# Decomposing Duration Dependence in a Stopping Time Model*

Fernando Alvarez
University of Chicago

Katarína Borovičková
Federal Reserve Bank of Richmond

Robert Shimer
University of Chicago

November 17, 2023

**Abstract**

We develop an economic model of transitions in and out of employment. Heterogeneous workers switch employment status when the net benefit from working, a Brownian motion with drift, hits optimally-chosen barriers. This implies that the duration of jobless spells for each worker has an inverse Gaussian distribution. We allow for arbitrary heterogeneity across workers and prove that the distribution of inverse Gaussian distributions is partially identified from the duration of two non-employment spells for each worker. We estimate the model using Austrian social security data and find that dynamic selection is a critical source of duration dependence.

**Keywords:** Unemployment, Duration Models, Job Finding Rate, Switching Costs, Unobserved Heterogeneity, Inverse Gaussian Distribution, Set Identification

*We are grateful for comments from Jaap Abbring, Stéphane Bonhomme, Jan Eberly, Bo Honoré, Rasmus Lentz, Andreas Mueller, Pedro Portugal, Andrea Rotnitzky, Quang Vuong, Josef Zweimüller, as well as the editor Aureo de Paula, four anonymous referees, and numerous participants in numerous seminars. This material is based in part on work supported by the National Science Foundation under grant numbers SES-1559225 and SES-1559459. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

# 1 Introduction

The hazard 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. This reflects a combination of two factors: structural duration dependence in the job finding rate for each individual worker, and changes in the composition of workers at different non-employment durations due to dynamic sorting (Lancaster, 1979). The goal of this paper is to combine an economically-grounded and flexible model of the job finding rate for an individual worker with arbitrary heterogeneity across workers, in order to understand the importance of these two factors.

Our economic model views the duration of non-employment spells as the outcome of a comparison between the time-varying costs and benefits of working at each instant to the cost of switching between employment and non-employment. In one interpretation of our model, each worker has two options at each instant of time: working at some time-varying wage or not working and receiving some time-varying income and utility from leisure. Due to a cost of switching employment status, a worker starts working when the difference between the wage and non-employment income is sufficiently large and stops working when the difference is sufficiently small. Formally, this gives us an optimal stopping problem for each worker. An alternative interpretation of the model has a firm with time-varying productivity which unilaterally decides when to employ a worker who receives a fixed wage.

Under either interpretation, we allow for arbitrary cross-worker heterogeneity in all the fundamental parameters describing these stopping problems. For example, some workers may expect their labor market productivity to increase the longer they stay out of work while others may expect it to fall. Some workers may have high switching costs, while others have low ones. We maintain two key restrictions: for each worker, the evolution of a latent variable, the net benefit from employment, follows a geometric Brownian motion; and the decision to work is made optimally. We allow the parameters of the Brownian motion to depend on the worker's employment status. These assumptions imply that the duration of a non-employment spell is given by the first passage time of a Brownian motion with drift, a random variable with an inverse Gaussian distribution. The inverse Gaussian distribution has two reduced-form parameters, which themselves are known functions of the fundamental parameters in the stopping time problem. The fundamental parameters are fixed over time for each worker, but we allow them to vary arbitrary across workers, so the reduced-form parameters of the inverse Gaussian distribution have some unknown distribution $G$.

Given this environment, we present three main theoretical results. First, we discuss which aspects of the distribution $G$ are identified using duration data. We start by noticing that the distribution of the duration of one non-employment spell does not identify $G$. We then proceed to develop a novel strategy for inverting the distribution of the duration of two completed spells in order to recover $G$. Our main result is Theorem 1, which states that the joint distribution of the duration of two completed non-employment spells identifies $G$ if we know the sign of the drift in the underlying Brownian motion. Even though this is an infinite dimensional problem, our proof does not require any other conditions on the underlying parameters.

The lack of identification of the sign of the Brownian motion's drift is closely related to the fact that for some reduced-form parameters of the inverse Gaussian distribution, a positive fraction of non-employment spells last forever, i.e. the non-employment duration distribution is defective; see Proposition 4 characterizing the identified set of distributions. When the non-employment duration distribution may be defective, we show that information on the fraction of first and second non-employment spells which are completed provides restrictions on the set of identified distributions. Finally, we consider the consequences of defective employment duration distributions and of censoring of all durations. These further enlarge the set of distributions that are consistent with the data. In particular, we define the set $\mathbb{G}_{\chi^c}$ of distributions that can be identified taking into account both defective duration distributions and censoring.

Second, we prove two results for mixture models, which we use to evaluate our model. The first is an exact statistical decomposition of the mixture hazard. The mixture hazard can be expressed as the product of "the structural hazard," whose change is equal to the cross sectional average change in the hazard among surviving types, and "the heterogeneity effect," whose change is equal to minus the cross sectional variance of the hazards among the surviving types. The heterogeneity effect, which accounts for the effect of dynamic selection on the level of the mixture hazard, is decreasing with duration. This result generalizes the decomposition in Lancaster (1992, Chapter 4) from the mixed proportional hazard (MPH) model to an arbitrary mixture model. The second result states that in the MPH model, the mixture hazard of the second spell at duration $t_2$ for those with first spell duration $t_1$, is decreasing in $t_1$ (Proposition 5). In contrast, our model does not imply this monotonicity property. This prediction can easily be tested in the data without estimating a full model.

Finally, we show that we can use duration data as well as information about wage dynamics to infer the size of the fixed cost of switching employment status. Dixit (1991) and Abel and Eberly (1994) show that even small fixed costs give rise to a large region of inaction (i.e. an infinite derivative of the range of inaction with respect to the cost evaluated at zero cost), which in turn affects the duration of non-employment spells. We also show how to invert this relationship to recover the fixed costs.

We illustrate our theoretical results by estimating the model using data from the Austrian social security registry from 1986 to 2007 on more than one million workers who experience at least two non-employment spells, as well as a number of other workers who experience only one spell. We postulate a distribution $G$ with a finite number of types, and estimate it using a small modification of the standard EM algorithm. We find that a few types account for most of the workers: 3 types account for two thirds, and ten types account for 99 percent. The estimated type distribution contrasts starkly with the MPH model, where all these hazards would be proportional to each other. While the model fit is good, we also highlight some limitations. For example, in the data many jobs start on the first day of the month and end on the last day in the data, while there is no scope for this in the model.

While our model is only set identified, in practice we find that the quantitative answers to the questions we posed above are almost the same for any distribution that is in the identified set. For example, our estimates uncover evidence of substantial heterogeneity across workers, and hence of dynamic sorting, which we quantify using our statistical decomposition of the mixture hazard. We find that while the mixture hazard is hump-shaped with a peak at around 10 weeks, the structural hazard increases until about 20 weeks and then declines by much less. The heterogeneity effect is substantial: it reduces the mixture hazard during the first half year of non-employment by about 65 percent, and by over 90 percent during the first two years. Although observed worker characteristics account for some of these compositional changes, the bulk are due to characteristics that are unobserved, at least in our data set.

We compute the mixture hazard for the second spell conditional on the duration of the first. We find clear evidence against the MPH model's predicted monotonicity. Instead, we find that our estimated parameters reproduce the empirical non-monotonic relationship quite well.

We also estimate small fixed costs. For the median worker, the total cost of taking a job and later leaving it are approximately equal to five to ten minutes of leisure time. As a result, the median newly employed worker leaves her job if she experiences a 1.3 percent drop in the wage. We find that not only are the fixed costs small, but so is the region of inaction. On the other hand, we also show that a model without fixed costs could not possibly fit our data.

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

Whitmore (1979) assumes that the duration of an employment spell is given by the first passage time of a Brownian motion with drift. Shimer (2008) makes a similar assumption about the duration of an unemployment spell. Neither paper allows for heterogeneity across workers. Buhai and Teulings (2014) propose an economic model of the duration of job spells characterized by the first passage time of a Brownian and estimate its parameters, allowing for a parametric distribution of observed and unobserved heterogeneity. 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). The bio-statistics literature has so far not introduced unobserved individual heterogeneity into the model, nor has it considered the identification of such a model.

De Paula (2009) studies a game with multiple players where each player's decision to leave a certain state is given by the first passage time of a Brownian motion with drift. The model captures interactions between players by allowing the drift to depend on how many other players have already left the state. De Paula (2009) shows how to use data on exit dates to test for the presence of endogenous interactions and studies identification of the model. He shows that the full parameter vector is not identified but parametric identification can be achieved if parameters are allowed to be functions of observables and there is "enough variability" in covariates. He applies it to data on desertion in the Union Army during the American Civil War. Our approach to identification is different. Like in De Paula (2009), the full set of fundamental parameters is not identified from data we use, but we argue that for duration analysis we only need to know two reduced form parameters for each worker type, and prove that the distribution of those is set identified.

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

Finally, some recent papers analyze duration dependence using models that are identified through assumptions on the extent of unobserved heterogeneity. For example, Krueger, Cramer and Cho (2014) argue that observed heterogeneity is not important in accounting for duration dependence and so conclude that unobserved heterogeneity must also be unimportant. Schmieder, von Wachter and Bender (2016) reach a similar conclusion using German data. Similarly, DellaVigna, Lindner, Reizer and Schmieder (2017) argue that selection on observable characteristics is muted, and hence selection on unobservables likely follows a similar pattern. Hornstein (2012) and Ahn and Hamilton (2020) both assume there are two types of workers with different job finding hazards at all durations and estimate the unobserved types using single-spell data.

The remainder of the paper proceeds as follows. In Section 2, we describe our economic model. Section 3 shows that our model generates an inverse Gaussian distribution of duration for each worker and contains our main theoretical results on duration analysis. This section contains our main results on identification: the case of two completed spells, the case with defective duration distributions, and the case with censored spells. In Section 4, we propose a multiplicative decomposition of the mixture hazard into the portion attributable to structural duration dependence and the proportion attributable to heterogeneity. Section 5 summarizes the Austrian social security registry data. Section 6 presents our empirical results, including estimates of the model, the decomposition of hazards, a comparison to the MPH model, and inference of the distribution of fixed costs. Finally, Section 7 briefly concludes.

# 2 Theory

## 2.1 Economic Model of an Individual Worker

We consider the problem of a risk-neutral, infinitely-lived worker with discount rate $r$ who can either be employed, $s(t) = e$, or non-employed, $s(t) = n$, at each instant in continuous time $t$. The worker earns a wage $e^{w(t)}$ when employed and gets flow utility $e^{b(t)}$ when non-employed. Both $w(t)$ and $b(t)$ follow correlated Brownian motions with drift. The drift and standard deviation of each, as well as the correlation between them, may depend on the worker's employment status:

$$db(t) = \mu_{b,s(t)}\, dt + \sigma_{b,s(t)}\, dB_b(t) \text{ and } dw(t) = \mu_{w,s(t)}\, dt + \sigma_{w,s(t)}\, dB_w(t).$$

$B_b(t)$ and $B_w(t)$ are correlated Brownian motions, and we use $\rho_s \in [-1, 1]$ to denote the instantaneous correlation between $dw$ and $db$ in state $s$,

$$\mathbb{E}\left[dw(t)\, db(t)\right] = \sigma_{w,s(t)}\sigma_{b,s(t)}\, \rho_{s(t)}\, dt.$$

At time 0, the worker starts in an initial state $(w(0), b(0), s(0))$. At any date $t \geq 0$ where the worker is non-employed, she can become employed by paying a fixed cost $\psi_e e^{b(t)}$ for a constant $\psi_e \geq 0$. Likewise, the worker can switch from employment to non-employment by paying a cost $\psi_n e^{b(t)}$ for a constant $\psi_n \geq 0$. The worker decides optimally whether to change her employment status at date $t$ given her information set $(w(\tau), b(\tau), s(\tau))$ for $\tau \in [0, t]$.

Let $\tilde{E}(w, b)$ and $\tilde{N}(w, b)$ be the value functions of an employed and non-employed worker with log wage $w$ and log non-employment utility $b$, respectively. The value functions satisfy a standard stopping time, i.e. state-dependent stopping rule, formulation; see Grossman and Laroque (1990) or the textbook treatment in Stokey (2008):

$$\tilde{E}(w, b) = \max_{\tau_e} \mathbb{E}\left[\int_0^{\tau_e} e^{-rt} e^{w(t)} dt + e^{-r\tau_e}\left(\tilde{N}(w(\tau_e), b(\tau_e)) - \psi_n e^{b(\tau_e)}\right) | w(0) = w, b(0) = b\right] \tag{1}$$

$$\tilde{N}(w, b) = \max_{\tau_n} \mathbb{E}\left[\int_0^{\tau_n} e^{-rt} e^{b(t)} dt + e^{-r\tau_n}\left(\tilde{E}(w(\tau_n), b(\tau_n)) - \psi_e e^{b(\tau_n)}\right) | w(0) = w, b(0) = b\right]. \tag{2}$$

An employed worker chooses a stopping time $\tau_e$ at which to switch to non-employment, described by equation (1). This will be the first $\tau \geq 0$ with $\tilde{E}(w(\tau), b(\tau)) \leq \tilde{N}(w(\tau), b(\tau)) - \psi_n e^{b(\tau)}$. Similarly in equation (2), a non-employed worker chooses the stopping time $\tau_n$ at which to change her status to employment. The expectation operators in equations (1) and (2) are taken with respect to the law of motion for $w(t)$ and $b(t)$.

To ensure the value functions are finite, we impose the following parameter restrictions:

$$r > \mu_{w,s} + \tfrac{1}{2}\sigma_{w,s}^2 \text{ and } r > \mu_{b,s} + \tfrac{1}{2}\sigma_{b,s}^2, \text{ for } s \in \{e, n\}. \tag{3}$$

The restriction that $r > \mu_{w,s} + \frac{1}{2}\sigma_{w,s}^2$ guarantees that the value of being employed (non-employed) forever is finite. Moreover, $r > \mu_{b,s} + \frac{1}{2}\sigma_{b,s}^2$ ensures that the value of being non-employed (employed) for $T$ periods and then switching to employment (non-employment) forever is also finite in the limit as $T$ converges to infinity. All four conditions hold with a sufficiently large discount rate.

This partial equilibrium model can capture a number of important phenomena. For example, if $\mu_{w,e} > \mu_{w,n}$, the model picks up on-the-job learning and off-the-job forgetting, emphasized by Ljungqvist and Sargent (1998), who explain the high duration of European unemployment using "...a search model where workers accumulate skills on the job and lose skills during unemployment." Similarly, differential drifts in the flow utility from non-employment, $\mu_{b,e} > \mu_{b,n}$, may capture declining wealth or unemployment income during non-employment.[1] Finally, we can pick up seasonal work and recalls through a particular pattern in $b$ and $w$. In an extreme example, if $w - b$ falls deterministically while employed and rises deterministically while unemployed, the worker will follow a deterministic cycle of employment and non-employment.

We have so far described a model of voluntary non-employment, in the sense that a worker optimally chooses when to work. But a simple reinterpretation of the objects in the model turns it into a model of involuntary unemployment. In this interpretation, the wage is $e^{b(t)}$, while a worker's productivity is $e^{w(t)}$. If the worker is employed by a monopsonist, it earns flow profits $e^{w(t)} - e^{b(t)}$. If the worker is unemployed, the firm may hire her by paying a fixed cost $\psi_e e^{b(t)}$, and similarly the firm must pay $\psi_n e^{b(t)}$ to fire the worker. In this case, the Bellman equations are interpreted as the firm's value and it is the relevant decision-maker. Bentolila and Bertola (1990) study the effect of hiring and firing costs on employment in a closely related model.

## 2.2 Dimension Reduction: The net benefit to work

Because both benefits and costs are homogeneous of degree 1 in $(e^w, e^b)$, and because $w$ and $b$ follow Brownian motions, the value functions satisfy the following homogeneity property: for any pair $(w, b)$ and any constant $a$,

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

By choosing $a = -b$, we get

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

which implicitly defines $E(\cdot)$ and $N(\cdot)$ as a function of the scalar $w - b$. This motivates us to define $\omega(t) \equiv w(t) - b(t)$, the log net benefit to work. This also follows a state-contingent Brownian motion,

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

where $\{B\}$ is a standard Brownian motion defined in terms of $\{B_b, B_w\}$, and the drift and the diffusion coefficient are given by

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

[1] Our model cannot capture finite duration unemployment benefits, which lead to a sudden and predictable drop in $b$. This in turn would lead to a jump in the reemployment hazard, which we do not see in our dataset.

Using this dimension reduction in the state space, we prove in Online Appendix H that the optimal decision of switching from employment to non-employment and vice versa is described by thresholds $\underline{\omega}$ and $\bar{\omega}$ such that a non-employed worker chooses to become employed if the net benefit from working is sufficiently high, $\omega(t) > \bar{\omega}$, and an employed worker switches to non-employment if the benefit is sufficiently low, $\omega(t) < \underline{\omega}$. Employment status stays constant for $\omega(t) \in (\underline{\omega}, \bar{\omega})$. We characterize the thresholds $\underline{\omega}, \bar{\omega}$ in terms of parameters of the model in the online appendix.

# 3 Duration Analysis

In this section, we examine the implications of this economic model for duration data. We first argue that in our model, the non-employment duration of any worker has an inverse Gaussian distribution with two time-invariant reduced-form parameters, $\alpha$ and $\beta$, which in turn are functions of the model's structural parameters. We then imagine a population which has an arbitrary mixture of heterogeneous workers described by an unknown joint distribution $G(\alpha, \beta)$. We are interested in whether we can use non-employment duration data to identify $G$. If the population is such that all employment and non-employment spells have finite duration and we observe all workers for infinitely long, we prove that $G$ is identified using data on the completed duration of two spells for each worker. If the non-employment duration distribution may be defective, we show that the same data, as well as information on the frequency of incomplete spells, partially identifies $G$. If the employment duration distribution may also be defective, our approach partially identifies a weighted distribution $G_\chi$, where weights $\chi(\alpha, \beta)$ correspond to the probability that an employment spell for a typical $(\alpha, \beta)$ worker ends in finite time. Finally, we show how this result generalizes to the case where we observe each worker for a finite amount of time, so the relevant data may be censored.

## 3.1 Duration Distribution for an Individual Worker

We use the economic model to determine the distribution of non-employment duration for any single worker. We assume that we observe a worker at the moment she becomes non-employed, and so know her state is $w(0) - b(0) = \underline{\omega}$ and $s(0) = n$. We then observe the worker forever, as she transitions back and forth between non-employment and employment. We let $t_1 \geq 0$ denote the duration of the initial non-employment spell, with $t_1 = \infty$ denoting a spell that doesn't end in finite time. Assuming $t_1$ is finite, we let $t^e \geq 0$ denote the duration of the subsequent employment spell, with $t^e = \infty$ again denoting a spell that doesn't end in finite time. And finally, if $t_1$ and $t^e$ are both finite, we let $t_2 \geq 0$ denote the duration of the second non-employment spell, with the analogous interpretation of $t_2 = \infty$. One could similarly define the duration of subsequent spells, but we will not use them in our analysis.

The structure of the model implies that, conditional on the parameters of the model, $t_1$ is a random variable given by the first passage time of a Brownian motion with drift. Moreover, if $t_1$ and $t^e$ are both finite, $t_2$ is an independent random variable with the same distribution as $t_1$. In particular, each has an inverse Gaussian distribution with density function at duration $t$

$$f(t; \alpha, \beta) \equiv \frac{\beta}{\sqrt{2\pi}\, t^{3/2}} e^{-\frac{(\alpha t - \beta)^2}{2t}}, \tag{4}$$

where $\alpha \equiv \mu_n/\sigma_n$ and $\beta \equiv (\bar{\omega} - \underline{\omega})/\sigma_n$ (see, for example, Karatzas and Shreve (1998), equation (5.12)). We let $F(t; \alpha, \beta)$ be the cumulative distribution function associated with $f(t; \alpha, \beta)$. Hence, even though each worker is described by a large number of structural parameters, only two reduced-form parameters, $\alpha$ and $\beta$, determine how long a worker stays without a job. Note $\beta$ is nonnegative by assumption, while $\alpha$ may be positive or negative. If $\alpha$ is nonnegative, $\int_0^\infty f(t; \alpha, \beta) dt = 1$, so a worker almost surely returns to work. But if $\alpha$ is negative, the probability of eventually returning to work is $e^{2\alpha\beta} < 1$, so the duration distribution is defective.[2] That is, a non-employed worker with $\alpha$ negative faces a risk of staying non-employed forever.

The inverse Gaussian distribution is flexible, but the model still imposes some restrictions on behavior. Figure 1 shows the hazard $h(t; \alpha, \beta) \equiv f(t; \alpha, \beta)/(1 - F(t; \alpha, \beta))$ for different values of $\alpha$ and $\beta$. It reveals that for the most part, $\beta$ controls the shape of the hazard and $\alpha$ controls its level. Assuming $\beta$ is strictly positive, the hazard always satisfies $h(0; \alpha, \beta) = 0$, achieves a maximum value at some finite time $t$ which depends on both $\alpha$ and $\beta$, and then declines to a long run limit of $\lim_{t\to\infty} h(t; \alpha, \beta) = \alpha^2/2$ if $\alpha$ is positive and 0 otherwise (Chhikara and Folks, 1977). If $\beta = 0$, the hazard is initially infinite and declines monotonically towards its long-run limit.

If $\alpha$ is positive, the expected duration of a completed non-employment spell is $\beta/\alpha$ and the variance of duration is $\beta/\alpha^3$, as can be confirmed directly from the functional form in equation (4). As the duration of a spell goes to infinity, the expected residual duration converges to $2/\alpha^2$, which may be bigger or smaller than the initial expected duration $\beta/\alpha$. The model is therefore consistent with both positive and negative duration dependence in

[2] The usual definition of an inverse Gaussian distribution imposes $\alpha \geq 0$. Whitmore (1978) appears to be the first to recognize that the case with $\alpha < 0$ may be of interest and coined the name *defective* inverse Gaussian distribution to handle this case. Whitmore (1979) shows that a defective inverse Gaussian distribution fits the duration of employment spells at one firm. In the interest of brevity, we refer to both positive and negative values of $\alpha$ as an inverse Gaussian distribution.

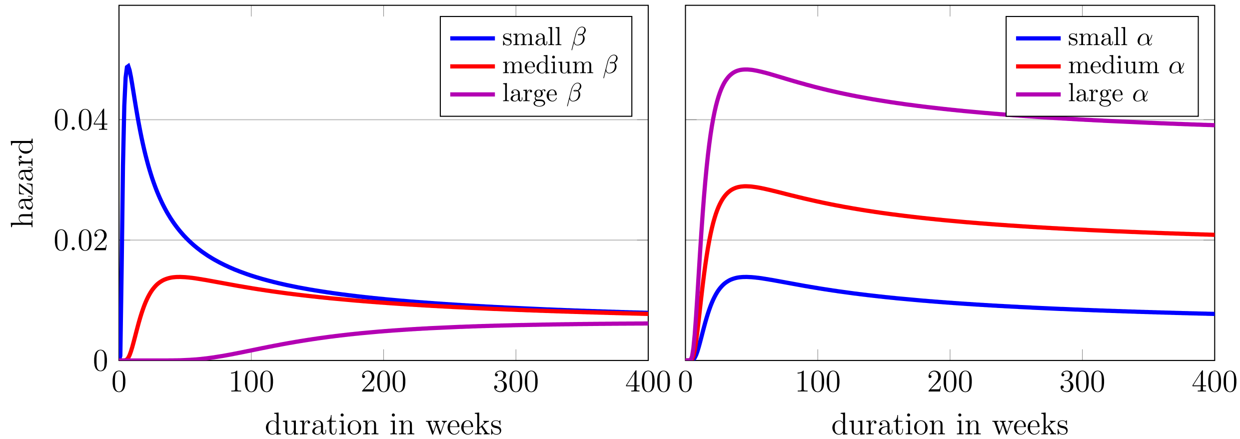



Figure 1: Hazards $h(t; \alpha, \beta)$ implied by the inverse Gaussian distribution for different values of $\alpha > 0$ and $\beta$. The left panel shows hazards for $\alpha = 0.1$ and three different values of $\beta$, 4, 10, and 30. The right panel shows hazards for three different values of $\alpha$, 0.10, 0.18, and 0.27. We also adjust the value of $\beta$ to keep the peak of the hazard at the same duration, which gives $\beta = 10$, 9.5, and 9.2, respectively.

the exit rate from non-employment.

Our model implies that the duration of the intervening employment spell, $t^e$, is also governed by the first passage time of a Brownian motion with drift. Recall that a worker becomes employed when her net benefit from working hits the upper threshold $\bar{\omega}$ and stops working when it hits the lower threshold $\underline{\omega}$. The net benefit from working when employed has a drift $\mu_e$ and standard deviation $\sigma_e$, implying that the distribution of employment spells is given by the inverse Gaussian distribution defined in (4) but with parameters $\alpha_e \equiv -\mu_e/\sigma_e$ and $\beta_e \equiv (\bar{\omega} - \underline{\omega})/\sigma_e$. The negative sign in the expression for $\alpha_e$ reflects the fact that an employment spell ends when $\omega$ travels from the upper barrier to the lower one.

## 3.2 Type Distribution $G$

We consider an economy populated by heterogeneous workers who all face the same economic environment except for the value of their structural parameters. An individual worker is described by a large number of structural parameters, including her discount rate $r$, her fixed costs $\psi_e$ and $\psi_n$, and all the parameters governing the joint stochastic processes for her wage and unemployment utility, both while the worker is employed and while she is non-employed. We denote the vector of these parameters as $\vartheta$,

$$\vartheta \equiv (\mu_{b,e}, \mu_{b,n}, \mu_{w,e}, \mu_{w,n}, \sigma_{b,e}, \sigma_{b,n}, \sigma_{w,e}, \sigma_{w,n}, \rho_e, \rho_n, r, \psi_e, \psi_n).$$

We impose the restrictions in (3), and also assume that $\sigma_n$ is positive. Besides these constraints, we allow for an arbitrary distribution of these structural parameters in the population. Since a worker's non-employment duration only depends on the reduced-form parameters $\alpha = a(\vartheta)$ and $\beta = b(\vartheta)$, we focus our analysis on the joint distribution of these parameters among workers who ever transition from employment to non-employment, $G(\alpha, \beta)$, and we let $Z(\vartheta|\alpha, \beta)$ denote the conditional distribution of $\vartheta$ among such workers, given $\alpha = a(\vartheta)$ and $\beta = b(\vartheta)$.[3] A major goal of this paper is to recover $G$ from data on the duration of non-employment spells.[4]

As is common in the literature (Honoré, 1993, for example), we start by proving identification in an ideal environment where we observe two completed non-employment spells for all workers. More precisely, we imagine an economy that runs forever and assume that $\alpha$ and $\alpha_e = a_e(\vartheta)$ are both nonnegative for all workers, so the completed duration of both non-employment and employment spells is finite. This guarantees that we can measure the completed duration of infinitely many non-employment spells for each worker. We prove in Section 3.4 that the distribution $G$ is identified using two such spells for each worker. We sequentially relax the assumptions that $\alpha$ and $\alpha_e$ are nonnegative in Sections 3.5 and 3.6, respectively, leading to a partial identification result. Finally, in Section 3.7, we relax the assumption that the economy runs forever to show how our approach handles censored data.

## 3.3 Intuition for Identification

We want to use data on the duration of completed non-employment spells to recover the type distribution $G$. Below we give intuition for why observing one non-employment spell per worker is insufficient, but observing two spells helps with identification.

Consider the following two data generating processes. In the first, there is a single type of worker $(\bar{\alpha}, \bar{\beta})$ with $\bar{\alpha} > 0$, giving rise to the duration density $f(t; \bar{\alpha}, \bar{\beta})$ in equation (4). In the second, there are many types of workers $i$ who share a common and large value of $\alpha_i = \alpha > 0$ but differ in their value of the other parameter $\beta_i$. Recall that the expected duration for worker $i$ is then $\beta_i/\alpha$. Assume that $\beta_i/\alpha$ has population density $f(t; \bar{\alpha}, \bar{\beta})$ and take the limit as $\alpha$ goes to infinity holding the distribution of $\beta_i/\alpha$ fixed. In the limiting economy, each worker $i$ has a deterministic duration for their $j^{th}$ spell $t^i_j = \beta_i/\alpha$ but the population duration density is, by construction, the same as with the first data generating process. There is no way to distinguish these two data generating processes using a single

[3]To say anything about workers who never exit employment or workers who never work, we would have to make some untestable assumptions about their unobserved characteristics. Instead, as is standard in the literature, we drop these workers from our analysis.

[4]Whitmore (1979, Section 3.2) fits a mixture of inverse Gaussian distributions to data on employment duration but does not discuss identification.

non-employment spell.[5]

With two completed spells for each worker, however, distinguishing these two data generating processes is trivial. With the first data generating process, the duration of a worker's first spell tells us nothing about the duration of her second spell. In particular, the correlation between the durations of the two spells is zero. With the second data generating process, the duration of a worker's two spells is identical and so the correlation between the durations of the two spells is one.

This simple example suggests the result in our main theorem, that the joint distribution of the duration of two completed spells identifies the joint distribution of $(\alpha, \beta)$ when $\alpha$ and $\alpha_e$ are non-negative.

## 3.4 Proof of Identification

In this section, we establish our main identification result. We maintain the assumptions that $\alpha_e \geq 0$ and time runs forever throughout the section. We initially allow $\alpha$ to be positive or negative, but later in the section we introduce the additional assumption $\alpha \geq 0$ with $G$-probability 1.

Let $\mathbb{T} \subseteq \mathbb{R}_+$ be a set of durations with a non-empty interior. Let $\phi_{\mathbb{T}} : \mathbb{T}^2 \to \mathbb{R}_+$ denote the joint density of completed durations for the subset of our sample who have two completed non-employment spells with durations $(t_1, t_2) \in \mathbb{T}^2$. We make two assumptions when deriving $\phi_{\mathbb{T}}$. First, $\alpha_e \geq 0$ so that all employment spells end in finite time with probability one. Second, time runs forever. For an arbitrary type distribution $G$, the joint density of two completed non-employment durations, conditional on these durations being in the set $\mathbb{T}^2$, is

$$\phi_{\mathbb{T}}(t_1, t_2) = \frac{\int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG(\alpha, \beta)}{\int_{\mathbb{T}^2} \int f(t_1'; \alpha, \beta) f(t_2'; \alpha, \beta) dG(\alpha, \beta)\, d(t_1', t_2')} \equiv V_{\mathbb{T}}(G | t_1, t_2), \tag{5}$$

where $f$ is the density function of the inverse Gaussian distribution, defined in equation (4). Performing this operation for all $(t_1, t_2) \in \mathbb{T}^2$, this defines a mapping $V_{\mathbb{T}}$ which, for the given choice of $\mathbb{T}$, takes a type distribution $G$ and returns the density of the duration of two completed spells $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G)$. Importantly for our analysis in Section 3.5, equation (5) holds even if we do not assume that $\alpha$ is nonnegative with $G$-probability 1. We are interested whether we can use $\phi_{\mathbb{T}}$, which we think of as observable to the econometrician, to uniquely

[5] Al-Hussaini and Ahmad (1981) prove that a finite mixture of inverse Gaussian distributions with non-negative $\alpha$ is identified using a single non-employment spell. Our example does not contradict this result as it requires continuum of types. Instead, our Theorem 1 establishes that an arbitrary mixture (including a continuum) of inverse Gaussian distributions with non-negative $\alpha$ is identified using two-spell data.

recover the unknown type distribution $G$.[6] We prove that we can do so in Theorem 1 below, which establishes that the mapping $V_{\mathbb{T}}$ is injective under the restriction that $\alpha$ is nonnegative.

We prove the identification result through a series of Propositions. The first shows that the partial derivatives of $V_{\mathbb{T}}(G)$ exist almost everywhere:

**Proposition 1** *Take any $(t_1, t_2) \in \mathbb{T}^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$. For any type distribution $G$, $V_{\mathbb{T}}(G)$ is infinitely many times differentiable at $(t_1, t_2)$.*

We prove all the results in this subsection in Appendix A. The proof verifies the conditions under which the Leibniz formula for differentiation under the integral is valid. This requires us to bound the derivatives in appropriate ways, which we accomplish by characterizing the structure of the partial derivatives of the product of two inverse Gaussian densities. Our bound uses that $t_1 \neq t_2$, and indeed Example 1 in Online Appendix C shows that $V_{\mathbb{T}}(G)$ is not necessarily differentiable at points where $t_1 = t_2$.

For the next Proposition, we look at the conditional distribution of $(\alpha, \beta)$ among workers whose two non-employment spells last exactly $(t_1, t_2)$ periods:

$$d\tilde{G}(\alpha, \beta | t_1, t_2) \equiv \frac{f(t_1; \alpha, \beta)\, f(t_2; \alpha, \beta)\, dG(\alpha, \beta)}{\int f(t_1; \alpha', \beta')\, f(t_2; \alpha', \beta')\, dG(\alpha', \beta')}, \tag{6}$$

where again $f$ is the density function of the inverse Gaussian distribution, defined in equation (4). For fixed $(t_1, t_2) \in \mathbb{R}_+^2$, this defines a mapping $W$ which takes a type distribution $G$ and returns another type distribution $\tilde{G} = W(G|t_1, t_2)$. Notably, for any $t_1 > 0$ and $t_2 > 0$, this mapping is bijective, with its inverse satisfying

$$dG(\alpha, \beta) = \frac{d\tilde{G}(\alpha, \beta | t_1, t_2)}{f(t_1; \alpha, \beta)\, f(t_2; \alpha, \beta)} \left( \int \frac{1}{f(t_1; \alpha', \beta')\, f(t_2; \alpha', \beta')} d\tilde{G}(\alpha', \beta' | t_1, t_2) \right)^{-1}. \tag{7}$$

We prove that the function $\phi_{\mathbb{T}}$ identifies all the *even* moments of $\tilde{G}$ for any $t_1 \neq t_2$. More precisely, the partial derivatives of $\phi_{\mathbb{T}}$ at $(t_1, t_2)$ all exist and they jointly identify the even moments of $\tilde{G}(\alpha, \beta | t_1, t_2)$.

**Proposition 2** *Take any $(t_1, t_2) \in \mathbb{T}^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$, and any strictly positive integer $m$. Fix a type distribution $G$ and let $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G)$. Any moment of the distribution $\tilde{G} = W(G|t_1, t_2)$ of the form $\int \alpha^{2i} \beta^{2j} d\tilde{G}(\alpha, \beta | t_1, t_2)$ for $i \in \{0, 1, \ldots, m\}$ and $j = m - i$ is a known function of the partial derivatives $\partial^{i'+j'} \phi_{\mathbb{T}}(t_1, t_2) / \partial t_1^{i'} \partial t_2^{j'}$ for all $i' \in \{0, 1, \ldots, m\}$ and $j' \in \{0, 1, \ldots, m - i'\}$.*

[6] The distribution $\phi_{\mathbb{T}}$ is a non-trivial function of $\mathbb{T}$, but one of the strengths of our identification result is that we prove that the distribution $G$ is identified for *any* choice of $\mathbb{T}$. We discuss how we choose $\mathbb{T}$ when we turn to estimation in Section 5.2.

For convenience, let $\mathbb{E}(\alpha^{2i}\beta^{2j}|t_1,t_2) \equiv \int \alpha^{2i}\beta^{2j} d\tilde{G}(\alpha,\beta|t_1,t_2)$. The statement of the proposition suggests a recursive structure, which we follow in our proof in Appendix A. In the first step, set $m = 1$. The two first partial derivatives $\partial\phi_{\mathbb{T}}(t_1,t_2)/\partial t_1$ and $\partial\phi_{\mathbb{T}}(t_1,t_2)/\partial t_2$ determine the two first even moments, $\mathbb{E}(\alpha^2|t_1,t_2)$ and $\mathbb{E}(\beta^2|t_1,t_2)$. In the second step, set $m = 2$. The three second partial derivatives and the results from first step then determine the three second even moments, $\mathbb{E}(\alpha^4|t_1,t_2)$, $\mathbb{E}(\alpha^2\beta^2|t_1,t_2)$, and $\mathbb{E}(\beta^4|t_1,t_2)$. In the $m^{th}$ step, the $m+1$ $m^{th}$ partial derivatives and the results from the previous steps determine the $m+1$ $m^{th}$ even moments of $\tilde{G}$. Our proof gives explicit functions at each of these steps. It also establishes that all the moments of $\tilde{G}$ exist and are finite for any type distribution $G$, even one that does not itself have any finite moments; see Example 2 in Online Appendix C.

We now seek to use the even moments of the distribution $\tilde{G} = W(G|t_1,t_2)$ to recover the entire distribution. Doing so requires the additional assumption that $\alpha$ is nonnegative, for reasons that we return to in Section 3.5:

**Proposition 3** *Take any $(t_1,t_2) \in \mathbb{T}^2$ with $t_1 > 0$, $t_2 > 0$ and $t_1 \neq t_2$. Let $G$ be any type distribution with $\alpha$ nonnegative with $G$-probability 1. Then $\tilde{G} = W(G|t_1,t_2)$, is uniquely identified by the set of moments $\int \alpha^{2i}\beta^{2j} d\tilde{G}(\alpha,\beta|t_1,t_2)$ for all $(i,j) \in \{0,1,\ldots\}^2$.*

The proof of the proposition establishes a version of the Stieljes moment problem, that the moments $\int \alpha^{2i}\beta^{2j} d\tilde{G}(\alpha,\beta|t_1,t_2)$, $(i,j) \in \{0,1,\ldots\}^2$ uniquely determine the joint distribution of $(\alpha^2,\beta^2)$ conditional on completed duration $(t_1,t_2)$. For this we establish that the moments do not grow too fast when $\tilde{G} = W(G|t_1,t_2)$. If $\alpha$ and $\beta$ are both nonnegative, we can trivially recover $\tilde{G}(\alpha,\beta|t_1,t_2)$ from the joint distribution of $(\alpha^2,\beta^2)$ conditional on completed duration $(t_1,t_2)$. The model implies $\beta$ is nonnegative, while Proposition 3 imposes that $\alpha$ is nonnegative with probability 1.

Our main identification result follows immediately from Propositions 1–3:

**Theorem 1** *Assume that $\alpha$ is nonnegative with $G$-probability 1. The function $V_{\mathbb{T}}$ is injective.*

**Proof of Theorem 1.** Proposition 1 shows that for any $G$, $\phi_{\mathbb{T}}$ is infinitely many times differentiable. Proposition 2 shows that for any $(t_1,t_2) \in \mathbb{T}^2$, $t_1 \neq t_2$, $t_1 > 0$, and $t_2 > 0$, there is one solution for the moments of $(\alpha^2,\beta^2)$ conditional on durations $(t_1,t_2)$, given all the partial derivatives of $\phi_{\mathbb{T}}$ at $(t_1,t_2)$. Proposition 3 shows that these moments uniquely determine the distribution function $\tilde{G}(\alpha,\beta|t_1,t_2)$ with the additional assumption that $\alpha \geq 0$ with $G$-probability 1. Finally, $G = W^{-1}(\tilde{G}|t_1,t_2)$, defined in equation (7). Thus for any $\phi_{\mathbb{T}}$, there is at most one such $G$ satisfying $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G)$, i.e. $V_{\mathbb{T}}$ is injective. ■

We stress that our result does not assume any auxiliary rank or completeness conditions (Newey and Powell, 2003; Canay, Santos and Shaikh, 2013), even though this is an infinite dimensional problem. These conditions instead follow from the structure of our model.

An interesting question is whether the mapping $V_{\mathbb{T}}$ is *bijective* when $\alpha$ is nonnegative with $G$-probability 1, so every data set $\phi_{\mathbb{T}}$ can be generated by some type distribution $G$. In Online Appendix D, we develop a set of non-trivial restrictions that our model imposes on $\phi_{\mathbb{T}}$, proving that $V_{\mathbb{T}}$ is not bijective. Indeed, if the data generating process is a MPH model with some restrictions on the frailty distribution, or if it is a mixture of log-normal duration distributions, the resulting $\phi_{\mathbb{T}}$ could not be generated by our stopping time model.

Finally, we note that Theorem 1 proves $V_{\mathbb{T}}$ is injective when $\alpha$ is always nonnegative, regardless of the sign of the corresponding parameter when employed, $\alpha_e$. The sign of $\alpha_e$ still plays a critical role in our analysis, however, since if $\alpha_e$ can be negative, the employment duration distribution may be defective; and if that were the case, the joint density of the duration of two completed non-employment spells, $\phi_{\mathbb{T}}$, would no longer satisfy equation (5). We address this issue in Section 3.6.

## 3.5 Bounds for the Share of Workers with Negative Drift

We now explore what happens if we allow for the possibility that $\alpha < 0$ but still assume that $\alpha_e$ is nonnegative, so the non-employment duration distribution may be defective, but the employment duration distribution is not.

When $\alpha < 0$, there is a positive probability that a non-employment spell never ends. Nevertheless, the joint density of completed durations for the subset of our sample who have two completed non-employment spells with durations $(t_1, t_2) \in \mathbb{T}^2$, $\phi_{\mathbb{T}}(t_1, t_2)$, still satisfies $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G)$, defined in equation (5), since all workers with $t_1 \in \mathbb{T}$ experience a second non-employment spell. The issue that arises when $\alpha$ may be positive or negative is that the mapping $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G)$ is not injective.

To understand the failure of invertibility, observe that the functional form of the (defective) inverse Gaussian distribution in equation (4) implies

$$\frac{f(t;\alpha,\beta)}{f(t;-\alpha,\beta)} = e^{2\alpha\beta} \tag{8}$$

for all $(\alpha, \beta, t) \in \mathbb{R}^3_+$. Proportionality of $f(t;\alpha,\beta)$ and $f(t;-\alpha,\beta)$ implies that the probability distribution over completed durations $\phi_{\mathbb{T}}$ is the same if a worker is described by reduced-form parameters $(\alpha, \beta)$, $\alpha > 0$, or if there are $e^{4\alpha\beta}$ times as many workers in the sample described by reduced-form parameters $(-\alpha, \beta)$; see equation (5). The consequence is that completed spell data can never help us recover the sign of $\alpha$:

**Proposition 4** *Take any two distributions $G_1$ and $G_2$. $V_{\mathbb{T}}(G_1) = V_{\mathbb{T}}(G_2)$ if and only if there* *exists functions* $\gamma_1, \gamma_2 : \mathbb{R}_+^2 \to [0,1]$ *and a distribution* $G^+ : \mathbb{R}_+^2 \to [0,1]$ *such that for all* $\alpha > 0$ *and* $\beta \geq 0$ *and* $i = 1, 2$*,*

$$dG_i(\alpha,\beta) = \frac{\gamma_i(\alpha,\beta)dG^+(\alpha,\beta)}{\int \big(\gamma_i(\alpha',\beta') + e^{4\alpha'\beta'}(1-\gamma_i(\alpha',\beta'))\big)dG^+(\alpha',\beta')}, \tag{9}$$

$$dG_i(-\alpha,\beta) = \frac{e^{4\alpha\beta}(1-\gamma_i(\alpha,\beta))dG^+(\alpha,\beta)}{\int \big(\gamma_i(\alpha',\beta') + e^{4\alpha'\beta'}(1-\gamma_i(\alpha',\beta'))\big)dG^+(\alpha',\beta')},$$

*and for all* $\beta \geq 0$ *and* $i = 1, 2$*,*

$$dG_i(0,\beta) = \frac{dG^+(0,\beta)}{\int \big(\gamma_i(\alpha',\beta') + e^{4\alpha'\beta'}(1-\gamma_i(\alpha',\beta'))\big)dG^+(\alpha',\beta')}.$$

The proof in Appendix A establishes the "if" part by showing that when $G_i$ is given by equation (9), the distribution of the duration of two completed spells $\phi_{\mathbb{T}}$, defined in equation (5), does not depend on $\gamma_i$. The "only if" part constructs the unique $\gamma$ and $G^+$ for any distribution $G$ such that $V_{\mathbb{T}}(G) = V_{\mathbb{T}}(G^+)$, proving that equation (9) must hold.

This limitation on identification could matter for economically interesting environments. The reduced-form parameter $\alpha$ is negative whenever the drift in the net benefit from employment while non-employed is negative. With the reweighting in equation (9), this does not affect the joint distribution of the duration of two completed non-employment spells, but it does affect the fraction of spells that are completed and hence the hazard of exiting non-employment. This insight motivates our approach to partially identifying $G$ when $\alpha$ can be negative: we use data on incomplete spells.

We proceed in two steps. First we use the distribution of the duration of two completed non-employment spells, $\phi_{\mathbb{T}}$, to identify a candidate type distribution $G^+(\alpha,\beta)$ with support $\mathbb{R}_+^2$. Theorem 1 tells us that if $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G^+)$ for some such $G^+$, then $G^+$ is unique. Proposition 4 and equation (9) then tell us how to construct all the type distributions $G$ such that $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G)$, one type distribution for each function $\gamma : \mathbb{R}_+^2 \to [0,1]$.

In the second step, we restrict the set of distributions $G$ by bringing in two additional pieces of information, which we again think of as observable data when the economy runs forever and $\alpha_e$ is nonnegative. The first, $p_{1,\mathbb{T}} \in [0,1]$, is the fraction of workers whose first non-employment spell has completed duration $t_1 \in \mathbb{T}$. According to the model, this satisfies

$$p_{1,\mathbb{T}} = \int_{\mathbb{T}} \int f(t_1;\alpha,\beta)dG(\alpha,\beta)\,dt_1 \tag{10}$$

The second, $p_{2,\mathbb{T}} \in [0,1]$, is the fraction whose second non-employment spell has completed duration $t_2 \in \mathbb{T}$ conditional on $t_1 \in \mathbb{T}$. According to the model, this satisfies

$$p_{2,\mathbb{T}} = \frac{\int_{\mathbb{T}^2} \int f(t_1;\alpha,\beta) f(t_2;\alpha,\beta) dG(\alpha,\beta)\, d(t_1,t_2)}{\int_{\mathbb{T}} \int f(t_1;\alpha,\beta) dG(\alpha,\beta)\, dt_1}. \tag{11}$$

We look for distributions $G$ that are consistent with equations (9), (10), and (11). We call the set of such distributions $\mathbb{G}$. We discuss the construction and interpretation of $\mathbb{G}$ at the end of the next subsection.

### 3.6 Weighted Type Distribution $G_\chi$

Now we relax the assumption that all employment spells have a finite duration, allowing for the possibility that $\alpha_e < 0$ (as well as $\alpha < 0$). Our model allows for an arbitrary relationship between the reduced-form parameters $(\alpha, \beta)$ and $(\alpha_e, \beta_e)$, creating a new issue for identification. For example, we cannot identify the joint distribution of $(\alpha, \beta)$ for those workers who find a job that they never lose, since we only observe a single non-employment spell. Nevertheless, we show how to build on our previous analysis to identify a set of weighted distributions $\mathbb{G}_\chi$, still for workers who ever transition from employment to non-employment, but now with weights $\chi(\alpha, \beta)$ corresponding to the probability that a typical employment spell ends in finite time for a worker with type $(\alpha, \beta)$.

We start by constructing the weights. For a worker with reduced-form parameters $(\alpha_e, \beta_e)$ while employed, the inverse Gaussian distribution implies that a typical employment spell ends in finite time, $t^e < \infty$, with probability $1 - \mathbb{1}_{\alpha_e<0}\big(1 - e^{2\alpha_e\beta_e}\big)$, where $\mathbb{1}$ is the indicator function. Now let $a_e(\vartheta)$ and $b_e(\vartheta)$ be functions which take the vector of parameters $\vartheta$ and return the reduced-form parameters $\alpha_e$ and $\beta_e$, respectively. Then the weight we attach to a typical $(\alpha, \beta)$ worker is

$$\chi(\alpha,\beta) \equiv 1 - \int \mathbb{1}_{a_e(\vartheta)<0} \left(1 - e^{2a_e(\vartheta)b_e(\vartheta)}\right) dZ(\vartheta|\alpha,\beta), \tag{12}$$

where $Z(\vartheta|\alpha,\beta)$ is the conditional distribution of $\vartheta$. This is the probability that an $(\alpha, \beta)$ worker has an employment spell that ends in finite time. We are then able to partially identify the weighted type distribution

$$dG_\chi(\alpha,\beta) \equiv \frac{\chi(\alpha,\beta) dG(\alpha,\beta)}{\int \chi(\alpha',\beta') dG(\alpha',\beta')}. \tag{13}$$

Note that if the type while employed $(a_e(\vartheta), b_e(\vartheta))$ is independent of the reduced-form parameters $(\alpha, \beta)$, equation (12) implies $\chi$ is constant and so $G_\chi = G$. Otherwise $G_\chi$ puts more weight onto individuals who are more likely to become non-employed. While we did not choose the weights $\chi$ for this reason, it seems natural to focus an analysis of non-employment duration on such individuals.

Identification of $G_\chi$ proceeds in two steps, similarly to Section 3.5. In the first step, we use $\phi_{\mathbb{T}}(t_1, t_2)$, the density of the duration of two completed non-employment spells with $(t_1, t_2) \in \mathbb{T}^2$, to identify a candidate type distribution $G_\chi^+(\alpha, \beta)$ with support $\mathbb{R}_+^2$. Our model implies that this satisfies a generalized version of equation (5),

$$\phi_{\mathbb{T}}(t_1, t_2) = \frac{\int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG_\chi(\alpha, \beta)}{\int_{\mathbb{T}^2} \int f(t_1'; \alpha, \beta) f(t_2'; \alpha, \beta) dG_\chi(\alpha, \beta)\, d(t_1', t_2')} = V_{\mathbb{T}}(G_\chi | t_1, t_2) \tag{14}$$

where we recognize that a necessary condition for observing two completed non-employment spells is that the duration of the intervening employment spell is finite, an event with probability $\chi(\alpha, \beta)$ for an $(\alpha, \beta)$ worker. It follows from Theorem 1 that using data on the distribution of the duration of two completed non-employment spells for only those workers who have two completed non-employment spells, we can uniquely recover a candidate distribution $G_\chi^+$ with support $\mathbb{R}_+^2$. Proposition 4 then tells us how to construct all the type distributions $G_\chi$ such that $\phi_{\mathbb{T}} = V_{\mathbb{T}}(G_\chi) = V_{\mathbb{T}}(G_\chi^+)$.

In the second step, we restrict the set of distributions by using additional information. According to the model, the fraction of workers whose first non-employment spell has completed duration $t_1 \in \mathbb{T}$ under the weighted distribution $G_\chi$ is

$$p_{1,\mathbb{T}} = \int_{\mathbb{T}} \int f(t_1; \alpha, \beta) dG_\chi(\alpha, \beta)\, dt_1 \equiv P_{1,\mathbb{T}}(G_\chi). \tag{15}$$

When $\alpha_e$ may be negative, we can no longer measure $p_{1,\mathbb{T}}$, since we do not know whether a worker whose first non-employment spell lasts forever would have had an employment spell that ended in finite time.[7] However, we can observe how many workers' first non-employment spell lasts forever and can then bound $p_{1,\mathbb{T}}$ through the restriction that the weight $\chi$ for such workers lies between 0 and 1. This gives us $p_{1,\mathbb{T}} \in [\underline{p}_{1,\mathbb{T}}, \bar{p}_{1,\mathbb{T}}]$.

The model also implies that the probability that a worker has two non-employment spells with completed durations $(t_1, t_2) \in \mathbb{T}^2$ conditional on having one such spell and starting a second spell (i.e. having a finite duration employment spell) is

$$p_{2,\mathbb{T}} = \frac{\int_{\mathbb{T}^2} \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG_\chi(\alpha, \beta)\, d(t_1, t_2)}{\int_{\mathbb{T}} \int f(t_1; \alpha, \beta) dG_\chi(\alpha, \beta)\, dt_1} \equiv P_{2,\mathbb{T}}(G_\chi). \tag{16}$$

[7] We could measure $p_{1,\mathbb{T}}$ satisfying equation (10) when $\alpha_e$ may be negative; however, this tells us about the unweighted distribution $G$ rather than the weighted distribution $G_\chi$ which appears in equation (14). Without auxiliary assumptions, e.g. on the relationship between $(\alpha_e, \beta_e)$ and $(\alpha, \beta)$, this is not useful for us.

We can measure $p_{2,\mathbb{T}}$ directly in the data using the distribution of the duration of non-employment spells.

We look for the set of distributions

$$\mathbb{G}_\chi \equiv \Big\{ G_\chi | V_{\mathbb{T}}(G_\chi) = V_{\mathbb{T}}(G_\chi^+), P_{1,\mathbb{T}}(G_\chi) \in [\underline{p}_{1,\mathbb{T}}, \bar{p}_{1,\mathbb{T}}], \text{ and } P_{2,\mathbb{T}}(G_\chi) = p_{2,\mathbb{T}} \Big\}. \tag{17}$$

This set is defined for a given distribution function $G_\chi^+$, and scalars $\underline{p}_{1,\mathbb{T}}$, $\bar{p}_{1,\mathbb{T}}$, and $p_{2,\mathbb{T}}$. In practice, we first find $G_\chi^+$ and then construct $G_\chi$ for an arbitrary function $\gamma$ using equation (9), which ensures that $V_{\mathbb{T}}(G_\chi) = V_{\mathbb{T}}(G_\chi^+)$. We then find the implied $p_{1,\mathbb{T}}$ and $p_{2,\mathbb{T}}$ from equations (15) and (16) and check whether $p_{1,\mathbb{T}} \in [\underline{p}_{1,\mathbb{T}}, \bar{p}_{1,\mathbb{T}}]$ and $p_{2,\mathbb{T}}$ has the correct value.

## 3.7 Censored Data

Finally, we relax the assumption that we observe all workers forever, allowing for censored data. We also do not restrict the sign of $\alpha$ or $\alpha_e$. We find that censoring changes the weights $\chi$ but otherwise does not affect our analysis.

Consider a worker with parameter vector $\vartheta$ and hence type $(\alpha, \beta) = (a(\vartheta), b(\vartheta))$. That is, the uncensored duration of the worker's $j^{th}$ non-employment spell $t_j$ has density $f(\alpha, \beta)$ defined in equation (4). When we first observe a worker, they may be employed or non-employed. We focus on the $T > 0$ periods after we observe the worker transitioning from employment to non-employment and before they exit the data set. As before, we let $G$ denote the distribution of $(\alpha, \beta)$ among such workers. There are also some workers whom we never observe becoming non-employed, either because they never work or never lose their job. We set $T = 0$ for those workers.

Since $T$ is finite, some employment or non-employment spell durations are censored. We place no restrictions on the joint distribution of $T$ and the parameter vector $\vartheta$ and let $Z(\vartheta, T|\alpha, \beta)$ denote their joint distribution conditional on $\alpha$ and $\beta$ and on $T > 0$.

Next, set the observation window for complete spells to $\mathbb{T} = [0, \bar{T}]$ for some fixed $\bar{T} > 0$. We weight type $(\alpha, \beta)$ workers by the probability that the difference between the censoring time $T$ and the uncensored duration of a typical employment spell $t^e$ is strictly more than $2\bar{T}$:

$$\chi^c(\alpha, \beta) \equiv \int \Pr(T - t^e > 2\bar{T}|\vartheta, T) dZ(\vartheta, T|\alpha, \beta), \tag{18}$$

where $t^e$ has a (possibly defective) inverse Gaussian distribution that depends on $a_e(\vartheta)$ and $b_e(\vartheta)$.[8] We are then able to partially identify the weighted type distribution $G_{\chi^c}$ defined as

[8] If we consider a sequence of economies with fixed $\vartheta$ and increasingly and unboundedly long observation times $T$, $\chi^c \to \chi$, the probability that a typical employment spell has finite duration for a type $(\alpha, \beta)$ worker. Thus $\chi^c$ is a natural extension of the weights $\chi$ in Section 3.6.

in equation (13). Note that if the type while employed $(a_e(\vartheta), b_e(\vartheta))$ and censoring time $T$ are independent of the reduced-form parameters $(\alpha, \beta)$ conditional on $T > 0$, equation (18) implies $\chi^c$ is constant and so $G_{\chi^c} = G$. Otherwise $G_{\chi^c}$ puts more weight onto individuals who have shorter employment spells and who stay in the sample for longer after we observe them becoming non-employed.

Identification proceeds in two steps as in Section 3.6. In both steps, we focus on workers with censoring time $T$ and employment duration $t^e$ satisfying $T - t^e > 2\bar{T}$. In the first step, we use the joint density of two completed non-employment spells with duration shorter than $\bar{T}$, $\phi_{[0,\bar{T}]}(t_1, t_2)$, to recover a candidate distribution $G^+_{\chi^c}$, following the logic in Theorem 1. In the second step, we use data on $p_{1,[0,\bar{T}]}$, the fraction of the population with $T - t^e > 2\bar{T}$ which has completed duration $t_1 < \bar{T}$, and $p_{2,[0,\bar{T}]}$, the fraction of the population with $T - t^e > 2\bar{T}$ and $t_1 \leq \bar{T}$ which has $t_2 < \bar{T}$, to set-identify distributions $\mathbb{G}_{\chi^c}$ consistent with available data on both complete and incomplete spells.

For any worker for whom we observe two completed non-employment spells with $(t_1, t_2) \in [0, \bar{T}]^2$, we can observe if $T - t^e > 2\bar{T}$. This means that we can measure $\phi_{[0,\bar{T}]}(t_1, t_2)$ and so invert equation (14) to recover $G_{\chi^+}$. Similarly, we can always observe if a worker with $t_1 \leq \bar{T}$ has $T - t^e > 2\bar{T}$; in this event, he must start a second non-employment spell and we can tell whether he has $t_2 \in [0, \bar{T}]$. This means that we can measure $p_{2,[0,\bar{T}]}$ and so use equation (16) to get a restriction on the set $\mathbb{G}_{\chi^c}$. However, censoring does affect our ability to observe if a worker with $t_1 > 2\bar{T}$ satisfies $T - t^e > 2\bar{T}$, either because we do not observe any employment spell ($T = t_1$) or because the employment spell is censored ($T - t^e \leq t_1$). We thus bound $p_{1,[0,\bar{T}]}$ through the restriction that the weight $\chi$ for such workers lies between 0 and 1, giving us $p_{1,[0,\bar{T}]} \in [\underline{p}_{1,[0,\bar{T}]}, \bar{p}_{1,[0,\bar{T}]}]$. Thus equation (15) also gives us a restriction on the set $\mathbb{G}_{\chi^c}$. Putting this together, we can use the distribution $G_{\chi^+}$ and the scalars $\underline{p}_{1,[0,\bar{T}]}$, $\bar{p}_{1,[0,\bar{T}]}$, and $p_{2,[0,\bar{T}]}$ to recover $\mathbb{G}_{\chi^c}$ from equation (17).

The fact that $\mathbb{G}_{\chi^c}$ can have multiple elements means that our model is generally partially identified. Whether this partial identification result is useful is an empirical issue and may depend on the question we are interested in answering.

# 4 Decomposition of the Hazard

Suppose we know the type distribution $G$. This section discusses how to use that information, together with the known functional form of the inverse Gaussian distribution, to understand duration dependence in the hazard of exiting non-employment. Towards that end, we propose a multiplicative decomposition of the mixture hazard into two components, one measuring the average hazard for a typical worker and the other measuring the impact of heterogeneity on the mixture hazard.

Our decomposition can be applied in any model in which we know distribution of types and hazard conditional on type; that is, nothing in the decomposition relies on the assumption of the inverse Gaussian distribution of non-employment duration. To emphasize this, in this section (with some abuse of notation) we denote a worker type as $\zeta$ with (possibly defective) duration distribution $F(t;\zeta)$ and duration density $f(t;\zeta)$. We also let $G$ denote the cumulative distribution of $\zeta$ in the newly non-employed population. In our structural model, $\zeta = (\alpha, \beta)$ and $F$ is an inverse Gaussian distribution. Here we only impose that $f$ is everywhere differentiable with respect to $t$.

Let $h(t;\zeta) \equiv f(t;\zeta)/(1 - F(t;\zeta))$ denote the hazard for type $\zeta$ at duration $t$. With some abuse of notation, let $G(\zeta;t)$ denote the type distribution among workers whose duration exceeds $t$ periods. The mixture hazard at duration $t$, $H(t)$, is an average of individual hazards weighted by their share among workers with duration $t$,

$$H(t) = \frac{\int f(t;\zeta)dG(\zeta)}{\int(1 - F(t;\zeta))dG(\zeta)} = \int h(t;\zeta)dG(\zeta;t), \tag{19}$$

where the integrals are taken with respect to $\zeta$. The second equality uses the definition of $h(t;\zeta)$ and the fact that duration-$t$ type distribution reflects the survival probability of each type:

$$dG(\zeta;t) \equiv \frac{(1 - F(t;\zeta))dG(\zeta)}{\int(1 - F(t;\zeta'))dG(\zeta')}. \tag{20}$$

We propose an exact multiplicative decomposition of the mixture hazard, $H(t) = H^s(t)H^h(t)$, based on a Divisia index:

$$\frac{d\log H^s(t)}{dt} \equiv \frac{\int \dot{h}(t;\zeta)dG(\zeta;t)}{H(t)} \text{ and } \frac{d\log H^h(t)}{dt} \equiv \frac{\int h(t;\zeta)d\dot{G}(\zeta;t)}{H(t)} \tag{21}$$

with the normalization $H^s(0) = H(0)$ and $H^h(0) = 1$. Here 'dots' represent time-derivatives, i.e. $\dot{h}(t;\zeta) \equiv \partial h(t;\zeta)/\partial t$ and $d\dot{G}(\zeta;t) \equiv \partial dG(\zeta;t)/\partial t$. That this is an exact decomposition follows immediately from time differentiating $H(t) = \int h(t;\zeta)dG(\zeta;t)$ and dividing through by $H(t)$:

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

We interpret $H^s(t)$ as the contribution of structural duration dependence, since its change is equal to the average change in the hazard of workers who are still non-employed at duration $t$. If each worker had a constant hazard, this term would be constant and we would conclude that there is no structural duration dependence. The remaining term $H^h(t)$ represents the heterogeneity effect, because it captures how the mixture hazard changes due to changes in the distribution of types at a particular duration.

The structural hazard $H^s(t)$ can be either increasing or decreasing, but the heterogeneity effect $H^h(t)$ is a non-increasing function:

$$\frac{d\log H^h(t)}{dt} = -\frac{\int (h(t;\zeta) - H(t))^2 dG(\zeta;t)}{H(t)} \leq 0. \tag{22}$$

See the proof in Appendix Online E. The numerator is just the cross-sectional variance of the hazard at duration $t$, and so this result generalizes the *fundamental theorem of natural selection* (Fisher, 1930), which states that "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time."[9] Intuitively, types with a higher than average hazard are always declining as a share of the population.

These results generalize equation (3.7) in Lancaster (1992, Chapter 4) beyond the MPH model, the special case where we can write $h(t;\zeta) \equiv \zeta\bar{h}(t)$ for all $\zeta$ and $t$. Then equation (21) implies that the structural hazard is $H^s(t) = \bar{h}(t)$, as in his book. Lancaster (1992, Chapter 4) also shows in his equation (3.5) that the term we call the heterogeneity effect is always decreasing with duration. Equation (22) generalizes that result to an arbitrary mixture model.

# 5 Austrian Data

We estimate our model and decompose duration dependence using data from the Austrian social security registry (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf and Buchi, 2009). The data set covers the universe of private sector workers over the years 1986–2007. It contains information on workers' employment, registered unemployment, maternity leave, and retirement, with the exact begin and end date of each spell.[10] Online Appendix J discusses relevant properties of the Austrian labor market in more detail.

## 5.1 Definition of Spell Duration

Our model features two labor market states, employment and non-employment, while the data have some other states, including maternity leave, registered unemployment, marginal jobs, out of the labor force, and retirement. We first explain how we map those states into

[9]We are grateful to Jörgen Weibull for pointing out this connection to us.
[10]We have data available back to 1972, but can only measure registered unemployment after 1986.

our two states. To start, we truncate the worker's labor market history to focus only on the time when she is at least 25 years old and less than 60 years old. The age restrictions mitigate issues related to schooling and retirement. We then define an *employment spell* to be the time spent working in a non-marginal job in a private sector firm. Next, we drop dates with employment, maternity leave, or retirement to get spells when the worker is potentially non-employed. If the worker is registered as unemployed for at least one day during such a spell, we call it a *non-employment spell*. Otherwise, we drop the spell.[11] This gives us non-overlapping spells of employment and non-employment with some breaks in between.

We use both complete and incomplete spell data. A non-employment spell starts when a worker transitions from employment to non-employment and is complete if it ends with worker starting another employment spell. Otherwise it is incomplete. We similarly measure the duration of each complete or incomplete employment spell. 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.[12]

Now consider a *segment* of the worker's history consisting of alternating non-employment and employment spells, with two segments separated by a period that we have labeled as neither employment nor as non-employment, e.g. by a maternity leave. For each segment, we compute the censoring time as the number of weeks from the worker's first transition from employment to non-employment until the end of the segment. If the worker never makes such a transition during that segment, the censoring time is zero. We then focus our analysis for each worker on the segment with the longest censoring time, and denote the censoring time as $T$.

## 5.2 Sample Selection and Estimation

We estimate our model in two steps, following our identification proof. In the first step, we use the joint distribution of two completed non-employment spells to estimate $G^+_{\chi^c}$, in line with the identification result established in Theorem 1. The weights $\chi^c$ are defined in equation (18) and the weighted distribution is defined in equation (13). We obtain the estimate $\hat{G}^+_{\chi^c}$ using the Expectation-Maximization (EM) algorithm,[13] taking the distribution

[11]If a worker spends less than two months between jobs and is never registered as unemployed, on maternity leave, or retired during those intervening days, we consider that to be part of an employment spell. This allows us to account for job-to-job transitions where, for example, one job ends on Friday and the new job starts on Monday.

[12]We measure spells in calendar weeks. A calendar week starts on Monday and ends on Sunday. If a worker starts and ends a spell in the same calendar week, we code it as duration of 0 weeks. Duration of 1 week means that the spell ended in the calendar week following the calendar week it has started, and so on.

[13]The EM algorithm is commonly used for estimating mixture models, see for example Heckman and Singer (1984) or Lindsay (1995). We modify the usual EM algorithm to our setting. In particular, we use

to be discrete with weight on a finite number $K$ of positive-valued pairs $(\alpha, \beta)$, where $K$ is determined by the Akaike Information Criterion. In the second step, we use data on incomplete spells to bound the distribution of workers with $\alpha < 0$, estimating the set of admissible distributions $\hat{\mathbb{G}}_{\chi^c}$. Details of estimation, including the procedure for choosing the number of types, are in Online Appendix M.

We choose $\bar{T} = 104$ weeks and so our estimates weight worker $i$ by the probability that $T^i - t^{e,i} > 208$, where $T^i$ is the censoring time and $t^{e,i}$ is the duration of the first employment spell; see equation (18).

Appendix B shows that the workers for whom we cannot tell whether $T^i - t^{e,i} > 208$ create a bound on the size of the population described by the estimated distribution $\hat{G}_{\chi^c}$, between 959,623 and 1,124,833 workers, out of a sample of 1,543,609 workers with $T^i > 208$. Most of the workers we lose from the sample had one non-employment spell with $t_1^i \leq 104$, returned to work, and never lost their job again.

The definition of $\chi^c$ in equation (18) clarifies the tradeoff in our choice of $\bar{T}$. If we set a larger value, we have more variation in the duration of completed spells, but a smaller fraction of our population has $T^i - t^{e,i} > 2\bar{T}$.[14] In Online Appendix K we show that our results are robust to setting $\bar{T} = 260$.

## 5.3 Description of Completed Spell Data

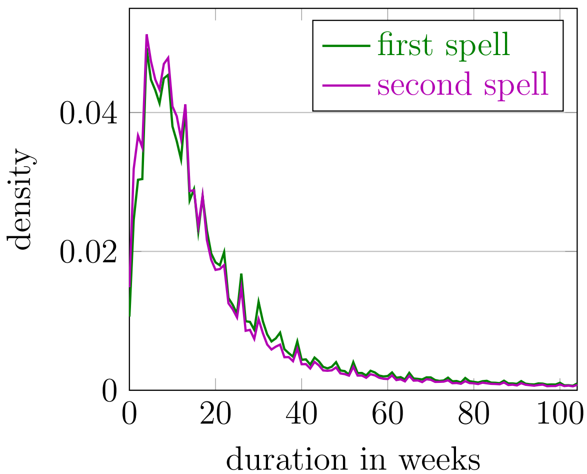
There are 783,810 individuals whose first two non-employment spells are completed in less than 104 weeks. For such workers, the average duration of a non-employment spell is 19 weeks, and the average employment duration between these two spells is 71 weeks. Figure 2 depicts the marginal densities of the duration of the first two completed non-employment spells for these workers. The densities are very similar, rising sharply during the first five weeks, hovering near 4.5 percent for the next seven weeks, and then gradually starting to decline. First spells last two weeks longer than second spells on average, a difference we suppress in our analysis.

The correlation between the duration of the first two spells is 0.24 for these workers. As we discussed in Section 3.3, this correlation is informative about the extent of heterogeneity, since a mixture model with no heterogeneity produces zero correlation, while a model with a degenerate duration for each worker produces a correlation of one. In the Online Appendix,

---

the likelihood of two spells for each type, and we condition on these two spells being in the set $\mathbb{T}^2$ as in equation (14).

[14] Additionally, at short durations, expiration of unemployment benefits at 20, 30, 39, and 52 weeks may influence the transition rate from non-employment to employment, which would mean that our model would be misspecified around those durations. We stress that we do not observe unusual spikes at these durations in our data; see Figure 2.



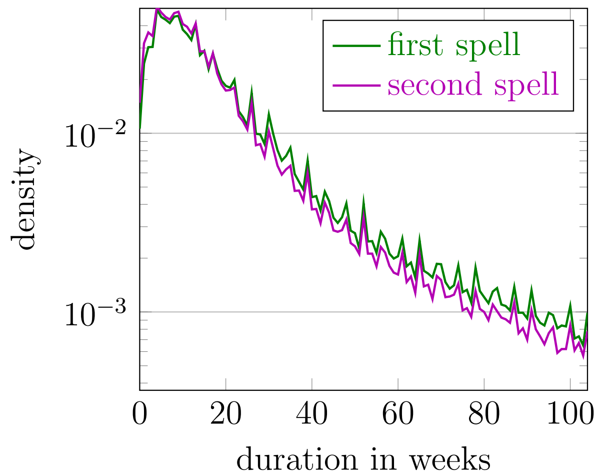




Figure 2: Density of non-employment spell duration. The figures show the density of the duration of the first and second spell in our sample of workers with two completed spells, both with duration less than $\bar{T} = 104$ weeks, and with enough time in the sample, $T^i - t^{e,i} > 2\bar{T}$, as discussed in the text. The two panels show the same data, but the right hand panel has a log scale.

we show additional aspects of the joint distribution. Figure 9 depicts the joint density of the duration of the first two spells, while Figure 10 shows the marginal density of the duration of the second non-employment spell conditional on the duration of the first non-employment spell. Figure 7 shows the hazard of the second spell conditional on the duration of the first spell. We observe that the hazard of workers with a short first spell peaks at a shorter duration than the hazard of workers with a longer first spell, consistent with different workers having different shaped hazards. This is something we will find in our estimated model.

# 6 Results

## 6.1 Estimation of $G^+_{\chi^c}$

We start by estimating the type distribution under the assumption that $\alpha$ is nonnegative, calling our estimate $\hat{G}^+_{\chi^c}$. Our parameter estimate places a positive weight on $K = 22$ different types $(\alpha, \beta)$. Figure 3 shows the weekly hazard for the ten types with the highest shares in $\hat{G}^+_{\chi^c}$, accounting for 99 percent of the population. The figure shows that we have uncovered a lot of heterogeneity, with the shape of hazards differing substantially across types. Compare, for example, the three most common types, depicted in the left panel. The most common type ($\alpha = 0.205$, $\beta = 3.638$) has a higher hazard than the other two

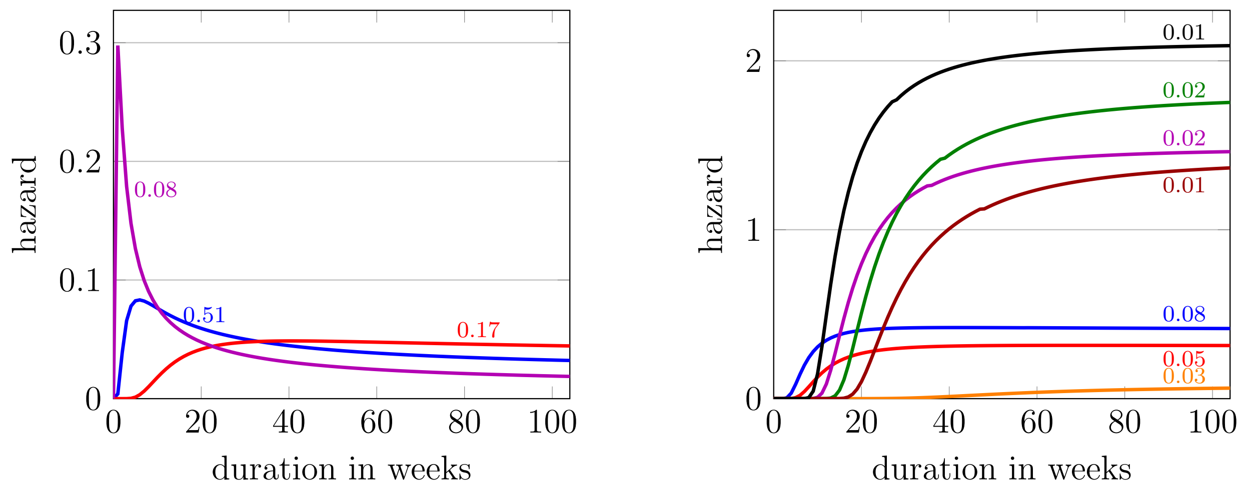



Figure 3: Hazards $h(t;\alpha,\beta)$ of the three types with the highest share in $\hat{G}^+$ (left panel) and next seven highest shares (right panel); numbers indicate the share of each type. Together these account for 99 percent of workers in $\hat{G}^+_{\chi^c}$.

at intermediate durations. The second most common type ($\alpha = 0.264$, $\beta = 8.746$) has the highest hazard at durations exceeding 33 weeks. The third most common type ($\alpha = 0.132$, $\beta = 1.441$) has a very high hazard at short durations but has the lowest of the three hazards after 23 weeks. Thus workers who are likely to find a job at short durations may be unlikely to find one at long durations if they were unsuccessful, while other workers predominately find a job at longer durations. The right panel depicts the next seven most common types. Many of these have a low hazard initially that then suddenly jumps up, resulting in little uncertainty about the duration of their non-employment spells. For example, the standard deviation of log duration is only 0.14 for the type with the highest hazard. Figure 3 contrasts with the MPH model, a common specification in the literature, where all these hazards would be proportional to each other.

Table 1 summarizes our estimates. We report the mean, minimum, and standard deviation of $\alpha$ and $\beta$, as well as the drift and standard deviation of the net benefit from employment relative to the width of the inaction region, $\mu_n/(\bar{\omega} - \underline{\omega}) = \alpha/\beta$ and $\sigma_n/(\bar{\omega} - \underline{\omega}) = 1/\beta$. All these parameters are measured on a weekly time scale. Our estimates uncover a considerable amount of heterogeneity. For example the cross-sectional standard deviation of $\alpha$ and $\beta$ are of the same order as their corresponding cross-sectional means. Moreover, $\alpha$ and $\beta$ are positively correlated in the cross-section, with correlation 0.823.

The last three rows of Table 1 show other moments. Recall that for positive $\alpha$, the expected duration of a completed non-employment spell is $\beta/\alpha$. In the data, the mean non-

| | mean | st. dev | min |
|---|---|---|---|
| $\alpha$ | 0.435 | 0.495 | 0.132 |
| $\beta$ | 7.888 | 8.953 | 1.441 |
| $\mu_n/(\bar{\omega}-\underline{\omega})$ | 0.060 | 0.023 | 0.015 |
| $\sigma_n/(\bar{\omega}-\underline{\omega})$ | 0.235 | 0.163 | 0.013 |
| $\beta/\alpha$ | 20.069 | 10.360 | 9.010 |
| $2/\alpha^2$ | 39.265 | 28.750 | 0.314 |
| $(\beta/\alpha)^3$ | 17,194 | 45,391 | 732 |

Table 1: Summary statistics for $\hat{G}^+_{\chi^c}$. Note that all parameters are measured in weekly time units. See the text for the interpretation of the different moments.

employment duration among workers with two completed spells shorter than 104 weeks is 19 weeks, which is in line with mean of expected duration of estimated types. For $\alpha > 0$, $2/\alpha^2$ is the expected residual duration for a worker with a long current non-employment duration. Our estimates indicate that the expected duration of an "average" worker at the beginning of the non-employment spell is 20 weeks (mean of $\beta/\alpha$), while the expected duration of such worker doubles to 39 weeks if she stays non-employed for a long time (mean of $2/\alpha^2$). Since expected residual duration increases with the current non-employment duration for an "average" worker, we expect that structural duration dependence will be important.

The last row of Table 1 shows the distribution of $(\beta/\alpha)^3$, an object which plays a role in estimating fixed costs of switching from non-employment to employment. We describe in detail how we estimate these costs in Section 6.6.

The smooth red line in Figure 4 shows the fitted marginal distribution of the duration of a non-employment spell conditional on completed duration $t \in [0, 104]$. The model matches the initial increase in the density during the first ten weeks, as well as the gradual decline the subsequent two years. We miss the distribution at long durations above 80 weeks, where there are very few observations, as well as the high-frequency wiggles in the duration distribution.[15]

Finally, we do a good job of matching the joint density of the duration of the first two spells. Let $\hat{\phi}_{\mathbb{T}} \equiv V_{\mathbb{T}}(\hat{G}^+_{\chi^c})$ be the model-implied distribution of two completed spells shorter than 104 weeks, while $\phi_{\mathbb{T}}$ is what we measure. Let

$$\mathcal{F}(\phi_{\mathbb{T}}, \hat{\phi}_{\mathbb{T}}) \equiv 1 - \frac{\text{var}\big(\phi_{\mathbb{T}} - \hat{\phi}_{\mathbb{T}}\big)}{\text{var}(\phi_{\mathbb{T}})}$$

measure the goodness of fit. We find that our model explains 95.1 percent of variation in

[15]The confidence interval for the estimated distribution is very tight, and so we do not show it here. We plot confidence interval once we turn to the decomposition of the hazard.

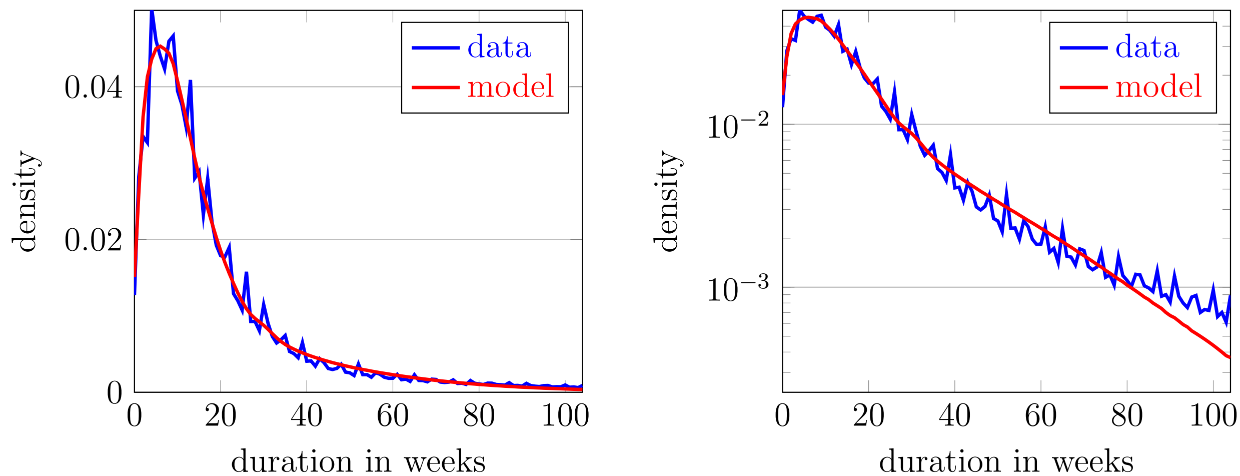



Figure 4: The blue line shows the marginal distribution of the duration of a non-employment spell for the sample of workers with two completed spells, both with duration less than $\bar{T} = 104$ weeks, and with enough time in the sample, $T^i \geq t^{e,i} + 2\bar{T}$, i.e. the average of the two lines in Figure 2. The red line shows the corresponding distribution from the model. The two panels show the same data, but the right hand panel has a log scale.

the joint density and we are thus confident that our model provides a useful tool to analyze the data. Part of the remaining noise has to do with the fact that months play a significant role in the data, but not in our model. Additionally, $\phi_{\mathbb{T}}$ has randomness coming from the fact that the data set is finite.

## 6.2 Estimation of the Set $\mathbb{G}_{\chi^c}$

In the second step, we use data on incomplete spells to estimate the set $\mathbb{G}_{\chi^c}$ defined in equation (17), calling our estimate $\hat{\mathbb{G}}_{\chi^c}$. We first check whether $\hat{G}^+_{\chi^c} \in \hat{\mathbb{G}}_{\chi^c}$. Using the notation from equations (15) and (16), we find that $P_{1,\mathbb{T}}(\hat{G}^+_{\chi^c}) = 0.99$ and $P_{2,\mathbb{T}}(\hat{G}^+_{\chi^c}) = 0.99$, while in the data $\hat{p}_{1,\mathbb{T}} \in [0.76, 0.89]$ and $\hat{p}_{2,\mathbb{T}} = 0.92$. That is, spells lasting longer than 104 weeks are far more common in the data than in the model where all workers have a positive drift in the net benefit to work while non-employed. We hence reject that model, $\hat{G}^+_{\chi^c} \notin \hat{\mathbb{G}}_{\chi^c}$.

We then look at distributions $\hat{G}_{\chi^c}$ where some workers have $\alpha < 0$, estimating $\hat{\mathbb{G}}_{\chi^c}$ using the numerical procedure described in Online Appendix M.6. We find that between 31.2 and 35.6 percent of workers have $\alpha < 0$, where the interval reflects our partial identification result. Allowing for workers with a defective non-employment duration distribution is critical to our hazard decomposition, which we turn to next.

## 6.3 Hazard Decomposition

We now use our estimated set of type distributions to evaluate the importance of heterogeneity in shaping the mixture hazard. We present our results for all distributions in the identified set $\hat{\mathbb{G}}_{\chi^c}$. Take a particular distribution function $\hat{G}_{\chi^c} \in \hat{\mathbb{G}}_{\chi^c}$. Since the hazard for each type $(\alpha, \beta)$ is known from the functional form of the inverse Gaussian distribution, we can construct the mixture hazard $\hat{H}$ using equation (19) and find the structural hazard $\hat{H}^s$ and the heterogeneity effect $\hat{H}^h$ by integrating equation (21). We do this decomposition for all such distributions $\hat{G}_{\chi^c} \in \hat{\mathbb{G}}_{\chi^c}$, which gives us a region for $\hat{H}(t)$, $\hat{H}^s(t)$, and $\hat{H}^h(t)$ for each $t$. We plot these as shaded regions in Figure 5.[16]

The purple region in the left panel of Figure 5 shows the set of estimated mixture hazards $\hat{H}(t)$. The mixture hazard peaks at between 4.5 to 5.0 percent after 9.5 weeks, declines to between 0.8 to 1.0 percent at a year's duration and to between 0.1 and 0.2 percent at two years' duration. In contrast, the blue region shows the corresponding structural hazard $\hat{H}^s(t)$. This increases for about 19 weeks, peaking at between 7.0 and 7.7 percent. It then falls to between 3.9 and 4.6 percent at a year's duration and further declines to between 1.4 and 1.7 percent at two years' duration. The non-employment duration of an individual worker thus has a significant effect on her future prospects for finding a job, but less than the mixture hazard indicates. After two years of non-employment, the structural hazard rate is less than a quarter of what it was at the peak.

The ratio of the structural and mixture hazards, $\hat{H}^h(t) \equiv \hat{H}(t)/\hat{H}^s(t)$, is the heterogeneity effect, shown in the right panel of Figure 5. Recall from equation (22) that the heterogeneity effect is necessarily decreasing in duration, since high hazard workers always find jobs faster than those with low hazards. We find strong evidence of this channel. The heterogeneity effect initially declines sharply, and after half a year of non-employment it is only between 34 to 35 percent of its initial value. Sorting continues at a slower, yet still high, rate thereafter. Even after a year of non-employment, the group of workers is still heterogenous and the heterogeneity effect deteriorates further. After two years, the heterogeneity effect is only 8 to 9 percent as high as at the start of a spell.

The choice of the type distribution from the identified set $\hat{\mathbb{G}}_{\chi^c}$ affects the hazard decomposition for two reasons. First, the mixture hazard directly depends on the sign of $\alpha$. The

[16] The dotted lines in Figure 5 show bootstrapped 95 percent confidence intervals. See Online Appendix N for details on their construction. Due to non-negativity constraints on the parameters, the bootstrap is not guaranteed to give valid confidence interval; see Andrews (2000) and Fang and Santos (2018). Our understanding of the literature is that there is no generally-accepted, computationally-feasible way of constructing valid confidence interval that can be applied to our setting. The applied literature proceeds to use bootstrap in these settings, for example Nevo, Turner and Williams (2016), and we follow that literature. We have also tried two other methods, subsampling and parametric bootstrap, and found a narrow confidence interval in all cases.

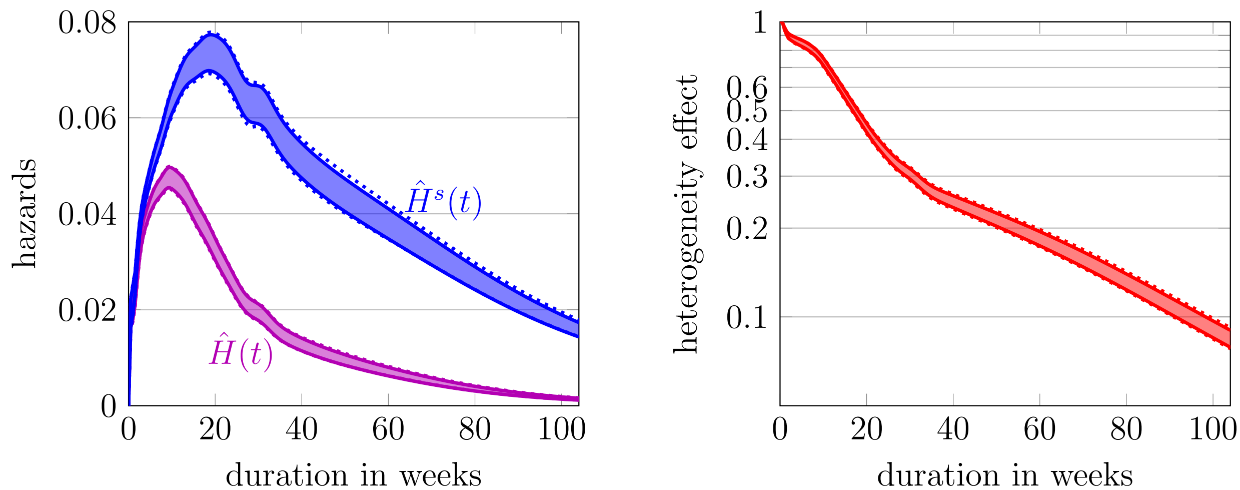



Figure 5: Hazard decomposition under distributions in $\hat{\mathbb{G}}_{\chi^c}$. The purple region in the left figure shows the mixture hazard $\hat{H}(t)$. The blue region in the left figure shows the structural hazard $\hat{H}^s(t)$. The ratio of them is the heterogeneity effect, plotted as the red region in the right figure. The dotted lines show bootstrapped 95% confidence interval.

mixture hazard for a given $\alpha > 0$ is higher than for $-\alpha < 0$, and only the latter converges to zero at long durations. Second, the weight we attribute to a type $(|\alpha|, \beta)$ depends on the sign of $\alpha$. The structural hazard thus tends to be lower for distributions which have a higher share of workers with negative $\alpha$. In general, there is no reason to think that one distribution will attribute a bigger role to heterogeneity than another. Despite this, we find that there is no economically interesting differences in the hazard decomposition across elements of the identified set $\hat{\mathbb{G}}_{\chi^c}$ for our estimated model.

## 6.4 Hazard Decomposition with Observables

Our results indicate that heterogeneity explains an important part of the decline in the job finding probability during the first year of non-employment. We further investigate whether some of this heterogeneity can be attributed to observable characteristics. In Online Appendix G.1, we propose a within-between decomposition of the heterogeneity effect, $H^h(t) = H^b(t)H^w(t)$. The between component $H^b(t)$ captures changes in the shares of workers with different observable characteristics and hazards; and the within component $H^w(t)$ captures the average change in the hazard for workers with different observable characteristics, i.e. it is a weighted average of $H^w_k(t)$, the hazard for workers with observable characteristic $k$ at duration $t$.

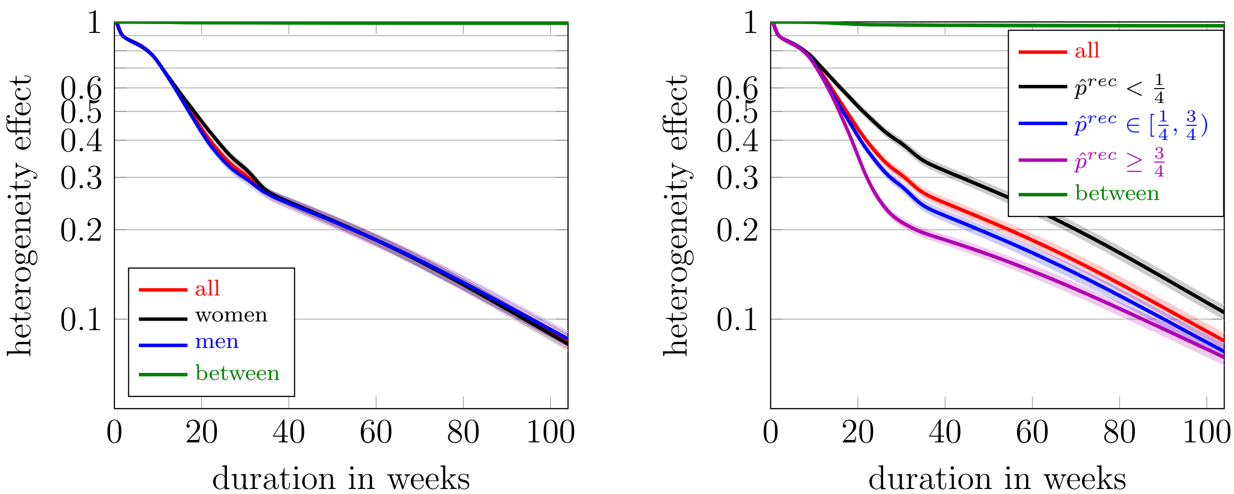



Figure 6: Decomposition of $\hat{H}^h(t)$ (red line) with observable characteristics. The green lines show the between-group heterogeneity effect, $\hat{H}^b(t)$, while other lines show the heterogeneity effect within a group of workers with observable characteristic $k$, $\hat{H}^w_k(t)$. The solid lines show the average of the upper and lower bound of the region, depicted in light colors. The left panel creates groups based on gender, the right panel based on probability of being recalled to the same employer after the first non-employment spell.

In Figure 6, we present hazard decompositions conditional on two different characteristics, gender and probability of being recalled. See Online Appendix G.2 for other characteristics. We again show results for the identified set $\hat{\mathbb{G}}_{\chi^c}$. To make figures legible, we plot the mean value of each component in the decomposition and show the corresponding region in light colors. We break the within component into that for each observable characteristic, $\hat{H}^w_k(t)$, as well as the between component, $\hat{H}^b(t)$.[17]

The left panel in Figure 6 shows the within heterogeneity effect $\hat{H}^w_k(t)$ for men and women, the between heterogeneity effect $\hat{H}^b(t)$, as well as the total heterogeneity effect, $\hat{H}^h(t)$. The lines for men and women are very similar, suggesting that there is a similar amount of heterogeneity within each group of workers. Moreover, between-group heterogeneity $\hat{H}^b(t)$ is essentially constant at 1, suggesting that gender itself does not contribute to dynamic sorting.

The right panel distinguishes workers based on their probability of being recalled back to

[17] For each worker with realized duration $(t_1, t_2)$, we use the distribution $\hat{G}_{\chi^c} \in \hat{\mathbb{G}}_{\chi^c}$ and the model to construct the posterior distribution of $(\alpha, \beta)$, say $\tilde{G}$, from equation (6). The type-contingent distribution $\hat{G}_k$ is then the average of these posterior distributions among all workers with characteristic $k$. This ensures that the decomposition is consistent with the aggregate data.

the same employer.[18] We estimate the probability of being recalled after a non-employment spell, $\hat{p}^{rec}$, as a function of gender, industry dummies, regional dummies, and calendar month dummies for the start of the non-employment spell using a logit specification.[19] We then create three groups based on the predicted recall probability: workers with $\hat{p}^{rec} \leq \underline{p}$, workers with $\underline{p} < \hat{p}^{rec} \leq \bar{p}$, and workers with $\hat{p}^{rec} > \bar{p}$. In the text, we set the thresholds at $\underline{p} = 0.25$ and $\bar{p} = 0.75$, but show results for other thresholds in Online Appendix G.2.

We observe some differences across groups, indicating that the distribution of unobservable characteristics among workers with a high probability of being recalled is different than among those who are unlikely to be recalled. The heterogeneity effect is stronger among those likely to be recalled, especially in the first 25 weeks of non-employment, possibly reflecting the fact that this group disproportionately consists of seasonal workers with different layoff durations. Another explanation is that an employer knows the quality of workers who are (temporarily) displaced and recalls back workers with the highest quality first. We further observe that there is substantial heterogeneity even among workers who are unlikely to be recalled. After one year of non-employment, the change in the heterogeneity effect is similar in all these groups. Between-group heterogeneity again does not contribute to the heterogeneity effect.

The results for other observable characteristics presented in Online Appendix G.2 send a similar message. We do not find evidence of dynamic selection based on observable characteristics. This is in line with the conclusion of other studies, such as Krueger, Cramer and Cho (2014), Schmieder, von Wachter and Bender (2016), and DellaVigna, Lindner, Reizer and Schmieder (2017). On the other hand, we argue that dynamic sorting based on unobservable characteristics is very important for the shape of the mixture hazard. Another conclusion we draw is that dynamic sorting on unobservables appears to be similar conditional on most observable characteristics. Two notable exceptions to this are workers with a high probability of being recalled to the same employer and workers older than 55 (see Figure 11).

## 6.5 Comparison to the Mixed Proportional Hazard Model

A large literature assumes a MPH structure, i.e. that individual hazards have the form $h(t;\zeta) = \zeta\bar{h}(t)$, where $\zeta$ is a fixed worker characteristic and $\bar{h}(t)$ is the unknown baseline hazard of a spell ending at duration $t$. The parameter $\zeta$, which may be unobserved, captures heterogeneity, while the baseline hazard $\bar{h}(t)$ is assumed to be identical across workers.

[18] We thank a referee for suggesting that we use the predicted probability of being recalled to classify workers, rather than the actual realization of returning back to the employer.

[19] Estimating probit instead of logit does not make any difference.

Lancaster (1992) analyzes the MPH model in detail. For a discussion of identification and estimation using multi-spell data, see Honoré (1993), Horowitz and Lee (2004) Alvarez, Borovičková and Shimer (2021).

While the MPH model is a convenient statistical representation of the data, this specification is restrictive in the sense that it has testable implications. In particular, let $H_2(t_2|[\underline{t}_1, \bar{t}_1])$ denote the mixture hazard during the second spell at duration $t_2$ for all workers whose first spell ends during the interval $[\underline{t}_1, \bar{t}_1]$. This is monotonic in first spell duration:

**Proposition 5** *Consider a large population of individuals described by the MPH model. Take two intervals $[\underline{t}_1, \bar{t}_1]$ and $[\underline{t}'_1, \bar{t}'_1]$ with $\underline{t}_1 \geq \underline{t}'_1$ and $\bar{t}_1 \geq \bar{t}'_1$, at least one of which is strict. For fixed $t_2$, $H_2(t_2|[\underline{t}_1, \bar{t}_1]) < H_2(t_2|(\underline{t}'_1, \bar{t}'_1])$.*

The proof is in Online Appendix L. First we prove that, due to dynamic selection, the distribution of $\zeta$ among workers with a shorter $t_1$ first order stochastically dominates the distribution among workers with a longer $t_1$. We then prove that this implies the desired property for the mixture hazard during the second spell. Intuitively, this conditional hazard rate dominance is a consequence of the proportionality of hazards, which in particular implies that all hazards peak at the same duration.

This condition can be tested in the data. We illustrate this with four non-overlapping intervals for the duration of the first spell, 0–4 weeks, 10–14 weeks, 20–24 weeks, 35–39 weeks, and plot the mixture hazard for the second spell for each interval in the left panel of Figure 7. We see a clear violation of monotonicity. For example, the hazard for the second spell for those who first spell lasts 10–14 weeks exceeds the hazard for the second spell for those whose first spell lasts 0–4 weeks at second spell durations 10–25 weeks. The right panel of Figure 7 shows that our estimated model captures the non-monotone conditional hazards, which is perhaps not surprising in light of the fact that our estimated model is very different from the proportional hazard structure; see Figure 3.

## 6.6 Estimated Switching Costs

In Online Appendix H.5, we prove that knowledge of $\alpha$ and $\beta$, together with four other parameters of the model, pins down the magnitude of the fixed costs of switching employment status. Here we use the estimated set $\hat{\mathbb{G}}_{\chi^c}$ as well as other parameter values of the model to find the implied distribution of the fixed costs in the population. For plausible parameter values, the implied fixed costs are small. Given that, our strategy is to choose parameters to make the fixed costs as large as possible while still staying within a range that can be supported empirically.

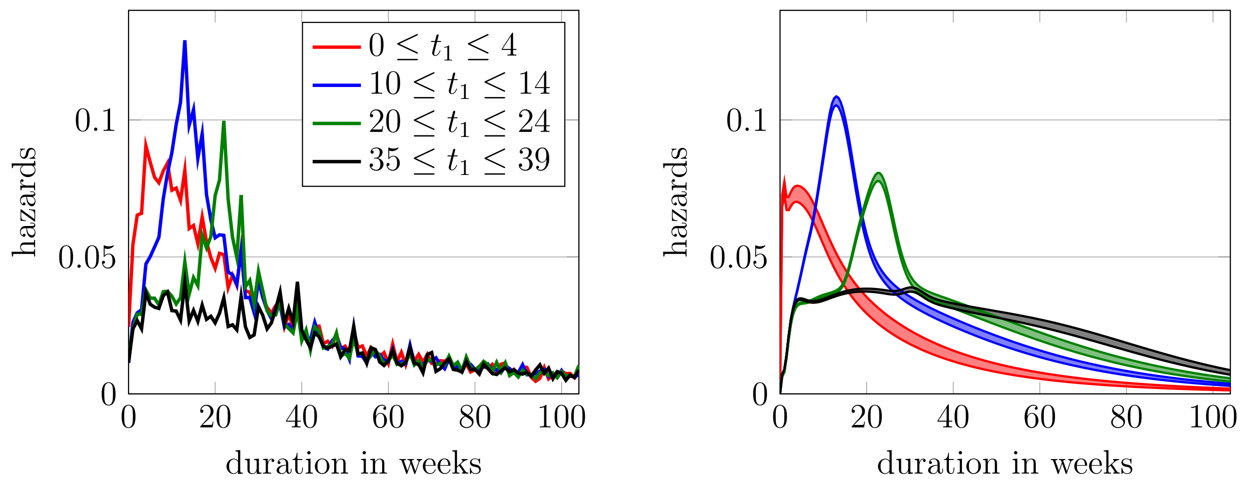



Figure 7: Mixture hazard for the second spell conditional on the duration of the first spell. The left panel shows the conditional hazard in the data for different values of the duration of the first spell. The right panel shows the conditional hazard rates under distributions $\hat{\mathbb{G}}_{\chi^c}$, for the same intervals of $t_1$ as in the left panel.

We assume that utility from non-employment is constant and equal to 1, i.e. $b(t) = 0$ for all $t$. Rewriting equation (O.37) in terms of $\alpha$ and $\beta$, we get an expression for the sum of fixed costs without reference to $\bar{\omega} - \underline{\omega}$ or $\sigma^2_{w,n}$:

$$\psi_e + \psi_n \approx \frac{(\mu_{w,e} + \sqrt{\mu^2_{w,e} + 2r\sigma^2_{w,e}})(-\alpha + \sqrt{\alpha^2 + 2r})\beta^3 \mu^2_{w,n}}{12\, r\, \alpha^2\, \sigma^2_{w,e}}. \tag{23}$$

To back out the magnitude of switching costs, we need to choose values for parameters $\mu_{w,n}$, $\mu_{w,e}$, $\sigma_{w,e}$, and $r$.

Equation (23) implies that for given value of $\alpha$ and $\beta$, higher $\mu_{w,e}$ and $|\mu_{w,n}|$ increase the implied fixed costs, while higher $\sigma_{w,e}$ and $r$ reduce the implied fixed cost. We keep that in mind and calibrate these parameters to find an upper bound on the fixed costs. First, we set the drift in employed workers' wages at $\mu_{w,e} = 0.01$ at an annual frequency. Estimates of the average wage growth of employed workers are often higher than one percent, but this is for workers who stay employed, a selected sample. The parameter $\mu_{w,e}$ governs wage growth for all workers without selection, and thus we view $\mu_{w,e} = 0.01$ as a large number. We set the standard deviation of log wages at $\sigma_{w,e} = 0.05$, again at an annual frequency. This is lower than typical estimates in the literature, which are closer to ten percent.

We cannot observe the drift of latent wages when non-employed, $\mu_{w,n}$, but we can infer its value relative to $\mu_{w,e}$ from the duration of completed employment and non-employment

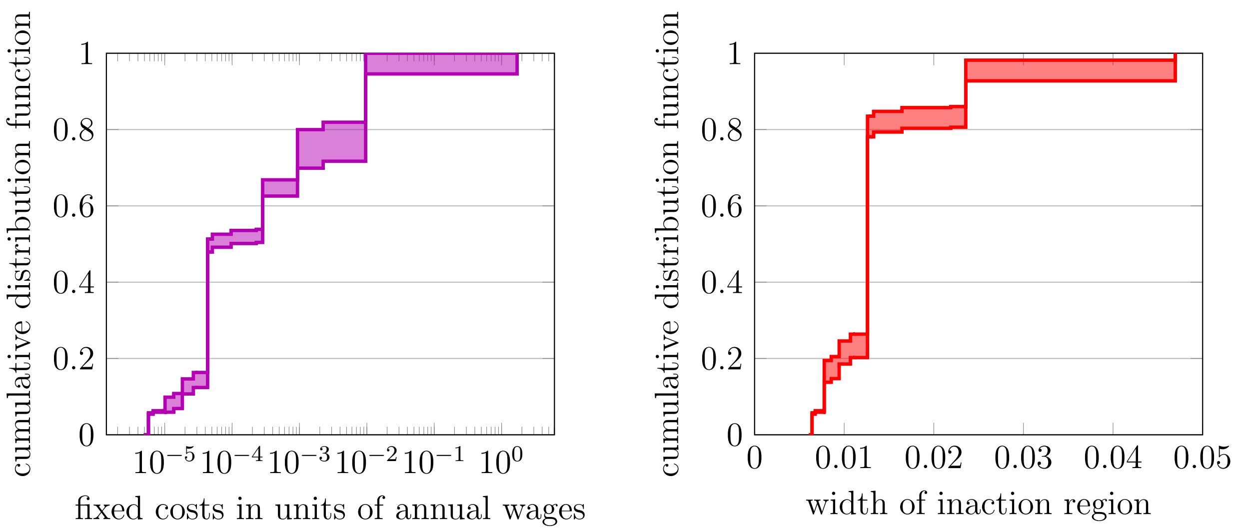



Figure 8: Distribution of estimated fixed costs and the width of inaction region. Fixed costs are expressed in units of annual wages, the width of inaction region is in log differences of wages.

spells. The expected duration of *completed* employment and non-employment spells are $(\bar{\omega} - \underline{\omega})/|\mu_{w,e}|$ and $(\bar{\omega} - \underline{\omega})/|\mu_{w,n}|$, respectively, and thus $|\mu_{w,n}|/|\mu_{w,e}|$ determines their relative expected duration. In our sample, the average duration of completed non-employment spells is 19.0 weeks, while the average duration of completed employment spells is 71 weeks, implying that $|\mu_{w,n}| = 3.7|\mu_{w,e}|$.

Finally, we choose a low value for $r$. Since workers in the model are infinitely lived, we think of this as the sum of workers' discount rate and their death probability. A lower bound on this is 0.02, consistent with no discounting and a fifty year working lifetime.

Given this calibration, we estimate the distribution of fixed costs under type distributions in $\hat{\mathbb{G}}_{\chi^c}$. The left panel of Figure 8 shows the region for the cumulative distribution function of fixed costs. Since we choose the parameter values to make fixed costs as high as possible, we again focus on the upper bound of the fixed costs distribution. There, the median value of the fixed costs are only 0.004–0.010 percent of the annual non-employment flow value, or about 4.8–10 minutes, assuming a 2,000 hours of work per year. The reported intervals are due to partial identification. The costs vary across types, ranging from 1.2–2.4 minutes in the bottom decile to 4.5–19.2 hours of time in the top decile. Silva and Toledo (2009), Table 1, row (6), report still higher costs, about 13 hours per new hire. Thus, our median costs are an order of magnitude lower.

To understand which moments of the estimated distribution drive the fixed costs distribution, it is useful to approximate equation (23) for a very small $r$. We get

$$\psi_e + \psi_n \sim \begin{cases} \frac{\mu_{w,e}\,\mu_{w,n}^2}{6\,\sigma_{w,e}^2} \frac{\beta^3}{|\alpha|^3} & \text{if } \alpha > 0 \\ \frac{\mu_{w,e}\,\mu_{w,n}^2}{3\,\sigma_{w,e}^2} \frac{\beta^3}{|\alpha|^3} \frac{\alpha^2}{r} & \text{if } \alpha < 0 \end{cases} \tag{24}$$

where $\sim$ means that the ratio of the two functions converges to one as $r$ converges to 0. Thus, for $\alpha > 0$, the distribution of $(\beta/\alpha)^3$ drives the results, while changes in the value of the parameters $\mu_{w,e}$, $\mu_{w,n}$, $\sigma_{w,e}$, and $r$ affect costs only proportionally.

Even though the magnitudes are very small, strictly positive switching costs are important for our results. If switching cost were zero for someone, their region of inaction would be degenerate. As the switching costs converge to zero we prove in Online Appendix I that the mean duration of spells in the interval $[1, 104]$ cannot exceed $\sqrt{104} \approx 10.2$ weeks. In the data, we find that the mean duration of these spells is 19 weeks, which requires that there be some switching cost.

Previous work by Mankiw (1985), Dixit (1991), Abel and Eberly (1994), and others has shown that even small fixed costs can generate large regions of inaction. To see the size of the region of inaction in our model, recall that $\alpha = \mu_{w,n}/\sigma_{w,n}$ and $\beta = (\bar{\omega} - \underline{\omega})/\sigma_{w,n}$, and hence $\bar{\omega} - \underline{\omega} = (\beta/|\alpha|)|\mu_{w,n}|$. This formula is intuitive: the distance between the barriers of the inaction region is the product of time (expected non-employment duration $\beta/|\alpha|$) and velocity (drift of the underlying Brownian motion $|\mu_{w,n}|$). The right panel of Figure 8 shows the distribution of the width of inaction region for distributions in $\hat{\mathbb{G}}_{\chi^c}$. The mean width is 0.014–0.015, and the median is 0.013. That is, the median worker who has just started working will quit if she experiences a 1.3 percent decline in her wage, holding fixed the value of non-employment. A similar wage increase will induce her to return to work.

We are unaware of other papers that study the cost of switching between employment and non-employment at the level of an individual worker. In other areas, empirical results on the size of fixed costs are mixed. Cooper and Haltiwanger (2006) find a large fixed cost of capital adjustment, around 4 percent of the average plant-level capital stock. Nakamura and Steinsson (2010) estimate a multisector model of menu costs and find that the cost of adjusting prices is less than 1 percent of firms' annual revenue. In a model of house selling, Merlo, Ortalo-Magné and Rust (2015) find a very small fixed cost of changing the listing price of a house, around 0.01 percent of the house value.

# 7 Conclusion

In this paper we propose a model where optimal behaviour gives rise to infrequent switches between employment and non-employment. We allow for an unrestricted distribution of the fundamental parameters for each worker and discuss how to use duration data to recover the distribution of related reduced-form parameters.

On the empirical side, we estimate substantial heterogeneity in the job finding hazard in Austria, as summarized by our decomposition of the mixture hazard rate. Some workers typically take a long time to find a job, but face little risk of permanent joblessness. Others typically find a job quickly, but when they do not, they face a heightened risk of never returning to work. This heterogeneity is important for the impact of unemployment benefits on unemployment duration, as well as for the distributional impact of the unemployment benefit system. For example, Shimer and Werning (2006) show that how we model heterogeneity and structural duration dependence affects optimal unemployment insurance, e.g. whether unemployment benefits should increase or decrease during an unemployment spell.

On the theoretical side, we have concentrated on identification of such model, and on its characterization and comparisons with other duration models. We leave for future work the development of estimation and inference with a continuum of types, something that our identification result permits. A second topic for future research is to extend the model, and the identification argument, to allow for additional realistic features such as aggregate fluctuations, finite duration unemployment benefits, and time varying individual parameters.

# 8 Data Availability

The replication package is available at `https://doi.org/10.5281/zenodo.8370851`.

# References

**Aalen, Odd O. and Håkon K. Gjessing**, "Understanding the Shape of the Hazard Rate: A Process Point of View (with comments and a rejoinder by the authors)," *Statistical Science*, 2001, *16* (1), 1–22.

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

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

**Ahn, Hie Joo and James D. Hamilton**, "Heterogeneity and Unemployment Dynamics," *Journal of Business & Economic Statistics*, 2020, *38* (3), 554–569.

**Al-Hussaini, E. and K. Ahmad**, "On the Identifiability of Finite Mixtures of Distributions (Corresp.)," *IEEE Transactions on Information Theory*, 1981, *27* (5), 664–668.

**Alvarez, Fernando, Katarína Borovičková, and Robert Shimer**, "Consistent Evidence on Duration Dependence of Price Changes," Working Paper 29112 July 2021.

**Andrews, Donald W. K.**, "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 2000, *68* (2), 399–405.

**Bentolila, Samuel and Giuseppe Bertola**, "Firing Costs and Labour Demand: How Bad is Eurosclerosis?," *Review of Economic Studies*, 1990, *57* (3), 381–402.

**Buhai, I. Sebastian and Coen N. Teulings**, "Tenure Profiles and Efficient Separation in a Stochastic Productivity Model," *Journal of Business & Economic Statistics*, 2014, *32* (2), 245–258.

**Canay, Ivan A., Andres Santos, and Azeem M. Shaikh**, "On the Testability of Identification in some Nonparametric Models with Endogeneity," *Econometrica*, 2013, *81* (6), 2535–2559.

**Chhikara, R.S. and J.L. Folks**, "The Inverse Gaussian Distribution as a Lifetime Model," *Technometrics*, 1977, *19* (4), 461–468.

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

**Cooper, Russell W and John C Haltiwanger**, "On the Nature of Capital Adjustment Costs," *Review of Economic Studies*, 2006, *73* (3), 611–633.

**De Paula, Aureo**, "Inference in a Synchronization Game with Social Interactions," *Journal of Econometrics*, 2009, *148* (1), 56–71.

**DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder**, "Reference-Dependent Job Search: Evidence from Hungary," *Quarterly Journal of Economics*, November 2017, *132* (4), 1969–2018.

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

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

**Fang, Zheng and Andres Santos**, "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 2018, *86* (1), 377–412.

**Feller, William**, *An Introduction to Probability Theory and its Applications. Vol. II*, second ed., New York: John Wiley & Sons Inc., 1966.

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

**Grossman, Sanford J. and Guy Laroque**, "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 1990, *58* (1), 25–51.

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

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

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

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

**Horowitz, Joel L. and Sokbae Lee**, "Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects," *Journal of Econometrics*, 2004, *119* (1), 155–198.

**Karatzas, I. and S. E. Shreve**, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1998.

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

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

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

— , "The Econometric Analysis of Transition Data," *Cambridge Books*, 1992.

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

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

— **and** — , "Proportional Hazards and Threshold Regression: Their Theoretical and Practical Connections," *Lifetime data analysis*, 2010, *16* (2), 196–214.

**Lindsay, Bruce G.**, "Mixture Models: Theory, Geometry and Applications," *NSF-CBMS Regional Conference Series in Probability and Statistics*, 1995, *5*, 1–163.

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

**Mankiw, N Gregory**, "Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly," *Quarterly Journal of Economics*, 1985, *100* (2), 529–537.

**Merlo, Antonio, François Ortalo-Magné, and John Rust**, "The Home Selling Problem: Theory and Evidence," *International Economic Review*, 2015, *56* (2), 457–484.

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

**Nevo, Aviv, John L. Turner, and Jonathan W. Williams**, "Usage-Based Pricing and Demand for Residential Broadband," *Econometrica*, 2016, *84* (2), 411–443.

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

**Newey, Whitney K. and James L. Powell**, "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 2003, *71* (5), 1565–1578.

**Petersen, L. C.**, "On the Relation between the Multidimensional Moment Problem and the One-Dimensional Moment Problem," *Mathematica Scandinavica*, June 1982, *51*, 361–366.

**Schmieder, Johannes F., Till von Wachter, and Stefan Bender**, "The Effect of Unemployment Benefits and Nonemployment Durations on Wages," *American Economic Review*, 2016, *106* (3), 739–777.

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

— **and Iván Werning**, "On the Optimal Timing of Benefits with Heterogeneous Workers and Human Capital Depreciation," April 2006. Mimeo.

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

**Stokey, Nancy L.**, *The Economics of Inaction*, Princeton University Press, 2008.

**Whitmore, G.A.**, "Discussion of the Paper by Professor Folks and Dr Chhikara," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1978, *40* (3), 285–286.

— , "An Inverse Gaussian Model for Labour Turnover," *Journal of the Royal Statistical Society: Series A (General)*, 1979, *142* (4), 468–478.

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

# Appendix

## A Proof of Identification

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

**Lemma 1** *Let $m$ be a nonnegative integer and $i = 0, \dots, m$. The partial derivative of the product of two inverse Gaussian distributions at $(t_1, t_2)$ is:*

$$\frac{\partial^m\big(f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta)\big)}{\partial t_1^i\, \partial t_2^{m-i}} = f(t_1;\alpha,\beta)\, f(t_2;\alpha,\beta) \left(\sum_{r,s=0}^{r+s\leq m} \kappa_{r,s}(t_1,t_2;i,m-i)\alpha^{2r}\beta^{2s}\right) \tag{25}$$

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

$$\kappa_{r,s}(t_1,t_2;i,m-i) = \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \theta_{k,\ell,r,s}(i,m-i) t_1^{-k}\, t_2^{-\ell}, \tag{26}$$

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

**Proof of Lemma 1.** The lemma holds trivially when $m = i = 0$, with $\kappa_{0,0}(t_1, t_2, 0, 0) = 1$. We now proceed by induction. Assume equation (25) holds for some $m \geq 0$ and all $i \in \{0, \dots, m\}$. We first prove that it holds for $m+1$ and all $i+1 \in \{1, \dots, m+1\}$, then verify that it also holds for $i = 0$. We start by differentiating the key equation:

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

or

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

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

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

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

for $m>0$ and $i\in\{0,\dots,m\}$. Let $B$ represent a bounded, non-empty open neighborhood of $(t_1,t_2)$ and let $\bar{B}$ denote its closure. Assume that there are no points of the form $(t,t)$, $(t_1,0)$, or $(0,t_2)$ in $\bar{B}$. In order to apply Leibniz's rule, we must check two conditions:

1. The partial derivative $\partial^m\big(f(t_1;\alpha,\beta)\,f(t_2;\alpha,\beta)\big)/\partial t_1^i\,\partial t_2^{m-i}$ exists and is a continuous function of $(t_1',t_2')$ for every $(t_1',t_2')\in B$ and $G$-almost every $(\alpha,\beta)$; and
2. There is a $G-$integrable function $h_{i,m-i}:\mathbb{R}_+^2\to\mathbb{R}_+$, i.e. a function satisfying

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

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

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

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

$$h_{i,m-i} \equiv \max_{(t_1,t_2)\in\bar{B}} \sum_{r,s=0}^{r+s\leq m} \sum_{k=0}^{2i} \sum_{\ell=0}^{2(m-i)} \frac{\left|\theta_{k,\ell,r,s}(i,m-i)\right|}{2\pi} t_1^{-k-\frac{3}{2}} t_2^{-\ell-\frac{3}{2}} \left(\frac{r+s+1}{\tau(t_1,t_2)}\right)^{r+s+1} e^{-(r+s+1)}, \tag{27}$$

where

$$\tau(t_1,t_2) = \frac{(t_1-t_2)^2}{2\big(t_1(1+t_2)^2+t_2(1+t_1)^2\big)}, \tag{28}$$

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

We now prove that $h_{i,m-i}$ is an upper bound on the magnitude of the partial derivative. Using Lemma 1, the partial derivative is the product of a polynomial function and an exponential function:

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

Only the constant terms $\theta$ may be negative.

To bound the partial derivative, first note that for any nonnegative numbers $\beta$, $r$, and $s$, and any $\alpha$,

$$(|\alpha|+\beta)^{2(r+s+1)} \geq |\alpha|^{2r}\beta^{2(s+1)} = \alpha^{2r}\beta^{2(s+1)} \tag{29}$$

as directly follows from the binomial formula, because the right hand side is smaller than one of the terms in the expansion, and all other terms are positive. Next note that

$$\exp\big(-(|\alpha|+\beta)^2\tau(t_1,t_2)\big) \geq \exp\left(-\frac{(\alpha t_1-\beta)^2}{2t_1} - \frac{(\alpha t_2-\beta)^2}{2t_2}\right) \tag{30}$$

This can be shown in two steps. For $\alpha>0$, we establish that

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

by finding a maximum of the right hand side of (30) with respect to $\alpha,\beta$ subject to the constraint that $\alpha + \beta = K$ for some $K > 0$. In the second step, we establish that

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

which follows from the fact that for $\beta \geq 0$ and $t > 0$, it holds $(|\alpha|t - \beta)^2 \leq (\alpha t - \beta)^2$. These two steps together imply (30).

Next, consider the function $a^x \exp(-ay)$ for $a$ and $x$ nonnegative and $y$ strictly positive. This is a single-peaked function of $a$ for fixed $x$ and $y$, achieving its maximum value at $a = x/y$. Letting $(|\alpha| + \beta)^2$ play the role of $a$, this implies in particular that

$$\left(\frac{r+s+1}{\tau(t_1,t_2)}\right)^{r+s+1} e^{-(r+s+1)} \geq (|\alpha| + \beta)^{2(r+s+1)} \exp\left(-(|\alpha| + \beta)^2 \tau(t_1,t_2)\right) \tag{31}$$

for all $\alpha$ and all nonnegative $r$, $s$, and $\beta$, as long as $\tau(t_1,t_2) \neq 0$, i.e. $t_1 \neq t_2$. Finally, combine inequalities (29)–(31) to verify the bound on the partial derivative,

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

where $h_{i,m-i}$ is defined in equation (27). ■

**Proof of Proposition 2.** Start with $m = 1$. Using the functional form of $f(t;\alpha,\beta)$ in equation (4), the partial derivatives satisfy

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

or

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

where

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

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

For higher moments, the approach is conceptually unchanged. First express the $(i,j)^{th}$

partial derivatives of $\phi(t_1,t_2)$ as

$$\begin{aligned}\frac{2^{i+j}t_1^{2i}t_2^{2j}}{\phi(t_1,t_2)}\frac{\partial^{i+j}\phi(t_1,t_2)}{\partial t_1^i \partial t_2^j} &= \mathbb{E}\big((\beta^2-\alpha^2 t_1^2)^i(\beta^2-\alpha^2 t_2^2)^j|t_1,t_2\big)+v_{ij}(t_1,t_2)\\ &= \sum_{x=0}^{i+j}\sum_{y=\max\{0,x-j\}}^{\min\{x,i\}}\frac{i!j!(-t_1)^y(-t_2)^{x-y}\mathbb{E}(\alpha^{2x}\beta^{2(i+j-x)}|t_1,t_2)}{y!(x-y)!(i-y)!(j-x+y)!}+v_{ij}(t_1,t_2), \quad (32)\end{aligned}$$

where $v_{ij}$ depends only on lower moments of the conditional distribution. The first line can be established by induction. Express $\frac{\partial^{i+j}\phi(t_1,t_2)}{\partial t_1^i \partial t_2^j}$ from the first line and differentiate with respect to $t_1$. One can realize that all terms except one contain conditional expected moments of order lower than $i+j$ and thus could be grouped into the term $v_{i+1,j}$. The only term of order $m+1$ has a form $\mathbb{E}\big((\beta^2-\alpha^2 t_1^2)^{i+1}(\beta^2-\alpha^2 t_2^2)^j|t_1,t_2\big)$ which follows directly from the derivative of $f(t_1;\alpha,\beta)$ with respect to $t_1$. The second line of (32) follows from the first by expanding the power functions.

Now let $i=\{0,\dots,m\}$ and $j=m-i$. As we vary $i$, equation (32) gives a system of $m+1$ equations in the $m+1$ $m^{th}$ moments of the joint distribution of $\alpha^2$ and $\beta^2$ among workers who find jobs at durations $(t_1,t_2)$, as well as lower moments of the joint distribution. These functions are linearly independent, which we show by expressing them using an LU decomposition:

$$\begin{pmatrix} \frac{2^m t_1^{2m}}{\phi(t_1,t_2)}\frac{\partial^m \phi(t_1,t_2)}{\partial t_1^m} \\ \frac{2^m t_1^{2(m-1)}t_2^2}{\phi(t_1,t_2)}\frac{\partial^m \phi(t_1,t_2)}{\partial t_1^{m-1}\partial t_2} \\ \frac{2^m t_1^{2(m-2)}t_2^4}{\phi(t_1,t_2)}\frac{\partial^m \phi(t_1,t_2)}{\partial t_1^{m-2}\partial t_2^2} \\ \vdots \\ \frac{2^m t_2^{2m}}{\phi(t_1,t_2)}\frac{\partial^m \phi(t_1,t_2)}{\partial t_2^m}\end{pmatrix} = L(t_1,t_2)\cdot U(t_1,t_2)\cdot \begin{pmatrix}\mathbb{E}(\alpha^{2m}|t_1,t_2)\\ \mathbb{E}(\alpha^{2(m-1)}\beta^2|t_1,t_2)\\ \mathbb{E}(\alpha^{2(m-2)}\beta^4|t_1,t_2)\\ \vdots \\ \mathbb{E}(\beta^{2m}|t_1,t_2)\end{pmatrix}+v_m(t_1,t_2), \quad (33)$$

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

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

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

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

for $0\le i\le j\le m$ and $U_{ij}(t_1,t_2)=0$ for $0\le j<i\le m$; and $v_m(t_1,t_2)$ is a vector that depends only on $(m-1)^{st}$ and lower moments of the joint distribution, each of which we

have found in previous steps.[20] It is easy to verify that the diagonal elements of $L$ and $U$ are nonzero if and only if $t_1 \neq t_2$. This proves that the $m^{th}$ moments of the joint distribution are uniquely determined by the $m^{th}$ and lower partial derivatives. The result follows by induction. ■

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

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

$$M_m = \max_{i=0,\ldots,m} \mathbb{E}(\alpha^{2i}\beta^{2(m-i)}). \tag{34}$$

*Assume that*

$$\lim_{m\to\infty} \frac{[M_m]^{\frac{1}{m}}}{m} = \lambda < \infty. \tag{35}$$

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

**Proof of Lemma 2.** We proceed in two steps. In the first step, we recall sufficient conditions in the one-dimensional Stieltjes moment problem for a non-negative variable. In the second step, we combine it with the Cramér-Wold theorem which states that the distribution of a random vector is determined by all its one-dimensional projections.[21]

Let's start with the first step and recall sufficient conditions for uniqueness in the Stieltjes moment problem. For a non-negative random variable $u \in \mathbb{R}^+$, its distribution is uniquely determined by its moments $\{\mathbb{E}[u^m]\}_{m=1}^{\infty}$ if the following condition holds:

$$\limsup_{m\to\infty} \frac{(\mathbb{E}[u^m])^{\frac{1}{m}}}{m} \equiv \lambda' < \infty, \tag{36}$$

as shown in the Appendix of Feller (1966) chapter XV.4.

Before we move to the second step, define variables $a \equiv \alpha^2, b \equiv \beta^2$ and denote their distribution $\check{G}(a, b)$. Since $\alpha, \beta$ are nonnegative, the mapping from $(\alpha, \beta)$ to $(a, b)$ is one-to-

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

[21]One can prove this lemma by applying a more elegant yet more abstract result in Petersen (1982). We

one and onto. Hence, the distribution $G(\alpha,\beta)$ uniquely determines $\check{G}(a,b)$ and vice-versa. Also, define $\check{M}_m$ as

$$\check{M}_m = \max_{i=0,\ldots,m} \mathbb{E}(a^i b^{m-i}), \tag{37}$$

and observe that $\check{M}_m = M_m$ for all $m = 0, 1, \ldots$

The Cramér-Wold theorem states that the distribution of a random vector, say $(a,b)$, is determined by all its one-dimensional projections. In particular, the distribution of the sequence of random vectors $(a_m, b_m)$ converges to the distribution of the random vector $(a,b)$ if and only if the distribution of the scalar $x_1 a_m + x_2 b_m$ converges to the distribution of the scalar $x_1 a + x_2 b$ for all vectors $(x_1, x_2) \in \mathbb{R}^2$.

We take a degenerate sequence $(a_m, b_m) = (a,b)$ for all $m \geq 0$ and verify that for any $(x_1, x_2)$ the distribution of $x_1 a + x_2 b$ is determined by its moments. For this we check the condition in equation (35) for $u(x) \equiv x_1 a + x_2 b$. We note that

$$\begin{aligned}
\mathbb{E}\left[u(x)^m\right] = \mathbb{E}\left[(x_1 a + x_2 b)^m\right] &= \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} x_1^i x_2^{m-i} \mathbb{E}\left[a^i b^{m-i}\right] \\
&\leq \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} |x_1|^i |x_2|^{m-i} \mathbb{E}\left[a^i b^{m-i}\right] \leq \check{M}_m \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} |x_1|^i \, |x_2|^{m-i} \\
&= \check{M}_m (|x_1| + |x_2|)^m,
\end{aligned}$$

where we use that $(a,b)$ are non-negative random variables and $\check{M}_m$ is defined in equation (37). Since $\check{M}_m = M_m$ for each $m = 0, 1, \ldots$, then assumption (35) implies that sufficient conditions in Stieljes's moment problem (36) are satisfied, and hence the distribution of each linear combination $x_1 a + x_2 b$ is determined. Hence, the joint distribution of $(a,b)$ is determined and so is the distribution of $(\alpha,\beta)$. ■

**Proof of Proposition 3.** The structure of the proof is as follows. We first find an upper bound for conditional moments $\mathbb{E}(\alpha^{2i}\beta^{2(m-i)}|t_1, t_2)$. We then apply of Lemma 2 to obtain the main result.

For the first step, we write the conditional moments as

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

where

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

thank our referee for pointing this out.

Using the definition of $f$, we have

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

where $\tau(t_1,t_2)$ is defined in equation (28) and the inequality uses the same steps as the proof of Proposition 1 to bound the function. In the language of Lemma 2, this implies

$$M_m = \frac{((m+1)/\tau(t_1,t_2))^{m+1}\exp(-(m+1))}{2\pi t_1^{3/2}t_2^{3/2}\int f(t_1;\alpha,\beta)f(t_2;\alpha,\beta)dG(\alpha,\beta)}. \tag{38}$$

We use this to verify condition (35) in Lemma 2.

Taking the log transformation of $(1/m)\,(M_m)^{1/m}$ and using the expression (38) we get:

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

where $\varphi$ is independent of $m$. We argue that the limit of this expression as $m\to\infty$ is at most $-\log(\tau(t_1,t_2))-1$. To see this, use that $\log(1+m)\leq\log(m)+1/m$ in the above expression,

$$\begin{aligned} \log\left(\frac{(M_m)^{\frac{1}{m}}}{m}\right) &\leq \frac{1}{m}\varphi(t_1,t_2)-\frac{m+1}{m}\log\tau(t_1,t_2) \\ &\quad +\frac{m+1}{m}\left(\log m+\frac{1}{m}\right)-\frac{m+1}{m}-\log m \\ &= \frac{1}{m}\varphi(t_1,t_2)-\frac{m+1}{m}\log\tau(t_1,t_2)+\frac{1}{m}\log m+\frac{m+1}{m^2}-\frac{m+1}{m} \\ \lim_{m\to\infty}\log\left(\frac{(M_m)^{\frac{1}{m}}}{m}\right) &\leq -\log\tau(t_1,t_2)-1. \end{aligned}$$

Hence, $\lim_{m\to\infty}(1/m)(M_m)^{1/m} < \exp(-\log\tau(t_1,t_2)-1)<\infty$, the desired result. ■

**Proof of Proposition 4.** "If" part: Let $\gamma_1,\gamma_2,G^+$ satisfy (9). We show that $G_1$ and $G_2$ generate the same $\phi_{\mathbb{T}}(t_1,t_2)$. To simplify the notation, denote the denominator of (9) as $K_i$,

$$K_i = \int\big(\gamma_i(\alpha',\beta')+e^{4\alpha'\beta'}(1-\gamma_i(\alpha',\beta'))\big)dG^+(\alpha',\beta'),$$

and the numerator of (5) as $Z_i(t_1, t_2)$ for some $t_1, t_2$:

$$Z_i(t_1, t_2) = \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG_i(\alpha, \beta)$$

Then we get

$$\begin{aligned}
K_i Z_i(t_1, t_2) =& K_i \int f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG_i(\alpha, \beta) \\
=& \int_{\alpha>0} f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \gamma_i(\alpha, \beta) dG^+(\alpha, \beta \\
& \quad + \int_{\alpha<0} f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) e^{4|\alpha|\beta} (1 - \gamma_i(|\alpha|, \beta)) dG^+(|\alpha|, \beta) \\
& \quad\quad + f(t_1; 0, \beta) f(t_2; 0, \beta) dG^+(0, \beta) \\
=& \int_{\alpha>0} f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) \gamma_i(\alpha, \beta) dG^+(\alpha, \beta) \\
& \quad + \int_{\alpha<0} e^{-4|\alpha|\beta} f(t_1; |\alpha|, \beta) f(t_2; |\alpha|, \beta) e^{4|\alpha|\beta} (1 - \gamma_i(|\alpha|, \beta)) dG^+(|\alpha|, \beta) \\
& \quad\quad + f(t_1; 0, \beta) f(t_2; 0, \beta) dG^+(0, \beta) \\
=& \int_{\alpha \geq 0} f(t_1; \alpha, \beta) f(t_2; \alpha, \beta) dG^+(\alpha, \beta),
\end{aligned}$$

where the first equation is the definition of $Z$, the second uses equation (9) to replace $G_i$, the third uses the property of $f$ noted in equation (8), and the fourth combines terms. This proves that $Z_i(t_1, t_2) K_i$ depends only on $dG^+$ but not $\gamma_i$ itself. It then follows the denominator of (5) is proportional to $K_i$ but otherwise does not depend on $\gamma_i$. Therefore, $G_1$ and $G_2$ generate the same $\phi_{\mathbb{T}}(t_1, t_2)$, and hence $V_{\mathbb{T}}(G_1) = V_{\mathbb{T}}(G_2)$.

"Only if" part. Suppose that $V_{\mathbb{T}}(G_1) = V_{\mathbb{T}}(G_2)$. We will find $\gamma_i$ and $dG^+$ such that equation (9) holds. For $\alpha > 0$ and $\beta \geq 0$ such that $dG_i(\alpha, \beta) > 0$ or $dG_i(-\alpha, \beta) > 0$, define

$$\gamma_i(\alpha, \beta) = \frac{dG_i(\alpha, \beta)}{e^{-4\alpha\beta} dG_i(-\alpha, \beta) + dG_i(\alpha, \beta)} \in [0, 1],$$

and

$$dG_i^+(\alpha, \beta) = K_i \frac{dG_i(\alpha, \beta)}{\gamma_i(\alpha, \beta)} = K_i \left( e^{-4\alpha\beta} dG_i(-\alpha, \beta) + dG_i(\alpha, \beta) \right), \tag{39}$$

with $\gamma_i(0, \beta) = 1$ and $dG_i^+(0, \beta) = K_i dG_i(0, \beta)$ for all $\beta$, where $K_i$ is a constant of proportionality which guarantees that $dG_i^+$ integrates to 1. It is straightforward to verify that $V_{\mathbb{T}}(G_i^+) = V_{\mathbb{T}}(G_i)$ by directly plugging the definition of $dG_i^+$ into equation (5). Since by assumption $V_{\mathbb{T}}(G_1) = V_{\mathbb{T}}(G_2)$, we have $V_{\mathbb{T}}(G_1^+) = V_{\mathbb{T}}(G_2^+)$, and so Theorem 1 implies

$G_1^+ = G_2^+ = G^+$. This means that we have found the distribution $G^+$ common to both $G_1$ and $G_2$. Equation (39) implies that for $\alpha > 0$ and $\beta \geq 0$,

$$dG_i(\alpha, \beta) = \frac{\gamma_i(\alpha, \beta) dG^+(\alpha, \beta)}{K_i},$$
$$dG_i(-\alpha, \beta) = e^{4\alpha\beta} \left( \frac{dG_i^+(\alpha, \beta)}{K_i} - dG_i(\alpha, \beta) \right) = e^{4\alpha\beta} (1 - \gamma_i(\alpha, \beta)) \frac{dG_i^+(\alpha, \beta)}{K_i},$$

and so equation (9) holds for all $\alpha > 0$ and $\beta \geq 0$. It is also immediate to show that it holds at $\alpha = 0$ and $\beta \geq 0$. ■

# B Data

Table 2 shows a detailed split of the data. We start with 2,303,698 workers with at least one non-employment spell which started after age of 25 but before age 60. We then drop those workers whom we observe for $2\bar{T}$ or fewer periods after they become non-employed, leaving us with 1,543,609 workers. For all these workers, let $T^i$ denote the amount of time that the worker is in the survey after becoming non-employed,[22] $t_1^i$ and $t_2^i$ denote the duration of the first two non-employment spells, and $t^{e,i}$ denote the duration of the intervening employment spell. Each of $t_1^i$, $t_2^i$, and $t^{e,i}$ may be censored in the data set. Additionally, $t^{e,i}$ and $t_2^i$ may be unobserved, e.g. if $t_1^i$ is censored.

Our data set includes 783,810 workers whose first two complete non-employment spells have duration less than $\bar{T}$, $(t_1^i, t_2^i) \in \mathbb{T}^2$, and whom we observe for sufficiently long, $T^i - t^{e,i} > 2\bar{T}$. We use these workers to estimate $G_{\chi^c}^+$. We also use these workers to estimate the shares $p_{1,\mathbb{T}}$ and $p_{2,\mathbb{T}}$.

Our data set also includes 71,844 workers whose first complete non-employment spell has duration less than $\bar{T}$, $t_1^i \in \mathbb{T}$, second non-employment spell has duration in excess of $\bar{T}$, $t_2^i \notin \mathbb{T}$, yet still has a sufficiently long observation window to see the worker for $\bar{T}$ periods after the second non-employment spell starts, $T^i - t^{e,i} > 2\bar{T}$. Note that for some of these workers, the second non-employment spell is censored, but for those we know that $t_2^i > \bar{T}$. We use these workers to estimate $p_{1,\mathbb{T}}$ and $p_{2,\mathbb{T}}$. In particular, $\hat{p}_{2,\mathbb{T}} = \frac{783{,}810}{783{,}810+71{,}844} = 0.916$ is the conditional probability of completing the second spell in under $\bar{T}$ weeks given $T^i - 2\bar{T} \geq t^{e,i}$ and $t_1^i \leq \bar{T}$.

Next, our data set also includes three groups of workers whom we use to bound our estimate of $p_{1,\mathbb{T}}$. First, there are 103,969 workers whose first non-employment spell has

[22] We cannot tell whether a worker's spell ends during the final week in the sample since we do not know her employment status in week $T^i + 1$.

| Description | Number | Used For |
|---|---|---|
| has non-employment spell, $T^i > 0$ | *2,303,698* | |
| I. long time in sample, $T^i > 2\bar{T}$ | *1,543,609* | |
| A. short first non-employment spell, $t_1^i \leq \bar{T}$ | *1,244,314* | |
| 1. short uncensored employment spell, $T^i - 2\bar{T} > \underline{t}^{e,i} = t^{e,i}$ | *855,654* | |
| a. short second non-employment spell, $t_2^i \leq \bar{T}$ | 783,810 | $\hat{G}^+_{\chi^c}$, $\hat{p}_{1,\mathbb{T}}$, $\hat{p}_{2,\mathbb{T}}$ |
| b. long second non-employment spell, $t_2^i > \bar{T}$ | 71,844 | $\hat{p}_{1,\mathbb{T}}$, $\hat{p}_{2,\mathbb{T}}$ |
| 2. long uncensored employment spell, $T^i - 2\bar{T} \leq \underline{t}^{e,i} = t^{e,i}$ | 113,608 | — |
| 3. censored employment spell, $T^i = t_1^i + \underline{t}^{e,i}$, $\underline{t}^{e,i} \leq t^{e,i}$ | 275,052 | — |
| B. long first non-employment spell, $t_1^i > \bar{T}$ | *299,295* | |
| 1. short uncensored employment spell, $T^i - 2\bar{T} > \underline{t}^{e,i} = t^{e,i}$ | 103,969 | $\hat{p}_{1,\mathbb{T}}$ |
| 2. long uncensored employment spell, $T^i - 2\bar{T} \leq \underline{t}^{e,i} = t^{e,i}$ | 4,625 | — |
| 3. censored employment spell, $T^i = t_1^i + \underline{t}^{e,i}$, $\underline{t}^{e,i} \leq t^{e,i}$ | *68,118* | |
| a. very long first non-employment spell $t_1^i > 2\bar{T}$ | 42,627 | $\hat{p}_{1,\mathbb{T}}$? |
| b. medium first non-employment spell $t_1^i \leq 2\bar{T}$ | 25,491 | — |
| 4. no employment spell, $T^i = t_1^i$, $0 = \underline{t}^{e,i} \leq t^{e,i}$ | 122,583 | $\hat{p}_{1,\mathbb{T}}$? |
| II. short time in sample, $T^i \leq 2\bar{T}$ | 760,089 | — |

Table 2: Number of workers with given non-employment spells, $\bar{T} = 104$. Italicized numbers are subtotals of the numbers below them. The last column indicates which workers are used for obtaining estimates $\hat{G}^+_{\chi^c}$, $\hat{p}_{1,\mathbb{T}}$, and $\hat{p}_{2,\mathbb{T}}$.

duration more than $\bar{T}$, $t_1^i \notin \mathbb{T}$, and who have an uncensored employment spell so we know that $T^i - 2\bar{T} \geq t^{e,i}$. These workers would have had enough time in the sample to complete two non-employment spells with duration up to $\bar{T}$. Second, there are 42,627 workers who find a job after more than $2\bar{T}$ periods, and the duration of their employment spell is censored at $\underline{t}^{e,i} \leq t^{e,i}$. And third, there are 122,583 workers who are non-employed for the entire $T^i > 2\bar{T}$ periods that they are in the sample. We can think of them as having an employment duration censored at $\underline{t}^{e,i} = 0 \leq t^{e,i}$. For both these groups $T^i - 2\bar{T} > \underline{t}^{e,i}$, but we cannot tell if $T^i - 2\bar{T} \gtreqless t^{e,i}$ because $\underline{t}^{e,i} \leq t^{e,i}$.

Using these groups, we can bound the probability of completing the first non-employment spell in under $\bar{T}$ weeks conditional on $T^i - 2\bar{T} \geq t^e$: $\hat{p}_{1,\mathbb{T}} = \frac{783{,}810+71{,}844}{783{,}810+71{,}844+103{,}969+(42{,}627+122{,}583)x}$, where $x \in [0,1]$ is the unknown probability that the last two groups of workers satisfy $T^i - 2\bar{T} > t^{e,i}$. This gives us the bounds $\underline{\hat{p}}_{1,\mathbb{T}} = 0.761$ and $\bar{\hat{p}}_{1,\mathbb{T}} = 0.892$.

Finally, there are four groups of workers whom we observe for more than $2\bar{T}$ periods after they become non-employed but are not used in our calculations because we know that $T^i - 2\bar{T} \leq t^{e,i}$. The largest, 275,052 workers, have a short non-employment spell, $t_1^i < \bar{T}$, and never lose their job again. Although we only observe a censored employment duration $\underline{t}^{e,i} \leq t^{e,i}$, we know $T^i - 2\bar{T} \leq \underline{t}^{e,i} \leq t^{e,i}$. The next biggest group, 113,608 workers, also have a short non-employment spell, $t_1^i < \bar{T}$, and do lose their job (so $t^{e,i}$ is uncensored), but have $T^i - 2\bar{T} \leq t^{e,i}$. And then there are two groups of workers who have a long first spell, $t_1^i > \bar{T}$, returned to work, but have $T^i - 2\bar{T} \leq t^{e,i}$. The first group, 4,625 workers, have a completed employment spell, so we can measure $t^{e,i}$ without censoring. The second group, 25,491 workers, never lost their job, but even though $t^{e,i}$ is censored at $\underline{t}^{e,i} \leq t^{e,i}$, we know $T^i - 2\bar{T} \leq \underline{t}^{e,i} \leq t^{e,i}$.