently large; therefore 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* (10) *and* (11) *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 *Suppose that each individual has a hazard rate equal to* $\theta\, h^b(t)$ *at times* $t \geq 0$*, where the baseline hazard* $h^b(\cdot)$ *is unrestricted and* θ *is an individual characteristic with distribution function again denoted by* G*. If* $h^b(t)$ *and* $|h^{b\prime}(t)/h^b(t)|$ *are both bounded as* t *converges to* 0*, the test implies* $\mathbb{E}(\alpha^2|t_1, t_2) < 0$ *for* (t_1, t_2) *sufficiently small. Online Appendix OA.C gives a more detailed description and proves this result.*

3.5 Single-Spell Data

The distribution of reduced-form parameters (α, β) is also identified using the duration of a single completed spell and auxiliary assumptions on the distribution function G. For example, we prove in Appendix C that the model is identified if every individual has the same expected duration $d = \beta/\alpha$. We also prove the model is identified if there are no switching costs, $\psi_e = \psi_n = 0$. Both of these economically-motivated restrictions reduce the unobserved type distribution to a single dimension.

Another approach would be to impose a known number of types. A finite mixture of inverse Gaussian distributions is identified by the distribution of the duration of a single spell. Conversely, a finite mixture model is sufficiently flexible so as to fit many realized single-spell duration distributions quite well. We show in Section 6.5 that a model with three types can capture the distribution of the duration of a single non-employment spell in our data set. The fundamental problem with this approach is that there is no good economic justification for the finite type assumption.[5] That is, the fact that we can fit the single-spell duration distribution does not imply that we have estimated an object of interest. We also show in Section 6.5 that estimates using single-spell data miss much of the heterogeneity that we uncover using multiple-spell data. Such estimates therefore understate the contribution of heterogeneity in explaining the shape of the hazard rate.

4 Decomposition of the Hazard Rate

Suppose we know the type distribution $G(\alpha, \beta)$. This section discusses how to use that information, together with the known functional form of the duration density $f(t; \alpha, \beta)$, to

[5] Heckman and Singer (1984b) pointed out a similar issue in the mixed proportional hazard model.

understand the relative importance of structural duration dependence and dynamic selection of heterogeneous individuals for the evolution of the hazard rate of exiting non-employment.

We propose a multiplicative decomposition of the aggregate hazard rate conditional on two completed spells into two components, one measuring a "structural" hazard rate and another measuring an "average type" among workers still nonemployed at a given duration. We interpret the structural hazard rate as the aggregate hazard rate that would prevail in the population if there were no heterogeneity.

The decomposition is based on a Divisia index. Let $h(t;\alpha,\beta)$ denote the hazard rate for type (α,β) at duration t,

$$h(t;\alpha,\beta) \equiv \frac{f(t;\alpha,\beta)}{1-F(t;\alpha,\beta)}, \tag{12}$$

where $F(t;\alpha,\beta)$ is the cumulative distribution function associated with the duration density f. With some abuse of notation, let $G(\alpha,\beta;t)$ denote the type distribution among individuals whose duration exceeds t periods. This depends on the initial type distribution and the functional form of the duration distribution for each type:

$$dG(\alpha,\beta;t) \equiv \frac{(1-F(t;\alpha,\beta))dG(\alpha,\beta)}{\int(1-F(t;\alpha',\beta'))dG(\alpha',\beta')}. \tag{13}$$

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

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

as can be confirmed directly from the definitions of $h(t;\alpha,\beta)$ and $dG(\alpha,\beta;t)$. For example, if a positive measure of individuals have $\alpha \leq 0$, the aggregate hazard rate converges to zero at sufficiently long durations, since those individuals dominate the population.

We propose an exact multiplicative decomposition of the aggregate hazard rate, $H(t) = H^s(t)H^h(t)$, where

$$H^s(t) \equiv H(0)e^{\int_0^t d\log H^s(t')} \text{ and } H^h(t) \equiv e^{\int_0^t d\log H^h(t')}$$

and

$$\frac{d\log H^s(t)}{dt} \equiv \frac{\int \dot{h}(t;\alpha,\beta)dG(\alpha,\beta;t)}{H(t)} \text{ and } \frac{d\log H^h(t)}{dt} \equiv \frac{\int h(t;\alpha,\beta)d\dot{G}(\alpha,\beta;t)}{H(t)}.$$

That this is an exact decomposition follows immediately from the product rule:

$$\frac{\dot{H}(t)}{H(t)} = \frac{\int \dot{h}(t;\alpha,\beta)dG(\alpha,\beta;t)}{H(t)} + \frac{\int h(t;\alpha,\beta)d\dot{G}(\alpha,\beta;t)}{H(t)} = \frac{d\log H^s(t)}{dt} + \frac{d\log H^h(t)}{dt}.$$

We interpret the term $H^s(t)$ as the contribution of structural duration dependence, since it is based on the change in the hazard rates of individual worker types. If each individual had a constant hazard rate, there would be no structural duration dependence, and so this term would be constant. The remaining term $H^h(t)$ represents the role of heterogeneity because it captures how the hazard rate changes due to changes in the distribution of workers in the non-employed population. We normalize this to equal 1 at duration 0.

One attractive feature of the multiplicative decomposition is that it nests the usual decomposition of the mixed proportional hazard model. That is, suppose it were the case that for each type (α, β), there is a constant θ such that $h(t;\alpha,\beta) \equiv \theta h^b(t)$ for some function $h^b(t)$. Normalizing the population average value of θ to one, our multiplicative decomposition would uncover that the structural hazard rate $H^s(t)$ is equal to the baseline hazard rate $h^b(t)$ and the heterogeneity portion $H^h(t)$ is equal to the average value of θ in the population of individuals whose spell lasts at least t periods.

The structural hazard rate $H^s(t)$ can either increase or decrease with duration, but the contribution of heterogeneity $H^h(t)$ is always decreasing with duration. It turns out that the change in the contribution of heterogeneity equals the minus the ratio of the cross-sectional variance and mean of the hazard rates:[6]

$$\frac{d\log H^h(t)}{dt} = -\frac{Var(h(t;\alpha,\beta))}{E(h(t;\alpha,\beta))} < 0. \tag{14}$$

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."[7] Intuitively, types with a higher than average hazard rate

[6]To prove this, first take logs and time-differentiate $dG(\alpha,\beta;t)$:

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

Substituting this result into the expression for $d\log H^h(t)/dt$ gives

$$\frac{d\log H^h(t)}{dt} = \frac{-\int h(t;\alpha,\beta)(h(t;\alpha,\beta)-H(t))dG(\alpha,\beta;t)}{H(t)}.$$

Since $\int(h(t;\alpha,\beta) - H(t))dG(\alpha,\beta;t) = 0$, we can add $H(t)$ times this to the numerator of the previous expression to get the formula in equation (14).

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

are always declining as a share of the population.

5 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 (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf, and Buchi, 2009). The data set covers the universe of private sector workers over the years 1986–2007. It contains information on individual's employment, registered unemployment, maternity leave, and retirement, with the exact begin and end date of each spell.[8]

5.1 Characteristics of the Austrian Labor Market

Austrian data are appropriate for our purposes. The Austrian labor market is flexible despite institutional regulations. 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 increases on the job, but do not directly restrict the hiring or firing decisions of employers. The main firing restriction is a severance payment, with size and eligibility determined by law. A worker becomes eligible for severance pay after three years of tenure if she does not quit voluntarily. The pay starts at two months salary and increases gradually with tenure. Depending on the details of how wages are set, this need not have any impact on employment or nonemployment duration (Lazear, 1990).

The unemployment insurance system in Austria is similar to the one in the United States. The potential duration of unemployment benefits depends on previous work history and age. If a worker has been employed for more than a year during the two years before a layoff, she is eligible for 20 weeks of the unemployment benefits. The potential duration of benefits increases to 30, 39, and 52 weeks for older workers with longer work histories.

Temporary separations and recalls are prevalent in Austria. Around 40 percent of nonemployment spells end with an individual returning to the previous employer. Our structural model already incorporates this possibility and so we do not break out recalls in our analysis.

Finally, we show in Online Appendix OA.D that the Austrian business cycle is moderate during this period. For example, the mean duration of in-progress non-employment spells varies very little over time. We therefore treat the Austrian labor market as a stationary environment and use pooled data for our analysis.

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

5.2 Sample Selection and Definition of Duration

Our data set contains the complete labor market histories of workers over a 21 year period, which allows us to observe multiple non-employment spells for many individuals. We use 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.[9]

A non-employment spell is incomplete 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 data set ends, 2) the worker retires, 3) the worker goes on a maternity leave, 4) the worker disappears from the sample or dies. We consider any of these as incomplete spells.

We consider only individuals who were younger than 46 in 1986 and older than 39 in 2007, and have at least one non-employment spell which started after the age of 25.[10] Imposing the age criteria guarantees that each individual has at least 15 years when she could potentially be at work. To estimate the distribution $G^+(\alpha, \beta)$, we will use information on two complete spells shorter than 260 weeks, which means that we are setting $T = [0, 260]$. This choice leads to further restrictions on the sample. We consider only spells which started before year 2003 to allow workers to have at least 260 weeks to complete their non-employment spell before the data set ends in 2007. Incomplete spells which end with retirement or death are included in our sample only if they are longer than 260 weeks.

We further restrict our population to workers who have at least two spells, with the first spell completed within 260 weeks and the second spell possibly incomplete. There are 751,125 individuals in this population. For estimating the type distribution G^+, we focus on the 704,794 individuals whose first two complete spells each have duration shorter than 260 weeks; however, we use the other workers to discipline the fraction of individuals with a negative drift in the net benefit from employment. Our final sample contains 59 percent of all workers who were younger than 46 in 1986, older than 39 in 2007, and started a non-employment spell between 1986 and 2003.

[9] 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.

[10] 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 $T = [0, 260]$.

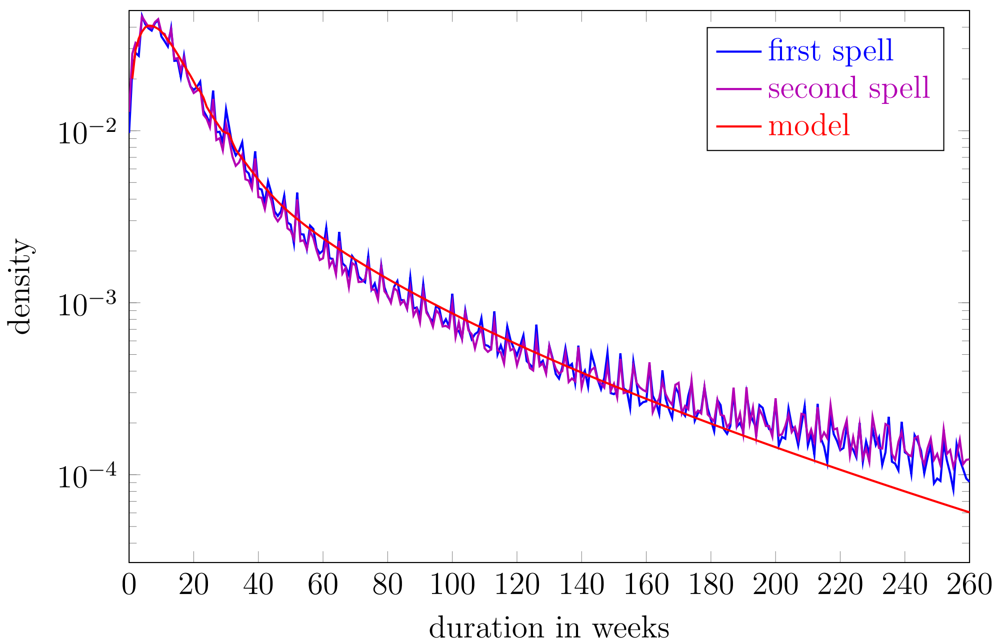

Figure 2: Marginal distribution of the first two non-employment spells, conditional on duration less than or equal to 260 weeks. First and second spell in data versus model.

In the subpopulation with two complete spells shorter than 260 weeks, the average duration of a completed non-employment spell is 25.9 weeks, and the average employment duration between these two spells is 85.7 weeks. Figure 2 depicts the marginal densities of the duration of the first two completed 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 first spells lasts 0.9 weeks longer than the second spell, a difference we suppress in our analysis.

Figure 3 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 more than 700,000 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 spikes in the marginal distribution of nonemployment spells every fourth or fifth week and, as Figure 2 shows, these spikes persist even at long durations.

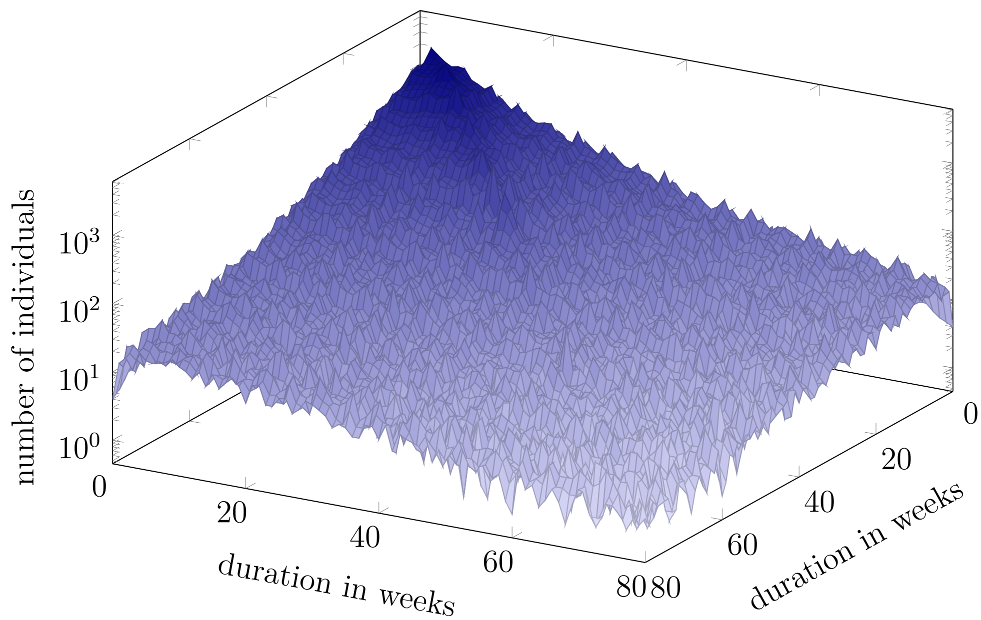

Figure 3: Non-employment exit joint density during the first two non-employment spells, conditional on duration less than or equal to 80 weeks.

6 Results

6.1 Test of the Model

We propose a test of the model inspired by the overidentifying restrictions in Section 3.4. We make three changes to accommodate the reality of our data. The first is that the data are only available in discrete time, and so we cannot measure the partial derivatives of the reemployment density ϕ. Instead, we propose a discrete time analog of equations (10)–(11):

$$a(t_1, t_2) \equiv \frac{t_2^2 \log\left(\frac{\phi(t_1,t_2+1)}{\phi(t_1,t_2-1)}\right) - t_1^2 \log\left(\frac{\phi(t_1+1,t_2)}{\phi(t_1-1,t_2)}\right)}{t_1^2 - t_2^2} - \frac{3}{t_1 + t_2} \geq 0 \tag{15}$$

$$\text{and } b(t_1, t_2) \equiv t_1 t_2 \left(\frac{t_1 t_2 \log\left(\frac{\phi(t_1,t_2+1)}{\phi(t_1,t_2-1)} \frac{\phi(t_1-1,t_2)}{\phi(t_1+1,t_2)}\right)}{t_1^2 - t_2^2} + \frac{3}{t_1 + t_2} \right) \geq 0, \tag{16}$$

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 2. We instead measure ϕ as $\frac{1}{2}(\phi(t_1, t_2) + \phi(t_2, t_1))$. The third is that the raw measure

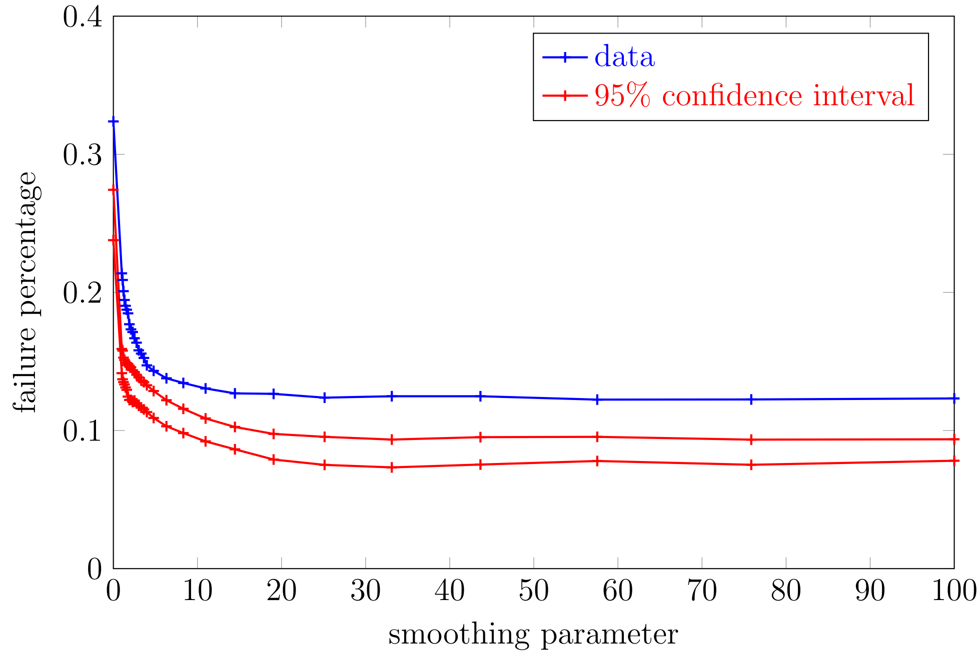

Figure 4: Nonparametric test of model. The blue circles show the percent of observations in the data with $a(t_1, t_2) < 0$ or $b(t_1, t_2) < 0$, weighted by share of workers with realized durations (t_1, t_2), for different values of the smoothing parameter. The red lines show a bootstrapped 95% confidence interval.

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 $\bar{\phi}$.[11] 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 E for more details on our filter.

Figure 4 displays our test results. We report the weighted fraction of points (t_1, t_2) with $0 \le t_1 < t_2 \le 260$ for which we compute either $a(t_1, t_2) < 0$ or $b(t_1, t_2) < 0$, weighting duration pairs by the density $\phi(t_1, t_2)$. Without any smoothing, we reject the model for 32 percent of workers in our sample. Setting the smoothing parameter to at least 7 reduces the rejection rate to thirteen percent, although further increases in the parameter do not significantly affect the rejection rate. The fact that the rejection rate declines with the smoothing parameter is not a trivial observation. In the limit as the smoothing parameter

[11] In practice we smooth the function $\log(\phi(t_1, t_2) + 1/n)$, rather than ϕ, where $n \approx 705{,}000$ is the number of individuals with two completed spells. This avoids taking $\log 0$.

becomes unboundedly large, the smoothed density converges to the one for the exponential hazard. In that limit, we reject the model at 44 percent of pairs; see Example 3 in Section 3.4.

To interpret the magnitude of the rejection rates, we show the bootstrapped 95% confidence interval in red in Figure 4.[12] The confidence interval is narrow and our test statistic lies above the upper bound for all values of λ. We think there are three reasons for this finding. First, the measured distribution of spells is not smooth both because of the finite sample of individuals and because of the role of months in measured durations. The model-generated data do not recognize the role of months. Second, the signal-noise ratio is low in our data at higher values of (t_1, t_2) because the number of observations declines quickly with duration. Third, our model does not describe the joint distribution of spells well at long durations. When we consider only pairs (t_1, t_2) such that $0 \leq t_1 < t_2 \leq 80$, the rejection rate in the data drops to 0.087 when $\lambda = 100$ and lies close to the corresponding 95% confidence interval, $[0.060, 0.073]$. We thus conclude that our data could have been generated by the proposed model at short durations, but we should be cautious in interpreting our results at long durations, where the inverse Gaussian assumption appears to break down. For example, the assumption that the net benefit from employment is a Brownian motion may not be valid for the longest spells.

6.2 Estimation

We estimate our model in several steps. To start, we assume that $\alpha \geq 0$ with G-probability 1, so all types have a positive drift in the net benefit from employment while non-employed. Using data on individuals with two completed spells, we obtain an estimate of the distribution function G^+. At the end, we allow for the possibility that some individuals have a negative drift and so use data on incomplete spells to bound the type distribution.

We start with estimating G^+. For a given type distribution $G(\alpha, \beta)$, the probability that any individual has duration $(t_1, t_2) \in T^2$ is

$$\frac{\int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\int_{T^2} \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\, dG(\alpha, \beta)\, d(t_1, t_2)}.$$

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

[12] To compute this statistic, we assume that the data generating process is a mixture of inverse Gaussians with the distribution G, which we estimate later in this section. We draw 500 samples of two non-employment spells for 850,000 individuals and keep individuals with two completed spells with duration between 0 and 260 weeks. We then proceed as in the data: we construct the empirical distribution $\phi(t_1, t_2)$, smooth it with our 2-dimensional HP filter for different values of the smoothing parameter λ, and apply our test. Confidence intervals for each λ is then the range of values which contain 95% of the rejection rates across samples.

	minimum distance estimate				EM estimate			
	mean	median	st. dev.	min	mean	median	st. dev	min
α	1.543	0.332	12.444	0.003	508.770	0.138	3123.989	0.096
β	16.878	4.768	94.010	0.667	3512.460	5.735	18390.764	1.465
$\frac{\mu_n}{\bar{\omega}-\underline{\omega}}$	0.045	0.037	0.043	0.005	0.052	0.043	0.033	0.018
$\frac{\sigma_n}{\bar{\omega}-\underline{\omega}}$	0.228	0.210	0.143	0.001	0.231	0.174	0.155	10^{-5}

Table 1: Summary statistics from estimation.

into a single term, we obtain that the log-likelihood of the data $\phi(t_1, t_2)$ is equal to

$$\sum_{(t_1,t_2)\in T^2} \phi(t_1, t_2) \log \left(\frac{\int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\int_{T^2} \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta)\, dG(\alpha, \beta)\, d(t_1, t_2)} \right).$$

Our basic approach to estimation chooses the distribution function G^+ to maximize this objective. More precisely, we follow a two-step procedure. In the first step, we constrain α and β to lie on a discrete nonnegative grid and use a minimum distance estimator to obtain an initial estimate. In the second step, we allow α and β to take nonnegative values off of the grid and optimize using the expectation-maximization (EM) algorithm. See Online Appendix OA.E for more details.

Our parameter estimates place a positive weight on 152 different types (α, β).[13] Table 1 summarizes our estimates. We report the mean, median, minimum, and standard deviation of α and β, as well as the drift and standard deviation of the net benefit from employment relative to the width of the inaction region, $\mu_n/(\bar{\omega} - \underline{\omega}) = \alpha/\beta$ and $\sigma_n/(\bar{\omega} - \underline{\omega}) = 1/\beta$. The first four columns summarize the estimates from the first estimation step, while the last 4 columns show results after refining the initial estimates using the EM algorithm. The mean and standard deviation of $\mu_n/(\bar{\omega} - \underline{\omega})$ and $\sigma_n/(\bar{\omega} - \underline{\omega})$ are similar in the two estimates, but the moments for α and β differ substantially. This is because the EM algorithm uncovers several types with a small $\sigma_n/(\bar{\omega} - \underline{\omega})$ (nearly deterministic duration), which implies a large value of α and β. The median values of α and β change by much less.

Our estimates uncover a considerable amount of heterogeneity. For example the cross-sectional standard deviation of α is six times its mean, while the cross-sectional standard deviation of β is around five times its mean. Moreover, α and β are positively correlated in the cross-section, with correlation 0.91 in the initial stage and 0.80 in the EM stage.

The smooth red line in Figure 2 shows the fitted marginal distribution of the duration

[13] A potential concern is that our model is over-parameterized, leading us to recover more types than are in the data. In the Online Appendix OA.F, we find that 77 types minimizes Akaike information criterion, but also document that our results are virtually identical with 77 and 152 types.

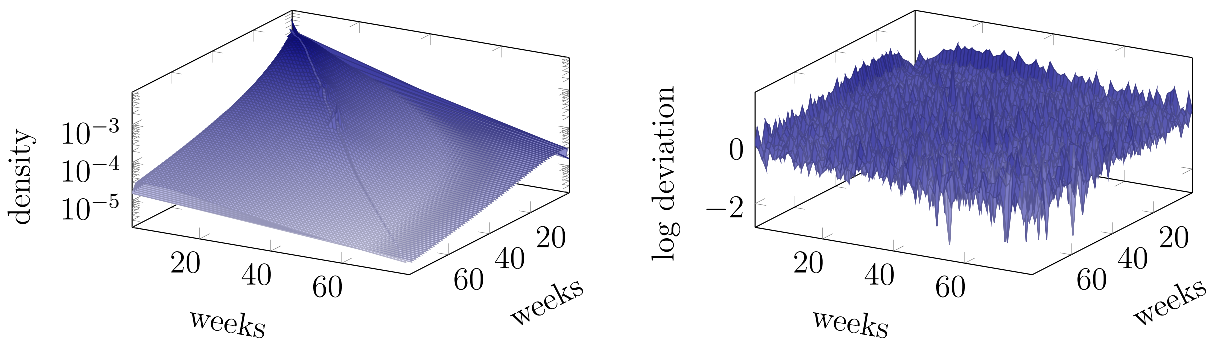

Figure 5: Nonemployment exit density: model (left) and log ratio of model to data (right)

of a single non-employment spell conditional on completed duration $t \in [0, 260]$. The model matches the initial increase in the density during the first ten weeks, as well as the gradual decline the subsequent five years. We miss the distribution at very long durations. There are two reasons for this. First, there are very few observations at long durations, so our procedure does not try to fit this data. Second, our model does not capture long durations very well, as we saw from the test results in Section 6.1. We could fit the marginal distribution better at long durations, but this would worsen the fit of the joint distribution ϕ.

Of course, it is not surprising that we can match the univariate hazard rate, since it 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 3. 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.

In the last step, we build on Section 3.3 to infer bounds on the fraction of the population with negative drift in the net benefit from employment while non-employed, and hence negative α. Our estimates of the distribution function G^+, which imposes that all individuals have a positive value of α, imply that 99 percent of individuals should have the second spell completed with duration less than 260 weeks. The corresponding number in the data is 94 percent. To fit this fact, we flip the sign of the drift for the smallest and largest possible fraction of individuals, constructing the distribution functions $\underline{G}$ and $\bar{G}$. Both of these distributions have identical predictions for completed spell data and are also able to match the prevalence of incomplete spells in our data set. The share of workers with negative α

lies between 6 and 16 percent.

6.3 Decomposition of the Hazard Rate

We now use our estimated type distribution to evaluate importance of heterogeneity in shaping the aggregate hazard rate. We consider three type distributions: G^+, where α is nonnegative with G-probability 1; and $\bar{G}$ and $\underline{G}$, where a positive fraction of types have negative α. The choice of the type distribution affects the hazard rate decomposition for two reasons. First, the weight we attribute to a type $(|\alpha|, \beta)$ depends on the sign of α. Ignoring the possibility that α is negative, we underestimate the number of individuals who start non-employment spells and so overestimate the hazard rate at long durations. Second, the hazard rate itself depends on the sign of α. The hazard rate for negative α is lower than for positive α, and only the former converges to zero at long durations. The structural hazard rate will thus be lower for $\bar{G}$ and $\underline{G}$ than for G^+, but so will be the aggregate hazard rate. In general, there is no a priori reason to think that one distribution will attribute a bigger role to heterogeneity than another.

We are interested in a decomposition of the hazard rate of the second spell in our population. The measured hazard rate of the first spell in our sample is biased because we constrain our population to be those workers whose first spell lasts ends within 260 weeks; we do not impose this constraint on the second spell. The type distributions $\bar{G}$ and $\underline{G}$ are suitable for the analysis of this hazard rate because they are consistent with the fraction of incomplete second spells in the sample. We focus on the distribution $\bar{G}$ but also compare our results to those obtained with $\underline{G}$ and G^+.

The purple line in the left panel of Figure 6 shows the raw hazard rate implied by the type distribution $\bar{G}$.[14] This peaks at 5.1 percent after 10 weeks, declines to 1.6 percent after a year and to 0.7 percent after two years. In contrast, the blue line shows the corresponding structural hazard rate $H^s(t)$. Most individuals have an increasing hazard for about 20 weeks. The structural hazard peaks at 7.6 percent, falls to 4.9 percent after a year and further declines to 2.8 percent after two years. The non-employment duration of an individual worker thus has a significant effect on her future prospects for finding a job, but less than the raw data indicates. After a worker is out of work for half a year, her chances of returning to work start to decline, possibly due to the loss of human capital. After two years of non-employment, the hazard of finding a job is only a third of what it was at the peak.

The ratio of the structural and aggregate hazard rate is attributed to heterogeneity and dynamic selection, measured by $H^h(t) \equiv H(t)/H^s(t)$ in the right panel of Figure 6. Recall

[14] The dotted lines in Figure 6 show bootstrapped 95 percent confidence intervals. See Appendix D for details on their construction.

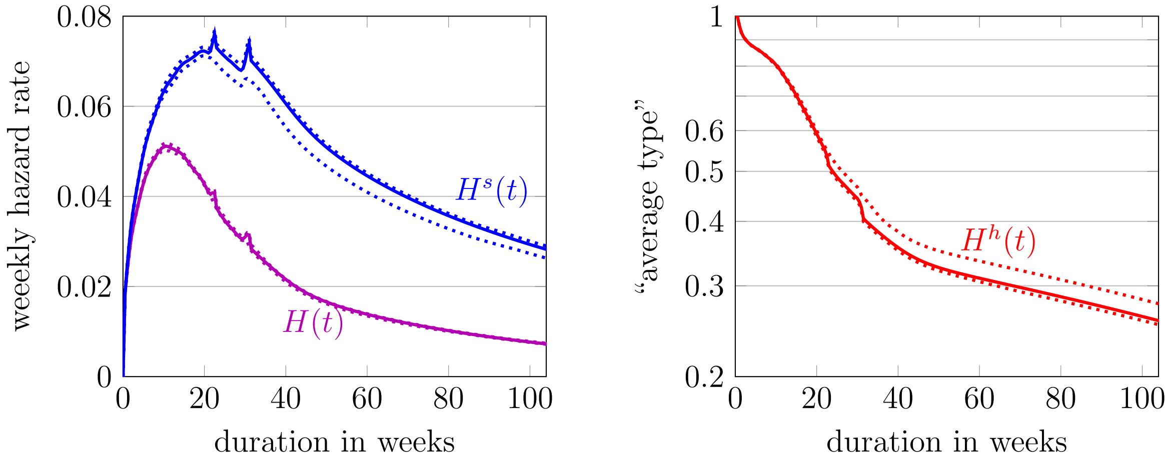

Figure 6: Hazard rate decomposition under $\bar{G}$. The purple line in the first figure shows the aggregate hazard rate $H(t)$. The blue line in the first figure shows the structural hazard rate $H^s(t)$. The ratio of them is the contribution of heterogeneity, plotted as the red line in the second figure. The dotted lines show bootstrapped 95 percent confidence bands. Note that these hazard rates do not condition on the spell ending within 260 weeks.

that selection necessarily pushes the hazard rate down, since high hazard individuals always find jobs faster than those with low hazard rates. We find very strong sorting during the first year of non-employment. The average type declines sharply, and after 39 weeks of non-employment it is only 36 percent of its initial value. The sorting continues even after a year, but at a slower rate than within the first year. After two years, the average type is only 26 percent as high as at the start of a spell.

Figure 7 replicates the decomposition using the type distributions G^+ and $\underline{G}$. The left hand panel confirms that the level of aggregate and structural hazard rate is lower with $\underline{G}$ and $\bar{G}$ than with G^+, a consequence of having types whose hazard rates are lower and converge to zero. The right hand panel indicates that the role of heterogeneity is similar during the first year under all three distributions; however, while with G^+, there is virtually no sorting after a year, dynamic selection continues to play an important role during the second year once we recognize that some spells will be defective. Under $\underline{G}$, dynamic selection reduces the hazard rate by 78 percent during the first two years of non-employment.

6.4 Comparison to the Mixed Proportional Hazard Model

A large literature assumes that the hazard rate of an individual has the form $h(t;\theta) = \theta h^b(t)$, where θ is an individual characteristic with an unknown distribution and $h^b(t)$ is the unknown baseline hazard of a spell ending at duration t. Heterogeneity is captured by the parameter

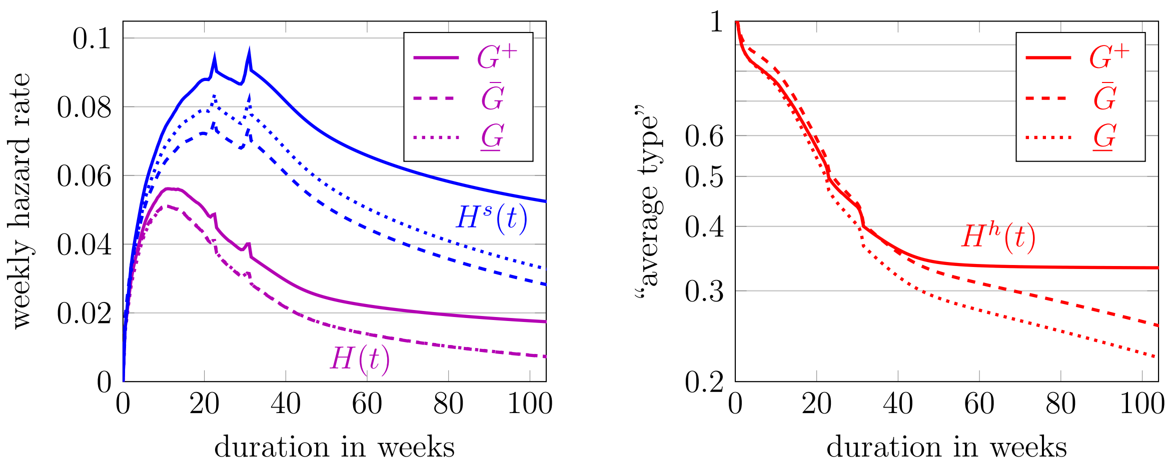

Figure 7: Decomposition of the hazard rate for distribution $G^+, \underline{G}, \bar{G}$. The blue lines show the structural hazard rate $H^s(t)$. The red lines show the contribution of heterogeneity, $H^h(t)$. The sum of the two is the raw hazard rate $H(t)$, shown as purple lines. Solid lines correspond to distribution G^+, dashed lines to $\bar{G}$ and dotted lines to $\underline{G}$.

θ, which may be partially unobserved, while structural duration dependence appears in the baseline hazard $h^b(t)$ and is assumed to behave identically across individuals. This is the mixed proportional hazard (MPH) model.

While we view the MPH model as a convenient reduced-form representation of the data, we argue in Alvarez, Borovičková, and Shimer (2015) that this specification is restrictive. In particular, the assumption that heterogeneity enters as a multiplicative constant on the common baseline hazard $h^b(t)$ can be rejected. Here we show that restricting heterogeneity in this way leads us to underestimate its importance.

We start by noting two restrictions imposed by the MPH model. First, the MPH model implies that the hazard rate for each type peaks at the same duration. We find that this is not the case in our estimated stopping time model. The hazard rate peaks within eighteen weeks for the most workers, but for more than ten percent of workers, the peak hazard rate does not occur until after one year. The ratio of the timing of the peak hazard for workers at the 90th and 10th percentile of this statistic is 182.

Second, the MPH model implies that the ratio of the hazard rate at any two durations (t_1, t_2) should be the same for all types. We analyze this ratio at $t_1 = 13$ and $t_2 = 52$. Again, there is considerable dispersion in this outcome. 85 percent of workers have a lower hazard rate after one year than after one quarter, but for nearly five percent of workers, the hazard rate has increased by a factor of ten.

Imposing homogeneity along these important dimensions leads us to underestimate the

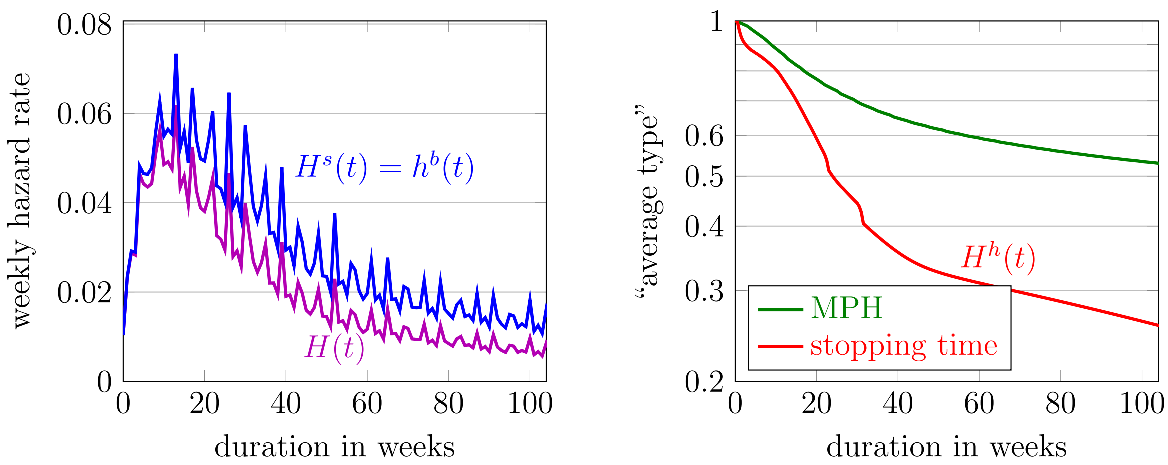

Figure 8: Decomposition of the hazard rate for the mixed proportional hazard model. In the left panel, the purple line shows the raw hazard rate, $H(t)$, and the blue line shows the baseline hazard rate, $h^b(t)$, which here is the same as the structural hazard rate, $H^s(t)$. The right panel shows the contribution of heterogeneity, $H^h(t)$. The green line shows the MPH model and the red line shows the stopping time model with distribution $\bar{G}$.

role of heterogeneity in determining duration dependence. To show this, we estimate the MPH model using maximum likelihood with a nonparametric baseline hazard.[15] The hazard rate decomposition is particularly simple in the MPH model: the structural hazard is the baseline hazard, $H^s(t) = h^b(t)$, and the contribution of heterogeneity $H^h(t)$ is the mean of θ among those still non-employed at duration t. Figure 8 shows our results. Our stopping time model implies much more dynamic selection. For example, we find that that the average type is only 35 percent as high after one year and 28 percent as high after two years as at the start of the spell. The comparable numbers in the MPH model are almost twice as large, 61 and 52 percent, severely understating the importance of heterogeneity.

6.5 Single-Spell Data

We comment briefly on what happens if we estimate the model using single-spell data. To do this, we construct a new data set consisting of all individuals with a single completed spell.

[15]We first modify the data set in a manner that makes it amenable to estimating the MPH model. We assume that the baseline hazard rate in the first and second spell is the same and that it integrates to infinity (i.e. there are no defective spells). Under these assumptions, we can extend our data set to include first spells that do not end within 260 weeks: for every worker in our sample whose second spell is longer than 260 weeks, we create a new observation that flips the durations of the first and second spells. This avoids biasing the estimated hazard rate by selecting a sample of individuals whose first spell ends within 260 weeks.

We then use the Stata command `streg` for estimation. We specify that the distribution of unobserved θ is gamma, but obtain similar results when we assume θ has an inverse Gaussian distirbution. We include a full set of dummy variables for each week of duration, which permits us to estimate a flexible baseline hazard.

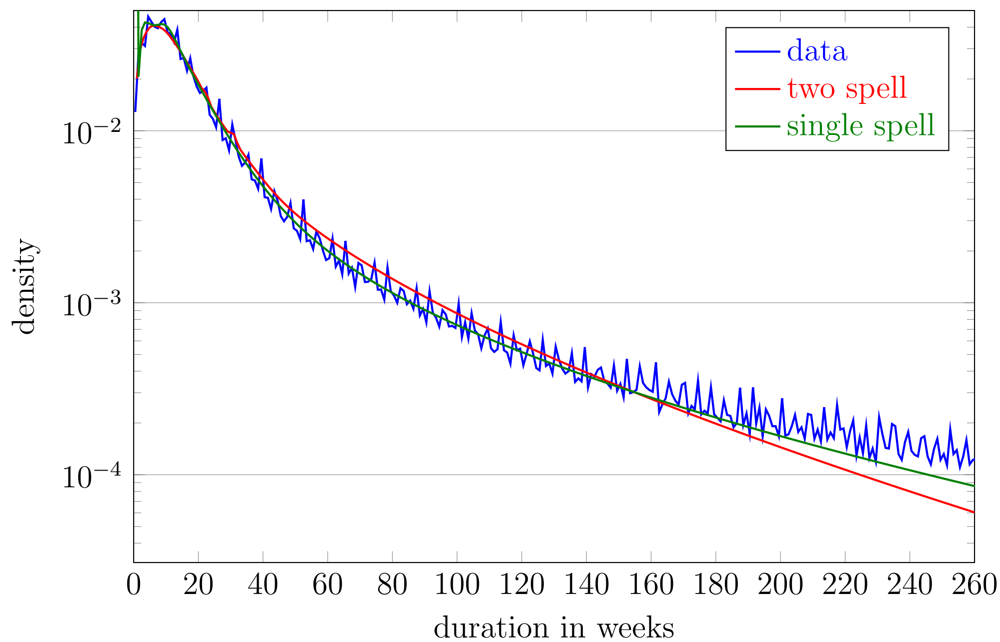

Figure 9: Distribution of nonemployment spells in the data and in the model estimated using the joint distribution of two spells (red, 152 types) and single-spell data (green, 5 types).

We assume that the data is generated by a mixture of five types of workers, each with an inverse Gaussian distribution. We estimate the mixing distribution using the EM algorithm to minimize the distance between the empirical and theoretical distribution of the duration of a single spell. Figure 9 shows that we can fit the data very well with only five types; indeed, we fit the single-spell density better with five types than we do in our preferred estimate with 152 types. Unsurprisingly, the single-spell estimates do worse at fitting joint density of the two-spell data. The question is whether this matters for our results.

We find that the five-type model substantially understates the importance of heterogeneity in duration dependence. The right hand panel of Figure 10 indicates that, after a steady 33 percent decline in the quality of job searchers during the first year, there is little subsequent change in the composition, so that after two years, the average type remains at 64 percent of its original value, more than twice as high as implied by our estimates with two-spell data.

The choice of the number of types is arbitrary and indeed our results are sensitive to it. With more than ten types, our single-spell estimates actually assign a bigger role to heterogeneity than our preferred estimates. Identification through prior knowledge of the number of unobserved types is tenuous. In any case, the fact that a finite mixture of inverse Gaussians can match the distribution of duration of a single spell does not imply that it

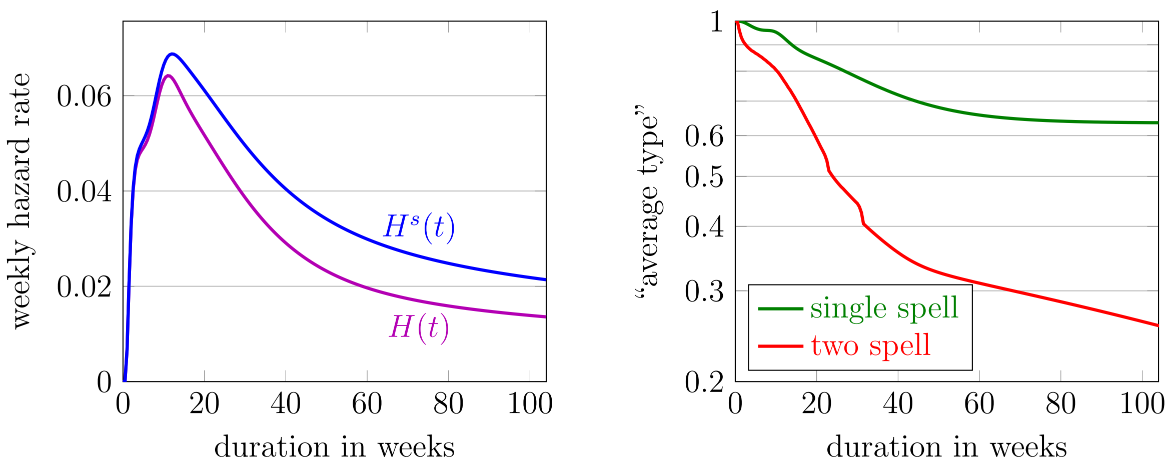

Figure 10: Decomposition of the hazard rate using single-spell data and an assumption that there are five types. In the left panel, the purple line shows the raw hazard rate, $H(t)$, and the blue line shows the structural hazard rate, $H^s(t)$. In the right panel, the green line shows the contribution of heterogeneity, $H^h(t)$. The red line corresponds to the contribution of heterogeneity in the stopping time model with distribution $\bar{G}$.

captures the real-world heterogeneity.

6.6 Estimated Switching Costs

In Section 2.3, we argued that knowledge of α and β, together with other four parameters of the model, pins down the magnitude of the fixed costs of switching employment status. Here we use the estimated distribution function G 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, ψ_e.[16]

Equation (4) implies that for given value of α and β, higher μ_e and $|\mu_n|$ increase the implied fixed costs, while higher σ_e and r reduce the implied fixed cost. With that in mind, we calibrate these parameters to find an upper bound on the fixed costs. First, we set the drift in employed workers' wages at $\mu_e = 0.01$ at an annual frequency. Estimates of the average wage growth of employed workers are often higher than one percent, but this is for workers who stay employed, a selected sample. The parameter μ_e governs wage growth for all workers without selection, and thus we view $\mu_e = 0.01$ as a large number. We set the standard deviation of log wages at $\sigma_e = 0.05$, again at an annual frequency. This is lower than typical estimates in the literature, which are closer to ten percent.

[16] Note that this is always expressed relative to the value of leisure.

mean	std	10^{th} per.	50^{th} per.	90^{th} per.
0.017%	0.024%	0.001%	0.005%	0.062%

Table 2: Summary statistics for the estimated switching costs, ψ_e, expressed as a percent of the annual flow value of leisure. Calculations assume $\mu_e = 0.01$, $\sigma_e = 0.05$, $|\mu_n| = 0.033$, $r = 0.02$, and the type distribution $\bar{G}$.

We cannot observe the drift of latent wages when non-employed, μ_n, but we can infer its value relative to μ_e from the duration of completed 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 their relative expected duration. In our sample, the average duration of non-employment spells is 25.9 weeks, while the average duration between two non-employment spells is 85.7 weeks, implying that $|\mu_n| = 3.3\mu_e$. Finally, we choose a low value for r. Since agents in the model are infinitely lived, we think of this as the sum of workers' discount rate and their death probability. A lower bound on this is 0.02, consistent with no discounting and a fifty year working lifetime.

Given this calibration, we estimate the distribution of fixed costs for the type distribution $\bar{G}$. Since α and μ_n have the same sign, we assume that workers with α positive have $\mu_n = 0.033$, while those with α negative have $\mu_n = -0.033$.

Table 6.6 summarizes our results. We again focus on the estimated costs for the distribution $\bar{G}$, but the costs for other two distributions are virtually the same. The median value of the switching costs is only 0.005 percent of the annual non-employment flow value, or about 6 minutes of time, assuming a 2000 hours of work per year. The costs vary across types, but even at the top decile, it amounts to 75 minutes of time. This is an order of magnitude smaller than the cost estimates in Silva and Toledo (2009).

Even though the magnitudes are very small, strictly positive switching costs are important for our results. If the switching cost were zero for someone, their region of inaction would be degenerate. And if the region of inaction were degenerate for everyone, the mean duration of spells in the interval $[1, 260]$ could not exceed $\sqrt{260} \approx 16$ weeks, as discussed in Section 2.3. Instead, we find that the mean duration of these spells is 26 weeks.

Previous work by Mankiw (1985), Dixit (1991), Abel and Eberly (1994), and others has shown that even small fixed costs can generate large regions of inaction. In our model, however, not only are the fixed costs small, but so is the region of inaction. The mean and median width of the inaction region are 1.2 and 1.6 log points, respectively. That is, the median worker who has just started working will quit if she experiences a 1.6 percent decline in her wage, holding fixed the value of nonemployment. A similar wage increase will induce her to return to work.

We are unaware of other papers that 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 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.

7 Conclusion

We develop a dynamic model of a worker's transitions in and out of employment. Our model features 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: what is the relative importance of heterogeneity versus structural duration dependence for explaining the evolution of the aggregate job finding rate; and how big are the fixed costs of switching between employment and non-employment. We 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. Workers differ not only in the average value of their job finding rate, but also in the timing of its peak. Finally, we find that fixed costs of switching employment status are small, but can also soundly reject any version of the model without fixed costs.

Our result that heterogeneity is an important driving force for duration dependence is in part a consequence of the stopping time model and its implied inverse Gaussian distribution. Our model allows for two dimensions of heterogeneity, while the MPH model, a common statistical model in applied research, allows for only a single dimension. Other statistical models, such as a mixture of log-normal distributions, has similar flexibility to the mixture of inverse Gaussian distributions. In fact, we have estimated a mixture of log-normals and found that it implies a similarly important role for heterogeneity; however, the mixture of log-normals has no structural interpretation and thus cannot be used to estimate the distribution of switching costs. The bottom line is that large data sets like the Austrian social security panel allow for a flexible treatment of heterogeneity; and that a flexible treatment of heterogeneity indicates an important role for dynamic selection of heterogeneous workers in driving the aggregate hazard rate.

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.

Ahn, Hie Joo, and James D. Hamilton, 2015. "Heterogeneity and Unemployment DYnamics." UCSD Mimeo.

Alvarez, Fernando, Katarína Borovičková, and Robert Shimer, 2015. "Testing the Mixed Proportional Hazard Model." University of Chicago Mimeo.

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.

Engl, Heinz W., Martin Hanke, and Andreas Neubauer, 1996. *Regularization of Inverse Problems*, Kluwer Academic Publishers.

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.

Hornstein, Andreas, 2012. "Accounting for Unemployment: The Long and Short of It." Federal Reserve Bank of Richmond Working Paper 12-07.

Krueger, Alan B, Judd Cramer, and David Cho, 2014. "Are the Long-Term Unemployed on the Margins of the Labor ¡arket?." *Brookings Papers on Economic Activity*. pp. 229–280.

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.

Lazear, Edward P., 1990. "Job Security Provisions and Employment." *Quarterly Journal of Economics*. 105 (3): 699–726.

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.

Silva, José Ignacio, and Manuel Toledo, 2009. “Labor Turnover Costs and the Cyclical Behavior of Vacancies and Unemployment.” *Macroeconomic Dynamics*. 13 (S1): 76–96.

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

In this section, we describe the structural model used in Section 2.1 in detail and show that the optimal worker's switching decision is described by two thresholds, $\underline{\omega} < \bar{\omega}$. We characterize these thresholds in the online appendix OA.A.

Let $s \in \{e, n\}$ denote employment status of a worker. We assume that $b(t)$ and $w(t)$ follow 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, and we use $\rho_s \in [-1, 1]$ to denote the instantaneous correlation between dw and db in state s,

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

The state of worker's problem is a triplet (s, w, b). Let $\tilde{E}(w, b)$ and $\tilde{N}(w, b)$ be the value functions of an employed and non-employed worker with state (w, b), respectively. It is technically convenient to denote the flow value of non-employment by $b_0 e^{b(t)}$; in the text we normalize $b_0 = 1$. The value functions satisfy

$$\tilde{E}(w, b) = \max_{\tau_e} \mathbb{E}\left[\int_0^{\tau_e} e^{-rt} e^{w(t)} dt + e^{-r\tau_e}\left(\tilde{N}(w(\tau_e), b(\tau_e)) - \psi_n e^{b(\tau_e)}\right) | w(0) = w, b(0) = b\right] \tag{17}$$

$$\tilde{N}(w, b) = \max_{\tau_n} \mathbb{E}\left[\int_0^{\tau_n} e^{-rt} b_0 e^{b(t)} dt + e^{-r\tau_n}\left(\tilde{E}(w(\tau_n), b(\tau_n)) - \psi_e e^{b(\tau_n)}\right) | w(0) = w, b(0) = b\right]. \tag{18}$$

An employed worker chooses the stopping time τ_e at which to switch to non-employment, described by equation (17). Similarly in equation (18), a non-employed worker chooses the first time τ_n at which to change her status to employment. The expectation in equations

(17) and (18) is taken with respect of the law of motion for $w(t)$ and $b(t)$ between $0 \leq t \leq \tau_s$.

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

$$r > \mu_{w,s} + \tfrac{1}{2}\sigma_{w,s}^2, \qquad \text{for } s \in \{e, n\} \tag{19}$$

$$r > \mu_{b,s} + \tfrac{1}{2}\sigma_{b,s}^2, \qquad \text{for } s \in \{e, n\} \tag{20}$$

The conditions in (19) guarantee that the value of being employed (non-employed) forever is finite. Moreover, if the conditions (20) 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.

Equations (17) and (18) imply 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 $E(\cdot)$ and $N(\cdot)$ as a function of the scalar $w - b$. We define $\omega(t)$, the log net benefit to work, as $\omega(t) \equiv w(t) - b(t)$. It also follows a state-contingent Brownian motion,

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

where $\{B\}$ is a standard Brownian motion defined in terms of $\{B_b, B_w\}$, and the drift and the diffusion coefficient are 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, $\omega(t) < \underline{\omega}$.

Figure 11 depicts value functions $E(\cdot)$ and $N(\cdot)$ for the case $\psi_n = 0$.

We characterize the thresholds $\underline{\omega}, \bar{\omega}$ in terms of parameters in of the model in the Online

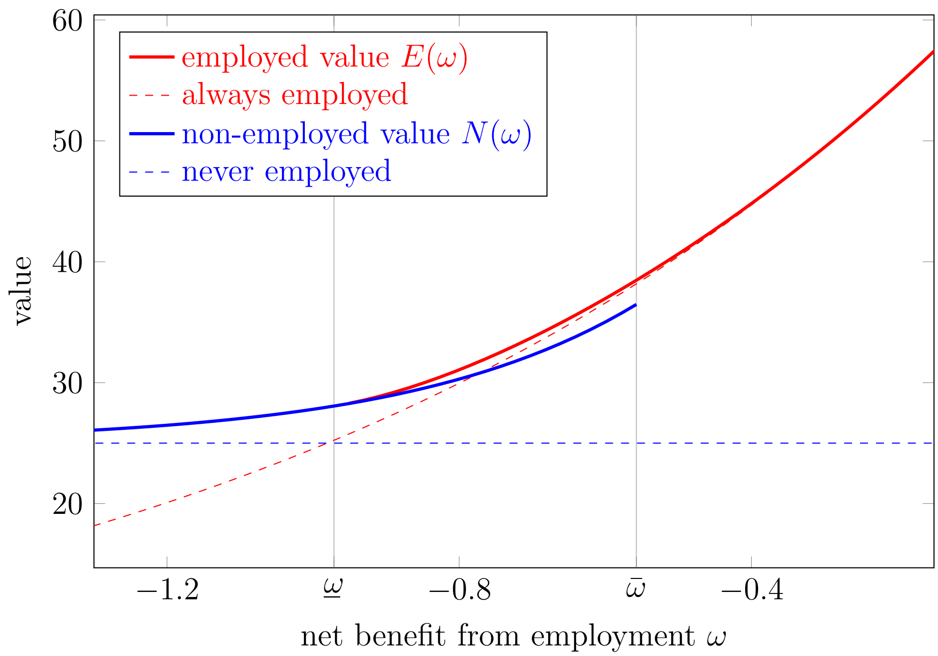

Figure 11: Value functions $E(\omega)$ and $N(\omega)$, together with thresholds $\underline{\omega} < \bar{\omega}$. The solid red line shows the value of being employed $E(\omega)$ for $\omega \in [\underline{\omega}, \infty)$. The solid blue line shows $N(\omega)$ for $\omega \in (-\infty, \bar{\omega}]$. The dotted red line indicates the value of being employed forever, $e^{\omega}/(r_e - \mu_e - \sigma_e^2/2)$, while the blue dotted line the value of being non-employed forever, b_0/r_n. The parameter 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 $\psi_n = 0$.

Appendix OA.A, here we only state an approximation for the distance between them. We use this result to infer the size of the fixed costs from the distance between the barriers and known values of the other parameters.

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

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

where λ_e*,* λ_n *and* r_n *are defined as*

$$\lambda_e = \frac{-\mu_e - \sqrt{\mu_e^2 + 2 r_e \sigma_e^2}}{\sigma_e^2} < -1, \qquad \lambda_n = \frac{-\mu_n + \sqrt{\mu_n^2 + 2 r_n \sigma_n^2}}{\sigma_n^2} > 1,$$
$$r_n = r - \mu_{b,n} - \frac{1}{2}\sigma_{b,n}^2.$$

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 1 *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:*

$$\frac{\partial^m \big(f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\big)}{\partial t_1^i\, \partial t_2^{m-i}} = f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta) \left(\sum_{r,s=0}^{r+s\leq m} \kappa_{r,s}(t_1, t_2; i, m-i) \alpha^{2r} \beta^{2s} \right) \quad (21)$$

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

$$\kappa_{r,s}(t_1, t_2; i, m-i) = \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \theta_{k,\ell,r,s}(i, m-i) t_1^{-k}\, t_2^{-\ell}, \quad (22)$$

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

Proof of Lemma 1. 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 (21) 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:

$$\begin{aligned}
&\frac{\partial^{m+1}\big(f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\big)}{\partial t_1^{i+1}\, \partial t_2^{m-i}} \\
&\quad = \frac{\partial}{\partial t_1}\left(\frac{\partial^{m}\big(f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\big)}{\partial t_1^{i}\, \partial t_2^{m-i}}\right) \\
&\quad = f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\left(\frac{\beta^2}{2t_1^2} - \frac{3}{2t_1} - \frac{\alpha^2}{2}\right)\left(\sum_{r,s=0}^{r+s\le m} \kappa_{r,s}(t_1,t_2;i,m-i)\alpha^{2r}\beta^{2s}\right) \\
&\qquad + f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\left(\sum_{r,s=0}^{r+s\le m} \frac{\partial \kappa_{r,s}\,(t_1,t_2;i,m-i)}{\partial t_1}\alpha^{2r}\beta^{2s}\right)
\end{aligned}$$

or

$$\begin{aligned}
&\frac{1}{f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)}\,\frac{\partial^{m+1}\big(f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\big)}{\partial t_1^{i+1}\, \partial t_2^{m-i}} \\
&\quad = -\frac{1}{2}\sum_{r,s=0}^{r+s\le m} \kappa_{r,s}(t_1,t_2;i,m-i)\alpha^{2(r+1)}\beta^{2s} \\
&\qquad + \frac{1}{2t_1^2}\sum_{r,s=0}^{r+s\le m} \kappa_{r,s}(t_1,t_2;i,m-i)\alpha^{2r}\beta^{2(s+1)} \\
&\qquad\quad + \sum_{r,s=0}^{r+s\le m}\left(-\frac{3}{2t_1}\kappa_{r,s}(t_1,t_2;i,m-i) + \frac{\partial \kappa_{r,s}\,(t_1,t_2;i,m-i)}{\partial t_1}\right)\alpha^{2r}\beta^{2s}.
\end{aligned}$$

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:

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

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

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

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

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

Existence of the partial derivatives follows from Lemma 1. The bulk of our proof establishes that the constant

$$h_{i,m-i} \equiv \max_{(t_1,t_2)\in \bar{B}} \sum_{r,s=0}^{r+s\le m} \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \frac{\big|\theta_{k,\ell,r,s}(i,m-i)\big|}{2\pi}\, t_1^{-k-\frac{3}{2}}\, t_2^{-\ell-\frac{3}{2}} \left(\frac{r+s+1}{\tau(t_1,t_2)}\right)^{r+s+1} e^{-(r+s+1)}, \tag{23}$$

where

$$\tau(t_1,t_2) = \frac{(t_1-t_2)^2}{2\big(t_1(1+t_2)^2 + t_2(1+t_1)^2\big)}, \tag{24}$$

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 1, the partial derivative is the product of a polynomial function and an exponential function:

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

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,

$$(\alpha+\beta)^{2(r+s+1)} \ge \alpha^{2r}\, \beta^{2(s+1)}. \tag{25}$$

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

$$\exp\left(-(\alpha+\beta)^2\tau(t_1,t_2)\right) \geq \exp\left(-\frac{(\alpha t_1-\beta)^2}{2t_1}-\frac{(\alpha t_2-\beta)^2}{2t_2}\right). \tag{26}$$

This can be verified by finding a maximum of the right hand side of (26) 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

$$\left(\frac{r+s+1}{\tau(t_1,t_2)}\right)^{r+s+1} e^{-(r+s+1)} \geq (\alpha+\beta)^{2(r+s+1)} \exp\left(-(\alpha+\beta)^2\tau(t_1,t_2)\right) \tag{27}$$

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 (25)–(27) 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 (23). ■

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

$$\frac{\partial\phi(t_1,t_2)}{\partial t_i} = \frac{\int\left(\frac{\beta^2}{2t_i^2}-\frac{3}{2t_i}-\frac{\alpha^2}{2}\right) f(t_1;\alpha,\beta)f(t_2;\alpha,\beta)dG(\alpha,\beta)}{\int_{T^2}\int f(t_1;\alpha,\beta)f(t_2;\alpha,\beta)dG(\alpha,\beta)\, d(t_1,t_2)}$$

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

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

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

$$\frac{2^{i+j}t_1^{2i}t_2^{2j}}{\phi(t_1,t_2)}\frac{\partial^{i+j}\phi(t_1,t_2)}{\partial t_1^i \partial t_2^j} = \mathbb{E}\big((\beta^2-\alpha^2t_1^2)^i(\beta^2-\alpha^2t_2^2)^j|t_1,t_2\big) + v_{ij}(t_1,t_2)$$
$$= \sum_{x=0}^{i+j}\sum_{y=\max\{0,x-j\}}^{\min\{x,i\}} \frac{i!j!(-t_1)^y(-t_2)^{x-y}\mathbb{E}(\alpha^{2x}\beta^{2(i+j-x)}|t_1,t_2)}{y!(x-y)!(i-y)!(j-x+y)!} + v_{ij}(t_1,t_2), \quad (28)$$

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 $\mathbb{E}\big((\beta^2-\alpha^2t_1^2)^{i+1}(\beta^2-\alpha^2t_2^2)^j|t_1,t_2\big)$ which follows directly from the derivative of $f(t_1,\alpha,\beta)$ with respect to t_1. The second line of (28) follows from the first by expanding the power functions.

Now let $i = \{0,\dots,m\}$ and $j = m-i$. As we vary i, equation (28) 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:

$$\begin{pmatrix} \frac{2^m t_1^{2m}}{\phi(t_1,t_2)}\frac{\partial^m\phi(t_1,t_2)}{\partial t_1^m} \\ \frac{2^m t_1^{2(m-1)}t_2^2}{\phi(t_1,t_2)}\frac{\partial^m\phi(t_1,t_2)}{\partial t_1^{m-1}\partial t_2} \\ \frac{2^m t_1^{2(m-2)}t_2^4}{\phi(t_1,t_2)}\frac{\partial^m\phi(t_1,t_2)}{\partial t_1^{m-2}\partial t_2^2} \\ \vdots \\ \frac{2^m t_2^{2m}}{\phi(t_1,t_2)}\frac{\partial^m\phi(t_1,t_2)}{\partial t_2^m} \end{pmatrix} = L(t_1,t_2)\cdot U(t_1,t_2)\cdot \begin{pmatrix} \mathbb{E}(\alpha^{2m}|t_1,t_2) \\ \mathbb{E}(\alpha^{2(m-1)}\beta^2|t_1,t_2) \\ \mathbb{E}(\alpha^{2(m-2)}\beta^4|t_1,t_2) \\ \vdots \\ \mathbb{E}(\beta^{2m}|t_1,t_2) \end{pmatrix} + v_m(t_1,t_2), \quad (29)$$

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

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

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

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

for $0 \le i \le j \le m$ and $U_{ij}(t_1,t_2) = 0$ for $0 \le j < i \le m$; 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.[17] 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 m^{th} moments of the joint distribution are uniquely determined by the m^{th} 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 2 *Let $\hat{G}(\alpha, \beta)$ denote the cumulative distribution of a pair of nonnegative random variables and let $\mathbb{E}(\alpha^{2i}\beta^{2j}) \equiv \int \alpha^{2i}\beta^{2j} d\hat{G}(\alpha, \beta)$ denote its $(i,j)^{th}$ even moment. For any $m \in \{1, 2, \ldots\}$, define*

$$M_m = \max_{i=0,\ldots,m} \mathbb{E}(\alpha^{2i}\beta^{2(m-i)}). \tag{30}$$

Assume that

$$\lim_{m\to\infty} \frac{[\, M_m \,]^{\frac{1}{2m}}}{2m} = \lambda < \infty\,. \tag{31}$$

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 2. 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:

$$\limsup_{m\to\infty} \frac{(\,\mathbb{E}\,[\,|u|^m\ ]\,)^{\frac{1}{m}}}{m} \equiv \lambda' < \infty, \tag{32}$$

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

$$\lim_{m\to\infty} \frac{(\,\mathbb{E}\,[\,u^{2m}\ ]\,)^{\frac{1}{2m}}}{2m} \equiv \lambda < \infty\,. \tag{33}$$

Note that if condition (33) holds, then condition (32) holds as well. To prove this, assume that $\lambda' = \infty$ and $\lambda < \infty$. Then there must be an odd integer m which is very large, and in particular

$$\frac{(\mathbb{E}\,[\,|u|^m\ ]\,)^{\frac{1}{m}}}{m} > (1+\varepsilon)\frac{m+1}{m}\,\lambda, \tag{34}$$

[17] 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.

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

$$\frac{(\mathbb{E}[\,|u|^m\,])^{\frac{1}{m}}}{m} \leq \frac{m+1}{m}\frac{(\mathbb{E}[\,|u|^{m+1}\,])^{\frac{1}{m+1}}}{m+1} \leq \frac{m+1}{m}\lambda(1+\varepsilon), \tag{35}$$

which is a contradiction with (34).

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 (33) for $u(x) \equiv (x_1\alpha + x_2\beta)$. We note that:

$$\begin{aligned}
\mathbb{E}\left[u(x)^{2m}\right] &= \mathbb{E}\left((x_1\alpha + x_2\beta)^{2m}\right) = \sum_{i=0}^{2m} \frac{2m!}{i!(2m-i)!} x_1^i\, x_2^{2m-i}\, \mathbb{E}\left(\alpha^{2i}\beta^{2(m-i)}\right) \\
&\leq \sum_{i=0}^{2m} \frac{2m!}{i!(2m-i)!} |x_1|^i\, |x_2|^{2m-i}\, \mathbb{E}\left(\alpha^{2i}\beta^{2(m-i)}\right) \leq M_m \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} |x_1|^i\, |x_2|^{m-i} \\
&= M_m\,(|x_1| + |x_2|)^{2m}
\end{aligned}$$

where we use that (α, β) are non-negative random variables, and M_m is defined in equation (30).

Now we check that the limit in equation (33) is satisfied given the assumptions in equations (30) and (31):

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

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

Proof of Proposition 3. Write the conditional moments as

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

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

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

where $\tau(t_1, t_2)$ is defined in equation (24) and the inequality uses the same steps as the proof of Proposition 1 to bound the function. In the language of Lemma 2, this implies

$$M_m = \frac{((m+1)/\tau(t_1, t_2))^{m+1} \exp(-(m+1))}{2\pi t_1^{3/2} t_2^{3/2} \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}. \tag{36}$$

We use this to verify condition (31) in Lemma 2.

Taking the log transformation of $(1/2m)\,(M_m)^{1/2m}$ and using the expression (36) we get:

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

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

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

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

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

$$\phi(t) = \int f(t; \mu_n \beta, \beta)\, g(\beta)\, d\beta. \tag{37}$$

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 k^{th} 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 k^{th} moment of the distribution $\phi(t)$ and $f(t; \mu_n \beta, \beta)$, respectively,

$$\begin{aligned} m(k, \mu_n \beta, \beta) &\equiv \int_0^\infty t^k f(t; \mu_n \beta, \beta) dt & (38) \\ M(k) &\equiv \int_0^\infty t^k \phi(t) dt = \int \left[\int_0^\infty t^k f(t; \mu_n \beta, \beta)\, dt \right] g(\beta)\, d\beta & (39) \\ &= \int m(k, \mu_n \beta, \beta) g(\beta)\, d\beta & (40) \end{aligned}$$

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

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

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

$$\begin{aligned} m(k, \mu_n \beta, \beta) &= \sum_{i=0}^{k-1} a(k, i, \mu_n) \beta^{-2i} \\ a(k, i, \mu_n) &\equiv 2^{-i} \frac{(k-1+i)!}{i!(k-1-i)!} \left(\frac{1}{\mu_n}\right)^{k+i} \end{aligned}$$

Then the k^{th} moment of the distribution ϕ is

$$\begin{aligned} M(k) &= \int \sum_{i=0}^{k-1} a(k, i, \mu_n) \beta^{-2i} g(\beta) d\beta \\ &= \sum_{i=0}^{k-1} a(k, i, \mu_n) \mathbb{E}\left[\beta^{-2i}\right] \end{aligned} \quad (41)$$

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 (41) can be solved to find $\mathbb{E}[\beta^{-2}]$. By induction, if $\mathbb{E}[\beta^{-2i}]$ are known for $i = 1, \ldots, k-1$, then equation (41) 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

$$f(t; \alpha, 0) = \frac{1}{\sigma_n \sqrt{2\pi} t^{3/2}} \exp\left(-\frac{1}{2}\alpha^2 t\right). \quad (42)$$

and thus the distribution of spells in the population is

$$\phi(t) = \int f(t; \alpha, 0) g(\alpha) d\alpha. \quad (43)$$

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

Let start by deriving the k^{th} 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

$$
\begin{aligned}
\frac{\partial^s}{\partial t^s}\left(t^{-3/2}\right) &= t^{-3/2}\prod_{i=0}^{s}\left(-\frac{3}{2}-i\right) \\
\frac{\partial^{m-s}}{\partial t^{m-s}}\exp\left(-\frac{1}{2}\alpha^2 t\right) &= \exp\left(-\frac{1}{2}\alpha^2 t\right)\left(-\frac{1}{2}\alpha^2\right)^{r-s},
\end{aligned}
$$

and thus we can write an equation for the m^{th} derivative of ϕ,

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

Finally, rearrange the terms

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

to find the m^{th} derivative of ϕ as a sum of m^{th} and lower moments of α^2.

D Standard Errors

We use a non-parametric resampling bootstrap to compute the standard errors for the hazard rate decomposition. We use our data to draw B samples of size N with replacement, where $N = 751,125$ is the number of people in our dataset. We treat each sample $b = 1, \dots B$ as our original data. That is, we select workers with two completed spells shorter than 260 weeks to construct the density $\phi_b(t_1, t_2)$. We smooth the density with a two-dimensional HP filter and use it to estimate the distribution $G_b^+(\alpha, \beta)$. We construct $\bar{G}_b(\alpha, \beta)$ and $\underline{G}_b(\alpha, \beta)$ using the share of completed spells in the sample b. Finally, we use $\bar{G}_b(\alpha, \beta), \underline{G}_b(\alpha, \beta), G_b^+(\alpha, \beta)$ to conduct the decomposition.

Let $H_b(t), H_b^s(t), H_b^h(t)$ be the aggregate hazard, structural hazard and average type at duration t under the distribution $\bar{G}_b(\alpha, \beta)$. To construct the standard errors for aggregate hazard $H(t)$, we choose $\underline{b}(t)$ and $\bar{b}(t)$ for each duration t, such that

$$
Prob[H_b(t) \leq H_{\underline{b}(t)}(t) = 0.025], \quad Prob[H_b(t) \geq H_{\bar{b}(t)}(t) = 0.975].
$$

Then, $H_{\underline{b}(t)}(t)$ and $H_{\bar{b}(t)}(t)$ represent error bands on our estimate of the hazard rate $H(t)$. We proceed in the same way to construct error bands on $H_b^s(t), H_b^h(t)$.

We choose $B = 200$. Figure 6 shows the decomposition together with the standard errors. The standard errors are tiny. We find similarly small standard errors when using non-parametric subsampling bootstrap or parametric bootstrap as well.

E 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 $\bar{\psi}(t_1, t_2)$ at $t_2 \geq t_1$ to solve

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

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.

Online Appendix

OA.A Characterization of the Thresholds

To characterize the thresholds $\underline{\omega}, \bar{\omega}$, it is useful to go back to value functions $\tilde{E}(w,b)$ and $\tilde{N}(w,b)$ and formulate the Hamilton-Jacobi-Bellman (HJB) equation for the worker's problem. If a worker is employed, then for all w and b such that $w-b \geq \underline{\omega}$, the HJB for $\tilde{E}(w,b)$ is

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

Similarly, if a worker is non-employed, then

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

for all w and b with $w-b \leq \bar{\omega}$. At the thresholds, a worker has to be indifferent between staying with her current status or switching. Using the homogeneity $\tilde{E}, \tilde{N}$, the boundary conditions are

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

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 $N(\omega)$:

$$r_e\, E(\omega) = e^\omega + \hat{\mu}_e\, E'(\omega) + \tfrac{1}{2}\,\sigma_e^2\, E''(\omega) \tag{OA.1}$$
$$r_n\, N(\omega) = b_0 + \hat{\mu}_n\, N'(\omega) + \tfrac{1}{2}\,\sigma_n^2\, N''(\omega) \tag{OA.2}$$

where the parameters are

$$
\begin{aligned}
r_s &\equiv r - \mu_{b,s} - \tfrac{1}{2}\sigma_{b,s}^2 \\
\hat{\mu}_s &\equiv \mu_{w,s} - \mu_{b,s} - \sigma_{b,s}^2 + \sigma_{w,s}\sigma_{b,s}\,\rho_s \\
\sigma_s^2 &\equiv \sigma_{w,s}^2 + \sigma_{b,s}^2 - 2\sigma_{w,s}\sigma_{b,s}\,\rho_s
\end{aligned}
$$

for $s \in \{e, n\}$. Conditions (19) and (20) reduce to

$$
r_s > \hat{\mu}_s + \tfrac{1}{2}\,\sigma_s^2 \text{ and } r_s > 0 \text{ for } s \in \{e, n\}. \tag{OA.3}
$$

We can also rewrite the boundary conditions as

$$
E(\underline{\omega}) = N(\underline{\omega}) - \psi_n \text{ and } E'(\underline{\omega}) = N'(\underline{\omega}) \tag{OA.4}
$$

$$
N(\bar{\omega}) = E(\bar{\omega}) - \psi_e \text{ and } N'(\bar{\omega}) = E'(\bar{\omega}). \tag{OA.5}
$$

The solution to equation (OA.1) and equation (OA.2) with boundary conditions equation (OA.4) and equation (OA.5) is of a form

$$
E(\omega) = \frac{e^{\omega}}{r_e - \hat{\mu}_e - \sigma_e^2/2} + c_{e,1}e^{\lambda_{e,1}\omega} + c_{e,2}e^{\lambda_{e_2}\omega} \tag{OA.6}
$$

$$
N(\omega) = \frac{b_0}{r_n} + c_{n,1}e^{\lambda_{n,1}\omega} + c_{n,2}e^{\lambda_{n,2}\omega} \tag{OA.7}
$$

where

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

are the roots of the equations $r_e = \lambda_e\,(\hat{\mu}_e + \lambda_e\sigma_e^2/2)$ and $r_n = \lambda_n\,(\hat{\mu}_n + \lambda_n\sigma_n^2/2)$. Hence we have six equations, (OA.4)–(OA.7), 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

$$
\lim_{\omega \to -\infty} N(\omega) = \frac{b_0}{r_n} \quad \text{and} \tag{OA.8}
$$

$$
\lim_{\omega \to +\infty} \frac{E(\omega)}{e^{\omega}} = \frac{1}{r_e - \hat{\mu}_e - \sigma_e^2/2} \tag{OA.9}
$$

Equation (OA.8) requires that the value function for arbitrarily small ω converges to the value of being non-employed forever. Likewise equation (OA.9) 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 (OA.6) and (OA.7) as:

$$E(\omega) = \frac{e^{\omega}}{r_e - \hat{\mu}_e - \sigma_e^2/2} + c_e e^{\lambda_e \omega} \text{ for all } \omega \geq \underline{\omega} \tag{OA.10}$$

$$N(\omega) = \frac{b_0}{r_n} + c_n e^{\lambda_n \omega} \text{ for all } \omega \leq \bar{\omega} \tag{OA.11}$$

with

$$\lambda_e = \frac{-\mu_e - \sqrt{\mu_e^2 + 2r_e\sigma_e^2}}{\sigma_e^2} < -1 \text{ and } \lambda_n = \frac{-\mu_n + \sqrt{\mu_n^2 + 2r_n\sigma_n^2}}{\sigma_n^2} > 1. \tag{OA.12}$$

Condition (OA.3) 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

$$\psi_n + \frac{e^{\underline{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e e^{\lambda_e \underline{\omega}} = \frac{b_0}{r_n} + c_n e^{\lambda_n \underline{\omega}} \tag{OA.13}$$

$$-\psi_e + \frac{e^{\bar{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e e^{\lambda_e \bar{\omega}} = \frac{b_0}{r_n} + c_n e^{\lambda_n \bar{\omega}} \tag{OA.14}$$

$$\frac{e^{\underline{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e \lambda_e e^{\lambda_e \underline{\omega}} = c_n \lambda_n e^{\lambda_n \underline{\omega}} \tag{OA.15}$$

$$\frac{e^{\bar{\omega}}}{r_e - \mu_e - \sigma_e^2/2} + c_e \lambda_e e^{\lambda_e \bar{\omega}} = c_n \lambda_n e^{\lambda_n \bar{\omega}} \tag{OA.16}$$

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 (OA.10) and equations (OA.11) by setting $c_e = 0$ and $c_n = 0$ respectively.

We conclude this section with a lemma on the units of the switching costs.

Lemma 3 *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 3, multiply the appropriate objects in equations (OA.13) and (OA.14) by k and then simplifying those equations as well as equations (OA.15) and (OA.16) using the expressions in the statement of the proof. The lemma implies that the size of the region of inaction, $\bar{\omega} - \underline{\omega}$, depends only on $(\psi_e + \psi_n)/b_0$. In the main text, we normalized $b_0 = 1$.

Thus, the switching costs we calculate in Section 6.6 are in units of the flow value of non-employment.

OA.B Constructing Distributions $\bar{G}$ and $\underline{G}$

We construct distributions $\bar{G}$ and $\underline{G}$ by flipping sign of α for the largest and smallest number of types from the distribution G^+. We have two restrictions: the distribution of the completed spells has to be the same as under G^+, and the share of workers with the second spell shorter than $\bar{t}$ has to equal c. The first restriction implies that the weight of a type with $(-\alpha, \beta)$ has to be given by $e^{4\alpha\beta} dG^+(\alpha, \beta)$.

Let $x(\alpha, \beta) \in [0, 1]$ be a fraction of the type (α, β) from G^+ that we will flip into the type $(-\alpha, \beta)$. Then we can formulate a maximization problem to construct $\bar{G}$ as follows:

$$\begin{aligned}
&\max_{x(\alpha,\beta)} \int x(\alpha,\beta) e^{4\alpha\beta} dG^+(\alpha,\beta) \\
&\text{s.t.} \quad c = \frac{\int \left[x(\alpha,\beta) e^{4\alpha\beta} F(\bar{t}; -\alpha, \beta) + (1 - x(\alpha,\beta)) F(\bar{t}; \alpha, \beta)\right] dG^+(\alpha,\beta)}{\int \left(1 - x(\alpha,\beta) + x(\alpha,\beta) e^{4\alpha\beta}\right) dG^+(\alpha,\beta)} \\
&\qquad x(\alpha,\beta) \geq 0, \quad x(\alpha,\beta) \leq 1
\end{aligned}$$

Assigning a Lagrange multipliers λ, $\lambda^0(\alpha, \beta)$ and $\lambda^1(\alpha, \beta)$ to the constraints, the first order condition for a given (α, β) is

$$0 = 1 + \lambda \left[c\left(1 - e^{-4\alpha\beta}\right) - \left(e^{-2\alpha\beta} - e^{-4\alpha\beta}\right) F(\bar{t}; \alpha\beta)\right] + \left[\lambda^0(\alpha,\beta) - \lambda^1(\alpha,\beta)\right].$$

It then follows that there will be at most one type (α, β) for which we have an interior solution $x(\alpha, \beta) \in (0, 1)$, while for other types either $x(\alpha, \beta) = 1$ or $x(\alpha, \beta) = 0$. The first order condition also defines the selection rule $p(\alpha, \beta) = c\left(1 - e^{-4\alpha\beta}\right) - \left(e^{-2\alpha\beta} - e^{-4\alpha\beta}\right) F(\bar{t}; \alpha, \beta)$ which we used in Section 3.3.

OA.C Power of the First Moment Test

We consider two interesting specifications 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. We find that in both cases the test for $\mathbb{E}[\alpha^2 | t_1, t_2]$ fails for $t_2 = 0$ and t_1 sufficiently small. We note that the first example has the property that ϕ is not differentiable at points where $t_1 = t_2$.

OA.C.1 A Model of Delayed Search

Consider an extension of a canonical search model where a non-employed individual starts actively searching for a job only after τ periods, and finds a job at the rate θ. Thus, the hazard rate of exiting non-employment is zero for $t \leq \tau$, and θ for $t \geq \tau$. Each worker is described by a pair (θ, τ) which is distributed in the population according to a cumulative distribution function G. The joint density of two spells is given by

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

We now 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:

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

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. $\mathbb{E}(\alpha^2|t_1,0) < 0$.

OA.C.2 Mixed Proportional Hazard Model

We consider a mixed proportional hazard (MPH) model, which specifies that worker's hazard of finding a job at duration t is given by $\theta h(t)$ where $h(t)$ is a so-called baseline hazard, and θ is an unobserved individual characteristic distributed according to G. The distribution of two spells t_1, t_2 is given by

$$\phi(t_1, t_2) = \int_0^{\infty} \theta^2 \, h(t_1) \, h(t_2) \, e^{-\theta \left(\int_0^{t_1} h(s)ds + \int_0^{t_2} h(s)ds \right)} \, dG(\theta) \tag{OA.17}$$

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:

$$\frac{\phi_i(t_1,t_2)}{\phi(t_1,t_2)} = \frac{h'(t_i)}{h(t_i)} - h(t_i)\,\mathbb{E}\left[\theta\,|\,t_1,t_2\right] \text{ where}$$
$$\mathbb{E}\left[\theta\,|\,t_1,t_2\right] \equiv \frac{\int_0^\infty \theta^3\, h(t_1)\, h(t_2)\, e^{-\theta\left(\int_0^{t_1} h(s)ds + \int_0^{t_2} h(s)ds\right)}\, dG(\theta)}{\int_0^\infty \theta^2\, h(t_1)\, h(t_2)\, e^{-\theta\left(\int_0^{t_1} h(s)ds + \int_0^{t_2} h(s)ds\right)}\, dG(\theta)}$$

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.

OA.D Selected Properties of the Austrian Data

In this section, we show business cycle properties of the Austrian data, and analyze the impact of our selection criteria on the final sample.

We start with the business cycle properties. During the analyzed period, Austrian annual GDP growth varied between 0.5 and 3.6 percent per year and never slipped negative. In Figure 12 we plot the mean duration of all in-progress nonemployment spells that are shorter than 5 years. We plot this between 1991 and 2006, where the delayed start date ensures that we have not artificially truncated the duration of any spells. There is virtually no cyclical variation. We smooth the data before we depict them but this is to remove weekly variation in the duration. In particular, there is always a peak in 4, 8, etc weeks which is due to the

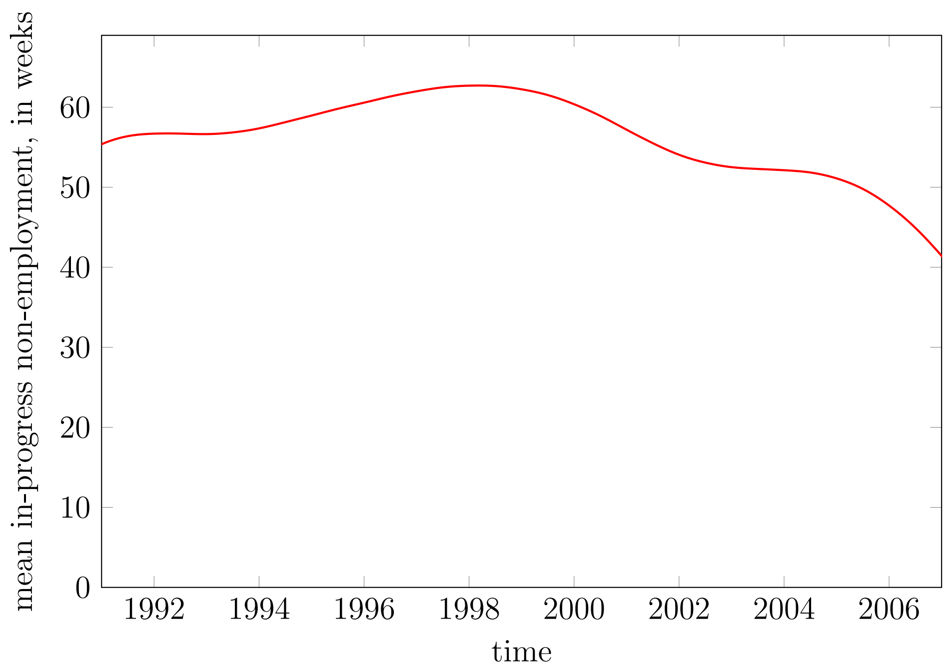

Figure 12: Mean duration of all in-progress non-employment spells with duration shorter than 260 weeks, smoothed using an HP filter with a smoothing parameter 14,400.

fact that employment spells tend to start in the first week of the month and end in the last week of the month.

We next analyze the impact of the selection criteria on our sample. To be included in our sample, a worker has to meet certain age criteria, and has to have at least two non-employment spells, with the first one longer than 260 weeks. We compare this sample to a sample which has the same age criteria but does not require a worker to have at least one spell.

Figure 13 shows a comparison of the distribution of non-employment spells in the full sample and in the sample we selected for estimation. For comparability, we condition on spells lasting no more than 260 weeks even in the full sample. We see that our selected sample has fewer very short spells and more longer spells.

OA.E Estimation

The link between the model and the data is given by equation (13). 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 (13) and solve it by minimizing the sum of squared errors between the data and the model-implied distribution of spells. In

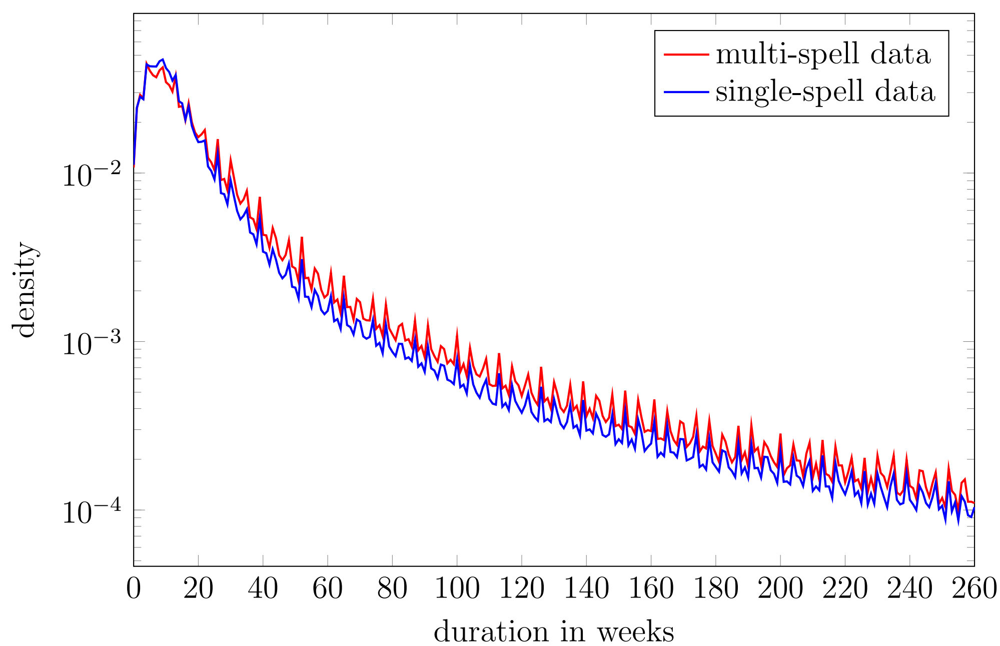

Figure 13: Comparison of the marginal distributions of non-employment spells with and without imposing restrictions of the number and duration of spells. The red line shows the distribution in our selected sample. The blue line shows the distribution of the single spell data where we impose the same age criteria as in our sample, but do not require a worker having at least two non-employment spells.

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.

OA.E.1 Step 1: Minimum Distance Estimator

To discretize equation (13), 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 $k = 1, \ldots, K$.[18] 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 (13) in the discretized form is

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

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

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

[18]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 $\mu_n/(\bar{\omega} - \underline{\omega}) = \alpha/\beta$.

In the first stage, we solve the minimization problem

$$\min_{\mathbf{g}} \left(\left(\mathbf{F} - \boldsymbol{\phi} \mathbf{H}' \right) \mathbf{g} \right)' \left(\left(\mathbf{F} - \boldsymbol{\phi} \mathbf{H}' \right) \mathbf{g} \right)$$
$$\text{s.t.} \quad \mathbf{g} \geq 0, \quad \sum_{k=1}^{K} g(k) = 1.$$

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 $\boldsymbol{\phi}$, and second, we stabilize the solution by replacing $\tilde{\mathbf{F}} \equiv \mathbf{F} - \boldsymbol{\phi} \mathbf{H}'$ 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.[19] This is called Tikhonov regularization method, see for example Engl, Hanke, and Neubauer (1996).

OA.E.2 Step 2: Maximum Likelihood using the EM algorithm

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 m^{th} 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 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.

[19] The L-curve is a graphical representation of the tradeoff between $|| (\mathbf{Fg} - \boldsymbol{\phi}) ||^2$ 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.

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

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

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

$$\log \ell(\mathbf{x}; \boldsymbol{\theta}, \mathbf{g}) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} h(x_i, \theta_k)\, g_k \right), \tag{OA.18}$$

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 i^{th} pair of spells comes from the type k, for all $i = 1, ..., N$ and $k = 1, ..., K$, as

$$z_{k,i}\,(x_i; \theta, g) \equiv \frac{h(x_i, \theta_k)\, g_k}{\sum_{k'=1}^{K} h(x_i, \theta_{k'})\, g'_k}. \tag{OA.19}$$

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

$$0 = \sum_{i=1}^{N} z_{k,i}\,(x_i;\theta,g)\;\;\frac{\partial \log h(x_i,\theta_k)}{\partial \theta_k} \tag{OA.20}$$

$$g_k = \frac{\sum_{i=1}^{N} z_{k,i}\,(x_i;\theta,g)}{\sum_{k'=1}^{K}\sum_{i=1}^{N} z_{k',i}\,(x_i;\theta,g)} \tag{OA.21}$$

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 $z^m_{k,i}$ as :

$$z^m_{k,i} = \frac{h(x_i,\theta^m_k)\,g^m_k}{\sum_{k'=1}^{K} h(x_i,\theta^m_{k'})\,g^m_{k'}} \quad \text{for all } k=1,...,K\,. \tag{OA.22}$$

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

$$0 = \sum_{i=1}^{N} z^m_{k,i}\;\frac{\partial \log h(x_i,\theta^{m+1}_k)}{\partial \theta_k}\,, \tag{OA.23}$$

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

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

$$g^{m+1}_k = \frac{\sum_{i=1}^{N} z^m_{k,i}}{\sum_{k'=1}^{K}\sum_{i=1}^{N} z^m_{k',i}}\,. \tag{OA.24}$$

OA.E.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 $t(i), i=1,\ldots N$ come from an inverse Gaussian distribution. A straightforward algebra leads to

$$\hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} t(i) = E[t], \quad \hat{\sigma^2}_{MLE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{t(i)} - \frac{1}{N}\sum_{i=1}^{N} t(i) = E[\frac{1}{t}] - E[t]. \tag{OA.25}$$

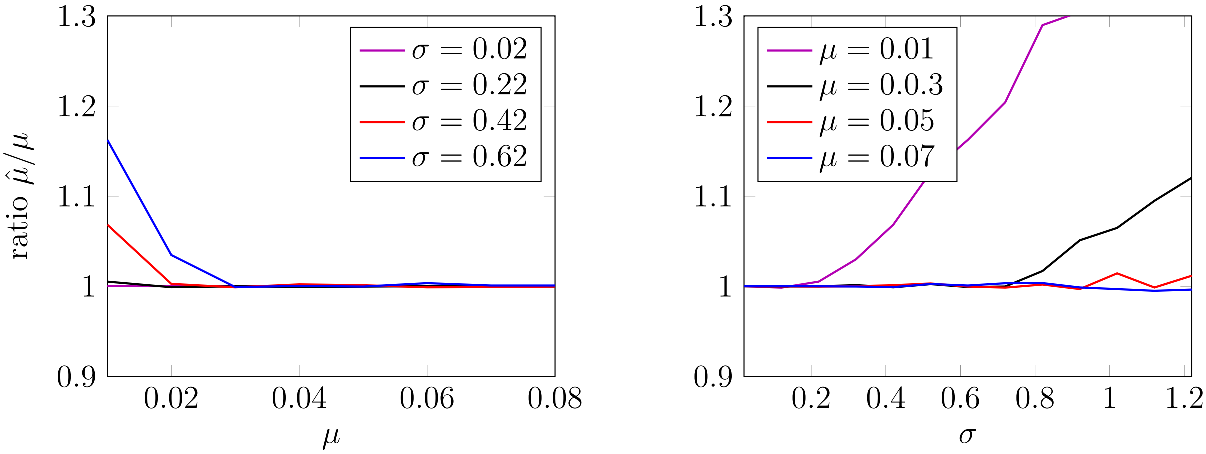

Figure 14: Maximum likelihood estimates of $\hat{\mu}$, relative to the true value of μ. The figure shows the ML estimate of μ using data on 1 million spells generated from a single type inverse Gaussian distribution with parameters $(\mu, \sigma) \in [0.01, 0.08] \times [0.02, 1.2]$. We show estimated $\hat{\mu}$ relative to the true μ as a function of μ in the left panel, or as a function of σ in the right panel.

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 14. 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

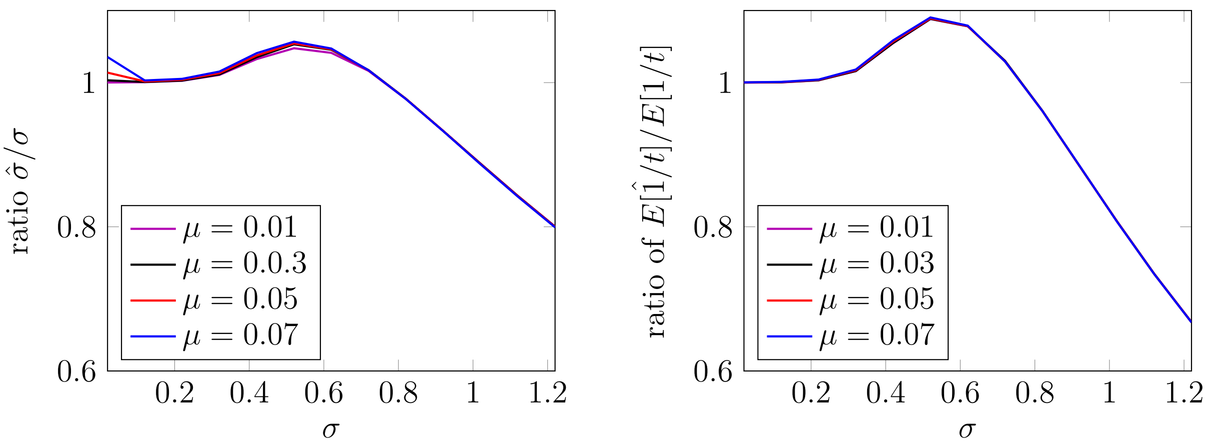

Figure 15: Maximum likelihood estimates of σ (left panel) and of the mean of $1/t$ (right panel) using discretized data, relative to their true values. We used 1,000,000 spells simulated from a single inverse Gaussian distribution with parameters $(\mu, \sigma) \in [0.01, 0.08] \times [0.02, 1.2]$. The ratios are plotted as a function of the true σ. Each line corresponds one value of $\mu \in [0.01, 0.08]$.

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 measurement introduces by treating the simulated data as if they were measured in discrete times. We find that this measurement affects estimates of σ, see the left panel of Figure 15, 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 larger for larger values of σ, see the right panel of Figure 15. 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.

	EM with 152 types				EM with 77 types			
	mean	median	st.dev.	min	mean	median	st.dev	min
α	508.770	0.138	3123.989	0.096	527.173	0.139	3146.488	0.097
β	3512.460	5.735	18390.764	1.465	3724.065	5.744	18916.972	1.466
$\frac{\mu_n}{\bar{\omega}-\underline{\omega}}$	0.052	0.043	0.033	0.018	0.052	0.044	0.033	0.018
$\frac{\sigma_n}{\bar{\omega}-\underline{\omega}}$	0.231	0.174	0.155	10^{-5}	0.230	0.174	0.155	10^{-5}

Table 3: Comparison of summary statistics with 152 and 77 types.

OA.F Akaike Information Criterion

One might worry that our model is over-parametrized and hence recovers too many types. We address this concern by using Akaike information criterion to reduce the number of types and compare the decomposition with fewer types to our main results.

We proceed as follows. Starting from the our estimated distribution G^+ with $K = 152$ types, characterized by pairs (α_k, β_k) and weights g_k, for $k = 1, \dots K$, we construct a new distribution with fewer types by eliminating $m > 0$ types with the lowest weights g_k. We use this new distribution as an initial guess for the EM algorithm, which further refines the estimates. We compute the Akaike information criterion for these estimates. We repeat this exercise for $m = 1, \dots K - 1$ and choose the value of m which minimizes the criterion. The minimum corresponds to a distribution with 77 types.

Table 3 and Figure 16 show compare summary statistics for the distribution and the decomposition using the distirbution with 152 or 77 types. The take away from this exercise is that the big role of heterogeneity which we recover does not hinge on the large number of types.

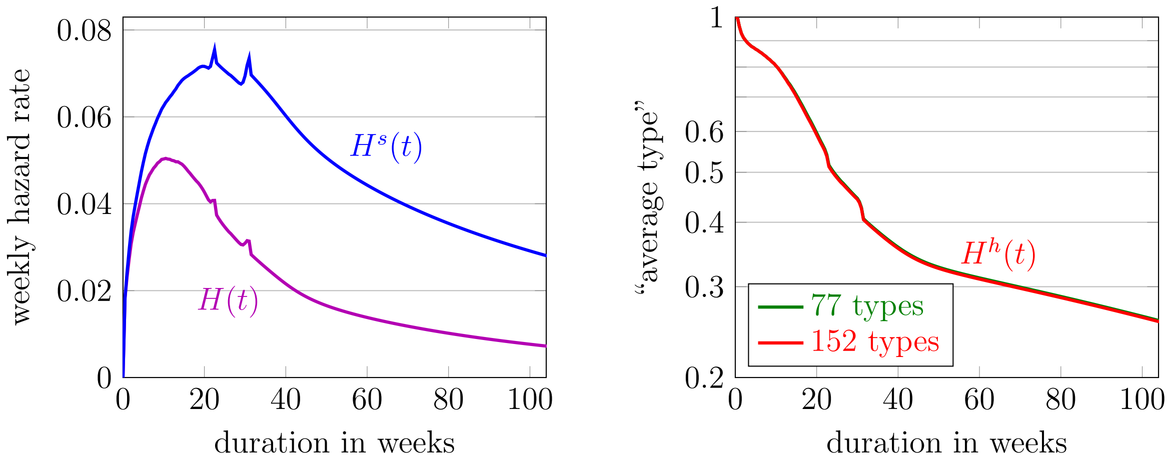

Figure 16: Decomposition of the hazard rate with 77 types. In the left panel, the purple line shows the raw hazard rate, $H(t)$, and the blue line shows the structural hazard rate, $H^s(t)$. The right panel shows the contribution of heterogeneity, $H^h(t)$. The green line shows the model with 77 types and the red line shows the model with 152 types, in both cases for the distribution $\bar{G}$.