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
June 24, 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 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 a nonparametric decompositions of these two factors. We conclude that 88 percent of the variance of realized nonemployment durations is structural while the remaining 12 percent occurs across heterogeneous workers.

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.\(^1\) 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 alterative 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 individual heterogeneity in preferences, fixed costs, and stochastic

\(^1\)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.
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 expect it to fall. We maintain two key restrictions: for each individual, the evolution of a latent variable, the net benefit from employment, follows a geometric Brownian motion with drift during a 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. This implies that the duration of a nonemployment spell is given by the first passage time of a Brownian motion with drift.

In this environment, we ask three key questions. First, does the model have testable implications? We show that an economist armed with data on the joint distribution of the duration of two nonemployment spells for each individual can potentially reject the model. For example, if the true data generating process is one in which each individual has a constant hazard of finding a job, possibly different across individuals, the economist will always reject our model. Second, is the model nonparametrically identified? Again, we show that an economist armed with data on the joint distribution of the duration of two nonemployment spells can identify a subset of the parameters of the model, statistics related to the stochastic process for the net benefit from employment during a nonemployment spell. The economist cannot disentangle the two alternative interpretations of the structural model, however. Third, can we use this partially-identified model to decompose the observed decline in the hazard of exiting nonemployment into the portion attributable to structural duration dependence versus unobserved heterogeneity? We offer a simple decomposition that relies on our nonparametric estimates of the distribution of individual characteristics.

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 counterfactual decomposition. Using data on over 2.25 million individuals who experience at least two nonemployment spells, we find that we cannot reject our model and we uncover substantial heterogeneity across individuals. Nevertheless, our estimates suggest that 88 percent of the variance in realized nonemployment durations occurs within workers of a given type, while 12 percent reflects worker at heterogeneity. Most of the increase in the residual duration of a nonemployment spell at long durations reflects the deteriorating condition of workers who remain out of work for a long time.

There are a few other papers that use the first passage time of a Brownian motion to model duration dependence. Lancaster (1972) shows that 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. 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) is the paper most closely related to ours. He 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. We allow for two dimensions of heterogeneity. He proves nonparametric identification in this framework in two cases, when an observable characteristic (as well as unobservable characteristics) shifts the distance between the barriers, and when each individual is observed for two nonemployment spells.

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 (1984), show that such a model is nonparametrically identified using a single spell of nonemployment and appropriate variation in the observable characteristics of individuals.

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 nonemployment 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
the model has testable implications if we observe at least two nonemployment spells for each
individual, that it is nonparametrically identified under the same conditions, and that we
can use the model to decompose changes in the hazard of exiting nonemployment into the
portion that is structural and the portion that is attributable to changes in the composition
of the nonemployment pool. Section 3 summarizes the Austrian social security registry
data. Section 4 presents our results. Our model is not rejected by the Austrian data and
our preliminary results indicate that the majority 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:

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

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

\[
\rho > \mu_s + \sigma_s^2/2 \text{ for } s = n, e.
\]

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, but we expect that welfare conclusions, e.g. about whether employed or nonemployed workers are better off, do depend on the assumption.

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 \( \bar{\omega} = \omega \), 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. The hazard rate of exiting nonemployment always starts at 0 when \( t = 0 \), achieves a maximum value at some finite time \( t \), and then declines to a long run limit of \( \alpha^2/2 \).

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. 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 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 second is 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 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 $\psi$.

### 2.2 Testable Implications

The model admits very flexible densities, and thus a natural question to ask is whether such a model can explain any data, or in other words, can this model be tested. Suppose we observe a population of individuals, each of whom has fixed structural parameters $(\rho, \psi, b, \mu_e, \mu_n, \sigma_e, \sigma_n)$ and hence fixed reduced-form parameters $(\alpha, \beta)$. 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 $\mu_n = (\bar{\omega} - \underline{\omega})/d$, which implies that both $\alpha$ and $\beta$ converge to infinity with $\beta/\alpha = d$.

With two (or more) completed nonemployment spells for each individual, the model is testable. Let $\phi(t_1, t_2)$ denote the density of the distribution of the duration of the two completed nonemployment spells:

$$ \phi(t_1, t_2) = \int \int f(t_1, \alpha, \beta)f(t_2, \alpha, \beta)g(\alpha, \beta)d\alpha d\beta. $$

Let $\phi_1(t_1, t_2)$ and $\phi_2(t_1, t_2)$ denote the partial derivatives of $\phi$ with respect to $t_1$ and $t_2$ respectively. Using the functional form of $f(t, \alpha, \beta)$ in equation (3), this satisfies

$$ \phi_i(t_1, t_2) = \int \int \left( \frac{\beta^2}{2t_i^2} - \frac{3}{2t_i} - \frac{\alpha^2}{2} \right) f(t_1, \alpha, \beta)f(t_2, \alpha, \beta)g(\alpha, \beta)d\alpha d\beta $$

But note that we do allow for learning-by-doing, since a worker’s wage may increase faster on average when employed than when nonemployed, $\mu_e > \mu_n$. 

3But note that we do allow for learning-by-doing, since a worker’s wage may increase faster on average when employed than when nonemployed, $\mu_e > \mu_n$. 

7
or

\[ \frac{2t_i^2 \phi_i(t_1, t_2)}{\phi(t_1, t_2)} = b(t_1, t_2) - 3t_i - a(t_1, t_2)t_i^2, \quad (4) \]

where the constants \( a \) and \( b \) depend on the durations \( t_1, t_2 \) but not on the spell \( i \):

\[
a(t_1, t_2) = \frac{\int \int \alpha^2 f(t_1, \alpha, \beta) f(t_2, \alpha, \beta) g(\alpha, \beta) \, d\alpha \, d\beta}{\int \int f(t_1, \alpha, \beta) f(t_2, \alpha, \beta) g(\alpha, \beta) \, d\alpha \, d\beta} \equiv \mathbb{E}(\alpha^2 | t_1, t_2),
\]

and

\[
b(t_1, t_2) = \frac{\int \int \beta^2 f(t_1, \alpha, \beta) f(t_2, \alpha, \beta) g(\alpha, \beta) \, d\alpha \, d\beta}{\int \int f(t_1, \alpha, \beta) f(t_2, \alpha, \beta) g(\alpha, \beta) \, d\alpha \, d\beta} \equiv \mathbb{E}(\beta^2 | t_1, t_2).
\]

That is, \( a(t_1, t_2) \) is the expected value of \( \alpha^2 \) among workers who find their jobs at durations \( (t_1, t_2) \), while \( b(t_1, t_2) \) is the expected value of \( \beta^2 \) for the same group of workers.

For any \( t_1 \neq t_2 \), equation (4) gives us two equations in the two unknowns \( a(t_1, t_2) \) and \( b(t_1, t_2) \). Since these are expected values of \( \alpha^2 \) and \( \beta^2 \), respectively, they are non-negative, yielding a nonparametric test of the model.

**Proposition 1** For any \( t_1 \neq t_2 \),

\[
a(t_1, t_2) \equiv \frac{2(t_2^2 \phi_2(t_1, t_2) - t_1^2 \phi_1(t_1, t_2))}{\phi(t_1, t_2)(t_1^2 - t_2^2)} - \frac{3}{t_1 + t_2} \geq 0
\]

and

\[
b(t_1, t_2) \equiv t_1 t_2 \left( \frac{2t_1 t_2 (\phi_2(t_1, t_2) - \phi_1(t_1, t_2))}{\phi(t_1, t_2)(t_1^2 - t_2^2)} + \frac{3}{t_1 + t_2} \right) \geq 0.
\]

To illustrate the power of this test, consider the canonical search model where the hazard of finding a job is a constant \( h \). The density of completed spells is \( \phi(t_1, t_2) = h^2 e^{-h(t_1 + t_2)} \).

Then applying the formulae in Proposition 1 gives

\[
a(t_1, t_2) \equiv 2h - \frac{3}{t_1 + t_2} \quad \text{and} \quad b(t_1, t_2) = \frac{3t_1 t_2}{t_1 + t_2}.
\]

In particular, \( a(t_1, t_2) < 0 \) whenever \( t_1 + t_2 < 3/2h \), where \( 1/h \) represents the mean duration of a nonemployment spell. 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 \( G \) in the population, so the density of completed spells is \( \phi(t_1, t_2) = \int h^2 e^{-h(t_1 + t_2)} G(h) \). Then \( b(t_1, t_2) \) is unchanged, while

\[
a(t_1, t_2) = 2 \int h^3 e^{-h(t_1 + t_2)} dG(h) \int h^2 e^{-h(t_1 + t_2)} dG(h) - \frac{3}{t_1 + t_2}.
\]

If the ratio of the third moment of \( h \) to the second moment is positive, \( a \) is always negative for sufficiently small \( t_1 + t_2 \) and hence the more general model is rejected.
Further differentiating $\phi$ yields additional tests of the model. Let $\phi_{ij}(t_1, t_2)$ represent the partial derivative of $\phi$ with respect to $t_i$ and $t_j$. Then this satisfies

$$
\frac{2t_1^4\phi_{ii}(t_1, t_2)}{\phi(t_1, t_2)} = E(\alpha^4|t_1, t_2)t_1^4 + E(\beta^4|t_1, t_2) - 2E(\alpha^2\beta^2|t_1, t_2)t_1^2 + 6a(t_1, t_2)t_1^3 - 10b(t_1, t_2)t_1 + 15t_1^2
$$

and for $i, j = 1, 2$

$$
\frac{2t_1^2t_2^2\phi_{ij}(t_1, t_2)}{\phi(t_1, t_2)} = E(\alpha^4|t_1, t_2)t_1^2t_2^2 + E(\beta^4|t_1, t_2) - E(\alpha^2\beta^2|t_1, t_2)(t_1^2 + t_2^2) + 3a(t_1, t_2)t_1t_2(t_1 + t_2) - 3b(t_1, t_2)(t_1 + t_2) + 9t_1t_2
$$

For any $t_1 \neq t_2$, this gives three equations in three additional unknowns, the expected values of $\alpha^4$, $\beta^4$, and $\alpha^2\beta^2$ conditional on realized duration $(t_1, t_2)$. These can easily be solved. Since each of these three unknowns are the expected value of nonnegative numbers, this gives three additional tests of this model.

We can continue this process indefinitely at any fixed $t_1 \neq t_2$. The $k^\text{th}$ partial derivatives give the $k^\text{th}$ moment of the joint distribution of $(\alpha^2, \beta^2)$, a nonnegative number. In theory, we can reject the model if any of those moments is negative. In practice, it is difficult to measure the higher moments accurately and so we do not pursue these higher-order test statistics.

### 2.3 Nonparametric Identification

We turn next to the question of nonparametric identification. Once again, 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 individuals, each of whom has unboundedly large values of $\alpha$ and $\beta$, with the ratio determined so as to ensure that each worker deterministically finds a job at some time $t$. 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

---

4To prove this formally, we need to show that each step has a unique solution. We have shown this using the Mathematica software package for $k = 1, 2, \ldots, 19$ and anticipate that this is true for arbitrary $k$. 

---
as the density is strictly positive at all durations \( t > 0 \). Instead, we again use data on the joint density of two nonemployment spells to nonparametrically identify the joint distribution of \( \alpha^2 \) and \( \beta^2 \). Recall that \( \beta = (\bar{\omega} - \omega)/\sigma_n \) is nonnegative according to our theory, but \( \alpha = \mu_n/\sigma_n \) may be negative or positive. Therefore nonparametric identification of the joint distribution of \( \alpha^2 \) and \( \beta^2 \) is equivalent to nonparametric identification of the joint distribution of \( |\alpha| \) and \( \beta \). On the other hand, we also shows that the sign of \( \alpha \) is not identified using nonemployment duration data.

One way to understand the identification result is by building on the model’s testable implications. Suppose we have computed the complete set of moments of \( (\alpha^2, \beta^2) \) among those workers who find a job at any fixed \( (t_1, t_2) \). This tells us the joint distribution of \( (\alpha^2, \beta^2) \) among those workers who find a job at duration \( (t_1, t_2) \), say \( \psi(\alpha^2, \beta^2, t_1, t_2) \). Using that, we can compute the joint distribution of \( (\alpha^2, \beta^2) \) in the initial population of workers, say \( \hat{g}(\alpha^2, \beta^2) \), since

\[
\psi(\alpha^2, \beta^2, t_1, t_2) = \frac{f(t_1, \alpha, \beta)f(t_2, \alpha, \beta)\hat{g}(\alpha^2, \beta^2)}{\iint f(t_1, \alpha', \beta')f(t_2, \alpha', \beta')\hat{g}(\alpha'^2, \beta'^2)d\alpha'd\beta'},
\]

which can easily be inverted to solve for \( \hat{g} \). If we knew the sign of \( \alpha \), we could then immediately recover the object of interest, \( g(\alpha, \beta) \).

Indeed, this approach to nonparametric identification suggests another test of the model. The function \( \hat{g}(\alpha^2, \beta^2) \) should not depend on the duration of a nonemployment spell \( (t_1, t_2) \) that we use for measurement and in particular should be the same across any two pairs of durations. In practice, we expect that the power of such a test will be minimal since it depends on all the partial derivatives of \( \phi \) and hence is likely measured with noise.

An alternative approach to nonparametric identification turns out to be more useful in practice. Suppose we knew the distribution of types \( g(\alpha, \beta) \). For each type, the model gives the density over durations \( f(t_1, \alpha, \beta)f(t_2, \alpha, \beta) \), yielding the density of realized durations \( \phi(t_1, t_2) = \iint f(t_1, \alpha, \beta)f(t_2, \alpha, \beta)g(\alpha, \beta)d\alpha d\beta \). If there were finitely many types \( N \) and finitely many combinations of durations \( T \), we could represent this as a linear system \( \phi = F \cdot g \), where \( F \) is a \( T \times N \) likelihood matrix with typical element \( f(t_1, \alpha, \beta)f(t_2, \alpha, \beta) \), \( g \) is an \( N \) vector of type shares, and \( \phi \) is a \( T \) vector of duration shares. Identification then comes down to invertibility of the likelihood matrix \( F \): Can we solve this equation uniquely for \( g = F^{-1} \cdot \phi \)? If \( N > T \), there may be many such solutions to this equation, while if \( T > N \)

---

5 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.

6 The number of individuals with incomplete nonemployment spells reveals some information about negative values of \( \alpha \) but is insufficient for identification.
there is generically no solution. We are interested in the case where \( N = T \), so identification comes down to whether the rows of the likelihood matrix are linearly independent.

This argument provides additional insight into why the sign of \( \alpha \) is not identified. For any \((t, \alpha, \beta)\),

\[
f(t, \alpha, \beta) / f(t, -\alpha, \beta) = e^{2\alpha \beta},
\]

and so if we allow for both positive and negative values of \( \alpha \), two rows of the likelihood matrix are proportional to each other and so the matrix is singular. On the other hand, if we restrict attention to positive values of \( \alpha \) (and, according to the theory, \( \beta \)), then the only solution \( G \) to the equation

\[
f(t_1, \alpha, \beta)f(t_2, \alpha, \beta) = \int \int f(t_1, \alpha', \beta')f(t_2, \alpha', \beta)dG(\alpha', \beta')
\]

for all \((t_1, t_2)\) and arbitrary \((\alpha, \beta)\) is the one in which \( G \) puts all its mass on \((\alpha, \beta)\). The density of realized durations for one individual is not simply a linear combination of the density of realized durations for other individuals and so our model is nonparametrically identified.

### 2.4 Decomposition of Changes in the Hazard Rate

We turn now to the relative importance of structural duration dependence and dynamic selection for the evolution of the hazard rate of exiting nonemployment. When we do this, we assume that \( \alpha \geq 0 \) for all individuals, which ensures that every nonemployment spell ends in finite time. The limitations of our estimation procedure prevent us from testing this restriction.

We propose two decompositions. The first is based on variances. For any individual \((\alpha, \beta)\), the mean of his realized nonemployment duration is

\[
\int_0^\infty t f(t, \alpha, \beta) dt = \frac{\beta}{\alpha}
\]

and the associated variance is

\[
\int_0^\infty (t - \beta/\alpha)^2 f(t, \alpha, \beta) dt = \frac{\beta}{\alpha^3}
\]

The within-individual variance of duration is just the population average of this,

\[
\int \int (\beta/\alpha^3) g(\alpha, \beta) d\alpha d\beta.
\]
The between-individual variance of duration is the variance in \( \beta/\alpha \) across individuals,

\[
\int \int (\beta/\alpha)^2 g(\alpha, \beta) \, d\alpha \, d\beta - \left( \int \int (\beta/\alpha) g(\alpha, \beta) \, d\alpha \, d\beta \right)^2.
\]

The total variance is the sum of these two terms. Then the share of the total variance in durations that occurs within individuals is

\[
\int \int (\beta/\alpha^3) g(\alpha, \beta) \, d\alpha \, d\beta - \left( \int \int (\beta/\alpha) g(\alpha, \beta) \, d\alpha \, d\beta \right)^2.
\]

If there is no variance in the shocks, so \( \omega(t) \) is deterministic, \( \alpha \) and \( \beta \) are unboundedly large but their ratio converges to a finite number. The within share of variance converges to 0. Conversely, if the cross-sectional distribution \( g \) is degenerate, the within share converges to 1 and the between share to 0. Thus our model does not presuppose any particular result from this decomposition.

We also propose a decomposition of the hazard rate. Let \( \tilde{g}(\alpha, \beta|t) \) denote the distribution of types \( (\alpha, \beta) \) among the workers whose nonemployment spell lasts at least \( t \) periods. This satisfies

\[
\tilde{g}(\alpha, \beta|t) = \frac{(1 - F(t, \alpha, \beta))g(\alpha, \beta)}{\int \int (1 - F(t', \alpha', \beta'))g(\alpha', \beta') \, d\alpha' \, d\beta'},
\]

where \( F(t, \alpha, \beta) = \int_{t}^{\infty} f(t', \alpha, \beta) \, dt' \) is the fraction of type \( (\alpha, \beta) \) workers whose spell lasts at least \( t \) periods. Naturally \( \tilde{g}(\alpha, \beta|0) = g(\alpha, \beta) \). Also let \( \tilde{f}(\tau, \alpha, \beta|t) \) denote the density of residual nonemployment duration for a type \( (\alpha, \beta) \) worker conditional on the nonemployment spell lasting at least \( t \) periods:

\[
\tilde{f}(\tau, \alpha, \beta|t) = \frac{f(\tau + t, \alpha, \beta)}{1 - F(t, \alpha, \beta)}.
\]

Naturally \( \tilde{f}(\tau, \alpha, \beta|0) = f(\tau, \alpha, \beta) \). Then the actual density of residual nonemployment duration for workers whose nonemployment spell lasts at least \( t \) periods is

\[
f^*(\tau|t) \equiv \int \int \tilde{f}(\tau, \alpha, \beta|t) \tilde{g}(\alpha, \beta|t) \, d\alpha \, d\beta.
\]

If we know \( g(\alpha, \beta) \), we can compute this directly using the structure of the model.

We can then consider two counterfactuals. First, suppose that the distribution of residual nonemployment duration at time \( t \) for each worker \( (\alpha, \beta) \) changes as in the data, but the distribution of \( (\alpha, \beta) \) remains fixed at its time 0 level. Then residual nonemployment duration
would satisfy
\[ f^s(\tau|t) \equiv \int \int \tilde{f}(\tau, \alpha, \beta|t) g(\alpha, \beta) d\alpha d\beta. \]

Changes in \( f^s \) represent the structural component of duration dependence, since it holds fixed the average composition of the labor force. If there were no heterogeneity, then \( f^r(\tau|t) = f^s(\tau|t) \) for all \( \tau \) and \( t \). Conversely, if there were no structural duration dependence, so \( \tilde{f}(\tau, \alpha, \beta|t) = f(\tau, \alpha, \beta) \) for all \( \tau, \alpha, \beta \), then \( f^s(\tau|t) \) would not depend on \( t \) and in particular would equal \( f^r(\tau|0) = f^s(\tau|0) \).

Alternatively, suppose the distribution of residual nonemployment duration at time \( t \) for each worker \( (\alpha, \beta) \) were the same as it was at time 0, but the distribution of \( (\alpha, \beta) \) evolves as in the data. Then residual nonemployment duration would satisfy
\[ f^h(\tau|t) \equiv \int \int f(\tau, \alpha, \beta) \tilde{g}(\alpha, \beta|t) d\alpha d\beta. \]

Changes in \( f^h \) represents the ex ante heterogeneity component of duration dependence, since it holds fixed the hazard of transiting into employment for each type of worker. If there were no structural duration dependence, then \( f^r(\tau|t) = f^h(\tau|t) \) for all \( \tau \) and \( t \). Conversely, if there were no heterogeneity, then \( f^h(\tau|t) \) would equal \( f^r(\tau|0) = f^h(\tau|0) \) for all \( t \).

We can summarize these distributions through their moments. For example, the mean residual nonemployment duration is
\[ D^r(t) = \int_0^\infty \tau f^r(\tau|t) d\tau, \]
while the component of residual nonemployment duration due to structural duration dependence and ex ante heterogeneity are
\[ D^s(t) = \int_0^\infty \tau f^s(\tau|t) d\tau \text{ and } D^h(t) = \int_0^\infty \tau f^h(\tau|t) d\tau, \]
respectively. Assuming \( D^r(t) > D^r(0) \) for some \( t \), we can decompose the increase in residual duration as follows: the portion accounted for by structural duration dependence is \( (D^s(t) - D^s(0))/(D^r(t) - D^r(0)) \), while the portion accounted for by ex ante heterogeneity is \( (D^h(t) - D^h(0))/(D^r(t) - D^r(0)) \).
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 (Zweimuller, 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. We measure spells in weeks, starting at 2 weeks and topcode the spells at 401 weeks. 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. Identification of the model requires two completed spells for each individual: we consider the first two spells in the data set after the age of 25.
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.

Table 1 shows how the sample size shrinks after imposing different criteria. There are 38,828,896 non-employment spells in the Austrian data. We exclude 810,589 spells that contain a maternity leave, 16,356,697 spells for workers less than 25 years old, and 5,894,395 spells that last less than two weeks. The sample then contains 15,767,215 spells for 3,330,067 workers. Out of these, 1,078,721 workers have a single spell; these are excluded. In the final sample we are left with 2,251,346 workers who have two or more spells on-employment spells. In fact, these workers have 4.4 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 48.6 weeks, and the average employment duration between these two spells is 56.9 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 variation, but rather reflects the fact that many jobs start during the first week of the month.

---

Table 1: Construction of the data

<table>
<thead>
<tr>
<th>Spell Type</th>
<th>Spells</th>
</tr>
</thead>
<tbody>
<tr>
<td>all spells</td>
<td>38,828,896</td>
</tr>
<tr>
<td>maternity leave</td>
<td>810,589</td>
</tr>
<tr>
<td>age less than 25</td>
<td>16,356,697</td>
</tr>
<tr>
<td>shorter than 2 weeks</td>
<td>5,894,395</td>
</tr>
<tr>
<td>single spell</td>
<td>1,078,721</td>
</tr>
<tr>
<td>third spell or more</td>
<td>10,185,802</td>
</tr>
<tr>
<td>final sample</td>
<td>4,502,692</td>
</tr>
</tbody>
</table>

---

In the top-coded data, the mean is 43.8 weeks.
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 Proposition 1. 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 measure 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 follow Proposition 1 and test whether

\[ a(t_1, t_2) = \frac{t_2^2 \log \left( \frac{\phi(t_1, t_2+1)}{\phi(t_1, t_2-1)} \right) - t_1^2 \log \left( \frac{\phi(t_1+1, t_2)}{\phi(t_1-1, t_2)} \right)}{t_1^2 - t_2^2} - \frac{3}{t_1 + t_2} \]

and \[ b(t_1, t_2) = t_1 t_2 \left( \frac{t_1 t_2 \log \left( \frac{\phi(t_1, t_2+1) \phi(t_1-1, t_2)}{\phi(t_1, t_2-1) \phi(t_1+1, t_2)} \right)}{t_1^2 - t_2^2} + \frac{3}{t_1 + t_2} \) \]

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 \( \tilde{\phi}(t_1, t_2) \) that minimizes the sum of squared deviations of \( \tilde{\phi}(t_1, t_2) \) from
Figure 3: Nonparametric test of model. The figure shows the percent of observations with $a(t_1, t_2) < 0$ or $b(t_1, t_2) < 0$ with $2 \leq t_1 < t_2 \leq 60$ for different values of the filtering parameter $\lambda$.

the data $\phi(t_1, t_2)$ subject to a penalty $\lambda$ that penalizes changes in the slope of $\overline{\phi}(t_1, t_2)$:

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

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 $(t_1, t_2)$ with $2 \leq t_1 < t_2 \leq 60$. Setting the smoothing parameter $\lambda$ 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 $(t_1, t_2)$, we reject the model more often, even in smoothed data. This may be due to a reduction in the

---

8In practice we smooth the function $\log(1 + \phi(t_1, t_2))$, rather than $\phi$, where $\phi$ is the number of individuals whose two spells have durations $(t_1, t_2)$.
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:

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

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:

$$
F_{i,j} = \Pr \{ t_1 \in (t_1(i), t_1(i) + dt], t_2 \in (t_2(i), t_2(i) + dt] \mid (\alpha, \beta) = (\alpha(j), \beta(j)) \}
$$
for a small positive $dt$. For small positive values of $dt$,

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

(6)

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:

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

(7)

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 (6) 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 2: Summary statistics from estimation

<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>E(β/α)</th>
<th>Var(β/α)</th>
<th>E(β/α³)</th>
<th>E(2/α²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43.6</td>
<td>916</td>
<td>6948</td>
<td>243</td>
</tr>
</tbody>
</table>

6. If for some $j = 1, ..., N$ 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 $i = 1, ..., M$, we assign $g(j) = 0$ to that parameter pair.

7. With large values for $M$ and $N$, solving the problem in equation (7) 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.
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 Counterfactual Exercises

Finally, we report the outcome of our counterfactual exercises. Figure 7 shows that the mean residual duration of an in progress nonemployment spell, $D^r(t)$, increases rapidly from 42 weeks for a newly nonemployed worker to over 9 years for a worker who has been out of work for 8 years.\(^9\) In contrast, the portion attributable to heterogeneity, $D^h(t)$, rises more slowly and then falls at longer durations. That is, the initial nonemployment duration of the workers who experience the longest nonemployment spells is not particularly long, just 65 weeks for workers who have been out of work for 8 years.

Figure 8 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, most of the increase in residual duration is accounted for by changes in the characteristics of the unemployed pop-

---

\(^9\)In our data set, the residual nonemployment duration of a worker who has been out of work for 400 weeks is 235 weeks, about half the number we generate from the model. We believe this difference reflects selection in the data. If our model were the data generating process, we would miss many of the longest spells because we include only individuals who complete two spells and our data set “only” covers 35 years.
Figure 8: 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.

ulation, but the role of this factor eventually falls below twenty percent. The importance of structural duration dependence, on the other hand, becomes more pronounced after the first few weeks of nonemployment, reaching sixty percent asymptotically. Correlation between these two factors explain the remaining increase in nonemployment duration.

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 2.2 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

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

for all $t_1$ and $t_2$. Honoré (1993) uses this expression to argue that the model is nonparametrically identified.

---

10It 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.
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 2** 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 9 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 9: 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:

$$\rho E(\omega) = \exp(\omega) + \mu_e E'(\omega) + \frac{\sigma^2_e}{2} E''(\omega) \text{ for all } \omega \in [\omega, \infty),$$ (8)

$$\rho N(\omega) = b + \mu_n N'(\omega) + \frac{\sigma^2_n}{2} N''(\omega) \text{ for all } \omega \in (-\infty, \bar{\omega}].$$ (9)

The solution to these equations is

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

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

where

$$\lambda_{e_1} < 0 < \lambda_{e_2} \text{ and } \lambda_{n_1} < 0 < \lambda_{n_2}$$

are the roots of the equations

$$\rho = \lambda_e (\mu_e + \lambda_e \sigma^2_e/2) \text{ and } \rho = \lambda_n (\mu_n + \lambda_n \sigma^2_n/2).$$

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

$$E(\omega) = N(\omega)$$ (10)

$$E(\bar{\omega}) = N(\bar{\omega}) + \psi$$ (11)

$$E'(\omega) = N'(\omega)$$ (12)

$$E'(\bar{\omega}) = N'(\bar{\omega})$$ (13)

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:

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

Equation (14) requires that for arbitrarily low \( \omega \) the value functions converges to the value of nonemployment forever. Likewise equation (15) 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 (10)–(15). We turn to their solution.

First, the no-bubble conditions (14) and (15) 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:

\[
\begin{align*}
E(\omega) &= \frac{\exp(\omega)}{\rho - \mu_e - \sigma_e^2/2} + e \exp(\lambda_e \omega) \quad \text{for all} \ \omega \in [\omega, \infty) \\
N(\omega) &= \frac{b}{\rho} + n \exp(\lambda_n \omega) \quad \text{for all} \ \omega \in (-\infty, \bar{\omega}]
\end{align*}
\]

with

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

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

\[
\frac{\exp(\omega)}{\rho - \mu_e - \sigma_e^2/2} + e \exp(\lambda_e \omega) = \frac{b}{\rho} + n \exp(\lambda_n \omega) \tag{20}
\]

\[-\psi + \frac{\exp(\bar{\omega})}{\rho - \mu_e - \sigma_e^2/2} + e \exp(\lambda_e \bar{\omega}) = \frac{b}{\rho} + n \exp(\lambda_n \bar{\omega}) \tag{21}\]

\[
\frac{\exp(\omega)}{\rho - \mu_e - \sigma_e^2/2} + e \lambda_e \exp(\lambda_e \omega) = n \lambda_n \exp(\lambda_n \omega) \tag{22}
\]

\[
\frac{\exp(\bar{\omega})}{\rho - \mu_e - \sigma_e^2/2} + e \lambda_e \exp(\lambda_e \bar{\omega}) = n \lambda_n \exp(\lambda_n \bar{\omega}) \tag{23}\]

Note that the values of \(e\) and \(n\) have to be positive, since it is feasible to chose to either be employed forever or nonemployed forever, and since the value of being employed forever and nonemployed forever are the obtained in equations (8) and equations (9) by setting \(e = 0\) and \(n = 0\) respectively.

Figure 10 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 \(N\) coincides, 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 an 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:

\[
\omega' = \log(k) + \omega, \quad \bar{\omega}' = \log(k) + \bar{\omega}, \quad e' = ek^{1-\lambda_e}, \quad n' = nk^{1-\lambda_n}. \tag{24}\]

The proof of Lemma 1 follows by multiplying equations (20), (21), (22), and (23), which characterize the solution of the value function by \(k\), and using the relationships in equation (24). 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 \(\bar{\omega} - \omega\).
Figure 10: 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 $\rho - \mu_e - \sigma^2_e/2 > 0$. For any pair of thresholds $\underline{\omega} < \bar{\omega}$ 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 $\underline{\omega}, \bar{\omega}$ is optimal.

**Proof.** We need to show that there is a unique positive 4-tuple $(\psi, b, e, n)$ that solves equations (20), (21), (22), and (23). First we note that for any pair $\bar{\omega} > \underline{\omega}$, under the assumptions that $\lambda_e < -1$ and $\lambda_n > 1$ then the smooth pasting conditions (22) and (23) define a linear system of equations with a unique positive solution $(e, n)$. We can solve these equations to yield:

$$e = \frac{\exp(\underline{\omega}(1 - \lambda_n)) - \exp(\bar{\omega}(1 - \lambda_n))}{\lambda_e(\rho - \mu_e - \sigma^2_e/2)\left[\exp((\lambda_e - \lambda_n)\underline{\omega}) - \exp((\lambda_e - \lambda_n)\bar{\omega})\right]} > 0 \quad (25)$$

and

$$n = \frac{\exp(\underline{\omega}(1 - \lambda_n) + \bar{\omega}(\lambda_e - \lambda_n)) - \exp(\bar{\omega}(1 - \lambda_n) + \underline{\omega}(\lambda_e - \lambda_n))}{\lambda_n(\rho - \mu_e - \sigma^2_e/2)\left[\exp((\lambda_e - \lambda_n)\underline{\omega}) - \exp((\lambda_e - \lambda_n)\bar{\omega})\right]} > 0 \quad (26)$$
Since $\bar{\omega} > \underline{\omega}$ 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 $\bar{\omega} > \underline{\omega}$ then denominator of $e$ is positive, and hence $e > 0$. The numerator of $n$ is negative if

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

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 (20) and equation (21) we get:

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

$$\frac{b}{\rho} = - \psi + \frac{\exp (\bar{\omega})}{\rho - \mu_e - \sigma_e^2/2} + e \exp (\lambda_e \bar{\omega}) - n \exp (\lambda_n \bar{\omega}) \tag{28}$$

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 (27) 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 $\bar{\omega} > \underline{\omega}$ 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:

$$b = \rho \left( \frac{\lambda_n (\lambda_e - 1) (-e^{\lambda_e \bar{\omega} + \lambda_n \omega} + (\lambda_n - 1)\lambda_e e^{\lambda_n \bar{\omega} + \lambda_e \omega} + (\lambda_e - \lambda_n) e^{\lambda_n \bar{\omega} + \lambda_e \omega})}{\lambda_n \lambda_e (\mu + \rho + \sigma_e^2/2) (e^{\lambda_e \omega + \lambda_n \bar{\omega}} - e^{\lambda_n \omega + \lambda_e \bar{\omega}})} \right) \tag{29}$$
and

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

\[
+ \frac{-(\lambda_n - 1)\lambda_e e^{\lambda_n \omega + \lambda_n \bar{\omega}} + \bar{\omega}}{\lambda_n \lambda_e (\mu - \rho + \sigma^2/2)(e^{\lambda_n \omega + \lambda_n \bar{\omega}} - e^{\lambda_n \bar{\omega} + \lambda_e \bar{\omega}})} > 0
\]

(30)

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 11 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 12 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, \(\tilde{\psi}(t_1, t_2) + \tilde{\psi}(t_1, t_2)\), where \(\tilde{\psi}\) is a smooth “trend” and \(\tilde{\psi}\) is the residual. Following Hodrick and Prescott.
Figure 11: 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 12: Hazard rate in different subperiods
1997), we define the trend as the solution to

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

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

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

The first 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.

### D Higher Moments of \( g(\alpha, \beta) \) Distribution

As shown in section 2.2, the \( k^{th} \) partial derivatives of the joint distribution of spells give a system of \( k + 1 \) linear equations for \( k + 1 \) conditional moments of the distribution. In a matrix notation, this can be expressed as \( \mathbb{E}[\alpha^i \beta^j | t_1, t_2] = A^k[\phi_{ij}(t_1, t_2)] \), where \( \mathbb{E}[\alpha^i \beta^j | t_1, t_2] \)
is a vector of conditional moments for \( i, j = 1, \ldots k \), \( [\phi_{ij}(t_1, t_2)] \), \( i, j = 1, \ldots k \) is a vector of partial derivatives, and \( \mathbb{A}^k \) is a \((k+1) \times (k+1)\) matrix of coefficients.

For \( k = 1 \) and \( k = 2 \), the matrices \( A^1 \) and \( A^2 \) are

\[
A^1 = \begin{pmatrix}
-1 & 1/t_1^2 \\
-1 & 1/t_2^2
\end{pmatrix},
A^2 = \begin{pmatrix}
1 & -2/t_1 & 1/t_1^4 \\
1 & -2/t_2 & 1/t_2^4 \\
1 & -(t_1 + t_2)/t_1 t_2 & 1/t_1^2 t_2^2
\end{pmatrix}.
\]

The matrices have a full rank and thus the solution for 1\(^{st}\) and 2\(^{nd}\) moments is unique.