The Proportional Hazard Model: Estimation and Testing using Price Change and Labor Market Data

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

June 1, 2016
Preliminary

Abstract

We use labor market data and data on price changes to examine the role of structural duration dependence and heterogeneity in shaping aggregate hazard rates. In line with an extensive literature, we examine this question through the lens of the proportional hazard model. We focus on environments where we observe two observations per individual, as this not only allows us to estimate the model non-parametrically, but also to test whether the true data-generating process is likely to have a structure imposed by a proportional hazard model. Formally, we can reject the null hypothesis that the proportional hazard model is the data generating process, but this is not surprising given the size of our data sets. We also estimate the model and find that the estimates are sensitive to assumptions on the distribution of unobserved heterogeneity. In addition, we show that, if the true data generating process for our labor market data is given by a structural stopping time model, estimates that impose the proportional hazard assumption understate the importance of heterogeneity in the economy.
1 Introduction

The longer a worker has been out of work, the less likely he is to return to work in the near future. The longer he has been working, the less likely he is to lose his job. The more time has passed since a firm has last changed its price, the less likely the firm is to change its price in the near future. These facts are well-known, as is the difficulty in interpreting them: even if every individual finds a job, loses a job, or changes a price at a constant rate, heterogeneity across individuals can give rise to the patterns described here. The composition of the “surviving” population, individuals who have not yet experienced such an event, changes endogenously as time passes.

This paper uses large labor market and price data sets to reexamine the role of structural duration dependence and ex ante heterogeneity in shaping aggregate hazard rates. Following much of the literature on duration models, we examine the data through the lens of the proportional hazard model. The model specifies that the hazard rate of exiting a state in a given period (finding a job, losing a job, or changing a price) is the product of two functions: a function of an individual’s observed and unobserved characteristics; and a function of the time already spent in the state. We view this as a convenient statistical representation of the data but recognize that the multiplicative structure is restrictive and potentially incorrect.¹

In much of our analysis, we focus on environments in which we observe two spells for each individual. We also assume that each individual’s characteristics are constant across the two spells. Honoré (1993) proves that the proportional hazard model is nonparametrically identified under these assumptions. In particular, consider a data set containing the duration of two completed spells for each individual and no other covariates. Assume that the probability that an individual’s spell ends in period $t$ conditional on not having ended prior that period can be expressed as $\theta h(t)$, where $\theta$ is a function of the individual’s characteristics and $h(t)$ is the common baseline hazard rate. Except for one necessary normalization, we can recover both the distribution of $\theta$ and the baseline hazard $h$ from this data set.

We use two large data sets, one containing price data from the United States and the other containing employment data from Austria, to test and estimate a proportional hazard model. We start by proposing a hazard rate dominance test. Suppose we observe two spells for each individual. The duration of each spell is determined by a proportional hazard model and the proportionality constant $\theta$ is the same for each individual in the two spells. Then we prove that the hazard rate during the second spell is a decreasing function of the duration of the first spell at all durations. We find that this hazard rate dominance condition is satisfied

¹Van den Berg (2001) surveys the literature using the proportional hazard model. In Section 4.3, he develops an economic model that generates non-constant but proportional hazard rates.
in our price setting data but violated in our labor market data.\footnote{This version of the paper focuses only on the job finding probability. In future versions, we plan to look at the probability of exiting employment as well.} The nature of the violation is telling: some workers simply take longer to find a job, but do not have a low hazard rate at long durations.

The hazard rate dominance test is just one overidentifying restriction of the proportional hazard model. Building on Honoré (1993), we find the full set of overidentifying restrictions. We can reject the overidentifying restrictions in both data sets.

The question remains what to make of this finding. Almost any model will be rejected in a sufficiently large data set, even if the model offers a close approximation to reality. We therefore still estimate the proportional hazard model on both data sets. We consider several procedures. First, we estimate the proportional hazard model imposing parametric restrictions on the distribution of individual characteristics $G$. Second, we use a nonparametric estimator proposed by Horowitz and Lee (2004). Third, we develop an alternative non-parametric estimator using quasi-maximum likelihood. The latter two estimators give similar results, but, as emphasized by Heckman and Singer (1984b), the parametric estimators are misleading in our application. Imposing the wrong parametric functional form leads us to understate the importance of heterogeneity and hence overstate how much the baseline hazard declines as duration increases.

Finally, we turn to quantitatively-reasonable synthetic data in order to evaluate the effect of imposing the proportional hazard assumption in environments where it is invalid. To do this, we develop a structural model that we believe may be reasonable approximation to the data generating process for both the labor market models.\footnote{We are currently exploring the possibility of doing this for the price setting data.} We assume individual durations are determined by a stopping time model, as in Alvarez, Borovičková, and Shimer (2016). If this model is in fact the data generating process, we conclude that the proportional hazard assumption leads us to miss much of the heterogeneity that is present in the data.

At the end of the paper, we also considers another approach to nonparametric identification in the proportional hazard model, using covariates as in Elbers and Ridder (1982) and Heckman and Singer (1984a). More precisely, assume that an individual’s hazard is the product of three terms, a baseline hazard, an unknown function of observed covariates, and an individual fixed effect which is orthogonal to the covariates. Then Elbers and Ridder (1982) and Heckman and Singer (1984a) prove that, under certain regularity conditions, the model is nonparametrically identified. As in the previous case, we prove that if the model is correctly specified, there are many equivalent nonparametric estimates of the model. In this case, our results are noisy because we cannot find covariates that are good at predicting the
duration of a spell. We cannot reject the model, but we also do not obtain precise estimates of the baseline hazard. This section of the paper is more preliminary.

2 The Proportional Hazard Model

We look at both continuous and discrete time versions of the proportional hazard model. In both cases, we consider a population with measure 1. Each individual has a fixed type $\theta$. Let $G(\theta)$ denote the cumulative distribution function and $g(\theta)$ denote the associated probability density function in the population. As we discuss below, the fixed type may be correlated with some observable individual characteristics, but in general we are interested in cases where the econometrician does not observe $\theta$ perfectly.

2.1 Continuous Time

In continuous time, the probability that a spell ends prior to period $t$ for an individual with type $\theta$ is

$$F(t; \theta) \equiv 1 - e^{-\theta \int_0^t h(\tau) d\tau}$$

for some nonnegative, integrable function $h$. Equivalently, $\theta h(t) \equiv F_t(t; \theta)/(1 - F(t; \theta))$ is the instantaneous hazard of a spell ending at time $t$. Thus the individual’s type $\theta$ scales up and down the baseline hazard rate $h$ at all durations. This is the proportional hazard assumption.

The population hazard rate at duration $t$ is

$$H(t) \equiv \frac{\int \theta h(t)e^{-\theta \int_0^t h(\tau) d\tau} g(\theta) d\theta}{\int e^{-\theta \int_0^t h(\tau) d\tau} g(\theta) d\theta}$$

This is the weighted average of individual hazards, where the weights are proportional to the probability that the individual’s spell lasts at least $t$ periods. While we imagine that we can observe the population hazard $H$ in the data, the baseline hazard rate $h$ and the type distribution $G$ are unknown.

2.2 Discrete Time

In discrete time, the probability that a spell ends prior to period $t$ for an individual with type $\theta$ is

$$F(t; \theta) = 1 - \prod_{\tau=1}^{t-1} (1 - \theta h(\tau)),$$
where $h$ is nonnegative and $\theta h(\tau) \leq 1$ for all $\theta$ and $\tau$. The hazard of the spell ending in period $t$ is $\theta h(t) \equiv (F(t+1; \theta) - F(t; \theta))/F(t; \theta)$.

The population hazard rate at duration $t$ is

$$H(t) \equiv \frac{\int \theta h(t) \prod_{\tau=1}^{t-1} (1 - \theta h(\tau)) g(\theta) d\theta}{\int \prod_{\tau=1}^{t-1} (1 - \theta h(\tau)) g(\theta) d\theta}$$

Again, this is the weighted average of individual hazards conditional on the individual’s spell lasting at least $t$ periods. Again, we assume that we can measure $H$ in the data, but we do not know $h$ or $G$.

3 Data

Our paper uses two data sets to test and estimate the proportional hazard model. Each contains information on a large number of individuals with multiple completed spells. We discuss key aspects of these data sets here and leave a more detailed discussion of our sampling framework for Appendix A.

3.1 Price Setting Data

We look at data on the duration of time between price changes for individual goods. We measure this using the Nielsen-IRI retail scanner data sets, which are available through the Kilts Center for Marketing at the University of Chicago. The data sets contain a large number of weekly price observations for many products in many retail outlets. In particular, the data sets include approximately 2.6 million distinct UPC codes. For each code, participating stores report weekly revenue and quantity sold.

We define a product as a particular UPC code in a particular location. For each such product, we use weekly revenue and quantity sold to compute the average weekly price of the product, which we in turn use to calculate the duration of time between regular price changes. We define a price spell as the elapsed time (in weeks) between two price changes, and consider only price spells longer than 2 weeks. The products are divided into 31

---

4 The data set includes an indicator for sales. We exclude sales prices to compute regular price changes. We also consider only price changes larger than 0.1 percent. This is because some changes in average prices are due to the fact that some customers shop with coupons, which we cannot directly observe. Imposing a lower bound on the price change is a way to exclude such price changes.

5 Since we only observe average weekly prices, a price change in the middle of a week would otherwise generate a spurious spell of duration one week. For example, suppose that the price of a product increases from $1 to $2 in the middle of a week. Then we would measure average price of $1 in week 1, $1.5 in week 2 and $2 in week 3, which would looks like as if there were two price changes.
categories, e.g. coffee and frozen entrées. The number of products differs across categories, ranging from about 50,000 for razors to 1.8 million for frozen dinners. At this point, we present detailed results for the 600,000 coffee products, but in future versions of the paper we hope to summarize results for all 31 product categories.

### 3.2 Labor Market Data

For our labor market application, we use 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.

We work with non-employment spells, defined as the time from the end of one full-time job to the start of the following full-time job. We require that the worker is registered as unemployed for at least one day during the spell to get rid of job-to-job transitions. We drop incomplete spells and spells involving a maternity leave. Although in principle we could measure non-employment duration in days, disproportionately many jobs start on Mondays and end on Fridays, and so we focus on weekly data. We code less than one calendar week (i.e. between Monday and Sunday) out of work as 0 weeks, more than one and less than two calendar weeks as 1 week, and so on. Our sampling framework consists of all individuals who were no older than 45 in 1986 and no younger than 40 in 2007, so that each individual has at least 15 years when he could potentially be at work. We consider all individuals with at least one non-employment spell which started after 1986 and require that the individual was at least 25 years old when the spell started.

Austrian data is well-suited to this application. 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 non-employment spells per individual. Additionally, 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. 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.

4 Hazard Rate Dominance

4.1 Theory

Suppose each individual has an observable characteristic $x$ which is correlated with their unobserved type $\theta$. Our leading example of the observable characteristic will be the duration of a prior spell, but this is only one possible interpretation. Individuals with different values of $x$ have different type distributions. Denote the cumulative distribution function of the unknown type by $G(\theta|x)$ and the associated probability density function by $g(\theta|x)$.

Let $H(t|x)$ denote the hazard rate conditional on the characteristic $x$. We first claim that if a higher characteristic $x$ raises the type density $g$ proportionately more at higher types $\theta$, then a higher characteristic implies a lower hazard rate at all durations:

**Lemma 1** Assume that for all $\theta_1 < \theta_2$ and $x_1 < x_2$, $g(\theta_1|x_1)g(\theta_2|x_2) < g(\theta_1|x_2)g(\theta_2|x_1)$. Then for all $t$, $H(t|x_1) > H(t|x_2)$.

**Proof.** In continuous time, the characteristic-conditional hazard rate is

$$H(t|x) \equiv h(t) \int \frac{\theta e^{-\theta \int_0^t h(\tau)d\tau} g(\theta|x)}{\int e^{-\theta \int_0^t h(\tau)d\tau} g(\theta|x)d\theta} d\theta.$$

Let $\tilde{g}(\theta|x,t) \equiv k(x,t)e^{-\theta \int_0^t h(\tau)d\tau} g(\theta|x)$ denote the probability density function among individuals with characteristic $x$ whose spell lasts at least $t$ periods, where $k(x,t)$ is chosen to ensure that $\tilde{g}(\theta|x,t)$ is a proper density function for all $x$ and $t$. Also let $\tilde{G}(\theta|x,t)$ denote the associated cumulative distribution function. Thus $H(t|x) = h(t) \int \theta \tilde{g}(\theta|x,t)d\theta$, the product of the baseline hazard and the average value of $\theta$ in the relevant population.

The assumptions on $g$ immediately pass through to $\tilde{g}$: for all $\theta_1 < \theta_2$, $x_1 < x_2$, and $t$,

$$\tilde{g}(\theta_1|x_1,t)\tilde{g}(\theta_2|x_2,t) < \tilde{g}(\theta_1|x_2,t)\tilde{g}(\theta_2|x_1,t)$$
That is, \( \tilde{g}(\theta|x,t) \) is a log-submodular function of \( \theta \) and \( x \). In particular, for all \( x_1 < x_2 \) and \( t \), there is a threshold \( \bar{\theta} \) such that \( \tilde{g}(\theta|x_1,t) \geq \tilde{g}(\theta|x_2,t) \) if \( \theta \geq \bar{\theta} \). It follows immediately that \( \tilde{G}(\theta|x_1,t) \leq \tilde{G}(\theta|x_2,t) \) for all \( \theta, x_1 < x_2, \) and \( t \). That is, \( \tilde{G}(\theta|x_1,t) \) first order stochastically dominates \( \tilde{G}(\theta|x_2,t) \). The definition of \( H \) then implies \( H(t|x_1) > H(t|x_2) \).

The proof in discrete time is identical, except that the relevant density is \( \tilde{g}(\theta|x,t) \equiv k(x,t) \prod_{\tau=1}^{t-1}(1 - \theta h(\tau))g(\theta|x) \).

Now suppose that we observe two (or more) spells for each individual. An individual’s type \( \theta \) is fixed across the two spells, although in principle the baseline hazard rate may differ across spells. As a result, the duration of the first spell, \( x \), serves as a noisy signal of the duration of the second spell. We claim that Lemma 1 applies to this environment:

**Proposition 1** Consider a large population of individuals with type \( \theta \sim G \). Assume the hazard rate during the \( j \)th spell for a type \( \theta \) individual is \( \theta h_j(t) \) at duration \( t \). Then the hazard rate during the second spell at duration \( t \) is a decreasing function of the realized duration of the first spell \( x \) for all \( x \) and \( t \).

**Proof.** We derive expressions for the type distribution conditional on the realized duration of the first spell. In continuous time,

\[
g(\theta|x) = \frac{\theta h_1(x)e^{-\theta \int_0^x h_1(y)dy}g(\theta)}{\int \theta e^{-\theta \int_0^x h_1(y)dy}g(\theta')d\theta'}.
\]

The expression in discrete time is analogous. It follows that

\[
\frac{g(\theta_1|x_1)g(\theta_2|x_2)}{g(\theta_1|x_2)g(\theta_2|x_1)} = e^{(\theta_1 - \theta_2) \int_{x_1}^{x_2} h_1(y)dy},
\]

which is smaller than 1 if \( \theta_1 < \theta_2 \) and \( x_1 < x_2 \). Thus lemma 1 applies in this case: for all \( t \), \( H(t|x_1) > H(t|x_2) \). Individuals with a shorter first spell have a higher hazard rate at all durations during the second spell.

It is straightforward to think of other applications of Lemma 1. For example, workers with higher types may have a proportionately lower hazard rate of losing their job. If this is the case, individuals who spent longer in their previous job should have a higher hazard rate at all durations. Alternatively, a worker’s type may be reflected in her past earnings, with a similar conclusion.

There is one important case for which Lemma 1 is inapplicable and we cannot establish hazard rate dominance. Assume that the type is the product of two nonnegative numbers, \( \theta = x\varepsilon \), where \( x \) is observable (or more generally a function of observables) and \( \varepsilon \) is an
unobserved random variable with a distribution that is independent of $x$. Then individuals with higher values of $x$ always have a higher hazard rate at short durations; however, dynamic selection of the unobserved type can reverse this at long durations, depending on the type distribution $G$.\footnote{This is the case, for example, if the type distribution is beta distribution parameters ($\alpha, 1/2$) for any $\alpha > 0$, so $g(\theta) \propto \theta^{\alpha - 1}/(1 - \theta)^{1/2}$, which is increasing near the upper bound of $\theta = 1$.} Thus, for example, even if age reduces the hazard rate conditional on $\theta$ and the distribution of $\theta$ is the same for individuals of all ages, older individuals need not have a lower hazard rate at all durations.

### 4.2 Evidence

We test the hazard rate dominance condition in Proposition 1 using our data on price changes and job search. Figure 1 shows the hazard rate of the second price change conditional on the duration of time until the first price change for coffee products. We consider four intervals for the duration of the first spell: 2–7, 8–13, 14–39, and more than 40 weeks. At short durations, the hazard rate in the second spell is clearly higher when the first spell is shorter, but the various hazard rates converge (and the bootstrapped standard errors get bigger) at long durations. In this case, we cannot formally reject the hazard rate dominance condition.

Figure 2 shows conditional hazards for the job finding rate. We again consider four
categories for the duration of the first spell, 0–13, 14–26, 27–52, and more than 53 weeks. In this case, the conditional hazard rate dominance condition is clearly violated. Workers’ whose first spell ends during the second quarter are significantly more likely to find a job during the second quarter of their second spell than workers’ whose first spell lasts only one quarter.

Our intuition is that the proportional hazard condition is violated in a particular way: there is a well-defined group of workers who are likely to find a job in the first quarter and another group of workers who are unlikely to do so, but likely to find a job in the second quarter. That is, hazard rates are not just high and low, as in the proportional hazard model, but instead some workers are quick to find jobs and others take longer.

5 Overidentifying Restrictions using Two Spells

The hazard rate dominance test that we proposed in Proposition 1 exploits one overidentifying restriction of the proportional hazard model with multiple spell data. This section discusses the full set of overidentifying restrictions. We look at both continuous and discrete time versions of the model and assume throughout that each individual has a fixed type $\theta$ across the two spells. Our basic approach follows Honoré (1993), who established that the model is nonparametrically identified.
5.1 Continuous Time

Our basic approach uses 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. The structure of the proportional hazard model implies that this is

\[
\Phi(t_1, t_2) = \int e^{-\theta(z_1(t_1) + z_2(t_2))} dG(\theta),
\]

where \( z_j(t) \equiv \int_0^t h_j(\tau) d\tau \) is the \textit{integrated} baseline hazard during the \( j^{th} \) spell. This formula takes advantage of the fact that the durations of the two spells are independent conditional on the individual characteristic \( \theta \).

Following Honoré (1993), differentiate this expression with respect to its two arguments and take ratios to obtain

\[
\frac{\Phi_1(t_1, t_2)}{\Phi_2(t_1, t_2)} = \frac{h_1(t_1)}{h_2(t_2)}
\]

for all \( t_1 \) and \( t_2 \), where \( \Phi_j \) represents the partial derivative of the survivor function with respect to its \( j^{th} \) argument. If we impose the restriction that the two baseline hazard functions are identical, we can use this to recover the shape of the baseline hazard function up to a constant of proportionality.\footnote{This constant is not identified. We can double each individual’s type and halve the baseline hazard function without changing any outcome. We will typically normalize the baseline hazard rate at zero duration to be equal to the population hazard.} Otherwise, we can take Honoré’s argument one step further to find the shape of the baseline hazard during the first spell. Evaluating equation (2) at \((t_1, t_2)\) and \((t'_1, t_2)\) gives

\[
\frac{h_1(t_1)}{h_1(t'_1)} = \frac{\Phi_1(t_1, t_2)\Phi_2(t'_1, t_2)}{\Phi_2(t_1, t_2)\Phi_1(t'_1, t_2)}
\]

for all \( t_1, t'_1, \) and \( t_2 \). Curiously, the left hand side does not depend on \( t_2 \), while the right hand side depends on \( t_2 \), a testable prediction of the proportional hazard model:

**Proposition 2** For any \( t_1 \) and \( t'_1 \),

\[
\Psi(t_1, t'_1; t_2) \equiv \frac{\Phi_1(t_1, t_2)\Phi_2(t'_1, t_2)}{\Phi_2(t_1, t_2)\Phi_1(t'_1, t_2)}
\]

\( \Psi(t_1, t'_1; t_2) \) does not depend on \( t_2 \).

Proposition 2 suggests a nonparametric test of the model. \( \Psi(t_1, t'_1; t_2) \) can be measured directly in a large dataset for a particular value of \( t_1 \) and \( t'_1 \), and different values of \( t_2 \). The proportional hazard model implies that it should be independent of \( t_2 \). This result is a stronger version of our hazard rate dominance condition. In general, the relative hazard at
durations \( t_1 \) and \( t'_1 \) during the first spell depends on individual’s characteristics and hence are correlated with the duration of the second spell. But this is not the case with the proportional hazard model, since everyone has the same relative hazard at durations \( t_1 \) and \( t'_1 \). We discuss the test, and the appropriate standard errors, in detail in Appendix B.

Assuming this restriction is satisfied, we can use it to recover the baseline hazard rate \( h_1(t_1) \). A similar procedure delivers the other baseline hazard \( h_2(t_2) \). Using these, we can easily recover the integrated baseline hazards \( z_1(t_1) \) and \( z_2(t_2) \). Finally, observe from equation (1) that

\[
\Phi(z_1^{-1}(s), 0) = \int e^{-\theta s} dG(\theta) = \mathcal{L}(s),
\]

the Laplace transformation of the distribution of individual characteristics \( \theta \). If \( z_1(s) \) grows without bound, we can invert the Laplace transformation, to recover the distribution function \( G \). Alternatively, if \( \Phi(z_1^{-1}(s), 0) \) is not a Laplace transformation, we can again reject the proportional hazard model.

5.2 Discrete Time

We obtain similar results in the discrete time version of the model. Now the survivor function is

\[
\Phi(t_1, t_2) = \int \left( \prod_{\tau=1}^{t_1-1} (1 - \theta h_1(\tau)) \right) \left( \prod_{\tau=1}^{t_2-1} (1 - \theta h_2(\tau)) \right) dG(\theta). \tag{5}
\]

If we let \( \Phi_1(t_1, t_2) \equiv \Phi(t_1 + 1, t_2) - \Phi(t_1, t_2) \) and \( \Phi_2(t_1, t_2) \equiv \Phi(t_1, t_2 + 1) - \Phi(t_1, t_2) \), we can immediately obtain versions of equations (2) and (3) in the discrete time as well. Thus Proposition 2 applies, albeit only with an appropriate definition of the slope of the survivor function.

5.3 Evidence

We now return to the data, starting again with price setting. In Figure 3, we plot \( \Psi(t_1, 2; t_2) \), showing a different value of \( t_1 \) in each panel. For example, the top left panel shows \( \Psi(3, 2; t_2) \).

According to the proportional hazard model, this should be constant and equal to the ratio of the baseline hazard rate at 3 weeks to the baseline hazard rate at 2 weeks. The dotted lines show the bootstrapped confidence intervals.\(^8\) The null hypothesis that the data coming from a proportional hazard model is rejected whenever \( \Psi(t_1, 2; t_2) \) lies outside these critical values. Altogether, for \((t_1, t_2) \in [2, 52]^2\), only 11 percent of values \( \Psi(t_1, 2; t_2) \) lie within the

---

\(^8\)We construct the bootstrapped confidence intervals using 100 random samples from an estimated version of the proportional hazard model, where the unobserved heterogeneity is distributed according to gamma distribution with mean 1 and variance .056, and the baseline hazard as shown in Figure 5.
thresholds. Not surprisingly, we reject the null hypothesis of a proportional hazard model, finding $W = 18.7$ while the 5-percent one-sided critical value is $W_\alpha = 3.6$.

Figure 4 shows similar results for the job finding hazard. We now plot $\Psi(t_1, 0; t_2)$, which should be equal to the ratio of baseline hazards $h(t_1)/h(0)$ and in particular be independent of $t_2$. The figure shows a systematic dependence. Each line initially increases and then starts declining at some $t_2 < t_1$. The highest implied relative baseline hazard is in each case at least twice the minimum. The computed statistic lies within the bootstrapped confidence intervals for only 47 percent of the values of $(t_1, t_2) \in \{0, 1, \ldots, 104\}^2$ and the joint test rejects the null hypothesis of a proportional hazard model finding $W = 19,219$ while the 5-percent one-sided critical value is $W_\alpha = 725$.

Of course, we already knew from Section 4 that the proportional hazard model is inconsistent with this data. These results reaffirm those findings in a more systematic fashion, but we believe that both results are driven by the same phenomenon. Many individuals experience two spells of a similar length. While the proportional hazard model admits a
Figure 4: Test statistic $\Psi(t_1, 0; t_2)$ for nonemployment exit for different values of $t_1$ and $t_2$, together with critical values at a 5% confidence level. If the data were generate by a proportional hazard model, the test statistic should lie within the red lines for each value of $t_1, t_2$. 
peak in the hazard rate, it implies that all individuals have a peak at the same duration. As is clear from Figure 2, this is not what we observe in the data.

6 Estimation with Two Spells

Although we can formally reject the proportional hazard model, this is perhaps not too surprising in a large data set. We therefore also estimate the model, hoping that it can tell us something about the shape of the typical hazard rate. We follow several different approaches to estimation, all based on maximum likelihood.

6.1 Parametric Unobserved Heterogeneity

We start by imposing a functional form on the type distribution $G$, either that it has a Gamma or an inverse Gaussian distribution, although we allow for arbitrary flexibility in the baseline hazard function. We estimate this model using maximum likelihood, taking advantage of commands built in to Stata. This is the approach followed by Nakamura and Steinsson (2010). The resulting estimates are consistent if the model, including both the proportional hazard assumption and the functional form restriction on $G$, is correctly specified.

6.2 Horowitz and Lee

Horowitz and Lee (2004) propose a nonparametric estimate of the baseline hazard rate,

$$h(t) = \int_T w(t_2) R(t, t_2) dt_2,$$

where $w(t)$ are weights satisfying $\int_T w(t) dt = 1$, $T$ is the support for $t$ and $R(t, t_2)$ is an estimate of the hazard rate using information in $t_2$. Since the baseline hazard is only identified up to a multiplicative constant, they require one normalization and so impose $\int_T w(t)/h(t) dt = 1$. Basically, their estimator is a weighted average of the baseline hazards estimated using information at different $t_2$. They do not provide guidance for the choice of weights $w(t)$ but show that the estimator is consistent for any choice of weights.

---

9 That paper contains a useful review of the literature on hazard rate estimates in price setting models.
6.3 Quasi-Maximum Likelihood

We propose a nonparametric quasi-maximum likelihood estimator of the proportional hazard model. See Appendix C for the derivation of the likelihood function. We propose that if the baseline hazard is the same during the two spells, a quasi-maximum likelihood estimate of it solves\(^{10}\)

\[
\sum_{t_2=1}^{\infty} \frac{\Phi_1(t_1, t_2)}{h(t_1)} = \sum_{t_2=1}^{\infty} \frac{\Phi_1(t_1, t_2) + \Phi_2(t_1, t_2)}{h(t_1) + h(t_2)} \text{ for all } t_1, t_2.
\]

Note that if equation (2) holds for all \((t_1, t_2)\), this equation holds trivially. Otherwise it suggests that the relative hazard rates are a weighted average of the different partial derivatives of the survivor function.

Our quasi-maximum likelihood estimator differs from the one estimator proposed by Horowitz and Lee (2004) in that the weights are determined by the number of individuals who have a given observed duration. A limitation of both estimators is that they treat the type distribution as a nuisance parameter and so do not recover it. While in principle we could find it using the inverse Laplace transformation, in practice we find that this does not work very well.

6.4 Estimates

Figure 5 shows the raw hazard rate for coffee and compares it to four estimates of the baseline hazard rate. We normalize the baseline hazard rates so that they are equal to the aggregate hazard rate at duration of 2 weeks. Our two parametric estimates are very similar to each other and suggest that unobserved heterogeneity is not an important part of the story. In contrast, the two non-parametric estimates indicate that the hazard rate is nearly flat, increasing slightly for about ten weeks before falling slowly over the course of the next year.

Figure 6 shows similar estimates for the proportional hazard model.\(^{11}\) We now normalize the baseline hazards so that they are equal to the empirical hazard at the start of a spell, \(t = 0\). We again find that the nonparametric estimates indicate a larger role for unobserved heterogeneity than the parametric ones, although in this case the two parametric estimates are significantly different than each other. In particular, while the raw hazard rate peaks after about ten weeks of job search, the nonparametric estimators indicate that the hazard

---

\(^{10}\)Equations (19) and (20) in Appendix C give the formula for the general case when the hazard rate differs across the two spells.

\(^{11}\)We smooth the estimated baseline hazard using an HP filter with parameter 100 to remove the effect of calendar time, such as the beginning and end of the month. We present estimate without smoothing in Appendix D.
Figure 5: Parametric and non-parametric estimate of the baseline hazard rate for price changes, food category coffee. The figure shows the raw hazard rate (purple line), the baseline hazard estimated under assumption that the unobserved heterogeneity is distributed either according to gamma (green line) or inverse Gaussian (black line), the non-parametric Horowitz-Lee estimator (blue line), and our non-parametric quasi-maximum likelihood estimate (red line). The baseline hazard rate is normalized such that it equals aggregate hazard at 2 weeks duration.
rate plateaus until about thirty weeks of job search and then declines more slowly.

Nevertheless, we believe that even the nonparametric estimates understate the role of heterogeneity in shaping the hazard rate of finding a job. The hazard rate dominance and Honoré inspired tests suggest that the proportional hazard model does not adequately describe the data, but this fact alone does not tell us whether the estimates under- or overstate the role of heterogeneity. To answer that question, we develop a structural model which we believe is a good description of the data, the stopping time model in Alvarez, Borovičková, and Shimer (2016).

We view the stopping time model as the true data generating process and use it to construct an object which captures the contribution of structural duration dependence, which we call the structural hazard rate. We explain in detail in Appendix E how it is constructed. We then compare the structural hazard rate to the baseline hazard rate. Such a comparison is sensible because if the structural model had a proportional hazard representation, the structural hazard rate and the baseline hazard rate would coincide. Figure 7 depicts our
Figure 7: Structural hazard and the non-parametric estimate of the baseline hazard rate for non-employment exit rate. The figure shows the aggregate hazard rate (purple line), non-parametric baseline hazard estimate using Horowitz and Lee, and structural hazard estimated from a stopping-time model of non-employment (black line). We also show the aggregate hazard rate implied by the stopping-time model (dashed black line). The baseline hazard rate is normalized such that it equals aggregate hazard at duration of 0 week. The aggregate and baseline hazard rates are HP-filtered with a smoothing parameter 100 to remove seasonal patterns.

preferred structural hazard rate and shows that it is very different than the Horowitz-Lee estimate of the baseline hazard. It peaks later and remains high for longer. We therefore conclude that the MPH model underestimates the role of heterogeneity in determining the shape of the raw hazard rate of finding a job.

7 Covariates

Elbers and Ridder (1982) show that covariates help identify the proportional hazard model. In particular, they show that if the mean of $G$ is finite, and function $\phi$ is non-negative, differentiable and non-constant on an open set in $R^k$, where $k$ is the number of covariates, then the functions $\phi$ and $\bar{h}$ and the distribution $G$ are uniquely determined up to a constant. Heckman and Singer (1984a) give another identifiability theorem, which relaxes the assumption of the finite mean, but instead require a condition on the rate of decay of the tail of $G$. 
We first show how the survivor function can be used to determine \( \phi(x) \) and \( \bar{h}(x) \), and how it can be used to derive a test of the model. Our approach here is close to Elbers and Ridder (1982). In fact, after some algebra it can be shown that our results for \( \phi(x) \) and \( \bar{h}(x) \) are equivalent to equations (10) and (15) in Elbers and Ridder (1982).

This section is written in a continuous time and continuous space, as is also the case in Elbers and Ridder (1982) and Heckman and Singer (1984a). We believe that it is not possible to conduct the same analysis in when time or space are discrete, as it is not possible to eliminate the distribution of \( \theta \).

Assume now that the hazard of an individual \( i \) is given by

\[
h_i(t) = \theta_i \psi(x_i) \bar{h}(t)
\]

where \( x_i \) is an observable characteristic of individual \( i \), as for example age.

Let \( S(t, x) \) be the share of individuals with characteristic \( x \) for whom the spell lasts at least \( t \) periods. Then,

\[
S(t, x) = \int \exp \left( -\theta \psi(x) \int_0^t \bar{h}(s) ds \right) g(\theta) d\theta.
\]

Differentiate with respect to \( t \),

\[
S_t(t, x) = -\psi(x) \bar{h}(t) \int \theta \exp \left( -\theta \psi(x) \int_0^t \bar{h}(s) ds \right) g(\theta) d\theta.
\]

Evaluate this expression at \( t = 0 \), \( S_t(0, x) = -\psi(x) \bar{h}(0) \int \theta g(\theta) d\theta \). As usually, the baseline hazard is identified up to a scale and thus we can normalize \( \bar{h}(0) = \int \theta g(\theta) d\theta = 1 \). Equation (9) then identifies \( \psi(x) \). This result is analogous to equation (10) in Elbers and Ridder (1982).

Differentiate equation (8) with respect to \( x \) to find

\[
S_x(t, x) = -\psi'(x) \int_0^t \bar{h}(s) ds \int \theta \exp \left( -\theta \psi(x) \int_0^t \bar{h}(s) ds \right) g(\theta) d\theta.
\]

Take the ratio of \( S_t(t, x) \) and \( S_x(t, x) \),

\[
\frac{S_t(t, x)}{S_x(t, x)} = \frac{\psi(x) \bar{h}(t)}{\psi'(x) \int_0^t \bar{h}(s) ds}.
\]

Define \( y(t) \equiv \int_0^t \bar{h}(s) ds \), and observe that equation (11) is an ordinary differential equation
for \( y(t) \), with the initial condition \( y(0) = \int_0^t \bar{h}(s) ds = 0 \). Rewriting (11) in terms of \( y(t) \),

\[
y'(t) = y(t) \frac{\psi'(x)}{\psi(x)} \frac{S_t(t, x)}{S_x(t, x)} = k(t, x)y(t), \tag{12}
\]
we find the solution for \( y(t) \),

\[
y(t) = Ct \exp \left( \int_0^t \left( k(\tau, x) - \frac{1}{\tau} \right) d\tau \right), \tag{13}
\]

where \( C \) is a constant to be determined. By taking a logarithmic transformation of the above expression and differentiating, it is straightforward to verify that this is indeed a solution.

The condition \( y(0) = 0 \) does not help to pin down the value of \( C \). However, notice that the normalization \( h(0) = 1 \) implies another condition for \( y(t) \), specifically that \( y'(0) = h(0) = 1 \).

Take the derivative of equation 13 with respect to \( t \),

\[
\bar{h}(t) = y'(t) = Ct k(t, x) = Ct \frac{t \psi(x) S_t(t, x)}{\psi'(x) S_x(t, x)}. \tag{14}
\]

The value at \( t = 0 \) can be found by taking a limit as \( t \to 0 \),

\[
\bar{h}(0) = \lim_{t \to 0} Ct \frac{t \psi(x) S_t(t, x)}{\psi'(x) S_x(t, x)}.
\]

Recall that \( S_t(0, x) = -\phi(x) \), and use L’Hospital rule to find the following limit,

\[
\lim_{t \to 0} \frac{S_x(t, x)}{t} = - \lim_{t \to 0} \frac{\psi'(x) \int_0^t \bar{h}(\tau) d\tau}{t} = -\psi'(x) \lim_{t \to 0} \frac{\bar{h}(0)}{1} = -\psi'(x).
\]

Therefore, it follows that \( C = 1 \). After some algebra, it can be shown that our solution for \( y(t) \) is equivalent to the solution implied by equations (11) and (15) in Elbers and Ridder (1982).

Observe that the right hand side is directly measurable in the data so this gives one way to find the baseline hazard rate \( \bar{h}(t) \).

It is possible to use equation 11 for find an overidentifying test. Take a ratio of equation (11) evaluated at \((t, x)\) and \((t', x)\),

\[
\frac{S_t(t, x) S_x(t', x)}{S_x(t, x) S_t(t', x)} = \frac{\bar{h}(t) \int_0^{t'} \bar{h}(s) ds}{\bar{h}(t') \int_0^{t} \bar{h}(s) ds}. \tag{15}
\]

The right hand side of (15) does not depend on \( x \) while the left-hand side does. The left-hand side is directly measurable in the data and thus for given values of \( t, t' \) one can test whether
the measured ratio is independent of $x$.

In the case of one spell and covariates we consider a data set that for each $x$ has spells with at least $T$ periods. For each $x$ we top code the spells at duration $T$, i.e. for any completed or uncompleted duration $t \geq T$ we record them as of length $T$. Then for each $x$ and $t \leq T$ we record the number of spells as $n(t, x)$. The survivor function $S(x, t)$ is obtained for each $x$ and $t \leq T$ as:

$$S(x, t) = \frac{\sum_{s \geq t} n(x, s)}{\sum_{s} n(x, s)}.$$ 

In our labor market application, we choose covariate $x$ to be workers’ age at the beginning of a spell. We measure age in years, and continue to measure non-employment duration in weeks, starting at 0. We define a statistic

$$\Psi(t, x; t') \equiv \frac{S_t(t, x) S_x(t', x)}{S_x(t, x) S_t(t', x)}.$$ (16)

If the data were generated by a proportional hazard model, then $\Psi(t, x; t')$ would not depend on $x$. We choose $t' = 0$ and plot $\Psi(t, x; 0)$ as a function of $x$ for different values of $t$.

8 Test Results with Covariates

We apply the test proposed in the previous section to our data. We choose age at the beginning of the spell as our covariate $x$.

8.1 Price Changes

To construct the dataset, we use a different sampling procedure than in case of two spells. We aim to measure the baseline hazard through some pre-specified duration $T$ and so we truncate measurement at $T$. If a product has multiple spells, we choose one spell at random. We consider separately sales and regular price changes, and measure spells only longer than 2 weeks. We define the age of a product as the number of periods a product has been in the sample. We consider only products for which we can determine their age. This means, that we consider only products which do not have any recorded transaction for the first year since the beginning of the dataset. Here we present results for the good category coffee, where measure hazard rate up to $T = 104$ weeks. The age of the product is measured in weeks, and varies between 1 and more than 500 weeks.

Figure 8 shows estimated $\psi(x)$. We see that there is some variation between 1 and 10 weeks of age, but that $\psi(x)$ is very flat after 10 weeks of age.
Figure 8: Estimated function $\psi(x)$ for price change hazard for coffee products, with product’s age at the beginning of the spell as a covariate $x$.

Figure 9 shows the test statistic together with bootstrapped standard errors. The test statistic appears flat, not depending on the age, which suggests that we will not reject the proportional hazard model. The standard errors are wide, which we believe is a reflection of $\psi(x)$ being flat and not containing enough information on the baseline hazard.

### 8.2 Labor Market Outcomes

To construct the dataset, we use a different sampling procedure than in case of two spells. As before, we define a non-employment spell as time between two full-time jobs, and impose that a worker has to be officially registered as unemployed for at least one day during the non-employment period. We aim to measure the baseline hazard through some pre-specified duration $T$ and so we truncate measurement at $T$. We do not impose any age restrictions on the sample. Many workers experience more than one non-employment spell and we randomly select one of them.

Figure 10 shows the estimate of $\psi(x)$, estimated as $\psi(x) = -S_t(0, x)$. The first thing to notice is that there is not much variability in $\psi(x)$. This is problematic because identification relies on changes in $\psi(x)$.

We present test results in Figure 11. We plot the value of $\Psi(t, x; t')$ defined in (16) as a function of $x$, for different values of $t$. We choose $t' = 0$ weeks. If the data were generated
Figure 9: Nonparametric test of the proportional hazard model with covariates for coffee products. The figure shows the test ratio at 3, 6, 25, and 36 weeks, compared to 2 weeks duration at different ages. According to the proportional hazard model, each line should be independent of age. Dashed red lines show bootstrapped standard errors.
Figure 10: Estimated function $\psi(x)$ for the nonemployment exit hazard, with worker’s age at the beginning of the spell as a covariate $x$.

from a proportional hazard model, each of the depicted lines would be constant with respect to age. We see that $\Psi(t, x; t')$ is noisy but does not vary much with $x$. The standard errors are large, so it would be difficult to reject the MPH model. However, the test reveals that there is not much information about the baseline hazard $\bar{h}(t)$. Recall that the test statistic $\Psi(t, x; t')$ has an interpretation of the relative hazard rates at two different durations $t$ and $t'$. Since it varies a lot, it is not informative about the shape of $\bar{h}(t)$.

9 Conclusion

The proportional hazard model has been a leading model in duration analysis, especially for separating out the role of unobserved heterogeneity from structural duration dependence. To estimate the model, most authors make parametric assumptions on the distribution of unobserved heterogeneity. However, these assumptions are not innocuous. As argued by Heckman and Singer (1984a), the choice of a particular distribution can dramatically affect estimates of the baseline hazard, yet economic theory usually does not offer any guidance for which parametric assumptions to impose.

In this paper we show how to test whether data admit a proportional hazard representation without imposing additional assumption on the baseline hazard or the distribution
Figure 11: Nonparametric test of the proportional hazard model with covariates. The figure shows the test ratio at 13, 26, 39, and 52 weeks, compared to 0 weeks duration at different ages. According to the proportional hazard model, each line should be independent of age. Dashed red lines show bootstrapped standard errors.
of unobserved heterogeneity. We also show how to estimate the baseline hazard using quasi maximum likelihood, treating the distribution of unobserved heterogeneity as a nuisance parameter. We consider two different cases: one in which we observe two spell per individual as in Honoré (1993), and one in which we observe one spell and a covariate for each individual, as in Elbers and Ridder (1982). We apply these tests to price change and non-employment duration data, and in both cases we reject the MPH specification. We use a structural model to illustrate that estimating a proportional hazard model on data which do not admit this representation may lead to an underestimate of the role of heterogeneity.
References


To estimate the proportional hazard model, we need data on the survivor function $\Phi(t_1, t_2)$ for a large variety of “individuals.” An issue that immediately arises is that in any real-world data set, we do not observe two completed spells for all individuals. For example, in a labor market context, there are individuals who never work and others who only stop working when they hit retirement.

Our methodology recognizes that if the proportional hazard model is correctly specified, then we can estimate it using any subset of observations in the data. For example, in the context of price changes, we can estimate it only using retailers who stock a good for at least a pre-specified amount of time. While this may bias any estimates of the distribution of characteristics $G(\theta)$, it should not affect estimates of the baseline hazard $h(t)$. Our methodology also takes advantage of the symmetry of our setup. This is important because we may not observe the second spell for individuals whose first spell is very long, but we can observe the first spell of individuals whose second spell is very long.

Our goal is to measure the baseline hazard through to some pre-specified duration $T$. For the case of price changes, our sampling frame is the set of all products that are in our data set for at least $2T-1$ periods after the initial price change. In practice, we set $T = 104$ weeks, which is not restrictive because average duration of a price spell is around 12 weeks. For each product, we let $t_1$ equal the duration of the first spell, top-coded at $T$. For each product with $t_1 < T$, we let $t_2$ equal the duration of the second spell, again top-coded at $T$. This is feasible because we have at least $2T-1$ observations and because we do not look at products whose first spell is top-coded. Denote the the number of products with durations $(t_1, t_2)$ by $n(t_1, t_2)$ for all $t_1 < T$ and $t_2 \leq T$. Let $n(T, \cdot)$ denote the number of products whose first spell has duration at least $T$.

For $t_1 < T$ and $t_2 < T$, our measure of the number of spells is simply $N(t_1, t_2) = (n(t_1, t_2) + n(t_2, t_1))/2$, where we take advantage of the symmetry of our model to effectively enlarge the data set. For $t < T$, we also let $N(t, T) = N(T, t) = n(t, T)$, again using symmetry, but now to impute values for individuals whose first spell is top-coded and hence second spell may be truncated. Finally, our measure of $N(T, T)$ is $n(T, \cdot) - \sum_{t<T} n(t, T)$, i.e. the remaining spells.

Once we have recovered $N(t_1, t_2)$ for all $(t_1, t_2) \in 1, 2, ..., T^2$, we can define the survivor function as

$$
\Phi(t_1, t_2) = \frac{\sum_{\tau_1 \geq t_1, \tau_2 \geq t_2} N(\tau_1, \tau_2)}{\sum_{\tau_1 \geq 1, \tau_2 \geq 1} N(\tau_1, \tau_2)},
$$

the fraction of individuals with spells lasting at least $(t_1, t_2)$ periods. This is an unbiased
estimate of the survivor function for the original set of products and hence we can use it to estimate the baseline hazard, recover the distribution of characteristics, and test the model.

Our sampling framework for the labor market application is slightly different. We still measure two consecutive spells, top coding each spell at duration $T = 260$ weeks. But spells are no longer consecutive, since a worker spends some time employed between unemployment spells and vice versa.\textsuperscript{12} We therefore restrict attention to individuals whom we observe for at least $4T$ periods, still top-coding both spells at $T$ and inferring the duration of the second spell for individuals whose first spell is top-coded from the individuals whose second spell is top-coded.

\section*{B Testing the Proportional Hazard Model}

In any real-world data set generated from a proportional hazard model, we would not expect $\Psi(t_1, t'_{1}; t_2)$ to be exactly independent of $t_2$ due to sample variability. We use bootstrapping to derive critical values for a static $\tilde{\Psi}(t_1, t'_{1}; t_2)$. The null hypothesis is that the data are generated by a proportional hazard model against the hypothesis that the data come from a model that does not admit a proportional hazard representation. In an infinite sample under the null hypothesis, $\Psi(t_1, t'_{1}; t_2)$ should be the same for all values of $t_2$. In a finite sample, these values can differ even under the null. We use bootstrapping to find thresholds $\underline{c}_\psi(t_1, t'_{1}; t_2), \bar{c}_\psi(t_1, t'_{1}; t_2)$ such that in a finite sample, if the null hypothesis is true, then

$$\text{Prob} \left[ \Psi(t_1, t'_{1}; t_2) \notin [\underline{c}_\psi(t_1, t'_{1}; t_2), \bar{c}_\psi(t_1, t'_{1}; t_2)] \right] = \alpha,$$

where a typical choice of $\alpha$ is 0.05. If the value of $\Psi(t_1, t'_{1}; t_2)$ measured in the data falls outside this interval, we reject the null hypothesis.

The above strategy tests each $(t_1, t'_{1}, t_2)$ separately, and it is very likely that in any real-world data, we reject null for some values of $(t_1, t'_{1}; t_2)$ but cannot reject for other values. To test a joint hypothesis that for any two values of $t_1$, call them $t^1_1$ and $t^2_1$, it holds that $\Psi(t^1_1, t'_{1}; t_2)) = \Psi(t^2_1, t'_{1}; t_2))$, we propose a test which is similar to a Wald test. We compute $W$ as

$$W = \sum_{t_1} \sum_{t_2} \hat{\Psi}(t_1, t'_{1}; t_2)^2. \quad (18)$$

Here $\hat{\Psi}(t_1, t'_{1}; t_2) \equiv \Psi(t_1, t'_{1}; t_2) - \overline{\Psi}(t_1, t'_{1}; \cdot)$ is a normalized test statistic, where $\overline{\Psi}(t_1, t'_{1}; \cdot)$ is the mean value of $\Psi(t_1, t'_{1}; \cdot)$ across $t_2$, for given values $t_1, t'_1$. In a standard Wald test, if each

\textsuperscript{12}A similar consideration applies if we consider regular price changes, rather than all price changes, since sales can occur in between regular price changes.
of \( \hat{\Psi}(t_1, t'_1; t_2) \) were independent and distributed according to a standard normal distribution, then \( W \) has a chi-squared distribution. Since we do not have results on the distribution of \( \hat{\Psi}(t_1, t'_1; t_2)^2 \), we again use bootstrapping to find a critical value at the significance level \( \alpha \), call it \( \bar{W}_\alpha \), so that we reject the null if \( W > \bar{W}_\alpha \).

The key step in bootstrap hypothesis testing is to sample under the null hypothesis. In our case it means that we need to sample data from a proportional hazard model, but the sampled data should nevertheless be a close description of the data at hand. We therefore proceed as follows. We estimate a proportional hazard model on our data, imposing a parametric distribution of unobserved heterogeneity; see Section 6.1. This procedure recovers parameters of the distribution of \( \theta \) and the baseline hazard rate \( h(t) \), which fully describes the model. We then create a large number of synthetic datasets of \( N \) products/individuals from this model, where \( N \) is the number of products/individuals in the data, and compute \( \hat{\Psi}(t_1, t'_1; t_2) \) for all several values of \( t_1, t'_1, t_2 \) for each synthetic dataset. We find \( c_\psi(t_1, t'_1; t_2) \) and \( \bar{c}_\psi(t_1, t'_1; t_2) \) by ordering \( \hat{\Psi}(t_1, t'_1; t_2) \) across samples and taking the value at \( \alpha/2 \) and \( 1 - \alpha/2 \) position in the ordered sample. In each synthetic dataset, we compute statistic \( W \) and choose \( \bar{W}_\alpha \) such that \( \alpha \) percent of the \( W \) values lie below \( \bar{W}_\alpha \). If the value of \( W \) measured in the real-world data lies above \( \bar{W}_\alpha \), we reject the null hypothesis.

\section*{C Non-parametric Estimation with Two Spells}

We propose a non-parametric estimator of the baseline hazard for a discrete time formulation of the model and a countable number of types, using data on two spells. We consider here a more general version of the model where we allow for different baseline hazard rates in the first and the second spell, denoting hazard in the \( j \)th spell as \( h_j(\cdot) \).

We index the types of individuals by \( i = 1, \ldots, I \), where \( I = \infty \) implies infinitely many types, and let \( n_i \) be the number of individuals of type \( i \). An individual of type \( i \) with duration \( t \) in his \( j \)th spell exits the state (finds a job) with probability \( \theta_i h_j(t) \) and fails to exit the state with complementary probability. These outcomes are independent over time. We let \( S(t_1, t_2) \) denote the number of individuals whose first spell lasts at least \( t_1 \) periods and whose second spell lasts at least \( t_2 \) periods.

Consider an individual of type \( i \). The probability that his \( j \)th spell lasts for at least \( t \) periods is

\[
\prod_{\tau=1}^{t-1} (1 - \theta_i h_j(\tau)).
\]

Therefore the expected number of type \( i \) individuals whose \( j \)th spell lasts for at least \( t \) periods
is
\[ \mu_i^j(t) \equiv n_i \prod_{\tau=1}^{t-1} (1 - \theta_i h_j(\tau)). \]

The actual number is a binomial random variable with parameters \( n_i \) and \( \prod_{\tau=1}^{t-1} (1 - \theta_i h_j(\tau)) \).

We approximate the binomial with a Poisson distribution with mean \( \mu_i^j(t) \). That is, the probability that \( k \) individuals of type \( i \) experience their \( j^{th} \) spell of at least length \( t \) is approximately

\[ \frac{e^{-\mu_i^j(t)} \mu_i^j(t)^k}{k!} \]

for all nonnegative integer \( k \).

A similar logic implies that the expected number of type \( i \) individuals whose first spell lasts at least \( t_1 \) periods and second spell lasts at least \( t_2 \) periods is

\[ \mu_i(t_1, t_2) \equiv n_i \left( \prod_{\tau=1}^{t_1-1} (1 - \theta_i h_1(\tau)) \right) \left( \prod_{\tau=1}^{t_2-1} (1 - \theta_i h_2(\tau)) \right). \]

The actual number is a Poisson random variable with mean \( \mu_i(t_1, t_2) \).

We now use the property that the sum of two Poisson random variables with means \( \mu_1 \) and \( \mu_2 \) is a Poisson random variable with mean \( \mu_1 + \mu_2 \). The expected number of individuals of any type whose first spell lasts at least \( t_1 \) periods and second spell lasts at least \( t_2 \) periods is a Poisson random variable with mean

\[ \mu(t_1, t_2) \equiv \sum_i n_i \left( \prod_{\tau=1}^{t_1-1} (1 - \theta_i h_1(\tau)) \right) \left( \prod_{\tau=1}^{t_2-1} (1 - \theta_i h_2(\tau)) \right). \]

The probability of observing \( S(t_1, t_2) \) conditional on the parameters of the model is

\[ \frac{e^{-\mu(t_1, t_2)} \mu(t_1, t_2)^{S(t_1, t_2)}}{S(t_1, t_2)!}, \]

and therefore the log-likelihood, up to a constant, is given by

\[ \sum_{t_1=1}^{\infty} \sum_{t_2=1}^{\infty} \left( S(t_1, t_2) \log \mu(t_1, t_2) - \mu(t_1, t_2) \right). \]

We do not attempt to maximize this directly because it involves maximizing over \( I \), \( n_i \), and \( \theta_i \), as well as the parameters of interest \( h_1(t) \) and \( h_2(t) \).

Instead, define \( S_1(t_1, t_2) \equiv S(t_1, t_2) - S(t_1 + 1, t_2) \) is the number of individuals whose first spell lasts exactly \( t_1 \) periods and their second spell lasts at least \( t_2 \) periods. This is a Poisson
random variable with mean

$$\mu(t_1, t_2) - \mu(t_1 + 1, t_2) = h_1(t_1) \sum_i \theta_i n_i \left( \prod_{\tau=1}^{t_1-1} (1 - \theta_i h_1(\tau)) \right) \left( \prod_{\tau=1}^{t_2-1} (1 - \theta_i h_2(\tau)) \right)$$

$$= h_1(t_1) \mu'(t_1, t_2).$$

Similarly, $S_2(t_1, t_2) \equiv S(t_1, t_2) - S(t_1, t_2 + 1)$ is a Poisson random variable with mean

$$\mu(t_1, t_2) - \mu(t_1, t_2 + 1) = h_2(t_2) \sum_i \theta_i n_i \left( \prod_{\tau=1}^{t_1-1} (1 - \theta_i h_1(\tau)) \right) \left( \prod_{\tau=1}^{t_2-1} (1 - \theta_i h_2(\tau)) \right)$$

$$= h_2(t_2) \mu'(t_1, t_2).$$

In these terms, the likelihood of observing $S_1(t_1, t_2)$ and $S_2(t_1, t_2)$ is

$$\frac{e^{-h_1(t_1) \mu'(t_1, t_2)} (h_1(t_1) \mu'(t_1, t_2))^{S_1(t_1, t_2)} e^{-h_2(t_2) \mu'(t_1, t_2)} (h_2(t_2) \mu'(t_1, t_2))^{S_2(t_1, t_2)}}{S_1(t_1, t_2)! S_2(t_1, t_2)!}$$

and so the log-likelihood is (up to a constant) given by

$$\sum_{t_1=1}^{\infty} \sum_{t_2=1}^{\infty} \left( S_1(t_1, t_2) \log h_1(t_1) + S_2(t_1, t_2) \log h_2(t_2) + (S_1(t_1, t_2) + S_2(t_1, t_2)) \log \mu'(t_1, t_2) - (h_1(t_1) + h_2(t_2)) \mu'(t_1, t_2) \right)$$

We treat $\mu'(t_1, t_2)$ as an unrestricted function, defined on $\mathbb{R}^2$. Since we may have countably many types, this seems to be unrestrictive. We maximize the likelihood by choosing $h_1(t_1)$, $h_2(t_2)$, and $\mu'(t_1, t_2)$. We do this in two steps. Conditional on any value of $h_1(t_1) + h_2(t_2)$, $\mu'(t_1, t_2)$ must equal $(S_1(t_1, t_2) + S_2(t_1, t_2))/(h_1(t_1) + p^2(t_2))$. Therefore the log-likelihood is (up to a constant)

$$\sum_{t_1=1}^{\infty} \sum_{t_2=1}^{\infty} \left( S_1(t_1, t_2) \log h_1(t_1) + S_2(t_1, t_2) \log h_2(t_2) - (S_1(t_1, t_2) + S_2(t_1, t_2)) \log (h_1(t_1) + h_2(t_2)) \right).$$

We choose $h_1(t_1)$ and $h_2(t_2)$ to maximize this. Note that multiplying all the $h_j(t)$’s by a constant does not change any term in the sum, the standard result from Honoré (1993).
The first order conditions for this problem are

\[
\sum_{t_2=1}^{\infty} \frac{S_1(t_1, t_2)}{h_1(t_1)} = \sum_{t_2=1}^{\infty} \frac{S_1(t_1, t_2) + S_2(t_1, t_2)}{h_1(t_1) + h_2(t_2)} \text{ for all } t_1, \quad (19)
\]

\[
\sum_{t_1=1}^{\infty} \frac{S_2(t_1, t_2)}{h_2(t_2)} = \sum_{t_1=1}^{\infty} \frac{S_1(t_1, t_2) + S_2(t_1, t_2)}{h_1(t_1) + h_2(t_2)} \text{ for all } t_2. \quad (20)
\]

We implement this formulas via iterations. We start with an initial guess for \(h_1(t), h_2(t)\). In the \(k^{th}\) iteration, we use \(h_{1,k}(t), h_{2,k}(t)\) in the right-hand side of the equation (19) to get new values of the baseline hazard rate in the first period, \(h_{1,k+1}(t)\). Similarly, we use equation (20) to find \(h_{2,k+1}(t)\). We continue until the estimate of the baseline hazard rates do not differ much across iterations.

\section*{D Estimate of the Baseline Hazard Rate}

Figure 12 shows the baseline hazard rate for non-employment exit rate, estimated parametrically and non-parametrically, without smoothing.

\section*{E General Decomposition of the Hazard Rate}

The proportional hazard model offers a natural decomposition of the raw hazard rate into the baseline hazard and the portion attributable to heterogeneity. This section shows how to perform a more general decomposition of the raw hazard rate into two portions, the average change in individuals’ hazard rates (the structural hazard rate) and the remaining portion attributable to changes in the composition of individuals in the population.

Consider a population composed of many individuals characterized by a characteristic \(\theta\). Let \(h(t; \theta)\) denote the hazard rate of type \(\theta\) and duration \(t\). In the proportional hazard model, this can be expressed as the product of \(\theta\) and the baseline hazard, but we relax that restriction here. The probability that a spell lasts at least \(t\) periods is

\[F(t; \theta) = 1 - e^{-\int_0^t h(\tau; \theta) d\tau}\]

for all \(t\) and \(\theta\). This implies the density of \(\theta\) in the population surviving to duration \(t\) is

\[g(\theta; t) = \frac{e^{-\int_0^t h(\tau; \theta) d\tau} g(\theta; 0)}{\int e^{-\int_0^t h(\tau; \theta') d\tau} g(\theta'; 0) d\theta'},\]
Figure 12: Non-parametric and parametric estimate of the baseline hazard rate for non-employment exit rate. The figure shows the aggregate hazard rate (purple line), non-parametrically estimated baseline hazard using our method (red line) or Horowitz,Lee method (blue line), the baseline hazard estimated under assumption that the unobserved heterogeneity is distributed either according to gamma (black line) or inverse Gaussian (green line). The baseline hazard rate is normalized such that it equals aggregate hazard at duration of 0 week.
where \( g(\theta; 0) \) is the initial population density of \( \theta \). This model then gives rise to a raw hazard rate

\[
H(t) = \int h(t; \theta) g(\theta; t) d\theta. \tag{22}
\]

We are interested in decomposing the evolution of \( H \) into two terms, the portion due to the evolution of the average value of \( h \) and the portion due to the evolution of \( g \). We propose a multiplicative decomposition here, \( H(t) = H_{\text{str}}(t)H_{\text{het}}(t) \), or equivalently \( \log H(t) = \log H_{\text{str}}(t) + \log H_{\text{het}}(t) \).

We start by differentiating equation (22):

\[
\dot{H}(t) = \int \dot{h}(t; \theta) g(\theta; t) d\theta + \int h(t; \theta) \dot{g}(\theta; t) d\theta.
\]

The first term (if negative) is the decrease in the raw hazard rate coming from the fact that the average individual has an decreasing hazard rate. The second term (if negative) is the decrease in the hazard rate coming from the fact that individuals with a high hazard rate become a small share of the population over time. To perform a multiplicative decomposition, we divide through by both sides by the hazard rate and write

\[
\frac{\dot{H}(t)}{H(t)} = \frac{\dot{H}_{\text{str}}(t)}{H_{\text{str}}(t)} + \frac{\dot{H}_{\text{het}}(t)}{H_{\text{het}}(t)},
\]

where

\[
\frac{\dot{H}_{\text{str}}(t)}{H_{\text{str}}(t)} = \frac{\int \dot{h}(t; \theta) g(\theta; t) d\theta}{H(t)} \quad \text{and} \quad \frac{\dot{H}_{\text{het}}(t)}{H_{\text{het}}(t)} = \frac{\int h(t; \theta) \dot{g}(\theta; t) d\theta}{H(t)}.
\]

Finally, we define

\[
\log H_{\text{str}}(t) = \int_{t_0}^t \frac{\dot{H}_{\text{str}}(s)}{H_{\text{str}}(s)} ds + \log H(t_0) \quad \text{and} \quad \log H_{\text{het}}(t) = \int_{t_0}^t \frac{\dot{H}_{\text{het}}(s)}{H_{\text{str}}(s)} ds \tag{23}
\]

for some carefully chosen value of \( t_0 \), e.g. \( t_0 = 0 \).

In the proportional hazard model, \( h(t; \theta) = \theta h(t) \), so

\[
\frac{\dot{H}_{\text{str}}(t)}{H_{\text{str}}(t)} = \frac{\dot{h}(t)}{h(t)}.
\]

Thus this decomposition recovers the baseline hazard rate, \( H_{\text{str}}(t) = h(t) \). More generally,

\footnote{With a general formulation of the hazard rate \( h \), there is no loss of generality in assuming that the initial distribution admits a density function.}

\footnote{See Alvarez, Borovičková, and Shimer (2016) for an additive decomposition. We use a multiplicative decomposition here because it has the same structure as the proportional hazard model.}
however, the growth in the structural portion of the hazard rate represents the increase in the average hazard rate relative to the average level of the hazard rate. In a structural model, we can compute the contribution of structural duration dependence $H^{str}(t)$ and compare it to what one would get by treating the data as if it comes from a proportional hazard model.