Decomposing Duration Dependence in the Job Finding Rate in a Stopping Time Model

Fernando Alvarez  Katarína Borovičková  Robert Shimer
University of Chicago  New York University  University of Chicago

November 20, 2014
Preliminary

Abstract

We develop a simple dynamic model of a worker’s transitions between employment and nonemployment. Our model implies that a worker finds a job at an optimal stopping time, when a Brownian motion with drift hits a barrier. The model has structural duration dependence in the job finding rate, in the sense that the hazard rate of finding a job changes during a nonemployment spell for a given worker. In addition, we allow for arbitrary parameter heterogeneity across workers, so dynamic selection also affects the average job finding rate at different durations. We show that our model has testable implications if we observe at least two completed nonemployment spells for each worker. Moreover, we can nonparametrically identify the distribution of a subset of our model’s parameters using data on the duration of repeated nonemployment spells. We use a large panel of social security data for Austrian workers to test and estimate the model. Our model is not rejected by the data, while a mixed proportional hazard model with arbitrary heterogeneity and an arbitrary baseline hazard rate is rejected using the same data set. Our preliminary parameter estimates indicate that dynamic selection is important for the decline in the job finding rate at short durations, while structural duration dependence drives most of the decline in the job finding rate at long durations.
1 Introduction

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

Our analysis is built around a structural model which views finding a job as an optimal stopping problem. One interpretation of our structural model is a classical theory of employment. All individuals always have two options, working at some wage $w(t)$ or not working and receiving some income and utility from leisure $b(t)$. The difference between these values is persistent but changes over time. If there were no cost of switching employment status, an individual would work if and only if the wage exceeds the value of not working.\footnote{In our model, we allow the evolution of the wage to depend on a worker’s employment status, which means that a worker may choose to work even when the static net benefit from employment is negative because working raises future wages. In this more general setup, an individual works whenever the net benefit from employment exceeds some threshold.} We add a switching cost to this simple model, so a worker starts working when the difference between the wage and the value of leisure is sufficiently large and stops working when the difference is sufficiently negative. Given a specification of the individual’s preferences, a level of the switching cost, and the stochastic process for the wage and nonemployment income, this theory generates a structural model of duration dependence for any individual worker. The theory is sufficiently general that the expected residual duration of a nonemployment spell may increase or decrease during the spell.

An alternative interpretation of our structural model is a classical theory of unemployment. According to this interpretation, a worker’s productivity $p(t)$ and her wage $w(t)$ follow a stochastic process. Again, the difference is persistent but changes over time. If the worker is unemployed, a monopsonist has the option of employing the worker, earning flow profits $p(t) - w(t)$, by paying a fixed cost. It pays the cost if flow profits are sufficiently positive and fires the worker if flow profits are sufficiently negative. Once again, 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 preferences, fixed costs, and stochastic processes. For example, some individuals may expect the residual duration of their nonemployment spell to increase the longer they stay out of work while others may
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 nonemployment spell; and each individual starts working when the net benefit exceeds some fixed threshold and stops working when it falls below some (weakly) lower threshold. In the first interpretation of our structural model, this threshold is determined by the worker while in the second interpretation it is determined by the firm. These assumptions imply that the duration of a nonemployment spell is given by the first passage time of a Brownian motion with drift, a random variable with an inverse Gaussian distribution. The parameters of the inverse Gaussian distribution are fixed over time for each individual but may vary arbitrarily across individuals.

In this environment, we ask three 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 nonemployment spells can identify all the even moments of the joint distribution of the parameters of the inverse Gaussian distribution. Our proof, which relies on the functional form of the inverse Gaussian distribution, establishes that the $m^{th}$ partial derivatives of the joint distribution of the spells identifies the $2m^{th}$ moments of the joint distribution of the parameters. Our inability to identify the odd moments implies that we cannot identify the sign of the drift of the underlying Brownian motion, but beyond this important limitation, we can identify the distribution of inverse Gaussian duration distributions.

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

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

We then use data from the Austrian social security registry from 1972–2007 to test our model, nonparametrically estimate the distribution of unobserved parameters, and evaluate the decomposition. Using data on 1.7 million individuals who experience at least two non-
employment spells, we find that we cannot reject our model and we uncover substantial heterogeneity across individuals. Our other results are still preliminary.

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

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

Abbring (2012) considers a more general model than ours, allowing that the latent net benefit from employment is 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, which implies that his model falls within the accelerated failure time class (Kalbfleisch and Prentice, 1980). In contrast, we allow for two dimensions of heterogeneity, taking us outside the class of accelerated failure time models, and so our approach to nonparametric identification is completely different.

Within economics, the mixed proportional hazard model (Lancaster, 1979) has received far more attention than the first passage time model. This model assumes that the probability of finding a job at duration $t$ is the product of three terms: a baseline hazard rate that varies depending on the duration of nonemployment, a function of observable characteristics of individuals, and an unobservable characteristic. Our model neither nests the mixed proportional hazard model nor is it nested by that model. A large literature, starting with Elbers and Ridder (1982) and Heckman and Singer (1984a), show that such a model is non-parametrically identified using a single spell of nonemployment and appropriate variation.
in the observable characteristics of individuals. Heckman and Singer (1984b) illustrates the
perils of parametric identification strategies in this context.

Closer to the spirit of our paper, Honoré (1993) shows that the mixed proportional hazard
model is also nonparametrically identified with data on the duration of at least two nonem-
ployment spells for each individual. Indeed, in Section 5 we build on the analysis in Honoré
(1993) to construct a nonparametric test of the mixed proportional hazard model using the
same Austrian social security registry data on the duration of two nonemployment spells.
While we cannot reject our model, we do reject the mixed proportional hazard model.

The remainder of the paper proceeds as follows. In Section 2, we describe our structural
model and show how to use the model to address the questions of interest. We prove that
a subset of the parameters is nonparametrically identified if we observe at least two nonem-
ployment spells for each individual, that the model has testable implications under the same
conditions, and that we can use the model to decompose changes in the hazard of exiting
nonemployment and the residual duration of a nonemployment spell into the portion that is
structural and the portion that is attributable to changes in the composition of the nonem-
ployment pool. Section 3 summarizes the Austrian social security registry data. Section 4
presents our results. Our model is not rejected by the Austrian data. Our preliminary results
indicate that most of the decline in the hazard rate and much of the increase in the residual
duration of a nonemployment spell at longer nonemployment durations is due to structural
duration dependence rather than changes in the composition of the nonemployed population.
Finally, Section 5 develops a nonparametric test of the mixed proportional hazard model and
uses the same Austrian data set to perform the test. That model is rejected by the data.

2 Theory

2.1 Structural Model

We consider the problem of a worker who can either be employed, \( s(t) = e \), or nonemployed,
\( s(t) = n \), at each instant in continuous time \( t \). A nonemployed worker gets a flow benefit \( b \)
while an employed worker earns a wage \( w(t) \). The natural logarithm of the potential wage,
\( \omega(t) \equiv \log w(t) \), follows a random walk with drift both when the worker is employed and
when the worker is nonemployed. The drift and standard deviation may depend on the
worker’s employment status:

\[
d\omega(t) = \mu_{s(t)}dt + \sigma_{s(t)}dB(t),
\]  

(1)
where $B(t)$ is a standard Brownian motion. The worker’s state at time $t$ is her employment status $s(t)$ and her (potential log) wage $\omega(t)$.

A nonemployed worker can become employed at $t$, obtaining a wage $e^{\omega(t)}$, by paying a fixed cost $\psi$. An employed worker earns $e^{\omega(t)}$ and can become nonemployed without paying any cost. The worker is risk-neutral and has a discount rate $\rho > 0$. She must decide optimally when to change her employment status $s$.

In order for the problem to be well-behaved, we impose

$$\rho > \mu_s + \sigma_s^2/2 \text{ for } s = n, e.$$  \hfill (2)

This ensures that the worker’s problem has finite value. If $\rho < \mu_e + \sigma_e^2/2$, the expected value of working forever would be infinite. If $\rho < \mu_n + \sigma_n^2/2$, the expected value of not working for $T$ periods and then working forever would grow without bound as $T$ increases.

The flow net benefit from employment is the difference between the potential wage $e^{\omega(t)}$ and the income and utility while nonemployed $b$. We assume for expositional purposes that fluctuations in the net benefit from employment are driven entirely by fluctuations in the potential wage, but expect that in reality changes in both the potential wage and the income and utility from nonemployment drive changes in the net benefit from employment. Our key results on testing and identification do not depend on this assumption. Nevertheless, we do not use wage data in our analysis because we believe that nonemployment duration is driven by fluctuations in both the potential wage and the utility from nonemployment. We also stress that any welfare conclusions, e.g. about whether employed or nonemployed workers are better off, are sensitive to the assumption that only the wage fluctuates over time.

Under condition (2), the worker’s optimal policy involves a pair of thresholds; see Appendix A. If $s(t) = e$ and $\omega(t) \geq \bar{\omega}$, the worker remains employed, while she stops working the first time $\omega(t) < \bar{\omega}$. If $s(t) = n$ and $\omega(t) \leq \bar{\omega}$, the worker remains nonemployed, while she takes a job the first time $\omega(t) > \bar{\omega}$. Assuming the fixed cost $\psi$ is strictly positive, the thresholds satisfy $\bar{\omega} > \omega$, while the thresholds are equal if the fixed cost is zero.

We have so far described a model of voluntary nonemployment, 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 fixed at $b$, while a worker’s productivity $p(t) = e^{\omega(t)}$ follows a geometric Brownian motion with drift. If the worker is employed by a monopsonist, it earns flow profits $p(t) - b$. If the worker is unemployed, a firm may hire her by paying a fixed cost $\psi$. In this case, the firm’s optimal policy involves the same pair of thresholds. If $s(t) = e$ and $\omega(t) \geq \bar{\omega}$, the firm retains the worker, while she is fired the first time $\omega(t) < \bar{\omega}$. If $s(t) = n$ and $\omega(t) \leq \bar{\omega}$, the worker
remains unemployed, while a firm hires her the first time \( \omega(t) > \bar{\omega} \).

This structural model is similar to the one in Alvarez and Shimer (2011) and Shimer (2008). In particular, setting the switching cost to zero \( (\psi = 0) \) gives a decision rule with \( \bar{\omega} = \bar{\omega}_0 \), as in the version of Alvarez and Shimer (2011) with only rest unemployment, and with the same implication for nonemployment 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 most important difference is that this paper allows for arbitrary time-invariant worker heterogeneity. An individual worker is described by seven structural parameters: her discount rate \( \rho \), her fixed cost \( \psi \), her nonemployment benefit \( b \), and the four parameters governing the stochastic process of the potential wage, \( \mu_e, \mu_n, \sigma_e, \) and \( \sigma_n \). These in turn determine two reduced-form parameters, the thresholds \( \bar{\omega} \) and \( \omega \). We allow for arbitrary distributions of the seven structural parameters in the population, subject only to the constraints in condition (2).

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

\[
f(t, \alpha, \beta) = \frac{\beta}{\sqrt{2\pi t^{3/2}}} \exp\left(-\frac{(\alpha t - \beta)^2}{2t}\right),
\]

where \( \alpha \equiv \mu_n/\sigma_n \) and \( \beta \equiv (\bar{\omega} - \omega)/\sigma_n \). Note \( \beta \geq 0 \) by assumption, while \( \alpha \) may be positive or negative. If \( \alpha \geq 0 \), \( \int_0^\infty f(t, \alpha, \beta) dt = 1 \), so a worker almost surely returns to work. But if \( \alpha < 0 \), the probability of eventually returning to work is \( e^{2\alpha\beta} < 1 \), so there is a probability the worker never finds a job.

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

\footnote{For expositional purposes, we will call \( b \) the nonemployment benefit and \( \omega(t) \) the log wage throughout the remainder of the paper, but the alternative interpretation of the model is equally valid.}
The model is therefore consistent with both positive and negative duration dependence in the structural exit rate from nonemployment.

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

We have three goals, which we describe in turn in the next three subsections. The first is to understand whether the joint distribution of \(\alpha\) and \(\beta\) in the population, \(G(\alpha, \beta)\), is nonparametrically identified using nonemployment duration data. The distribution of these reduced-form parameters reflects the underlying joint distribution of the seven structural parameters, but only these two reduced form parameters affect nonemployment duration and so only the distribution of these two parameters can possibly be identified using nonemployment duration data. The second is understand whether our model is testable. If we allow for an arbitrary distribution of the seven structural parameters in the model, are there nonemployment duration data that are inconsistent with our theory? The third is to examine how the estimated joint distribution of \(\alpha\) and \(\beta\) can be used to decompose the overall evolution of the hazard of exiting nonemployment into two components: the portion attributable to changes in the hazard for each individual worker as nonemployment duration changes, and the portion attributable to changes in the population of nonemployed workers at different durations.

In performing this analysis, we assume that the reduced-form parameters \(\alpha\) and \(\beta\) are fixed over time for each worker. In principle, variation in these parameters across workers may reflect some time-invariant observable characteristics of the workers or it may reflect time-invariant unobserved heterogeneity. We do not attempt to distinguish between these two possibilities. Our analysis precludes the possibility of time-varying heterogeneity. For example, a worker’s experience cannot affect the stochastic process for the potential wage while nonemployed \((\mu_n, \sigma_n)\), nor can it affect the search cost \(\psi\). Note, however, that our model does allow for learning-by-doing, since a worker’s wage may increase faster on average when employed than when nonemployed, \(\mu_e > \mu_n\).

### 2.2 Nonparametric Identification

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

Instead, our approach to identification uses the joint density of the duration of two nonemployment spells. In the example in the previous paragraph, the true model would generate no correlation between the realized duration of two spells, while the alternative model would imply that the two spells are of equal duration. Thus repeated spells opens up the possibility of identification.

Let \(\phi(t_1, t_2)\) denote the density of the distribution of the duration of the two completed nonemployment spells:

\[
\phi(t_1, t_2) = k^{-1} \int \int f(t_1, \alpha, \beta)f(t_2, \alpha, \beta)dG(\alpha, \beta),
\]

where \(k\) is the fraction of individuals who complete two spells,

\[
k \equiv \int \int e^{4 \min\{0, \alpha\} \beta} dG(\alpha, \beta).
\]

Note that \(f(\cdot, \alpha, \beta)\) is a smooth function, so \(\phi\) is smooth as well even if \(G\) has mass points.

Before proceeding with our identification result, we note a limitation of it: the sign of \(\alpha\) is not identified. That is, take any distribution \(G\) that puts positive mass on negative values of \(\alpha\). Construct a new density \(G^+\) by taking the absolute value of \(\alpha\):

\[
G^+(\alpha, \beta) = G(\alpha, \beta) - G(-\alpha, \beta)
\]

for all \(\alpha \geq 0\) and \(G^+(\alpha, \beta) = 0\) for all \(\alpha < 0\). Then \(\phi(t_1, t_2)\) is identical under \(G\) and \(G^+\) since \(f(t, \alpha, \beta) = e^{2\alpha \beta}f(t, -\alpha, \beta)\) for all \((\alpha, \beta)\).

Our main identification result is that for any \(t_1 \neq t_2\), the joint density of spell lengths \(\phi\) identifies all the moments of the joint distribution of characteristics \((\alpha^2, \beta^2)\). Since \(\beta\) is always positive and we have just noted that we cannot identify the sign of \(\alpha\), this effectively means all the moments are identified.

Our approach is constructive. For any \(m \in \{1, 2, \ldots\}\), the joint distribution \(\phi\) has \(m + 1\) partial derivatives. Using linear algebra and induction, we first prove that those partial

\[3\]If the nonemployment duration distribution is ever 0, it must be the case that \(\alpha\) and \(\beta\) are infinite for all workers, pinning down the distribution. In any empirical application with a finite sample of data, the realized density may be zero at some durations even if \(\alpha\) and \(\beta\) are finite.
derivatives uniquely identify the $m + 1$ $m^{th}$ moments of the joint distribution of $(\alpha^2, \beta^2)$ among individuals whose two durations are $(t_1, t_2)$. We then use those conditional moments to find the moments of the unconditional joint distribution of $(\alpha^2, \beta^2)$.

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

$$\frac{\partial \phi(t_1, t_2)}{\partial t_i} = k^{-1} \int \int \left( \frac{\beta^2}{2t_i^2} - \frac{3}{2t_i} - \frac{\alpha^2}{2} \right) f(t_1, \alpha, \beta)f(t_2, \alpha, \beta) dG(\alpha, \beta)$$

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 the constants $\mathbb{E}(\alpha^2|t_1, t_2)$ and $\mathbb{E}(\beta^2|t_1, t_2)$ are the expected values of $\alpha^2$ and $\beta^2$ among all workers who find their jobs at durations $(t_1, t_2)$:

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

and

$$\mathbb{E}(\beta^2|t_1, t_2) = \frac{\int \int \beta^2 f(t_1, \alpha, \beta)f(t_2, \alpha, \beta) dG(\alpha, \beta)}{\int \int f(t_1, \alpha, \beta)f(t_2, \alpha, \beta) dG(\alpha, \beta)}.$$

For any $t_1 \neq t_2$, we can solve these equations for these two expected values:

$$\mathbb{E}(\alpha^2|t_1, t_2) = \frac{2(t_2^2 \phi_2(t_1, t_2) - t_1^2 \phi_1(t_1, t_2))}{\phi(t_1, t_2)(t_1^2 - t_2^2)} - \frac{3}{t_1 + t_2}$$

(4)

and

$$\mathbb{E}(\beta^2|t_1, t_2) = t_1 t_2 \left( \frac{2t_1 t_2 (\phi_2(t_1, t_2) - \phi_1(t_1, t_2))}{\phi(t_1, t_2)(t_1^2 - t_2^2)} + \frac{3}{t_1 + t_2} \right).$$

(5)

Moreover, we can evaluate this at $t_1 = t_2$ by taking limits.\footnote{This limit is correct unless there are some individuals with deterministic nonemployment duration, i.e. $\sigma_n = 0$ and $\bar{\omega} - \underline{\omega}/\mu_n = t_1 = t_2$. This would imply $\beta = \infty$ for such an individual, a possibility that our notation precludes.}

By integrating over the density $\phi$, we can recover the unconditional first moments of the distribution of $\alpha^2$ and $\beta^2$:

$$\mathbb{E}(\alpha^2) = \int \int \mathbb{E}(\alpha^2|t_1, t_2) \phi(t_1, t_2) \, dt_1 \, dt_2$$

$$\mathbb{E}(\beta^2) = \int \int \mathbb{E}(\beta^2|t_1, t_2) \phi(t_1, t_2) \, dt_1 \, dt_2.$$
Next consider \( m = 2 \). The second partial derivatives satisfy

\[
\frac{4t^4_i}{\phi(t_1, t_2)} \frac{\partial^2 \phi(t_1, t_2)}{\partial t_i^2} = \mathbb{E}(\alpha^4|t_1, t_2)t_i^4 + \mathbb{E}(\beta^4|t_1, t_2) - 2\mathbb{E}(\alpha^2\beta^2|t_1, t_2)t_i^2 \\
+ 6\mathbb{E}(\alpha^2|t_1, t_2)t_i^3 - 10\mathbb{E}(\beta^2|t_1, t_2)t_i + 15t_i^2
\]

and for \( i = 1, 2 \) and

\[
\frac{4t_1^2t_2^2}{\phi(t_1, t_2)} \frac{\partial^2 \phi(t_1, t_2)}{\partial t_1 \partial t_2} = \mathbb{E}(\alpha^4|t_1, t_2)t_1^2t_2^2 + \mathbb{E}(\beta^4|t_1, t_2) - \mathbb{E}(\alpha^2\beta^2|t_1, t_2)(t_1^2 + t_2^2) \\
+ 3\mathbb{E}(\alpha^2|t_1, t_2)t_1t_2(t_1 + t_2) - 3\mathbb{E}(\beta^2|t_1, t_2)(t_1 + t_2) + 9t_1t_2.
\]

For any \( t_1 \neq t_2 \), this gives three equations in three unknowns, the expected values of \( \alpha^4 \), \( \beta^4 \), and \( \alpha^2\beta^2 \) conditional on realized duration \((t_1, t_2)\), in addition to the known moments \( \mathbb{E}(\alpha^2|t_1, t_2) \) and \( \mathbb{E}(\beta^2|t_1, t_2) \). Again assuming \( t_1 \neq t_2 \), these equations can be solved explicitly for the three second moments. Integrating over the density \( \phi \) gives us the three unconditional second moments.

As \( m \) increases, this algebra becomes increasingly messy, but the approach conceptually remains unchanged. There are \( m + 1 \) \( m^{th} \) partial derivatives of \( \phi(t_1, t_2) \). These are linear functions of the \( m + 1 \) \( m^{th} \) moments of the joint distribution of \( \alpha^2 \) and \( \beta^2 \) among workers who find jobs at durations \((t_1, t_2)\), as well as lower moments of the joint distribution. Moreover, the linear functions can be expressed as an LU decomposition:

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

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

\[
L_{ij}(t_1, t_2) = \frac{(-1)^j(m - i)!}{(m - j)!(j - i)!} t_1^{2(j-i)}(t_1^2 - t_2^2)^{i/2}
\]

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

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

for \(0 \leq i \leq j \leq m\) and \(U_{ij}(t_1, t_2) = 0\) for \(0 \leq j < i \leq m\); and \(v_m(t_1, t_2)\) is a vector that depends only on \((m - 1)^{st}\) and lower moments of the joint distribution, each of which we have found in previous steps. It is easy to verify that the diagonal elements of \(L\) and \(U\) are nonzero if and only if \(t_1 \neq t_2\) and so it is possible to invert this expression to solve for the \(m^{th}\) moments of the joint distribution. Proceeding by induction, we can thus characterize all the moments of the joint distribution of \((\alpha^2, \beta^2)\) conditional on realized duration \((t_1, t_2)\). And at each step we can integrate over the distribution \(\phi(t_1, t_2)\) to recover the unconditional \(m^{th}\) moments of the joint distribution. Thus all the moments are identified.

### 2.3 Testable Implications

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

Our approach to identification yields the model’s overidentifying restrictions. First, we can reject the model if \(\phi\) is not infinitely differentiable at all \((t_1, t_2)\) with \(t_1 \neq t_2\). Second, we can reject the model if any of the estimated even-powered moments is negative, \(\mathbb{E}(\alpha^{2m}\beta^{2n}|t_1, t_2) < 0\) for any \(m, n\), and \(t_1 \neq t_2\). For this overidentifying restriction, we measure each of these moments using the LU decomposition equation (6), which again requires data on \(\phi\) and its partial derivatives. Third, all the moments are related to each other, as in the Stieltjes moment problem. For example, if \(\mathbb{E}(\alpha^2|t_1, t_2) = 0\) at some \((t_1, t_2)\), then \(\mathbb{E}(\alpha^4|t_1, t_2) = 0\) as well.

In practice, measuring higher moments can be difficult and so we focus on the simplest restriction that comes from the model, \(\mathbb{E}(\alpha^2|t_1, t_2) > 0\) and \(\mathbb{E}(\beta^2|t_1, t_2) > 0\) for all \(t_1 \neq t_2\). Even this test has considerable power. We illustrate this with two examples.

First, consider the canonical search model where the hazard of finding a job is a constant \(h\) and so the density of completed spells is \(\phi(t_1, t_2) = h^2e^{-h(t_1 + t_2)}\). Then applying equations (4) and (5) gives

\[ \mathbb{E}(\alpha^2|t_1, t_2) = 2h - \frac{3}{t_1 + t_2} \quad \text{and} \quad \mathbb{E}(\beta^2|t_1, t_2) = \frac{3t_1t_2}{t_1 + t_2}. \]
In particular, \( \mathbb{E}(\alpha^2|t_1, t_2) < 0 \) whenever \( t_1 + t_2 < 3/2h \), where \( 1/h \) represents the mean duration of a nonemployment spell, a contradiction. Our model cannot generate this density of completed spells for any joint distribution of parameters.

More generally, suppose the constant hazard \( h \) has a distribution \( \tilde{G} \) in the population, so the density of completed spells is \( \phi(t_1, t_2) = \int h^2 e^{-h(t_1+t_2)} d\tilde{G}(h) \). Then
\[
\mathbb{E}(\alpha^2|t_1, t_2) = 2\int \frac{h^3 e^{-h(t_1+t_2)} dG(h)}{\int h^2 e^{-h(t_1+t_2)} d\tilde{G}(h)} - \frac{3}{t_1 + t_2},
\]
while \( \mathbb{E}(\beta^2|t_1, t_2) \) is unchanged. If the ratio of the third moment of \( h \) to the second moment is finite—for example, if the support of the distribution \( \tilde{G} \) is compact, 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 \( \phi \) is decreasing, while the density of a random variable with an Inverse Gaussian distribution is hump-shaped. This is not the case. Our second example shows that this is not the case. Suppose that the density of durations is log-normally distributed with mean \( \mu \) and standard deviation \( \sigma \). For each individual, we observe two draws from this distribution and test the model using equations (4) and (5). Then our approach implies
\[
\mathbb{E}(\alpha^2|t_1, t_2) = \frac{2}{\sigma^2(t_1+t_2)} \left( \frac{t_1 \log t_1 - t_2 \log t_2}{t_1 - t_2} - \left( \mu + \frac{1}{2} \sigma^2 \right) \right),
\]
and
\[
\mathbb{E}(\beta^2|t_1, t_2) = \frac{2t_1 t_2}{\sigma^2(t_1+t_2)} \left( \frac{t_2 \log t_1 - t_1 \log t_2}{t_1 - t_2} + \left( \mu + \frac{1}{2} \sigma^2 \right) \right).
\]

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 small. Therefore, for any \( \mu \) and \( \sigma > 0 \), \( \mathbb{E}(\alpha^2|t_1, t_2) \) is negative at some values of \((t_1, t_2)\). Similarly, \( \frac{t_2 \log t_1 - t_1 \log t_2}{t_1 - t_2} \) is decreasing in \( t_1 \) and \( t_2 \), converging to minus infinity when \( t_1 \) and \( t_2 \) are sufficiently large. Therefore, for any \( \mu \) and \( \sigma > 0 \), \( \mathbb{E}(\beta^2|t_1, t_2) \) is also negative for some values of \((t_1, t_2)\).

Finally, the same logic implies that the mixture of log-normally distributed random variables generates a joint density \( \phi \) 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 equations (4) and (5) would never confuse a mixture of log normal distributions with a mixture of inverse Gaussian distributions.
2.4 Decomposition of Duration Dependence

We turn now to the relative importance of structural duration dependence and dynamic selection for the evolution of the hazard rate of exiting non-employment. We assume that \( \alpha \geq 0 \) for all individuals, or alternatively that if \( \alpha < 0 \) for some individuals, we only focus on the subset who complete two non-employment spells. The limitations of our identification prevent us from relaxing this restriction. In addition, it will be convenient for notation (but not necessary) to assume that the type distribution \( G \) has a density \( g \).

2.4.1 Decomposition of the Hazard Rate

We decompose the aggregate hazard rate of separation using a Divisia index. Let \( h(t, \alpha, \beta) \) denote the hazard rate for type \((\alpha, \beta)\) at duration \( t \),

\[
h(t, \alpha, \beta) = \frac{f(t, \alpha, \beta)}{1 - F(t, \alpha, \beta)}. \tag{7}
\]

Also let \( g(t, \alpha, \beta) \) denote the density of the type distribution among individuals who are out of work after \( t \) periods,

\[
g(t, \alpha, \beta) = \frac{(1 - F(t, \alpha, \beta))g(\alpha, \beta)}{\iint (1 - F(t, \alpha', \beta'))dG(\alpha', \beta')}.
\]

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

\[
H(t) = \frac{\iint f(t, \alpha, \beta)dG(\alpha, \beta)}{\iint (1 - F(t, \alpha', \beta'))dG(\alpha', \beta')} = \iint \frac{f(t, \alpha, \beta)}{(1 - F(t, \alpha, \beta))} \frac{(1 - F(t, \alpha, \beta))g(\alpha, \beta)}{\iint (1 - F(t, \alpha', \beta'))dG(\alpha', \beta')} d\alpha d\beta = \iint h(t, \alpha, \beta)g(t, \alpha, \beta) d\alpha d\beta.
\]

Taking a derivative with respect to \( t \),

\[
\dot{H}(t) = \dot{H}^s(t) + \dot{H}^h(t),
\]
where

\[
\begin{align*}
\dot{H}^s(t) &= \int\int \dot{h}(t, \alpha, \beta) g(t, \alpha, \beta) d\alpha d\beta \\
\dot{H}^h(t) &= \int\int h(t, \alpha, \beta) \dot{g}(t, \alpha, \beta) d\alpha d\beta.
\end{align*}
\]

We interpret the term \(\dot{H}^s(t)\) as the instantaneous contribution of structural duration dependence since it is based on the change in the hazard rates of individual worker types. Observe that if the hazard rate were constant (and thus there were not structural duration dependence), this term would be zero. The second term \(\dot{H}^h(t)\) captures the instantaneous role of heterogeneity because it captures how the distribution of worker types changes with unemployment duration.

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

\[
\dot{H}^h(t) = - \int\int (h(t, \alpha, \beta) - H(t))^2 g(t, \alpha, \beta) d\alpha d\beta < 0. \tag{8}
\]

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.” Intuitively, types with a higher than average hazard rate are always declining as a share of the population.

To derive this result, first take logs and differentiate \(g(t, \alpha, \beta)\):

\[
\frac{\dot{g}(t, \alpha, \beta)}{g(t, \alpha, \beta)} = - \frac{f(t, \alpha, \beta)}{1 - F(t, \alpha, \beta)} + \int\int 
\frac{f(t, \alpha', \beta') dG(\alpha', \beta')}{(1 - F(t, \alpha', \beta')) dG(\alpha', \beta')} = -h(t, \alpha, \beta) + H(t).
\]

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

\[
\dot{H}^h(t) = - \int\int h(t, \alpha, \beta)(h(t, \alpha, \beta) - H(t)) dG(\alpha, \beta).
\]

Since \(\int\int (h(t, \alpha, \beta) - H(t)) dG(\alpha, \beta) = 0\), we can add \(H(t)\) times this to the previous expression to get the formula in equation (8).

\[\text{We are grateful to Jörgen Weibull for pointing out this connection to us.}\]
2.4.2 Decomposition of the Expected Residual Duration

We can perform a similar exercise for expected residual nonemployment duration. Conditional on surviving until \( t \), the expected residual duration for the type \((\alpha, \beta)\) is

\[
r(t, \alpha, \beta) = \frac{\int_t^{\infty} (s-t) f(s, \alpha, \beta) ds}{1 - F(t, \alpha, \beta)}.
\]

Residual duration is effectively the integral over future hazard rates. To see this, use integration by parts to rewrite this as

\[
r(t, \alpha, \beta) = \frac{\int_t^{\infty} (1 - F(s, \alpha, \beta)) ds}{1 - F(t, \alpha, \beta)}.
\]

Moreover, the definition of the hazard rate (7) implies

\[
F(t, \alpha, \beta) = 1 - e^{-\int_0^t h(\tau, \alpha, \beta) d\tau}.
\]

Combining these last two equations gives

\[
r(t, \alpha, \beta) = \int_t^{\infty} e^{-\int_t^{\tau} h(s, \alpha, \beta) ds} d\tau
\]

Thus if all future hazard rates are lower for one type, that type has a longer residual duration.

The aggregate expected residual duration at time \( t \) integrates type-specific residual durations across types at duration \( t \),

\[
R(t) = \int\int r(t, \alpha, \beta) g(t, \alpha, \beta) d\alpha d\beta.
\]

We again apply a Divisia index to get

\[
\dot{R}(t) = \dot{R}^s(t) + \dot{R}^h(t),
\]

where

\[
\dot{R}^s(t) = \int\int \dot{r}(t, \alpha, \beta) g(t, \alpha, \beta) d\alpha d\beta
\]

\[
\dot{R}^h(t) = \int\int r(t, \alpha, \beta) \dot{g}(t, \alpha, \beta) d\alpha d\beta.
\]

Similarly as before, the term \( \dot{R}^s(t) \) captures the role of the structural duration dependence, measuring the change in the expected nonemployment duration in response to a
marginal increase in the current duration. If there were no duration dependence, then $\dot{R}_s(t) = 0$ for all $t$. The term $\dot{R}_h(t)$ expresses the role of heterogeneity as it shows how the expected residual duration changes due to changes in the composition of the non-employment pool. If there were no heterogeneity, then $\dot{R}_h(t) = 0$ for all $t$.

In contrast to the hazard rate, $\dot{R}_h(t)$ may be positive or negative. Types with a shorter than average residual duration may in principle have a lower than average hazard rate at the current moment (but not in the future), and hence may be increasing as a share of the population.

In practice, residual duration may be extremely long for some types because their asymptotic hazard rate may be very low. We therefore also consider a measure of discounted residual duration,

$$r_\delta(t, \alpha, \beta) = \int_0^\infty e^{-\delta(s-t)}(s-t)f(s, \alpha, \beta)ds$$

$$= \int_t^\infty e^{\int_t^s(h(\tau, \alpha, \beta)+\delta)d\tau}d\tau.$$  

When $\delta = 0$, this reduces to residual duration, while higher values of $\delta$ are associated with more discounting. In the limit as $\delta$ converges to infinity, residual duration is effectively equivalent to the reciprocal of the current hazard rate. One economically sensible way to think of $\delta$ is as a probability of death. If an individual dies before finding a job, we do not measure the observation in our data set.

We can decompose aggregate discounted residual duration,

$$R_\delta(t) = \int \int r_\delta(t, \alpha, \beta)g(t, \alpha, \beta) d\alpha d\beta,$$

into its components, $\dot{R}_\delta(t) = \dot{R}_s(t) + \dot{R}_h(t)$. Note that discounting does not affect the population shares $g(t, \alpha, \beta)$, since the risk of “death” falls equally on everyone.

### 3 Austrian Data

We test our theory, estimate our model, and evaluate the role of structural duration dependence using data from the Austrian social security registry. The data set covers the universe of private sector workers over the years 1972–2007 (Zweimüller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf, and Buchi, 2009). It contains information on individual’s employment, registered unemployment, maternity and retirement, with the exact begin and end date of each spell.

The use of the Austrian data is compelling for two reasons. First, the data set contains the complete labor market histories of the majority of workers over a 35 year period, which allows
us to construct multiple nonemployment spells per individual. Second, the labor market in Austria remains flexible despite institutional regulations, and responds only very mildly to the business cycle. Therefore, we can treat the Austrian labor market as a stationary environment and use the pooled data for our analysis. Some robustness checks are done in the Appendix. We discuss the key regulations below.

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

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

Temporary separations and recalls are prevalent in Austria. Around 40 percent of non-employment spells end with an individual returning to the previous employer.

We work with nonemployment spells, defined as the time from the end of one full-time job to the start of the following full-time job. We drop incomplete spells and spells involving a maternity leave. Although in principle we could measure nonemployment duration in days, disproportionately many jobs start on Mondays and end on Fridays, and so we focus on weekly data. We code 0 to 6 days out of work and 0 weeks, 7 to 13 days as 1 week, and so on.

Our sample consists of all individuals who were no older than 45 in 1972 and no younger than 40 in 2007. Thus each individual has at least 15 years when he could potentially be at work. For each individual, we take the first two completed nonemployment spells satisfying the following criterion: the age at the start of the spell must be at least 25, the duration of the spell must be between two and 260 weeks. We drop all individuals who do not have two such spells.

We do not consider spells shorter than 2 weeks because we believe that these are mainly driven by planned job-to-job movements, which are not in our model (or which are subsumed by the stochastic process for wages while employed). Indeed, the hazard of starting a new job within two weeks of leaving the previous one is very high, 13.5 percent per week, and then it falls sharply to around five or six percent. We drop longer spells so that we have a
consistent sample across all years. Workers with few years in the data set are unlikely to have very long spells.

Table 1 shows how the sample size shrinks after imposing different criteria. There are 40,017,411 non-employment spells in the Austrian data. We exclude 837,633 spells that contain a maternity leave, 16,911,328 spells for workers less than 25 years old at the beginning of the nonemployment spell. We further exclude 1,019,749 spells for workers older than 45 in 1972, and 3,416,253 spells for workers younger than 40 in 2007. Out of these spells, 4,706,740 are shorter than 2 weeks and 394,648 are longer than 260 weeks. The sample then contains 12,731,060 spells for 2,383,306 workers. Out of these, 704,724 workers have a single spell; these are excluded. In the final sample we are left with 1,678,582 workers who have two or more spells non-employment spells lasting between 2 and 260 weeks. In fact, these workers have 3.6 non-employment spells on average, but we only use information on the first two spells.

In this sample, the average duration of a completed nonemployment spell is 31.3 weeks, and the average employment duration between these two spells is 36.8 weeks. Figure 1 depicts the nonemployment exit hazard rate during each of the first two nonemployment spells for all workers who experience at least two spells. The two hazard rates are very similar. They fall from five or six percent during the first six weeks of nonemployment to around two percent by the end of the first year of nonemployment, then drops sharply to just over one percent by early in the second year. The decline thereafter is slow and steady.

Figure 2 depicts the joint density $\phi(t_1, t_2)$ for $t_1, t_2 \in \{2, \ldots, 80\}^2$. Several features of the joint density are notable. First, it is generally convex, which reflects the declining hazard rate of finding a job at long durations. Second, it has a noticeable ridge at values of $t_1 \approx t_2$. Many workers experience two spells of similar durations. Third, the joint density is noisy, even with 2.25 million observations. This does not appear to be primarily due to sampling
Figure 1: Nonemployment exit probability during the first two nonemployment spells.

variation, but rather reflects the fact that many jobs start during the first week of the month and end during the last one. There are notable spikes in the probability of finding a job every fourth or fifth week and, as Figure 1 shows, these spikes persist even at long durations.

We discuss sensitivity of the sample and the resulting hazard rate to other criteria in Appendix B.

4 Results

4.1 Test of the Model

We propose a test of the model inspired by Section 2.3. We make three changes to accommodate the reality of our data. The first is that our model implies that the reemployment density $\phi(t_1, t_2)$ is symmetric, while this is not exactly true in the real world data; see the difference between the two lines in Figure 1. We instead estimate $\phi$ as $\frac{1}{2}(\phi(t_1, t_2) + \phi(t_2, t_1))$. The second is that the data are only available with weekly durations, and so we cannot mea-
sure the slope of the reemployment density $\phi$. Instead, we propose a discrete time analog:

$$
\frac{2\phi_1(t_1, t_2)}{\phi(t_1, t_2)} \approx \log \left( \frac{\phi(t_1 + 1, t_2)}{\phi(t_1 - 1, t_2)} \right)
$$

and

$$
\frac{2\phi_2(t_1, t_2)}{\phi(t_1, t_2)} \approx \log \left( \frac{\phi(t_1, t_2 + 1)}{\phi(t_1, t_2 - 1)} \right).
$$

Then we test whether

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

and

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

are nonnegative.

Finally, the raw measure of $\phi$ is noisy, as we discussed in the previous section. This makes estimates of the slope $\log \left( \frac{\phi(t_1 + 1, t_2)}{\phi(t_1 - 1, t_2)} \right)$ and $\log \left( \frac{\phi(t_1, t_2 + 1)}{\phi(t_1, t_2 - 1)} \right)$ noisy. 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 empirical density using a multidimensional Hodrick-Prescott filter with parameter $\lambda$. More precisely, we find the function $\bar{\phi}(t_1, t_2)$ that minimizes the sum of squared deviations of $\bar{\phi}(t_1, t_2)$ from

![Figure 2: Nonemployment exit density: data](image)
the data φ(t1, t2) subject to a penalty λ that penalizes changes in the slope of ˜φ(t1, t2): ⁶

\[
\min_{\{\phi(t_1, t_2)\}} \left( \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} (\phi(t_1, t_2) - ˜\phi(t_1, t_2))^2 
+ \lambda \sum_{t_1=2}^{T-1} \sum_{t_2=1}^{T} (\tilde{\phi}(t_1 + 1, t_2) - 2\tilde{\phi}(t_1, t_2) + \tilde{\phi}(t_1 - 1, t_2))^2 
+ \lambda \sum_{t_1=1}^{T} \sum_{t_2=2}^{T-1} (\tilde{\phi}(t_1, t_2 + 1) - 2\tilde{\phi}(t_1, t_2) + \tilde{\phi}(t_1, t_2 - 1))^2 \right). 
\]

See Appendix C for more details.

Figure 3 displays our test results. Without any smoothing, we reject the model for nearly 40 percent of pairs (t1, t2) with 2 ≤ t1 < t2 ≤ 60. Setting the smoothing parameter λ to at least 7 reduces the rejection rate below five percent. Setting it to at least 20 reduces the rejection rate below one percent. When we look at higher values of (t1, t2), we reject the model more often, even in smoothed data. This may be due to a reduction in the

⁶In practice we smooth the function log(1+φ(t1, t2)), rather than φ, where φ is the number of individuals whose two spells have durations (t1, t2).
signal-to-noise ratio in our data set.

### 4.2 Estimation

Let $D(\mathbb{R}_+^2)$ be the set of density functions with domains in pairs $(\alpha, \beta)$, so that $g \in D(\mathbb{R}_+^2)$ means that $g(\alpha, \beta) \geq 0$ for all $(\alpha, \beta)$ and that $\int \int g(\alpha, \beta) d\alpha d\beta = 1$. Note that we impose $\alpha > 0$ since we cannot identify the sign of $\alpha$ from nonemployment duration data.

Our model connects the data, $\phi \in D(\mathbb{R}_+^2)$, with the distribution of parameters, $g \in D(\mathbb{R}_+^2)$, as follows:

$$
\phi(t_1, t_2) = \int \int \prod_{i=1}^{2} f(t_i, \alpha, \beta) g(\alpha, \beta) d\alpha d\beta 
$$

(9)

for all $(t_1, t_2) \in \mathbb{R}_+^2$, or more compactly $\phi = Fg$. Note that $F$ is a linear positive operator.

**Analogy with linear algebra in finite dimensions.** In this section we develop the notation for a finite dimension representation of the model. We view $\phi$ as a vector in a finite dimensional space. In particular we consider the set $T \subset \mathbb{R}_+^2$ of pair durations $(t_1, t_2)$. We refer to the typical elements as $(t_1(i), t_2(i)) \in T$ with $i = 1, \ldots, M$. In this case we can write the distributions of spells as:

$$
\phi \in \Delta^M, \text{ where } \Delta^M \equiv \left\{ \phi \in \mathbb{R}_+^M : \sum_{i=1}^{M} \phi_i = 1 \right\}.
$$

The fact that the vector is normalized to one means that we are considering the conditional distribution of spells in set $T$.

We also view $g$ as a vector in a finite dimensional space. In particular we consider the set $\Theta \subset \mathbb{R}_+^2$ of pairs of parameters $(\alpha, \beta)$. We refer to the typical element $(\alpha(j), \beta(j)) \in \Theta$ with $j = 1, \ldots, N$. In this case we can write the distribution of types as

$$
g \in \Delta^N, \text{ where } \Delta^N \equiv \left\{ g \in \mathbb{R}_+^N : \sum_{j=1}^{N} g_j = 1 \right\}.
$$

The likelihood $F$ can be thus viewed as a $M \times N$ positive matrix with nonnegative entries $F_{ij}$ and columns that add up to 1, $\sum_{i=1}^{M} F_{ij} = 1$ for all $j = 1, \ldots, N$. The interpretation of $F_{i,j}$ is:

$$
F_{i,j} = \Pr \{ t_1 \in (t_1(i), t_1(i) + dt] \text{, } t_2 \in (t_2(i), t_2(i) + dt] \mid (\alpha, \beta) = (\alpha(j), \beta(j)) \}
$$

22
for a small positive $dt$. For small positive values of $dt$,

$$\Pr\{t \in (t + dt) \mid (\alpha, \beta)\} \approx f(t, \alpha, \beta) \, dt,$$

(10)

where $f$ is the density of the spells given by equation (3). Then taking limits as $dt$ converges to 0 gives

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

for each pair of spells $i = 1, ..., M$ and parameters $j = 1, ..., N$.

**Minimum distance estimator.** Our benchmark estimator is a simple minimum distance estimator. We let $N < M$, so that $F$ can be regarded as a $M \times N$ stochastic matrix, and solve the following quadratic problem:

$$\min_{g \in \Delta^N} ||F \, g - \phi||$$

(11)

We briefly discuss some details of the implementation.

1. Given the symmetry of the likelihood, we symmetrize the data, by setting $\phi(t_1, t_2)$ to be the average of the observed densities at $(t_1, t_2)$ and $(t_2, t_1)$.

2. As discussed elsewhere, we use a grid of points for $\alpha$ with strictly positive values, since with completed spells we can only identify $|\alpha|$. We use an equal spaced grid for $\alpha$ and $\beta$.

3. As a consequence of the symmetrization, we only use a grid of values for $\mathbb{T}$ for which $t_1 \leq t_2$.

4. We use both the approximation in equation (10) and the exact formula for the CDF, which is also known in close form. The reason we use both is that for extreme values of $t, \mu$, and $\sigma$ the PDF is numerically more stable.

5. We exclude the observations for weeks 0 and 1, that is, those that are employed after a nonemployment spell shorter than one week ($t = 0$) or two weeks ($t = 1$). We do so because a disproportionate number of very short spells corresponds to planned job-to-job movements where the leaves the job in order to pursue a job in another firm, an aspect not captured by our model. Otherwise we use all the weekly data for durations up to $T = 80$ weeks.
<table>
<thead>
<tr>
<th></th>
<th>E(α)</th>
<th>Std(α)</th>
<th>E(β)</th>
<th>Std(β)</th>
<th>Corr(α, β)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38.6</td>
<td>270</td>
<td>35.6</td>
<td>161</td>
<td>0.85</td>
</tr>
<tr>
<td>E(β/α)</td>
<td>Var(β/α)</td>
<td>E(β/α)</td>
<td>E(2/α²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>43.6</td>
<td>916</td>
<td>6948</td>
<td>243</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Summary statistics from estimation

6. If for some \( j = 1, \ldots, N \) we have a combination of \((\alpha(j), \beta(j))\) in the grid for which, due to numerical accuracy, all values \( F_{ij} = 0 \), i.e. for which \((t_1(i), t_2(i))\) for all \( i = 1, \ldots, M \), we assign \( g(j) = 0 \) to that parameter pair.

7. With large values for \( M \) and \( N \), solving the problem in equation (11) directly is infeasible. Instead, we iterate on a relaxed problem which imposes neither the non-negativity of \( g \) nor that the elements of \( g \) sum to one. Between iterations, we take the positive elements of \( g \) and rescale them to sum one. After one iteration, we have a solution where \( g \) almost sums to one and satisfies the Kuhn-Tucker condition exactly, so we think we have an appropriate solution of the problem for a given grid.

8. We fit the density \( \phi \) for \((t_1, t_2) \in \{2, 3, \ldots, 160\}^2 \) with \( t_1 \leq t_2 \), at a total of \( M = 12,720 \) points. We allow for \( N = 6400 \) types \((\alpha, \beta)\). We choose values of \( \alpha \) and \( \beta \) so that \( \alpha/\beta \) takes 80 proportionately spaced values between 0.001 and 0.5 while \( 1/\beta \) takes 80 proportionately spaced values between 0.005 and 1.5. This concentrates the density in the most relevant part of the parameter space.

9. We throw away pairs of \((\alpha, \beta)\) with an estimated density below 1 basis point. We then use the EM algorithm to refine our estimate of \((\alpha, \beta)\) and the associated density. That is, when using this local minimizer, we do not constrain \((\alpha, \beta)\) to lie on a grid.

Our preferred parameter estimates place positive weight on 58 different types \((\alpha, \beta)\). Table 2 summarizes our estimates. We find that there is a considerable amount of heterogeneity. For example the cross-sectional standard deviation of \( \alpha \) is seven times its mean, while the cross-sectional standard deviation of \( \beta \) is four and half times its mean. Moreover, \( \alpha \) and \( \beta \) are positively correlated in the cross-section. Perhaps more useful is to report the average value of \( \beta/\alpha = 43.6 \). This is the initial unconditional duration of a nonemployment spell. The cross-sectional variance of the mean duration of a nonemployment spell \( \beta/\alpha \) is 916 weeks, while the average of the variance in the duration of a nonemployment spell, \( \beta/\alpha^3 \), is 6958 weeks. Thus within-worker variation in realized durations account for 88 percent of the total variation in duration.
Figure 4: Nonemployment exit probability in the data and in the model

Another way to look at this is through the mean value of $2/\alpha^2$, the asymptotic duration of a nonemployment spell. This is 243 weeks. If the average newly nonemployed worker didn’t find a job for sufficiently long, the residual duration of his nonemployment spell would approach 5 years. Thus the average worker exhibits a substantial amount of negative duration dependence in the hazard of exiting nonemployment.

Figure 4 shows the fitted hazard rate during the first 400 weeks of a nonemployment spell. Our estimates use the data from the first 160 weeks, while the fit during the last 240 weeks comes from the structure of the model. In particular, the model is able to capture many of the wiggles in the univariate hazard rate, including the sharp decline in the reemployment probability at 52 weeks, while also matching the gradual decline in the hazard at longer durations.

Figure 5 shows the theoretical joint density of the duration of the first two nonemployment spells, the theoretical analog of Figure 2. Figure 6 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 $\phi$, with the model able to match the major features of the empirical joint density $\phi$, leaving primarily the high frequency fluctuations that we previously indicated we would not attempt to match.
Figure 5: Nonemployment exit density: model

Figure 6: Nonemployment exit density: Log ratio of model to data
4.3 Estimation of Switching Costs and the Degenerate Range of Inaction

We can use the estimated distribution of \(|\alpha|, \beta\) to compute the distribution of the switching costs \(\psi/b\) and the standard deviation \(\sigma_n\) of the process for latent log-wages, \(\sigma_n\).

Recall that \(\alpha = \mu_n/\sigma_n\) and \(\beta = (\bar{\omega} - \omega)/\sigma_n\), that parameters \(\mu_e, \sigma_e, \mu_n, \sigma_n\) govern processes of log-wages for employed and non-employed and that \(\bar{\omega}, \omega\) are optimal thresholds. We show in Corollary 1, that if parameters \(\mu_e, \sigma_e, \mu_n, \sigma_n\) satisfy (2), then for any width of the inaction region \(\bar{\omega} - \omega\), there exists a unique normalized switching costs \(\psi/b\) for which \(\bar{\omega} - \omega\) is optimal. In other words, if we know parameters for the wage process and the width of the inaction region, we can find the switching costs. However, we estimated \(\alpha, \beta\) and not \(\mu_n, \sigma_n, \bar{\omega} - \omega\) separately, and thus we need to impose additional assumptions. In particular, we assume that all types face the same values of \(\mu_e, \sigma_e, \mu_n, \rho\) which together with estimated values of \(\alpha, \beta\) allows us to recover parameter \(\sigma_n\) from \(\alpha\), the width of the inaction region \(\bar{\omega} - \omega\) from \(\beta\), and normalized switching costs \(\psi/b\) from the system of equations 14–17 for each worker type.

At the annual frequency, we choose \(\rho = 0.04\), \(\mu_e = 0.02\) and \(\sigma_e = 0.1\). The wage of an employed worker grows at 2% on average per year, with the variance of 0.01. The choice of these values is justified by Heathcote, Perri, and Violante (2010). In Figure 14 they show that the life-cycle variance of log earnings increases by 0.3 over the period of 30 years, governing our choice of annual variance to be 0.01.

We use two targets for \(\mu_n^* \in \{-0.01, +0.01\}\), both of them smaller than \(\mu_e\) so that non-employment decreases worker’s human capital at the rate \(\mu_e - \mu_n\). We use \(\mu_n^*\) equal to the target value as long as the implied \(\mu_n, \sigma_n\) satisfy the constraint \(\rho > \mu_n + \sigma_n^2\). Otherwise, we choose the value of \(\mu_n\) to be the closest value to the target \(\mu_n^*\) consistent with the above constraint.

For both values of \(\mu_n\), the estimated width of the range of inaction as well as the corresponding values of \(\psi/b\) are large for a few types but the median values are very small. The median value of the switching costs for \(\mu_n^* = +0.01\) is only 0.001% of the annual non-

<table>
<thead>
<tr>
<th>Avg (\psi/b) (\times 100)</th>
<th>Std (\psi/b) (\times 100)</th>
<th>Median (\psi/b) (\times 100)</th>
<th>Avg (\sigma_n)</th>
<th>Std (\sigma_n)</th>
<th>Avg (\bar{\omega} - \omega)</th>
<th>Fraction (\mu_n &lt; \mu_n^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu_n^* = +0.01)</td>
<td>3.26</td>
<td>11.4</td>
<td>0.00101</td>
<td>0.051</td>
<td>0.121</td>
<td>0.0334</td>
</tr>
<tr>
<td>(\mu_n^* = -0.01)</td>
<td>3.35</td>
<td>11.4</td>
<td>0.018</td>
<td>0.051</td>
<td>0.121</td>
<td>0.0334</td>
</tr>
</tbody>
</table>

Table 3: Summary statistics for the estimated switching cost.
employment flow value. The costs vary a lot across types: the highest cost is 43.8% of the annual non-employment flow value, while the lowest is of the order of $10^{-8}$ and $10^{-6}$ for the positive and negative $\mu_n^*$, respectively. The switching costs for $\mu_n^* < 0$ are a bit larger than for $\mu_n^* > 0$. Intuitively, a lower $\mu_n$ reduces human capital during non-employment even more, and thus to rationalize the same data, it must be that workers are deterred from switching to employment by higher costs.

Figure 7 displays the distribution of the estimated cost for the two values of $\mu_n^*$ on a log scale. The distribution of the implied cost for $\mu_n^* = -0.01$ first-order stochastically dominates the one for $\mu_n^* = 0.01$. Only about 6.5% of the estimated types have costs higher than 1% of the annual nonemployment flow value, and these type represent 7.5% of the population.

The estimated values for the range of inaction and the switching costs are very small which leads us to consider a special case of no switching costs. Besides it being a limiting case of our model, it also serves as a useful benchmark for comparison to the existing literature where such a model has been theoretically and empirically analyzed, see (Alvarez and Shimer (2011) and Shimer (2008)).

If there are no switching costs, $\psi = 0$, the decision of switching between employment and non-employment is static, and it is optimal to become employed whenever the wage is higher than the value of non-employment and vice-versa. Thus, $\bar{x} = \underline{x} = \log(b)$, and we verify in the Appendix that these thresholds are optimal.
With a degenerate range of inaction, the hazard rate of re-employment at the very low non-employment durations \( t \) is approximately \( 1/2t \), making the re-employment hazard at \( t = 0 \) infinite, as shown in Alvarez and Shimer (2011) and Shimer (2008). The reason is that starting at a threshold, a random walk crosses this threshold infinitely many times. This is a nice feature of this version of the model since it produces realistically high hazard rates at very low durations, as explored in Shimer (2008). Yet here we concentrate on the implications for longer durations by examining the distribution of spells conditional on duration larger than two weeks, \( t > \bar{t} = 2 \).

It turns out that this model cannot rationalize the data because it imposes a very tight restriction on the expected non-employment duration. It is even tighter once we condition on the duration being between some bounds \( t \) and \( \bar{t} \), as we did earlier.

The switching costs are very small in magnitude yet they matter a lot economically.

### 4.4 Counterfactual Exercises

Finally, we report the outcome of our counterfactual exercises: 1) variance decomposition, 2) decomposition of the expected residual duration and 3) decomposition of the hazard rates.

Figure 8 shows the increase in the expected residual duration \( R(t) - R(0) \) and its two components, \( \bar{R}^h(t) \) and \( \bar{R}^s(t) \). We see that the expected residual duration increases rapidly by more 6000 weeks in the course of 10 years of non-employment. This very high expected residual duration is driven by a few types with a small \( \alpha/\beta \), whose expected residual duration is then 10,000 weeks. This number surely is very high, but we need to point out that this is an unconditional residual duration. If we were to condition on the spell lasting at most 35 years, then this number drops to 218 weeks.

Figure 9 shows the share of the increase in mean residual nonemployment duration attributable to changes in the composition of the nonemployed population and to changes in residual duration for a fixed population. During the first year of non-employment, most of the increase in the residual duration is accounted for by changes in the composition of non-employed workers. The role of heterogeneity gradually declines to about 50% at long durations.

It turns out these few types with a long expected unconditional duration play an important role in the decomposition. To see this more clearly, we conduct the decomposition exercise conditioning on spells being shorter than 35. It is worth noting that the interpretation of the role of duration dependence is now less straightforward. In the unconditional case, we argued that if there is no duration dependence and the hazard rate of finding a job is constant, then the portion attributable to duration dependence is zero, \( \bar{R}^s(t) = 0 \) for all
Figure 8: An increase in the mean residual nonemployment duration in weeks as a function of realized nonemployment duration. The purple line shows the entire increase in mean residual duration, $D^r(t) - D^r(0)$. The red line shows the portion of the increase accounted for by changes in the composition of the nonemployed population, $D^h(t)$. The blue line shows the portion of the increase accounted for by the structure of the model for a fixed composition, $D^s(t)$.
Figure 9: The red line shows the share of the increase in residual nonemployment duration attributable to changes in the composition of the nonemployed population. The blue line shows the share attributable to structural duration dependence.

However, this would no longer be true if we condition on spells being shorter than certain duration since this necessarily generates some duration dependence even in the case of a constant hazard. We conduct the exercise nevertheless, and show the components $R(t) - R(0)$, $\bar{R}^s(t)$ and $\bar{R}^h(t)$ in Figure 10. The expected residual duration increases by about 310 weeks in the course of 10 years on non-employment, much less than in the unconditional case. The contribution of heterogeneity is much lower in this case. In particular, the contribution of heterogeneity increases from 28 to 46 percent during the first 20 weeks of non-employment, stays at about 46% till the end of the first year, and the declines to about 38%, see Figure 11.

Finally, we show the decomposition of the hazard rates. Figure 12 shows the change in the aggregate hazard rate $H(t) - H(0)$, together with changes attributable to heterogeneity $\bar{H}^h(t)$ or structural duration dependence $\bar{H}^s(t)$. This figure suggests that after one year of duration, the contribution of heterogeneity and structural duration dependence is approximately the same - the decline in the hazard due to heterogeneity is $-0.087$ while increase in the hazard due to duration dependence is 0.09. These two parts almost offset each other and lead to the hazard of 0.003 at long durations. During the first year, heterogeneity plays a bigger role than the duration dependence. Notice that if there was not heterogeneity, the hazard
Figure 10: Same as Figure 8 but conditional on $t < 35$ years.

Figure 11: The lines show the share of the increase in residual nonemployment duration attributable to changes in the composition of the nonemployed population. The red line represents an unconditional residual duration, the blue line conditional on spells lasting at most 35 years.
rate would have increased. The fact that is decreases is due to the fact that the changes in the composition more than offset the this increase and lead to a declining hazard rate.

The advantage of looking at the decomposition of the hazard rate as opposed to residual non-employment duration is that the this decomposition is not sensitive to the types with a very low $\alpha/\beta$, and that the hazard rate is more common in the literature.

5 Mixed Proportional Hazard Model

The standard approach to duration models with unobserved heterogeneity is the mixed proportional hazard model (Lancaster, 1979). An economy consists of a large number of individuals, each with a fixed characteristic $\theta$ distributed with distribution $G$ in the population. If an individual $\theta$ is nonemployed with duration $t$, she finds a job at hazard rate $\theta h(t)$, so her probability of experiencing a nonemployment spell lasting at least $t$ periods is $\exp\left(-\theta \int_0^t h(\tau)d\tau\right)$.

This model neither nests our model nor is nested by it. The mixed proportional hazard
model implies that the ratio of the job finding hazards for any two individuals is constant during a spell of nonemployment, while the hazards in our model are either identical (if the two individuals have the same reduced-form parameters $\alpha$ and $\beta$) or not proportional. Conversely, a special case of the mixed proportional hazard model is one in which all workers are identical, $\theta = 1$, and the baseline job finding hazard $h(t)$ is constant. We proved in Section 2.3 that our model cannot generate the same joint distribution of two nonemployment spells as this special case, regardless of the joint distribution of $\alpha$ and $\beta$.

In this section, we argue that the same Austrian data set is inconsistent with any version of the mixed proportional hazard model. Our approach is based on Honoré (1993), who develops results on nonparametric identification. We work directly in discrete time since our data are in discrete time, but we could derive a similar test in continuous time. We assume that there are many individuals, each with a fixed type $\theta$ distributed according to $G$ in the population. We observe exactly two nonemployment spells for each individual and assume that the probability of exiting nonemployment depends on the individual’s type, the duration of nonemployment, and the spell number, $i \in 1, 2$. More precisely, an individual with type $\theta$ and with nonemployment duration $t$ during spell $i$ finds a job during period $t$ with probability $\theta p_{i,t}$ and otherwise remains nonemployed into period $t + 1$.

A key object for this analysis is the survivor function. Let $\Phi(t_1, t_2)$ denote the fraction of individuals whose first spell lasts at least $t_1$ periods and second spell lasts at least $t_2$ periods. Using the law of large numbers, the structure of the mixed proportional hazard rate model implies that this is

$$
\Phi(t_1, t_2) = \int \left( \prod_{\tau=0}^{t_1-1} (1 - \theta p_{1,\tau}) \prod_{\tau=0}^{t_2-1} (1 - \theta p_{2,\tau}) \right) dG(\theta)
$$

Now define the first differences of this function, $\Phi_1(t_1, t_2) \equiv \Phi(t_1 + 1, t_2) - \Phi(t_1, t_2)$ and $\Phi_2(t_1, t_2) \equiv \Phi(t_1, t_2 + 1) - \Phi(t_1, t_2)$. Simple algebra implies

$$
\Phi_1(t_1, t_2) = -p_{1,t_1} \int \theta \left( \prod_{\tau=1}^{t_1-1} (1 - \theta p_{1,\tau}) \prod_{\tau=1}^{t_2-1} (1 - \theta p_{2,\tau}) \right) dG(\theta)
$$

and

$$
\Phi_2(t_1, t_2) = -p_{2,t_2} \int \theta \left( \prod_{\tau=1}^{t_1-1} (1 - \theta p_{1,\tau}) \prod_{\tau=1}^{t_2-1} (1 - \theta p_{2,\tau}) \right) dG(\theta).
$$

In particular, taking ratios of these two numbers, we get

$$
\frac{\Phi_1(t_1, t_2)}{\Phi_2(t_1, t_2)} = \frac{p_{1,t_1}}{p_{2,t_2}}
$$

It may seem natural to impose the restriction that $p_{i,t} = p_{j,t}$ for all $i$ and $j$, but our test allows for this more relaxed version of the model.
for all $t_1$ and $t_2$. Honoré (1993) uses this expression to argue that the model is nonparametrically identified.

We take this one step further. Compute this ratio for $(t_1, t_2)$ and $(t'_1, t_2)$ where $t_1 \neq t'_1$. Taking ratios gives

$$\frac{p_{1,t_1}}{p_{1,t'_1}} = \frac{\Phi_1(t_1, t_2)\Phi_2(t'_1, t_2)}{\Phi_2(t_1, t_2)\Phi_1(t'_1, t_2)}$$

for all $t_1$, $t'_1$, and $t_2$. According to the mixed proportional hazard model, the left hand side does not depend on $t_2$, while the right hand side, which can be measured in the data, depends on $t_2$. This yields a nonparametric test of the model.

**Proposition 1** For any $t_1$ and $t'_1$,

$$\frac{\Phi_1(t_1, t_2)\Phi_2(t'_1, t_2)}{\Phi_2(t_1, t_2)\Phi_1(t'_1, t_2)}$$

does not depend on $t_2$.

We implement this test using the same Austrian data set. Note that the second difference of the survivor function $\Phi$ is simply the density function $\phi$ that we used throughout our analysis of the net benefit from employment. We are therefore using exactly the same data to test the two models.

Figure 13 shows a subset of the results, the relative probability of finding a job at durations 13, 26, 39, and 52 weeks, compared to 2 weeks. According to the theory, these probabilities should not depend on the choice of $t_2$, and so should give accurate estimates of the relative baseline hazard $p_{1,t}/p_{1,2}$, but the figure shows a systematic dependence. Each line initially increases and then starts declining at some $t_2 < t_1$. The maximum implied relative baseline hazard is in each case at least twice the minimum. Monte Carlo simulations suggest to us that this is driven by the large number of individuals who have two spells of similar long lengths, an observation that cannot be accommodated by the mixed proportional hazard model.

Given the failure of the nonparametric test, we do not attempt to estimate the mixed proportional hazard model using our data set. Instead, we believe that our model of the net benefit from nonemployment offers a better description of the nonemployment duration data.

6 Conclusion

To be added
Figure 13: Nonparametric test of the mixed proportional hazard model. The figure shows the baseline probability of finding a job at 13, 26, 39, and 52 weeks, compared to 2 weeks duration for different values of the second spell duration $t_2$. According to the mixed proportional hazard model, each line should be independent of $t_2$.

References


Appendix

A Worker’s Problem

Let $E(\omega)$ denote the expected present value of income for an employed worker with log wage $\omega$ and $N(\omega)$ denote the expected present value of income for a nonemployed worker with log latent wage $\omega$. These satisfy standard Bellman-Jacobi-Hamilton equations:

$$\rho E(\omega) = \exp(\omega) + \mu_e E'(\omega) + \frac{\sigma_e^2}{2} E''(\omega) \quad \text{for all } \omega \in [\omega, \infty), \quad (12)$$

$$\rho N(\omega) = b + \mu_n N'(\omega) + \frac{\sigma_n^2}{2} N''(\omega) \quad \text{for all } \omega \in (-\infty, \bar{\omega}]. \quad (13)$$

The solution to these equations is

$$E(\omega) = \frac{\exp(\omega)}{\rho - \mu_e - \frac{\sigma_e^2}{2}} + e_1 \exp(\lambda_{e_1} \omega) + e_2 \exp(\lambda_{e_2} \omega)$$

$$N(\omega) = \frac{b}{\rho} + n_1 \exp(\lambda_{n_1} \omega) + n_2 \exp(\lambda_{n_2} \omega)$$

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

$$\rho = \lambda_e (\mu_e + \lambda_e \sigma_e^2/2) \quad \text{and} \quad \rho = \lambda_n (\mu_n + \lambda_n \sigma_n^2/2).$$

We look for a solution with two barriers $\omega < \bar{\omega}$ that satisfies the following equations:

$$E(\omega) = N(\omega) \quad (14)$$

$$E(\bar{\omega}) = N(\bar{\omega}) + \psi \quad (15)$$

$$E'(\omega) = N'(\omega) \quad (16)$$

$$E'(\bar{\omega}) = N'(\bar{\omega}) \quad (17)$$

The first two equations require that the value function is continuous at each of the two boundaries (“value matching”). The last two equations require that the value function is differentiable at each to the two boundaries (“smooth pasting”). We also have the two
no-bubble conditions, i.e. that:

\[
\lim_{\omega \to -\infty} N(\omega) = \frac{b}{\rho} \quad \text{and} \quad \lim_{\omega \to +\infty} \frac{E(\omega)}{\exp(\omega)} = \frac{1}{\rho - \mu_e - \sigma_e^2/2}
\]

Equation (18) requires that for arbitrarily low \( \omega \) the value functions converges to the value of nonemployment forever. Likewise equation (19) requires that for arbitrarily high \( \omega \) the value function converges to the value of employment forever. Hence we have six unknowns: \( (e_1, e_2, n_1, n_2, \omega, \bar{\omega}) \) and six equations, namely (14)–(19). We turn to their solution.

First, the no-bubble conditions (18) and (19) imply that \( e_2 = n_1 = 0 \). Otherwise if \( e_2 \) or \( n_1 \) will be different from zero, \( E \) (or \( N \)) will diverge relative to the value of employment (nonemployment) forever as \( \omega \) becomes very large (small). Abusing notation we then let:

\[ e = e_1 > 0, \quad n = n_2 > 0, \quad \lambda_e = \lambda_{e_1} < 0, \quad \text{and} \quad \lambda_n = \lambda_{n_2} > 0 \]

and hence we can rewrite the value functions as:

\[
E(\omega) = \frac{\exp(\omega)}{\rho - \mu_e - \sigma_e^2/2} + e \exp(\lambda_e \omega) \quad \text{for all} \quad \omega \in [\omega, \infty) \quad (20)
\]

\[
N(\omega) = \frac{b}{\rho} + n \exp(\lambda_n \omega) \quad \text{for all} \quad \omega \in (-\infty, \bar{\omega}] \quad (21)
\]

with

\[
\lambda_e = -\frac{\mu_e - \sqrt{\mu_e^2 + 2\rho \sigma_e^2}}{\sigma_e^2} < -1 \quad \text{and} \quad \lambda_n = -\frac{\mu_n + \sqrt{\mu_n^2 + 2\rho \sigma_n^2}}{\sigma_n^2} > 1
\]

where the inequalities follow from the assumptions in condition (2).

Thus we have four equations—two value matching and two smooth pasting—in the four
variables \((e, n, \omega, \bar{\omega})\) which can be written as

\[
\frac{\exp(\omega)}{\rho - \mu_e - \sigma_e^2/2} + e \exp(\lambda_e \omega) = \frac{b}{\rho} + n \exp(\lambda_n \bar{\omega})
\] (24)

\[
-\psi + \frac{\exp(\bar{\omega})}{\rho - \mu_e - \sigma_e^2/2} + e \exp(\lambda_e \bar{\omega}) = \frac{b}{\rho} + n \exp(\lambda_n \bar{\omega})
\] (25)

\[
\frac{\exp(\omega)}{\rho - \mu_e - \sigma_e^2/2} + e \lambda_e \exp(\lambda_e \omega) = n \lambda_n \exp(\lambda_n \bar{\omega})
\] (26)

\[
\frac{\exp(\bar{\omega})}{\rho - \mu_e - \sigma_e^2/2} + e \lambda_e \exp(\lambda_e \bar{\omega}) = n \lambda_n \exp(\lambda_n \bar{\omega})
\] (27)

Note that the values of \(e\) and \(n\) have to be positive, since it is feasible to choose to either be employed forever or nonemployed forever, and since the value of being employed forever and nonemployed forever are obtained in equations (12) and equations (13) by setting \(e = 0\) and \(n = 0\) respectively.

Figure 14 displays an example of the value functions \(E(\cdot)\) and \(N(\cdot)\). In the horizontal axis we plot the log wage and log-latent wage, and indicate the thresholds \(\omega < \bar{\omega}\). The domain of the employment value function \(E\) is \([\omega, +\infty)\), and the domain of the nonemployment value function is \((-\infty, \bar{\omega})\). Besides the functions \(E(\cdot)\) and \(N(\cdot)\) we also plot the value of nonemployment forever, i.e. \(b/\rho\), and the value of employment forever, i.e. \(\exp(\omega)/(\rho - \mu_e - \sigma_e^2/2)\). It is readily seen that as \(\omega \to -\infty\), the value function \(N(\omega)\) converges to the value of nonemployment forever, and that as \(\omega \to \infty\), the value function \(E(\omega)\) converges to the value of employment forever. Additionally it can be seen that at \(\bar{\omega}\) the value as well as the slopes of \(N\) and \(E\) coincide. Instead at \(\omega\) the slopes of \(E\) and \(N\) coinicides, but the value of \(E\) is \(\psi\) higher than \(N\), since a nonemployed worker must pay the fixed cost to become employed.

Now we turn to the analysis of the implications of the model for switching cost. We start with a result about the units in which switching cost are measured in the model.

**Lemma 1** Fix \(\lambda_n > 1, \lambda_e < -1\) and \(\rho - \mu_e - \sigma_e^2/2 > 0\). Suppose that \((e, n, \omega, \bar{\omega})\) solve the value function for fixed cost and flow benefit of nonemployment \((\psi, b)\). Then for any \(k > 0\), \((e', n', \omega', \bar{\omega}')\) solve the value function for flow benefit on nonemployment \(b' = kb\) and fixed cost \(\psi' = k\psi\) with:

\[
\omega' = \log(k) + \omega, \quad \bar{\omega}' = \log(k) + \bar{\omega}, \quad e' = ek^{1-\lambda_e}, \quad n' = nk^{1-\lambda_n}.
\] (28)

The proof of Lemma 1 follows by multiplying equations (24), (25), (26), and (27), which characterize the solution of the value function by \(k\), and using the relationships in equation (28). Lemma 1 implies that, keeping \(\lambda_n, \lambda_e\) and \(\rho - \mu_e - \sigma_e^2/2\) fixed, the scaled value of the fixed cost \(\psi/b\) determines the width of the range of inaction \(\omega - \bar{\omega}\).
Figure 14: Example of Value Functions. The parameters values are $\rho = 0.04, \mu_e = 0.02, \sigma_e = 0.1, \mu_n = 0.01, \sigma_n = 0.04, b = 1$, and $\psi = 2$.

Now we turn to the implication of observations on the width of the range of inaction to determine the size of the normalized fixed cost.

**Lemma 2** Consider a problem with $\rho > 0$, $\lambda_n > 1$, $\lambda_e < -1$ and $\rho - \mu_e - \sigma_e^2/2 > 0$. For any pair of thresholds $\omega < \bar{\omega}$ there is a unique pair of the fixed cost and flow benefit of nonemployments $(\psi, b)$ with $\psi > 0$ and $b > 0$, for which the pair of thresholds $\omega, \bar{\omega}$ is optimal.

**Proof.** We need to show that there is a unique positive 4-tuple $(\psi, b, e, n)$ that solves equations (24), (25), (26), and (27). First we note that for any pair $\omega > \bar{\omega}$, under the assumptions that $\lambda_e < -1$ and $\lambda_n > 1$ then the smooth pasting conditions (26) and (27) define a linear system of equations with a unique positive solution $(e, n)$. We can solve these equations to yield:

$$e = \frac{\exp(\omega(1 - \lambda_n)) - \exp(\bar{\omega}(1 - \lambda_n))}{\lambda_e (\rho - \mu_e - \sigma_e^2/2) [\exp((\lambda_e - \lambda_n)\omega) - \exp((\lambda_e - \lambda_n)\bar{\omega})]} > 0$$  \hspace{1cm} (29)$$

and

$$n = \frac{\exp(\omega(1 - \lambda_n) + \bar{\omega}(\lambda_e - \lambda_n)) - \exp(\bar{\omega}(1 - \lambda_n) + \omega(\lambda_e - \lambda_n))}{\lambda_n (\rho - \mu_e - \sigma_e^2/2) [\exp((\lambda_e - \lambda_n)\omega) - \exp((\lambda_e - \lambda_n)\bar{\omega})]} > 0$$  \hspace{1cm} (30)$$
Since $\bar{\omega} > \omega$ and $\lambda_e - \lambda_n < 0$ the denominators of $n$ is negative and, since $\lambda_e < 0$ the one of $e$ positive. Since $\lambda_n > 1$ and $\bar{\omega} > \omega$ then denominator of $e$ is positive, and hence $e > 0$. The numerator of $n$ is negative if

$$\omega (1 - \lambda_n) + \bar{\omega} (\lambda_e - \lambda_n) < \bar{\omega} (1 - \lambda_n) + \omega (\lambda_e - \lambda_n)$$

which is equivalent to

$$(\bar{\omega} - \omega) (\lambda_e - \lambda_n) < (\bar{\omega} - \omega) (1 - \lambda_n)$$

which holds since $\lambda_e < 1$.

Now we find the values of $b$ and $\psi$. Rewriting the value matching conditions in equation (24) and equation (25) we get:

\[
\begin{align*}
\frac{b}{\rho} &= \frac{\exp (\omega)}{\rho - \mu_e - \sigma_e^2 / 2} + e \exp (\lambda_e \omega) - n \exp (\lambda_n \omega) \\
\frac{b}{\rho} &= -\psi + \frac{\exp (\bar{\omega})}{\rho - \mu_e - \sigma_e^2 / 2} + e \exp (\lambda_e \bar{\omega}) - n \exp (\lambda_n \bar{\omega})
\end{align*}
\]  

(31)

(32)

Hence these two equations have a unique solution. Subtracting the smooth pasting condition for each boundary in each of them we obtain:

$$0 < (1 - \lambda_e) e \exp (\lambda_e \omega) - (1 - \lambda_n) n \exp (\lambda_n \omega)$$

$$< (1 - \lambda_e) e \exp (\lambda_e \bar{\omega}) - (1 - \lambda_n) n \exp (\lambda_n \bar{\omega})$$

The first inequality holds since $\lambda_e < 0$, $\lambda_n > 1$, $e > 0$ and $n > 0$. Thus equation (31) ensures that $b > 0$. The second inequality can be written as:

$$(1 - \lambda_e) e [\exp (\lambda_e \omega) - \exp (\lambda_e \bar{\omega})] < (1 - \lambda_n) n [\exp (\lambda_n \omega) - \exp (\lambda_n \bar{\omega})]$$

Note that $e > 0$, $1 - \lambda_e > 0$ and since $\lambda_e < 0$ and $\bar{\omega} > \omega$ then the left-hand side is the product of two negative numbers, and hence negative. Note that $n > 0$, $1 - \lambda_n < 0$, and $\lambda_n > 0$ so the right-hand side is the product of two negative numbers, and hence positive. This establishes the second inequality, and thus $\psi > 0$. For future reference, substituting the expression for $e$ and $n$ we obtain expressions for $b$ and $\psi$ as:

\[
b = \rho \left( \lambda_n (\lambda_e - 1) (\exp(\lambda_e \omega + \lambda_n \bar{\omega}) - \exp(\lambda_e \omega + \lambda_n \omega)) + (\lambda_n - 1) \lambda_e \exp(\lambda_n \omega + \lambda_n \bar{\omega}) + (\lambda_e - \lambda_n) \exp(\lambda_n \omega + \lambda_e \omega) - \exp(\lambda_n \omega + \lambda_e \omega) / \lambda_n \lambda_e (\mu - \rho + \sigma_e^2 / 2) (e^{\lambda_e \omega + \lambda_n \bar{\omega}} - e^{\lambda_e \omega + \lambda_e \omega}) \right)
\]  

(33)

42
and
\[
\psi = \frac{(\lambda_n - \lambda_e) e^{\omega (\lambda_n + \lambda_e) + \hat{\omega}} + (\lambda_n - \lambda_e) e^{\omega (\lambda_n + \lambda_e) + \lambda_n (\lambda_e - 1)} e^{\lambda_n \hat{\omega} + \lambda_e \hat{\omega} + \lambda_n \omega + \lambda_e \omega}}{\lambda_n \lambda_e (\mu - \rho + \sigma^2/2) (e^{\lambda_n \omega + \lambda_e \omega} - e^{\lambda_n \hat{\omega} + \lambda_e \hat{\omega}})} > 0
\]

(34)

□

The previous two lemmas can be used to find the normalized fixed cost as a function of the width of the range of inaction, as stated in the next corollary:

**Corollary 1** Lemmas 1 and 2 imply that for fixed values of the parameters \(\lambda_n > 1\), \(\lambda_e < -1\) and \(\rho - \mu_e - \sigma_e^2/2 > 0\), and any width of the range of inaction \(\bar{\omega} - \omega > 0\), there is a unique normalized fixed cost \(\psi/b > 0\) with \(\psi > 0\) and \(b > 0\), for which the width of inaction is optimal.

Figure 15 illustrates Corollary 1, by plotting the implied normalized fixed cost \(\psi/b\) as a function of the width of the range of inaction \(\bar{\omega} - \omega\). The figure displays four curves, each one for a different value of \(\lambda_n\). All the curves were drawn for the same value of \(\lambda_e\) and \(\rho - \mu_e - \sigma_e^2/2\).

**B Robustness**

**B.1 Hazard Rate in Subperiods**

Figure 16 shows the hazard rates in different periods. To construct the figures, we look at completed spells that last at most \(3 \times 52\) weeks, since otherwise the later periods would have many incomplete spells. We consider 8 subperiods, 1972–1975, 1976–1979, ... , 2000-2004, and count spells based on when they start (they do not have to end in the same subperiod). Except for the very low durations early in the sample, the hazard rates move together but the gap between the lowest and the highest one is around 1 percentage point.

**C Multidimensional Smoothing**

Take a data set \(\psi(t_1, t_2)\) which is noisy. We think of this as the sum of two terms, \(\bar{\psi}(t_1, t_2) + \tilde{\psi}(t_1, t_2)\), where \(\bar{\psi}\) is a smooth “trend” and \(\tilde{\psi}\) is the residual. Following Hodrick and Prescott
Figure 15: Implied normalized switching cost $\psi/b$ for different values of $\lambda_n$ and width of inaction range. The other parameters values are $\lambda_e = -5.46$, $\rho = 0.04$ and $\rho - \mu_e - \sigma_e^2/2 = 0.015$.

Figure 16: Hazard rate in different subperiods
(1997), we define the trend as the solution to

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

The first order conditions to this problem define \( \bar{\psi} \) as a linear function of \( \psi \) and so can be solved immediately.

Suppose in particular that \( \psi(t_1, t_2) = \zeta(t_1) + \zeta(t_2) \) for all \((t_1, t_2)\). Let \( \bar{\zeta}(t) \) be the usual Hodrick-Prescott filter of \( \zeta \) with parameter \( \lambda \). Then \( \bar{\psi}(t_1, t_2) = \bar{\zeta}(t_1) + \bar{\zeta}(t_2) \). This is a precise sense in which this filter is a multidimensional extension of the usual HP filter.

To prove this, first rewrite the objective function under the restriction that \( \bar{\psi} \) is additive:

\[
\min_{\{\bar{\zeta}(t)\}} \left( \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} (\zeta(t_1) - \bar{\zeta}(t_1) + \zeta(t_2) - \bar{\zeta}(t_2))^2 + 2\lambda \sum_{t=2}^{T-1} (\bar{\zeta}(t) - 2\bar{\zeta}(t) + \bar{\zeta}(t - 1))^2 \right).
\]

Now expand the first term in the minimization problem. We can always choose \( \bar{\zeta} \) so that \( \sum_{t=1}^{T} (\zeta(t) - \bar{\zeta}(t)) = 0 \) at no cost, so

\[
\sum_{t_1=1}^{T} \sum_{t_2=1}^{T} (\zeta(t_1) - \bar{\zeta}(t_1))(\zeta(t_2) - \bar{\zeta}(t_2)) = \sum_{t_1=1}^{T} (\zeta(t_1) - \bar{\zeta}(t_1)) \sum_{t_2=1}^{T} (\zeta(t_2) - \bar{\zeta}(t_2)) = 0.
\]

This means the minimization problem reduces to

\[
\min_{\{\bar{\zeta}(t)\}} 2 \left( \sum_{t=1}^{T} (\zeta(t) - \bar{\zeta}(t))^2 + \lambda \sum_{t=2}^{T-1} (\bar{\zeta}(t) - 2\bar{\zeta}(t) + \bar{\zeta}(t - 1))^2 \right),
\]

the usual Hodrick-Prescott filtering problem.