TSE WP working paper series

11-262

Research Group: *Econometrics and Statistics*

December 2011

A Duration Model with Dynamic Unobserved Heterogeneity

IRENE BOTOSARU

Ecole d'économie de Toulouse (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31000 TOULOUSE
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.TSE-fr.eu** - contact@TSE-fr.eu

Toulouse School of Economics

A Duration Model with Dynamic Unobserved Heterogeneity*

Irene Botosaru†

December 16, 2011

Abstract

The paper considers a new class of duration models in which unobserved heterogeneity changes with time. The class addresses two main questions: How does the exit probability from a state vary when unobserved heterogeneity evolves through time? And do changes in unobserved heterogeneity have a timing effect? We show the non- and semi-parametric identification of the new class by solving a nonlinear integral equation with unknown kernel. Both the function of observed covariates and the mean of the distribution of unobserved heterogeneity are nonparametrically identified. Identifying timing effects and the distribution of unobserved heterogeneity requires stronger assumptions on either one of the two. An extension to the case when unobserved heterogeneity is a function of observed covariates is also identified. We show that sieve maximum likelihood estimators are consistent and present Monte Carlo simulations for both correct specification and misspecification. The paper also presents an empirical model of unemployment duration in which individuals exit unemployment when total accumulated losses due to unemployment cross over a self-imposed spending limit.

This paper considers a new class of duration models in which individual unobserved heterogeneity changes with time in an uncertain way and in which unobserved heterogeneity is allowed to have timing effects on the probability of exiting a specific state. The main questions addressed by the new class are: How does the exit probability from a state vary when unobserved heterogeneity evolves through time? Do changes in heterogeneity that happen earlier on in the spell have the same effect on the exit probability as changes that happen later on?

The new class, henceforth known as dynamic heterogeneity (DH), is relevant to applied economics, especially to labor, health, and, potentially, industrial organization. For example, when analyzing the

*This paper is a revision of my dissertation. I am deeply thankful to Donald Andrews, Donald Brown, Xiaohong Chen, Yuichi Kitamura, and Peter Phillips for their suggestions, advice, and support. I have benefited from discussions with Jaap Aabring, Jean-Pierre Florens, James Heckman, Bo Honore, Lena Janys, Simon Lee, Franco Peracchi, Christoph Rothe, and Francis Vella. I thank Jeremy Lise for making the data available and seminar participants at Yale, Mannheim, Montreal, Toulouse, Georgia, Brandeis, ESEM Oslo 2011, and MOOD 2010. Financial support of Cowles Foundation via Carl Arvid Anderson Prize is gratefully acknowledged.

†Toulouse School of Economics, 21 allee de Brienne, 31000 Toulouse, France; Email: irene.botosaru@tse-fr.edu

probability of promotion, it is important to account for the fact that skills accumulate during the spell of employment and that skills that are learned earlier on in the spell may have a different effect on the probability of promotion than skills learned later on. As another example, consider the probability of early retirement. During the spell of employment, health may depreciate, or there may be exposure to a toxic substance, and the timing of depreciation or exposure may have different effects on the probability of exiting employment early. Likewise, in an IO type setting, when considering the probability of restocking an inventory, it may be more realistic to think about consumption as a process that depletes the initial inventory in an uncertain way and that the timing of consumption shocks may affect the restocking probability.

In contrast, standard duration models assume unobserved heterogeneity does not change with time. As such, there is no possibility of studying possible timing effects unobserved heterogeneity may have on the probability of exit. By modeling unobserved heterogeneity as a time invariant random variable, it is implicitly assumed that unobserved heterogeneity is realized when individuals enter the state and then kept constant at its initial level throughout the duration of the state. In the examples above this would mean that skills, health, and consumption remain unchanged during the duration of the initial states of employment and not restocking, respectively.

The main framework considered in this paper models the instantaneous probability of exit, or the hazard function, as:

$$h\left(t|x, \{Z(u)\}_0^t\right) = \phi(x) \int_0^t f(u)\, dZ(u) \tag{1}$$

where $t \in \mathbb{R}_+$ is time spent in a state, $x \in \mathbb{R}$ are observed covariates[1], $\phi(x)$ is a function of observed covariates that acts as a weight on the hazard function, and $\{Z(u)\}_0^t$ is unobserved heterogeneity, modeled as a stochastic process. The weight function $\{f(u)\}_0^t \in \mathbb{R}_+$ allows changes in unobserved heterogeneity to have a permanent effect on the hazard function. The shape of $\{f(u)\}_0^t$ facilitates inference about possible timing effects of $\{Z(u)\}_0^t$. The parameters of interest are the function of observed covariates $\phi(x)$, the weight function $\{f(u)\}_0^t$, and the distribution of $\{Z(u)\}_0^t$.

To illustrate the interpretation of the elements entering (1), consider the hazard of promotion. Let x be any observed covariate that may affect the probability of upward mobility on the job. Then $\phi(x)$ models the effect of such a covariate. Let $\{Z(u)\}_0^t$ be unobserved skills that accumulate during the spell of employment. Then $\{f(u)\}_0^t$ models the timing effect of accumulating skills. The function $\{f(u)\}_0^t$ does not have to be monotonic, but suppose it is decreasing. Then skills accumulated earlier on in the spell have a higher weight and, ceteris paribus, they increase the probability of promotion. If the function is increasing, then skills accumulated later on increase the probability of exit, while if the function is flat, there are no timing effects.

In order to preserve the positivity of the hazard function, unobserved heterogeneity in this new class is

[1] For the moment, x is assumed to be time invariant. However, x can be time varying as well as k-dimensional. Appendix A.12 shows the identification of $\phi(.)$ when observed covariates are time varying.

modeled as a positive Lévy process[2]. This implies that unobserved heterogeneity accumulates in jumps that are independently and identically distributed and that happen at random times. The size and the rate of the jumps can be small or large depending on the distribution of the Lévy process. For example, jumps can accumulate gradually, in tiny and frequent increments, if the process is the gamma process. On the other hand, if the process is the compound Poisson process, jumps are large with the total number of jumps in a spell being a random variable. By considering only positive Lévy processes, the DH class is applicable to cases where there is positive duration dependence at the individual level. At the sample level, the hazard function can show negative duration dependence.

A by product of modeling unobserved heterogeneity as a stochastic process is that the relative risk of two individuals of exiting the sample, also known as the proportionality of hazards, is not required even with time invariant observed covariates. This is an advantage since the proportionality assumption in standard duration models is controversial as it excludes transition rates that are converging, diverging, or crossing during the duration of a spell.

The first contribution of the paper is to show the identification of the parameters of interest entering (1). The identification strategy is new to duration literature. It is based on formulating the probability of no exit, or the survival function, in terms of the Laplace transform of unobserved heterogeneity. Setting equal the survival function obtained from the model to that observed in the data obtains a nonlinear integral equation of the first kind with unknown kernel and a variable upper limit of integration. Identification is based on solving this integral equation for the parameters of interest.

The identification results are as follows: Under the assumption of thin tails on the distribution of unobserved heterogeneity, both $\phi(x)$ and the mean of the distribution of unobserved heterogeneity are identified nonparametrically; that is, when the distribution of unobserved heterogeneity is not known. The joint identification of the weight function $\{f(u)\}_0^t$ and of the distribution of unobserved heterogeneity is not possible in this framework since the product of these two elements enters under an integral operator, which smooths out the variation due to each of the two sources[3]. However, two separate identification results are possible by making additional assumptions. First, the weight function $\{f(u)\}_0^t$ is identified semiparametrically by assuming the distribution of unobserved heterogeneity is known up to its mean. Second, the distribution of unobserved heterogeneity is identified when $\{f(u)\}_0^t$ is assumed known. These two separate results can be applied as follows. When one may have prior information or beliefs about the distribution of $\{Z(u)\}_0^t$ and when interest lies in the timing function $\{f(u)\}_0^t$, then one is able to identify $\{f(u)\}_0^t$. On the other hand, if one may have prior beliefs about $\{f(u)\}_0^t$ (or if the function is not relevant to the model and thus set equal

[2] This restriction could potentially be relaxed by formulating the hazard function as

$$h\left(t|x, \{Z(u)\}_0^t\right) = \int_0^t f(u, x)\, d\exp(Z(u))$$

In this case, the resulting stochastic process loses the independence of the increments (for example, if Z is the Brownian motion, $\exp(Z)$ would be the geometric Brownian motion). The survival function in this case would have a different form than the one presented in this paper. Identification of such a model would require controlling the quadratic variation of $\exp(Z)$.

[3] The failure of joint identification is shown in Appendix A.11.

to one), one may identify the distribution of unobserved heterogeneity.

The second contribution is to provide semiparametric consistent estimators for $\phi(x)$, $\{f(u)\}_{u>0}^{t}$, and the mean of the distribution of unobserved heterogeneity when unobserved heterogeneity is known up to its mean.

Finally, the paper presents an empirical illustration to unemployment duration inspired by behavioral economics, first hitting-time models, and labor supply decision models with reference points. The reduced form model is evaluated on NLSY79 data on the first spell of unemployment.

The present framework is not a generalization of the standard model used in duration analysis, the mixed proportional hazard (MPH) model. The MPH models the hazard function as

$$h_M(t|x,\nu) = \phi_M(x)\lambda(t)z \tag{2}$$

where $\phi_M(x)$ is a function of observed covariates, usually parametrized as $\exp(x'\beta)$, $\lambda(t)$ is a deterministic function of time known as the baseline hazard, and z is a random variable modeling unobserved heterogeneity. When unobserved heterogeneity is a stochastic process, at each point in time, there is an infinity of sample paths that unobserved heterogeneity can follow. In the MPH, there is an infinity of sample paths only when the individual enters the sample. Time effects in the MPH come through the baseline hazard which is a deterministic function of time, while in the DH model, time effects are stochastic. The resulting stochastic process in the DH hazard function, i.e. $\int_0^t f(u)\,dZ(u)$, cannot be reduced to a deterministic function of time $\lambda(t)$ multiplied by a random variable, z.

In the framework proposed in this paper there is no equivalent to the baseline hazard function from standard duration models. It is possible to introduce such a function in the present framework, however the function would have to be assumed known in order for the identification results presented in this paper to remain unchanged.

The paper is organized as follows. Section 1 presents a brief literature review of duration models. Section 2 presents the identification results, while Section 3 introduces the estimation procedure and derives the consistency of the estimators. Section 4 presents a model of unemployment and an empirical application. Section 5 concludes. All proofs, further identification results of possible interest, and Monte Carlo simulation results, as well as a brief summary of mathematical properties of Lévy processes and the description of the data are found in the Appendix.

Remark 1 *One advantage of modeling unobserved heterogeneity as a stochastic process is that the heterogeneity can be a function of observed covariates. That is, the process $\{Z(u)\}_0^t$ can be time deformed by a function $g(.)$ that depends on time varying covariates $y(t)$. The hazard function in this case takes the form*

$$h\left(t|x,\{y(u)\}_0^t,\{Z(u)\}_0^t\right) = \phi(x)\int_0^t dZ(g(y(u))) \tag{3}$$

As (J.H. Stock 1988) notes, the process $\{Z(u)\}_0^t$ *evolves in calendar time. In contrast, the process* $\{Z(g(y(u)))\}_0^t$ *evolves on a data driven time scale* $\{g(y(u))\}_0^t$*. The function* g *controls the speed at which the jumps in* $\{Z(u)\}_0^t$ *happen: If* g *is an increasing function, for greater values of* $y(t)$*, the faster the jumps will accumulate. In the skill example mentioned above, suppose the individual is paid a bonus,* $y(t)$*, for a successfully finished project. Then the bonus paid may affect the speed at which skills accumulate. Supposing* g *was an increasing function, the higher the bonus paid, the faster the accumulation of skills. The parameters of interest are* (ϕ, g) *and the mean of the distribution distribution of* $\{Z(u)\}_0^t$*. Identification of these parameters is presented in Appendix A.13.*

The time deformed framework is presented separately from (1) *since the two frameworks are both mathematically and conceptually different. Mathematically, in* (1) *the process is stationary, while the time deformed process in* (3) *is not. Conceptually, when unobserved heterogeneity takes place in calendar time, i.e. when it is modeled as* $\{Z(u)\}_0^t$*, the heterogeneity evolves for reasons related to time but independent from current observed covariates. This is not the case when unobserved heterogeneity is time deformed by observed covariates. In certain applications one framework may be more appropriate than the other. Appendix A.13 also presents an application of* (3) *that may be of interest.*

Notation 1 *In this paper, the following notation for derivatives is used. Let* $H(.,x)$ *and* $\Psi(.,.)$ *be differentiable functions.*

$$\begin{aligned} H_{tt}(t,x) &= \frac{\partial^2}{\partial t^2} H(t,x) \\ \Psi_1(\lambda,k) &= \frac{\partial}{\partial \lambda}\Psi(\lambda,k)\,;\ \Psi_2(\lambda,k) = \frac{\partial}{\partial \lambda}\Psi(\lambda,k) \\ \Psi_{11}(\lambda,k) &= \frac{\partial^2}{\partial \lambda^2}\Psi(\lambda,k)\,;\ \Psi_{12}(\lambda,k) = \frac{\partial^2}{\partial \lambda \partial k}\Psi(\lambda,k) \end{aligned} \tag{4}$$

1 Literature Review

The paper belongs to the literature on duration models. Regression analysis of duration data is usually based on the MPH model, where the hazard function is defined by (2) and where z is assumed to be distributed according to $G(z)$. The MPH has been studied extensively. Two of the first papers dealing with the nonparametric identification of $(\beta, \lambda(t), G(z))$ are (C. Elbers & G. Ridder 1982) and (J.J. Heckman & B. Singer 1984). These papers show the identification of the triplet $(\beta, \lambda(t), G(z))$ entering (2) under assumptions on either the mean of $G(z)$ or the tail of $G(z)$, respectively. The identification of the MPH with time varying observed covariates has been studied by (J.J. Heckman & C.R. Taber 1994), under the assumption that covariates are deterministic functions of t and by (B.E. Honoré 1993), under the assumption that observed covariates are step functions. (Heckman & Taber 1994) generalized the results of (Honoré 1993) to the case when observed covariates are realizations from a stochastic process with continuous sample paths almost surely. Along the same lines, (B.P. McCall 1993) presents conditions under which an MPH model with time-varying coefficients is identified. More recently, (C.N. Brinch 2007) shows the identification of the standard MPH when the baseline hazard in (2) is a function $\lambda(t, x)$.

The identification of the MPH rests on the multiplicative structure of the hazard function. When time invariant z enters the hazard multiplicatively, the survival function has the form of a Laplace transform of the probability density function of z, which allows the identification of $G(z)$.

Another class of duration models is formed by mixed hitting time (MHT) models, which are first passage models. In MHT, agents leave the initial state as soon as a spectrally negative Lévy process, $Y(t)$, hits a barrier, $\phi(x) z$, where $\phi(x)$ is a function of observed covariates and z is unobserved heterogeneity. Duration is defined as:

$$T = \inf\{t \geq 0 : Y(t) > \phi(x) z\} \tag{5}$$

The barrier is time invariant. As in the MPH, unobserved heterogeneity is determined at the beginning of the spell and kept constant throughout the duration of the spell.

The MHT has been studied in a series of papers by (M-L.T. Lee & G.A. Whitmore 2004) and (M-L.T. Lee & G.A. Whitmore 2006). (J.H. Abbring 2011) shows the nonparametric identification of the distribution of $Y(t)$, $\phi(x)$, and z under a strategy similar to that used by (Heckman & Singer 1984).

Rather than having an exogenously given barrier determine the timing of the event of interest as in the MHT, and in an attempt to allow unobserved factors change through time, unobserved heterogeneity can be modeled as a stochastic process. One of the first papers to mention the importance of having stochastic unobserved heterogeneity in duration models was (Y. Kebir 1991). An overview of models with stochastic randomness is given in (N.D. Singpurwalla 1995) and models with Lévy frailties are introduced in (O. Aalen & N.L. Hjört 2002) and (O. Aalen, H.K. Gjessing & N.L. Hjört 2003). (N.D. Singpurwalla 2006) presents a nice overview of models in which the hazard rate is viewed as a stochastic process.

A model similar to (1) was introduced in (Aalen, Gjessing & Hjört 2003). In their paper, (Aalen, Gjessing & Hjört 2003) do not allow dependence of the hazard function on observed covariates. They also model state dependence explicitly by a baseline hazard function. (Aalen, Gjessing & Hjört 2003) present certain properties of the survival function, without studying the identification of the model or the asymptotic properties of their estimators.

2 Identification

The section explains the identification of (ϕ, f), and the distribution of $\{Z(u)\}_0^t$ entering (1).

The identification strategy requires the existence and finiteness of the mean of the distribution of the stochastic process $\{Z(u)\}_0^t$, which implies the distribution has thin tails. This assumption excludes processes such as the stable process.

The identification strategy proceeds in several steps. First, under normalization and regularity assumptions, the function of observed covariates, $\phi(x)$, and the mean, k_1, of the distribution of the stochastic process

$\{Z(u)\}_0^t$ are nonparametrically identified; that is, they are identified when the distribution of $\{Z(u)\}_0^t$ is unknown. The identification holds in the limit as the time index approaches zero. This result is similar to identification at infinity, and may run into problems similar to irregular thin set identification[4]. In order to avoid this issue, the function $\phi(x)$ is shown to be identified for all values of time, $t > 0$, but when the distribution of the stochastic process is parametrized up to its mean. Thus, there is a trade-off in the identification of $\phi(x)$.

Second, because of the special interaction between the weight function $\{f(u)\}_0^t$ and the stochastic process, $\{Z(u)\}_0^t$, the two cannot be identified jointly. By assuming the distribution of the stochastic process to be known up to its mean, the weight function is identified for all $t > 0$. The identification of the weight function proceeds by the Banach fixed-point theorem. As a separate result, assuming the weight function $\{f(u)\}_0^t$ to be known, the distribution of $\{Z(u)\}_0^t$ is identified.

Misspecification of the mixing distribution generally results in inconsistent estimates. In practice, when choosing an appropriate distribution for unobserved heterogeneity, one can use mathematical properties of stochastic processes that explain real-life phenomena, as well as results from biostatistics, operations management, engineering, and psychology. Despite this limitation, there is scope to allow unobserved heterogeneity to develop in an uncertain manner as this specification may capture dynamic operating conditions in a more realistic way.

The intuition for the identification results begins with the transition rate out of the initial state among those surviving up to some t in the initial state, defined as:

$$\phi(X)\, E_Z\left[\int_0^t f(u)\, dZ(u) \middle| T \geq t, X\right] \tag{6}$$

For $t > 0$, the variation in the transition rate is due to variation in both observed and unobserved heterogeneity.

Near the start of the spell, as t approaches zero, individuals in the sample are similar in terms of unobserved heterogeneity up to the mean of the distribution of unobserved heterogeneity, k_1. That is so because under independence of $\{Z(u)\}_0^t$ and X, for all t, and under both a normalization assumption on $\{f(u)\}_0^t$ and the existence and finiteness of k_1, the expectation term in expression (6) collapses to k_1 as $t \downarrow 0$. Then variation in (6) is due to both the heterogeneity at the beginning of the spell, captured by k_1, and the variation in observed heterogeneity, captured by $\phi(x)$. The mean, k_1, is identified by imposing a normalization assumption on $\phi(x)$, which eliminates the variation in observed heterogeneity.

As time elapses, and given $\phi(x)$, variation in the transition rate is due to variation that comes in from two sources: the weight function $\{f(u)\}_0^t$ and unobserved heterogeneity $\{Z(u)\}_0^t$. For identification, the variation in $\{f(u)\}_0^t$ needs to be separated from that in $\{Z(u)\}_0^t$. However, the integral operator smooths out, or lumps together, the variation from each source. To recover the individual variation due to each

[4] See (S. Khan & E. Tamer 2010).

source, stronger restrictions on either $\{f(u)\}_0^t$ or $\{Z(u)\}_0^t$ need be imposed. To identify $\{f(u)\}_0^t$, the mechanism that generates the variation in $\{Z(u)\}_0^t$, i.e. the distribution of unobserved heterogeneity, needs to be assumed known up to its mean. On the other hand, the distribution of $\{Z(u)\}_0^t$ is identified assuming $\{f(u)\}_0^t$ to be known.

2.1 Identification of k_1 and $\phi(x)$

Consider the hazard function (1). Under certain assumptions listed in Appendix A.2.1, the survival function associated to (1) is defined by applying the usual exponential formula:

$$S(t|x) = E_Z \exp\left[-\int_0^t h(s|x, Z(s))\, ds\right] \tag{7}$$

By using the independence and stationarity of the increments of $\{Z(u)\}_0^t$ as well as the definition of the Laplace exponent of the distribution of $\{Z(u)\}_0^t$ (see A.1), (7) can be written as a function of the Laplace transform of the distribution of $\{Z(u)\}_0^t$. The survival function has the following form:

$$S(t|x) = \exp\left[-\int_0^t \Psi(\phi(x) f(u)(t-u))\, du\right] \tag{8}$$

where $\Psi(.)$ is the Laplace exponent of the subordinator and $S(t|x) \in (0,1)$. The detailed derivation of (8) is shown in Appendix A.2.2.

Let the true survival function be noted as $S_0(t|x)$. The true survival function equals $S(t|x)$, the survival function associated to (1), i.e.

$$S_0(t|x) = S(t|x) \tag{9}$$

Taking log of both sides of (9) and rearranging obtains:

$$H(t,x) = \int_0^t \Psi(\phi(x) f(u)(t-u))\, du \tag{10}$$

where $H(t,x) = -\log S_0(t|x)$.

The identification of the parameters of interest is based on solving (10) for $\phi(x), f(t)$, and for the moments of the distribution of $\{Z(u)\}_0^t$ for all $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$. Equations similar to equation (10) are known as nonlinear Volterra integral equations of the first kind[5]. Note that the kernel in (10) is unknown.

In general, Volterra integral equations of the first kind do not have unique solutions. However, if they

[5] An integral equation of the form below is a nonlinear Volterra integral equation of the first kind:

$$h(y) = \int_a^y K(\pi(s), y, s)\, ds$$

where $y \geq a$, where h is a given function and π is the solution to find. $K(z, y, s)$ is the nonlinear kernel.

can be transformed into Volterra integral equations of the second kind[6], then it is possible to show they have unique solutions. Under certain regularity assumptions, it is possible to reduce a Volterra integral equation of first kind to one of the second kind by differentiating the former with respect to its upper limit of integration until one obtains a free term added to the integral. This term usually depends on the functions of interest, and it allows one to solve for the unknown functions by the method of contraction mappings. Once the existence and uniqueness of the solution to the Volterra integral equation of the second kind has been shown, the solution also solves uniquely the Volterra integral equation of the first kind by the Fundamental Theorem of Calculus. This is the solution strategy adopted in this paper: First, the Volterra integral equation of the first kind, equation (10), is reduced to an integral equation of the second kind by differentiating (10) twice with respect to t. The differentiation is possible by Property A.1 (i), and by assumptions ID1(i) and ID2 below, where $H_{tt}(t,0) = -\frac{\partial^2}{\partial t^2}\log S_0(t|x)$.

Assumption ID1 (i) $H(t,x)$ is twice continuously differentiable in $t \in \mathbb{R}_+$ for all $x \in \mathbb{R}$; (ii) $\lim_{t\to 0} H_{tt}(t,0)$ is well-defined.

Assumption ID2 The Laplace exponent is differentiable at zero.

Assumption ID3 (i) $\phi(0) = 1$; (ii) $\lim_{t\to 0} f(t) = 1$.

Assumption ID2 implies that k_1 exists and is a nonzero finite number. ID2 excludes subordinators for which the mean does not exist, such as the stable processes. This assumption can be thought of as the analogue of the finite mean assumption in (Elbers & Ridder 1982), used to show the identification of the MPH model. Assumption ID3 are normalization assumptions. ID3(i) is standard in duration analysis, while ID3(ii) is needed in order to completely identify (ϕ, k_1). Else, the two parameters ϕ and k_1, would be nonparametrically identified up to $\lim_{t\to 0} f(t)$.

Lemma 1 *Let unobserved heterogeneity* $\{Z(u)\}_0^t$ *entering the hazard function* (1) *be described by a positive Lévy process with unknown distribution. By Property A.1 (i) and (ii), and under assumptions ID1(i) and ID2, the Volterra integral equation of the first kind with unknown kernel* (10) *is reduced to a Volterra integral equation of the second kind*

$$f(t) = \frac{1}{k_1\phi(x)}\left[H_{tt}(t,x) - \phi^2(x)\int_0^t f^2(u)\,\Psi_{11}(\phi(x) f(u)(t-u))\,du\right] \tag{11}$$

where $\Psi_{11}(.) = \frac{\partial^2}{\partial t^2}\Psi(.)$.

[6] An integral equation of the second kind has the following form:

$$h'(x) = K(\pi(y), y, y) + \int_a^y K_y(\pi(s), y, s)\,ds$$

where $y \geq a$, where the mapping $K(z, y, y)$ is invertible in z for all y and $K(\pi(y), y, y) \neq 0$.

Proof. See Appendix A.3. ■

Theorem 2 *Let the distribution of the positive Lévy process* $\{Z(u)\}_0^t$ *be unknown and let assumptions ID1, ID2, and ID3 hold. Then the mean of the distribution is identified and given by*

$$k_1 = \lim_{t \to 0} H_{tt}(t, 0)$$

Proof. See Appendix A.4. ■

Theorem 3 *Let the distribution of the positive Lévy process* $\{Z(u)\}_0^t$ *be unknown. Under assumptions ID1, ID2, and ID3, the function* $\phi(x)$ *is identified and given by*

$$\phi(x) = \frac{\lim_{t \to 0} H_{tt}(t, x)}{\lim_{t \to 0} H_{tt}(t, 0)}$$

Proof. See Appendix A.5. ■

2.2 Identification of the Weight Function

The identification of the weight function $\{f(u)\}_0^t$ proceeds by first assuming that the distribution of $\{Z(u)\}_0^t$ is known up to k_1. Then the true survival function depends on the true parameter k_1^0, i.e. $S_0\left(t|x, k_1^0\right)$. The survival function associated to (1) is $S(t|x, k_1)$. Property A.1 and assumptions ID1 and ID2 hold for all k_1, so that the integral equation of the first kind that is obtained by setting $S_0\left(t|x, k_1^0\right) = S(t|x, k_1)$ can be reduced to an integral equation of the second kind just as before.

The mean, k_1, is a nonlinear function of possibly d elements, such as the rate of the jumps, the scale of the jumps, the expected number of jumps, and so on[7]. If interest lies not only in $\{f(u)\}_0^t$ but also in one of the d elements of k_1, it is shown that by normalizing all but the element of interest in k_1, the free parameter of k_1 is identified. This result is shown for both the gamma process and the compound Poisson processes.

Let $f(t) \in C_w^s(\mathbb{R}_+)$, where C_w^s is the Banach space of s-times continuously differentiable functions endowed with the appropriate norm, weighted by a continuous, positive weight function, $w(t)$, defined by (13). The function, $\{f(u)\}_0^t$, is identified if (11) has a unique solution in C_w^s, which is established via the Banach Fixed Point Theorem by showing the integral operator defined below is an inclusion and a contraction:

$$(Tf)(t) = \frac{1}{k_1}\left[H_{tt}\left(t, x, k_1^0\right)\big|_{x=0} - \int_0^t f^2(u)\,\Psi_{11}(f(u)(t-u), k_1)\,du\right] \tag{12}$$

where $H_{tt}\left(t, 0, k_1^0\right) = -\frac{\partial^2}{\partial t^2}\log S_0\left(t|x=0, k_1^0\right)$.

Let the following assumptions hold:

[7] See examples in Appendix A.1.

Assumption ID4 (i) $f(t) \in C_w^s(\mathbb{R}_+)$; (ii) $\lim_{t\to 0} \frac{\partial}{\partial t} f(t)$ is well-defined; (iii) $0 < f(t) \leq M < \infty$, $\forall$ $t \in \mathbb{R}_+$.

Assumption ID5 $\left|\frac{\partial^3}{\partial \lambda^3} \Psi(\lambda, k_1)\right| \leq B$, $\forall$ $\lambda > 0$ and $k_1 \in R_+$.

Assumption ID6 (i) $H_{tt}\left(t, x, k_1^0\right) \in C_w^s$; and (ii) There exists a constant $\delta > 0$ such that

$$\lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right) \geq \delta$$

Assumptions ID4(iii) and ID5 coupled with property A.1(i), imply the second partial derivative of the Laplace exponent with respect to its first argument is Lipschitz continuous with Lipschitz constant B:

$$|\Psi_{11}(\lambda_1, k_1) - \Psi_{11}(\lambda_2, k_1)| \leq B|\lambda_1 - \lambda_2|$$

Lipschitz continuity is needed in order guarantee the kernel of (11) is Lipschitz continuous, which will further reflect into the operator T being Lipschitz continuous with a bounded Lipschitz constant. Assumption ID6(ii) implies the mean of the distribution of $\{Z(u)\}_0^t$ is bounded from below by a positive number δ. ID6(ii) is needed in order to guarantee that the operator T is a contraction. The Lipschitz constant of the operator is shown to be $\frac{\delta}{k_1}$. Since $k_1 = \lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right)$, the operator is a contraction if $\frac{\delta}{k_1} \in (0, 1)$. Examples 1 and 2 in Appendix A.6 show assumptions ID2, ID4 to ID6 are satisfied for both the gamma and the compound Poisson processes.

Theorem 4 *Assume the distribution of the stochastic process is known up to its mean, k_1. Under assumptions ID1 to ID6, the function $f(t) \in C_w^s$ is identified. The weight function for the norm is given by*

$$w(t) = \exp\left(\frac{3BM^2}{\delta} t\right) \tag{13}$$

The solution is found by the successive approximation method. That is, for $n \geq 1$:

$$\begin{aligned} f_0(t) &= \frac{1}{\lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right)} H_{tt}\left(t, 0, k_1^0\right) \\ f_n(t) &= \frac{1}{\lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right)} \left[H_{tt}\left(t, 0, k_1^0\right) - \int_0^t f_{n-1}^2(u)\, \Psi_{11}\left(f_{n-1}(u)(t-u), k_1\right) du\right] \end{aligned}$$

Proof. See Appendix A.9. ∎

Theorem 2 identifies the mean of the distribution of $\{Z(u)\}_0^t$ when the distribution is unknown. Once the distribution is parametrized up to k_1, the mean becomes a nonlinear function of several parameters such as the scale and the rate of the jumps. If interest lies in identifying the rate parameter ρ, then the other parameters entering the formulation of k_1 need to be normalized.

Assumption R1 The scale of the gamma process is normalized to one: $\nu = 1$.

According to properties of the gamma process, if Z is a gamma process with scale parameter ν, then νZ is a gamma process with scale parameter 1, see (N. Tsilevich, A. Vershik & M. Yor 2001). As such, a normalization assumption on the scale of the gamma process cannot be avoided. By normalizing the scale of the process, the jumps of the process have a magnitude of 1.

Lemma 2 *Let the conditions of Lemma 2 and Theorem 4 hold. Let the process be the gamma process and assume R1 holds. Then the rate of the jumps, ρ, is identified.*

Proof. For the gamma process, the mean is given by:

$$k_1 = \frac{\rho}{\nu} = \lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right) = \lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right)$$

by assumption R1. ■

Consider now the case of the compound Poisson process for which the following extra assumption is imposed:

Assumption R2 The scale and rate of the Poisson process are normalized to one: $\nu = \kappa = 1$.

The Lévy measure of the compound Poisson process is proportional to the jump distribution by the factor κ, as such a normalization assumption on κ cannot be avoided. The intensity parameter κ represents the expected number of events that occur per unit time. For example, if the expenditure process is modeled as the compound Poisson process, the expected number of expenditure shocks that happen in the interval $(t, t+\tau]$ is $\kappa\tau$. Normalizing κ implies that the expected number of expenditure shocks that happen in an interval of time equals the length of the interval.

Lemma 3 *Let the conditions of Lemma 2 and Theorem 4 hold. Let the process be the compound Poisson process and assume R2 holds. The rate parameter ρ and the weight function are identified.*

Proof. For the compound Poisson process, the mean is given by:

$$k_1 = \frac{\kappa\rho}{\nu} = \lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right) = \lim_{t\to 0} H_{tt}\left(t, 0, k_1^0\right)$$

by Assumption R2. ■

Consider now the identification of $\phi(x)$ when the distribution of $\{Z(u)\}_0^t$ is known up to k_1.

Theorem 5 *Let all the assumptions necessary for the identification of $f(t)$ and of k_1 hold, and let the distribution of the stochastic process be known up to k_1. Then the covariate function $\phi(x)$ is identified and given by*

$$\phi(x) = \frac{1}{k_1 f(t)}\left[H_{tt}\left(t, x, k_1^0\right) - \phi^2(x)\int_0^t f^2(u)\,\Psi_{11}\left(\phi(x) f(u)(t-u), k_1\right) du\right]$$

Proof. See Appendix A.10. ■

2.3 Identification of the Marginal Distribution of the Process

Suppose either one had prior information about the weight function $\{f(u)\}_{u>0}^{t}$ or the weight function was not relevant in the model, i.e. $\{f(u)\}_{u>0}^{t} = 1$, $\forall t$. Then all moments of the distribution of $\{Z(u)\}_0^t$ can be identified as $t \downarrow 0$. This result should be regarded as separate from those above as the weight function $\{f(u)\}_{u>0}^{t}$ and the distribution of $\{Z(u)\}_0^t$ cannot be jointly identified.

Condition A Let $f(t)$ be analytic and such that for all $t > 0$

$$\left\{\lim_{t\to 0} \frac{\partial^n}{\partial t^n} f(t)\right\}_{n\geq 1} = 1$$

Condition B (i) $H(t,x) \in C^{\infty}$ in $t > 0$ for all $x \in \mathbb{R}$. (ii) $\Psi(.)$ is infinitely many times differentiable at zero.

Condition A implies that $f(t)$ has the following form:

$$f(t) = f(0) + f'(0)\,t + \frac{f''(0)}{2!}t^2 + ... = 1 + \sum_{i=1}^{\infty} \frac{t^i}{i!} = e^t$$

Condition B(ii) implies all moments of the distribution of $\{Z(u)\}_0^t$ exist and are finite. The assumption implies that Carleman's condition holds. Carleman's condition is a sufficient condition for the uniqueness of the moment sequence. If the moment sequence is unique, then the moment generating function determines the distribution. Carleman's condition states that a sequence of moments $\{M_n\}_{n\in\mathbb{N}}$ that satisfy $\sum_j (M_{2j})^{-1/2j} = \infty$ uniquely determines a random variable with moments M_n. See (W. Feller 1970). For the problem, Carleman's condition takes the form (14) below:

$$\sum_{r=1}^{\infty} \left(\frac{1}{k_{2r}}\right)^{1/2r} = \infty \tag{14}$$

where k_r is the r^{th} moment of the distribution of $\{Z(u)\}_0^t$. A sufficient condition for (14) to hold is B(ii).

Theorem 6 *Let conditions A and B hold. Then the distribution of the stochastic process $\{Z(u)\}_0^t$ is uniquely determined by its positive order integer moments.*

Proof. See A.11. ■

3 Sieve Maximum Likelihood

The section considers the estimation of (1), where the distribution of $\{Z(u)\}_0^t$ is known up to the jump rate, ρ. The parameters of interest are (ϕ, f, ρ).

The identification strategy presented above is constructive. However, it is based on the second partial derivative of the true survival function with respect to the time index. Translating this into an estimation procedure means numerically differentiating a consistent nonparametric estimator of the conditional survival function, which would introduce numerical error in the estimation of the functions of interest. In order to overcome possible estimation issues resulting from numerical differentiation, the estimation procedure proposed in this paper does not follow the identification strategy. Since the weight function $\{f(u)\}_0^t$ is identified by parametrizing the distribution of $\{Z(u)\}_0^t$ up to ρ, the estimation procedure is semiparametric maximum likelihood (ML).

Parametrizing the distribution of $\{Z(u)\}_0^t$ up to ρ obtains the conditional density function of duration given observed covariates. Since interest lies in estimating two infinite dimensional parameters, ϕ and f, under shape restrictions and which enter the criterion function nonlinearly, the estimation procedure proposed in this section is semiparametric sieve ML. The basic picture is that sieve estimation avoids possible theoretical and computational issues associated with optimization over infinite dimensional spaces. Theoretically, the MLE may be inconsistent if the estimation procedure is carried out over an infinite dimensional space[8], while computationally, working with finite-dimensional spaces significantly reduces the dimensionality of the optimization problem.

The proposed estimation method is to approximate the unknown positive smooth functions by positive transformations of a linear span of known basis functions[9]. The basis functions chosen in this paper are polynomial splines since splines are known to approximate smooth functions well and have been widely applied in nonparametric estimation. Both the finite dimensional parameter and the coefficients in the linear expansions are then simultaneously estimated by maximizing the likelihood over a sequence of approximating spaces with the approximating error decreasing to zero as the sample size increases.

3.1 The Model and the Estimator

Let $\{(X_i, T_i)\}_{i=1}^n$ be iid draws from the distribution of (X, T) with bounded support $\mathcal{X} \times \mathcal{T}$ where $\mathcal{X} = [0,1]$ and $\mathcal{T} = (0,1]$.[10]. The survival function associated with the hazard function considered is given by

$$S(t|x;\phi,f,\rho) = \exp\left[-\int_0^t \Psi(\phi(x) f(u)(t-u), \rho)\, du\right] \tag{15}$$

[8] See (S. Geman & C. Hwang 1982), (A.R. Gallant & D.W. Nychka 1987), (X. Shen & W.H. Wong 1994), (X. Shen 1995), (C. Ai & X. Chen 2003), (X. Chen 2007).

[9] See (C. de Boor 1978), (L.L. Schumaker 1981), (C.J. Stone, M.H. Hansen, C. Kooperberg & Y.K. Truong 1997), (Chen 2007).

[10] As (M. Crowder 2001) discusses, the covariates of interest are more likely to have bounded supports in applications. Bounded support is usually assumed in the biostatistics literature on the MPH model. See (J.A. Wellner & Y. Zhang 2007), (J. Huang 1996). The result of this section can be easily extended to the case of $\mathcal{X} \subset \mathbb{R}^d$ with $\phi(x)$ approximated by a tensor product.

where $\Psi(\lambda,\rho)\,,\ \lambda \geq 0$ is the Laplace exponent of $\{Z_\rho(u)\}_0^t$. The conditional distribution of $T|X$ is given by

$$p(t|x;\phi,f,\rho) = -\frac{\partial}{\partial t} S(t|x;\phi,f,\rho)$$

The true value $\alpha_0 = (\phi_0, f_0, \rho_0) \in \mathcal{A} = \Phi \times \mathcal{F} \times \Theta$ solves

$$\begin{aligned} \alpha_0 &= \arg \max_{(\phi,f,\rho)\in\mathcal{A}} Q(\phi,f,\rho) \\ &= \arg \max_{(\phi,f,\rho)\in\mathcal{A}} E_{x,t} \log p(t|x;\phi,f,\rho) \end{aligned} \tag{16}$$

where the finite dimensional parameter $\rho \in \Theta$, a compact subset of $\mathbb{R}_+ - \{0\}$, while the two functions are assumed to belong to the following spaces:

$$\Phi = \{\phi(x) \in C^{s_1}(\mathcal{X}, \mathbb{R}_+):\ \phi(0) = 1\} \tag{17a}$$

$$\mathcal{F} = \left\{f(u) \in C^{s_2}(\mathcal{T}, \mathbb{R}_+):\ \lim_{t\to 0} f(t) = 1\right\} \tag{17b}$$

A sieve ML estimator is proposed for $\alpha_0 \in \mathcal{A}$ by replacing $\mathcal{A}$ by a sieve space $\mathcal{A}_n$ that is compact, linear, finite dimensional space and that becomes dense in $\mathcal{A}$ as $n \to \infty$. Let $B_j(.)$ be a sequence of known univariate basis functions. Then $\mathcal{A}_n$ is a linear span of finitely many $B_j(.)$. For sieve approximation, we consider the functions ϕ and f in finite dimensional spaces Φ_n and $\mathcal{F}_n$, respectively, defined as:

$$\Phi_n = \left\{\phi_n(x) \in \Phi : \phi_n(x) = \exp \sum_{j=1}^{m_n} a_j B_j(x)\right\} \tag{18a}$$

$$\mathcal{F}_n = \left\{f_n(t) \in \mathcal{F} : f_n(t) = \exp \sum_{j=1}^{m_n} b_j B_j(t)\right\} \tag{18b}$$

where m_n is the dimension of the sieve spaces, such that $m_n \to \infty$ with $\frac{m_n}{n} \to 0$. The exponential transformation serves to impose the positivity of the functions. The sieve spaces are open and convex, with approximation rate of order $O(n^{-s_1})$ and $O(n^{-s_2})$, respectively.

The sieve ML estimator $\widehat{\alpha}_n = \left(\widehat{\phi}_n, \widehat{f}_n, \widehat{\rho}_n\right) \in \mathcal{A}_n = \Phi_n \times \mathcal{F}_n \times \Theta$ maximizes the sample analog of (16) with α restricted to the sieve space $\mathcal{A}_n$:

$$\begin{aligned} \widehat{\alpha}_n &= \arg \max_{\alpha\in\mathcal{A}_n} \widehat{Q}_n(\phi,f,\rho) \\ &= \arg \max_{\alpha\in\mathcal{A}_n} \frac{1}{n} \sum_{i=1}^{n} \log p(t_i|x_i;\phi,f,\rho) \end{aligned}$$

Then, the sieve ML estimator satisfies

$$\widehat{Q}_n(\widehat{\alpha}_n) \geq \sup_{\alpha \in \mathcal{A}_n} Q_n(\alpha) - O_p(\eta_n), \ \eta_n = o(1)$$

3.2 Consistency

The following assumption is made on the parameter space, $\mathcal{A}$.

Assumption C0. (i) $\mathcal{A}$ is connected in the sense that for any $\alpha_1, \alpha_2 \in \mathcal{A}$, there exists a continuous path $\{\alpha(\tau) : \tau \in [0,1]\}$ in $\mathcal{A}$ such that $\alpha(0) = \alpha_1$ and $\alpha(1) = \alpha_2$. (ii) The parameter space is convex at α_0, such that for any $\alpha \in \mathcal{A}$, $(1-\tau)\alpha_0 + \tau\alpha \in \mathcal{A}$ for small $\tau > 0$. (iii) For almost all (X,T), $p(t|x,(1-\tau)\alpha_0 + \tau\alpha)$ is continuously differentiable at $\tau = 0$.

The consistency of the estimators is established under metric $||.||_\infty$ defined below. For any $\alpha \in \mathcal{A}$:

$$||\alpha - \alpha_0||_\infty = \sup_x |(\phi - \phi_0)(x)| + \sup_t \left| \int_0^t (f - f_0)(u)\, du \right| + ||\rho||_E \tag{19}$$

where $||.||_E$ is the Euclidean norm. To establish the consistency of the estimators, it is assumed that:

Assumption C1. (i) The functions $\phi(x)$ and $f(t)$ are such that (17a) and (17b) hold. (ii) $f(t)$ is bounded from above. (iii) $\phi(x)$ and $f(t)$ are bounded away from zero for all x and all t, respectively.

Assumption C2. Let $\Psi(\lambda,\rho)$ be such that for all $\lambda > 0$ and ρ, the following partial derivatives are bounded below and above:

$$\begin{aligned} 0 &< m_1 \leq \Psi_1(\lambda,\rho) \leq M_1 < \infty \\ -\infty &< m_{11} \leq \Psi_{11}(\lambda,\rho) \leq 0 \\ -\infty &< m_{12} \leq \Psi_{12}(\lambda,\rho) \leq M_{12} < \infty \\ 0 &< m_2 \leq \Psi_2(\lambda,\rho) \leq M_2 < \infty \end{aligned}$$

where the partial derivatives are evaluated at $\lambda = \phi(x) f(u)(t-u)$.

Let $\mathcal{A}_o$ be an open and convex space such that

$$\mathcal{A}_o = \Phi_o \times \mathcal{F}_o \times \Theta_o = \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_\infty = o(1)\}$$

Assumption C3. $\Psi(\lambda,\rho)$ is pathwise differentiable with respect to $\lambda \in \mathcal{A}_o$ for all $t \in \mathcal{T}$ and for all $\rho \in \Theta$ and continuously differentiable in $\rho \in \Theta_o$ for all $\lambda \in \mathcal{A}$ in the norm $||.||_w$ defined in 19.

Assumption C4. $\Psi(\lambda,\rho)$ is monotonic in ρ.

Assumption C1(ii) implies that the hazard function is bounded away from zero. As noted by (D.M. Dabrowska 2005), this assumption holds if the covariates are bounded and the regression coefficients vary over a bounded neighborhood of the true parameter, conditions which hold by construction in this paper. The uniform boundedness assumption on the functions of interest is used to verify the continuity of the sample criterion function in the consistency norm. The assumption controls the behavior of a term that explodes as the product of the functions $\phi(x)$ and $f(t)$ approaches zero.

Assumption C2 coupled with the differentiability property of the Laplace exponent with respect to λ and ρ, implies the Laplace exponent as well as its first partial derivative with respect to λ and ρ are Lipschitz continuous in λ and ρ. Define the following infima and suprema: $m_\phi = \inf \phi$, $M_\phi = \sup \phi$, $m_f = \inf f$, and $M_f = \sup f$. In the problem, $\lambda = \phi(x) f(u)(t-u)$, where $\phi : \mathcal{X} \to [m_\phi, M_\phi] \subset \mathbb{R}_+$, $f : \mathcal{T} \to [m_f, M_f] \subset \mathbb{R}_+$, and $t \in \mathcal{T}$. Although the partial derivatives of Ψ are continuous on $[m_\phi, M_\phi] \times [m_f, M_f] \times \mathcal{T}$, the range of λ is not closed, so that the partial derivatives are not bounded unless C2 holds. Note that C2 holds for both gamma and compound Poisson processes.

Assumption C4 is needed in order to derive the bracketing number of the class of functions indexing the criterion function. For the gamma and compound Poisson processes, assumption C4 holds automatically, see A.1.

Theorem 7 *Under Assumptions C0-C4 above*

$$||\widehat{\alpha}_n - \alpha_0||_\infty = o_p(1) \text{ as } n \to \infty$$

Proof. The proof can be found in Appendix A.14. ■

Monte Carlo studies for three separate studies have been performed; simulation results can be found in Appendix B. In the first study, the data was generated and estimated by the DH model. In the second study, the data was generated by the DH specification, but the model was estimated by the MPH with gamma heterogeneity. In the third study, the data was generated by the MPH with gamma heterogeneity, but estimated with the DH model. A summary of data generating processes (DGP) and estimating models is given below:

Study	DGP Hazard	Hazard of Estimating Model
1	(DH) $\phi(x)\int_0^t f(u)\,dZ_\rho(u)$	(DH) $\phi(x)\int_0^t f(u)\,dZ_\rho(u)$
2	(DH) $\phi(x)\int_0^t f(u)\,dZ_\rho(u)$	(MPH) $\phi_M(x)\lambda(t)z$
3	(MPH) $\phi_M(x)\lambda(t)z$	(DH) $\phi(x)\int_0^t f(u)\,dZ_\rho(u)$

where $\{Z(t)\}_{t>0} \sim Gamma(\rho t, \nu)$ and $z \sim Gamma(\rho, \nu)$.

Under correct specification, the estimators for (ϕ, f, ρ) perform quite well on average. There is individual variation in the estimators, particularly for the weight function $\{f(u)\}_0^t$. The standard error for $\widehat{\rho}$ is relative small to $\widehat{\rho}$.

Under the first misspecification analysis, the true DGP is the DH model with stochastic gamma heterogeneity with the model being estimated by the MPH with gamma distributed heterogeneity. The MPH estimator for the function $\phi(x)$ performs very well on average when the baseline hazard is estimated by a flexible function, i.e. by a polynomial spline. When the baseline hazard is parametrized as the Weibull function (which is usually done in practice), the MPH estimator for $\phi(x)$ is performing much worse, with the estimator showing positive bias in the tail. The MPH estimates of the survival function perform very poorly, showing a large negative bias. This may be because all the variation in t is picked up by the integrated baseline hazard and by the rate of the distribution of the gamma random variable; both of these estimators are very large. The bias is smaller when the baseline hazard is flexible rather than when it is parametrized.

For the second misspecification analysis, the true DGP is the MPH with gamma heterogeneity while the estimating model is the DH model with gamma heterogeneity. The DH estimators of $\phi(x)$ perform well on average, but they have high individual variance (relative to when the model is not misspecified). The estimate of the survival function shows a negative bias, but much smaller than that of the first misspecification analysis. In this case, both the rate parameter of the gamma process and the weight function $\{f(u)\}_0^t$ pick up the variation in the baseline hazard.

4 Unemployment Duration with Sunk Costs

4.1 The Model

This section presents a one-sided empirical job search model in which unemployment duration is defined as the first time a stochastic process hits a random threshold. Individuals exit unemployment when accumulating losses due to unemployment cross over a self-imposed spending limit on how many losses they are able and willing to sustain during the unemployment spell. The self-imposed limit can be thought of as a reference point. Under certain conditions, the survival function resulting from this model is that associated with the hazard of (1).

Let $\{Z(t)\}_{t\geq 0}$, with $Z(0) = 0$[11], denote monetary and non-monetary losses that accumulate in an uncertain way during the spell of unemployment. The process $\{Z(t)\}_{t\geq 0}$ is an adapted, increasing, right-continuous process with independent and stationary increments on a filtered probability space. Losses can be thought of as damage accumulating over time in a sequence of increments, whose size is governed by the Lévy measure of the stochastic process. It is assumed that losses have the same distribution for all individuals and that losses are unobserved by the econometrician.

In this set-up, there is one job offer per period and jobs are of different types. Rejecting an offer that may be preferred to another has a higher cost than rejecting the preferred offer. Rejection costs contribute to the process of losses due to unemployment.

[11] For now, it is assumed individuals have the same initial condition: They begin the unemployment spell with zero monetary and non-monetary losses due to unemployment.

At the beginning of the unemployment spell, individuals are endowed with a resource, R, denoting the total amount of losses individuals are willing (or able) to sustain during the spell of unemployment. The resource is a non-negative random variable, independent of accumulating losses, with a continuous distribution function G_R. It is a self-imposed spending limit, which can be thought of as a reference point.

Total losses (are perceived to) accumulate during the spell of unemployment at the rate $\int_0^t f(u)\, dZ(u)$. The weight function $\{f(u)\}_{u>0}^{t}$ is positive, right-continuous, and bounded away from zero. It accounts for how individuals weigh past losses and it can be used to test whether the sunk cost effect may weaken over time. For example, if $\{f(u)\}_{u>0}^{t}$ is increasing, individuals put more weight on more recent losses, and the sunk cost effect weakens over time. The rate of accumulation is further weighted by a function of observed covariates[12], $\phi(x)$. The rate at which perceived losses accumulate is modeled as (1) so that so that $\phi(x)$ has a scaling effect on the rate.

Total losses during the spell of unemployment are given by:

$$H\left(t|x, \{Z(t)\}_{t\geq 0}\right) = \int_0^t \left[\phi(x) \int_0^s f(u)\, dZ(u)\right] ds$$

Individuals exit unemployment when perceived total losses are greater than the threshold R, that is when $H\left(t|x, \{Z(t)\}_{t\geq 0}\right) > R$. The survival function for this model is given by:

$$\begin{aligned} P(T > t|x) &= P\left(H\left(t|x, \{Z(t)\}_{t\geq 0}\right) < R\right) \\ &= \int_0^\infty P(R > h)\, dF_t(h) \\ &= E_Z\left[1 - G_R\left(H\left(t|x, \{Z(t)\}_{t\geq 0}\right)\right)\right] \end{aligned}$$

where $F_t(.)$ is the distribution function of $\{Z(u)\}_0^t$. When R is assumed to be exponentially distributed with parameter one[13], the survival function becomes:

$$P(T > t|x) = E \exp\left[-H\left(t|x, \{Z(t)\}_{t\geq 0}\right)\right] \tag{20}$$

which is the expression for the survival function introduced in Section 2.

The resource level is based on financial wealth. Empirical studies have shown that the distribution of financial wealth among the low and middle class in the USA (or the lower 95% of the population) follows the exponential distribution, see, among many others, (J. Gruber 2001), (A. Dragulescu & V.M. Yakovenko 2001), (R. Lopez-Ruiz, J.L. Lopez & X. Calbet 2011). Note that in the empirical study, the sample is composed of

[12] Such as age, education level, marital status, etc.

[13] Note that the distribution can be allowed to be exponential with parameter λ. In this case, λ cannot be identified as it can be absorbed by the mean of the distribution of $\{Z(u)\}_0^t$. In order to identify λ, one has to parametrize fully the distribution. Else, in order to keep the previous identification results intact, one has to assume λ to be known.

low and middle class individuals. Thus, assuming R is exponentially distributed is justified.

In conclusion, unemployment duration is given by the first time perceived sunk losses cross over a random threshold with the standard exponential distribution.

The model introduced departs from existing models of unemployment duration in several ways:

(1) Sunk costs affect the decision of individuals to continue searching.

(2) Individuals weigh their losses differently. For example, some individuals may emphasize earlier losses or may be more prone to become more sensitized to further losses by earlier losses[14] (see (R. Thaler & E. Johnson 1990)). This type of behavior has been observed when individuals (firms) set self-imposed spending limits (budgets) (see (C. Heath 1995), (R. Thaler 1980), (J. Brockner & J.Z. Rubin 1985)).

(3) The probability of exiting unemployment is driven by three elements: the reaction of individuals to accumulating losses, their self-imposed limits on losses to be sustained, and their history of losses. In contrast, the exit probability in standard models is driven by the exogenous distribution of offers evaluated at a reservation wage.

(4) Long term unemployment results because individuals do not perceive their losses as "pressing enough," or because they do not have a high enough realization of the loss due to unemployment. In the framework of this paper it could be that individuals do not put enough effort in finding a job because they do not have a high enough realization of the loss due to unemployment.

4.2 Theoretical Background

The motivation for the framework presented in this section comes from several sources: A recent experimental study by (M. Brown, C. Flinn & A. Schotter 2011) (henceforth BFS), prospect theory, and the literature on labor supply with reference dependence.

In an experimental study, BFS find that reservation wages are declining even in a stationary environment[15] for which they propose two behavioral explanations[16]. First, when facing accumulating monetary costs, searchers may fail to recognize previously incurred costs as sunk. Second, searchers may experience nonstationary subjective costs of search, such as discouragement or the uncertainty of waiting; these costs accumulate during the spell of unemployment. The first explanation proposed by BFS has been previously proposed by (Thaler 1980) under the name of the sunk cost fallacy. The second explanation has been partially studied by the behavioral economics literature dealing with how individuals respond to sunk costs when they had invested time rather than money in an endeavor (see (D. Soman 2001)). The idea that unhappiness

[14] Those for whom earlier losses loom more important may quit searching faster for nothing more but frustration at what has already been lost. Such individuals may be regarded as being more averse to loss: They may become more easily frustrated at losses incurred and may want to stop incurring additional losses.

[15] A stationary environment is one in which one of the following exogeneous parameters is constant through time: the unemployment benefit, the arrival rate of job offers, or the probability of drawing a new wage offer. The reservation wage is declining if either one of the three exogeneous parameters is decreasing during unemployment. See Theorem 2 in (G.J. Van den Berg 2001).

[16] The effect of either of the two explanations is that individuals may accept lower wage offers in response to higher costs and thus exit unemployment earlier than it is predicted by standard search models.

and stress accumulate during the spell of unemployment and that they affect decisions has been documented by the literature on happiness and unemployment (see (A. E. Clark & A.J. Oswald 1994)).

The idea that sunk costs affect decisions under risk was first formalized by (D. Kahneman & A. Tversky 1979) in their prospect theory. Prospect theory asserts that sunk costs affect behavior by either strengthening commitment to an activity or by increasing the probability that the activity will be abandoned[17]. In prospect theory, the timing of sunk costs matters. For example, (Arkes & Blumer 1985) and (J.T. Gourville & D. Soman 1998) found that the further in the past the cost occurred, the least likely it is for it to influence present decisions[18]. Prospect theory also mentions the importance of self-imposed limits. The existence of self-imposed spending limits has been documented by a large number of papers in behavioral economics and in the literature studying mental budgeting (Heath 1995). Empirically, it has been noted that these self-imposed limits are arbitrary and do not follow any decision rule from economic theory (H. Shefrin & M. Statement 1985).

First hitting-time models in which individuals solve an optimal stopping problem driven by a stochastic process, and models in which labor supply decisions depend on a reference point are not new. For example, (R. Shmer 2008) models the difference in lifetime utility in the best available job and lifetime utility in unemployment as a Brownian motion. The duration of unemployment is determined as the first time this Brownian motion hits the zero threshold. Likewise, in the literature on labor supply, (H.S. Faber 2008) models a taxi driver's continuation value of driving as depending on whether the accumulated income is greater than a reference level, which is assumed to be a daily income target.

4.3 Data, Empirical Specification, and Results

The model introduced is evaluated empirically on NLSY79 data on the first spell of unemployment. The data is described in detail in Appendix B.4.

The NLSY79 was the preferred data set since it contains complete work history information for a specific cohort, available on a weekly basis regardless of the period of non-interview. As the model assumes continuous time and uncensored spells, the relatively high frequency of the data as well as being able to construct a complete work history for each surveyed individual are useful characteristics of the data set.

Completed spells are defined as transitions out of employment to unemployment and then back to employment. Unemployed workers are those who did not work at all during the survey week but have searched for a job in the four weeks prior to the survey and during the survey week. Out-of-labor force individuals and individuals for whom there is no information about whether or not they were actively searching during unemployment are dropped from the sample.

[17] Further discussion can be found in (Thaler 1980), (K. McKean 1985), (H. Arkes & C. Blumer 1985), and (C.A. Kogut 1990).

[18] (Arkes & Blumer 1985) tested in a real-money study whether a family was more likely to go to a basketball game if they purchased the tickets a week before rather than a year before. The study found that the sunk cost effect disappeared during a six to nine month period.

The first spell was selected for two reasons: First, it has been documented that future spells of unemployment are significantly affected by the first unemployment spell. Thus it seems appropriate to study what drives the first unemployment spell. Additionally, including future spells into the analysis would mean allowing for dependence between survival times, which cannot be dealt with by the framework presented in this paper. Second, the model presented requires uncensored survival times. Since individuals are observed when they become unemployment for the first time until they exit unemployment, the first unemployment spell is not censored.

4.3.1 Empirical Specification

For the estimation procedure, 1 is modeled as

$$h\left(t|x_{ind}, x_w, \{Z(t)\}_{t\geq 0}\right) = \phi_1(x_{ind})\,\phi_2(x_w)\int_0^t f(u)\,dZ(u)$$

where x_w is initial net wealth and x_{ind} is a vector of observed variables such as age, number of children, and either marital status or education group, depending on the analysis performed. The function $\phi_1(x_{ind})$ is parametrized as:

$$\log\phi_1(x_{ind}) = \beta_0 age + \beta_1 age^2 + \beta_2 kids + \beta_3 y$$

where $y \in \{marital\ status,\ education\ group\}$. Functions $\log\phi_2(x_w)$ and $\log f(t)$ are approximated by second degree splines with polynomial basis. $\log\phi_2(x_w)$ has one knot located at the 0.5 quantile, while $\log f(t)$ has nine knots located at the $\{0.1,\ 0.2,\ 0.3,\ ...,0.8,\ 0.9\}$ quantiles.

The process modeling costs of unemployment is assumed to be the gamma process with a scale parameter[19] $\nu = \left(\frac{1}{2}\right)10^{-6}$. Although the distribution of unobserved heterogeneity does not follow from economic theory, the gamma process was chosen for two reasons. First, in standard duration models, unobserved heterogeneity is usually parameterized as a gamma distributed variable. As such, the present analysis assumes that loss increments have a gamma distribution. Second, the gamma process implies losses are increasing gradually, in small and frequent increments rather than in large jumps that happen at relatively distant points in time.

The estimation procedure is performed on the following groups: (i) married men, (ii) men who are not married; (iii) married women, (iv) women who are not married; (v) men with more than twelve years of education (vi) men with less than twelve years of education; (vii) women with more than twelve years of education and (viii) women with less than twelve years of education. The optimization procedure used is Matlab fmincon[20].

[19] The scale parameter, which models the magnitude of the costs, was not normalized to one in order to be consistent with the scaling of the wealth data.

[20] Since Matlab approximates $\exp(c)$ by infinity whenever c is large,both the wealth data (which is of order 10^6) and the duration data (which is of order 10^2) are scaled down in order to avoid the approximation by infinity issue. The wealth data is scaled down by a factor of $\left(\frac{1}{2}\right)10^{-6}$ and the duration data is scaled down by a factor of 10^{-2} such that net wealth, $x \leq 2$ and

For comparison purposes, the MPH with gamma heterogeneity is estimated on the same data. The hazard function for the MPH is specified as

$$h^M(t|x) = \phi^M(x)\,\lambda(t)\,\eta$$

where $\phi^M(x) = \exp(x'\beta)$, with $x = [x_{ind}\ x_w\ y]$. The variables entering x are the same as those used in the DH specification. The random variable η is assumed to be gamma distributed with scale equal to that of the DH model and rate ρ, to be estimated. The baseline hazard is modeled in two different ways. For Case S, $\log\lambda(t)$ is modeled as a spline polynomial of second degree, while for Case P, $\lambda(t)$ is parametrized as the generalized Weibull function, (70). The parameters of interest are (β, λ, ρ). The baseline hazard is parametrized in order to illustrate the difference between the survival functions obtained when the following models are fitted to the data: DH, MPH with spline baseline hazard, and MPH with generalized Weibull baseline hazard.

Finally, a nonparametric estimator of the survival function is also fitted to the data to provide a benchmark for the comparison of the survival functions from the three models. The nonparametric estimator is the following generalized additive estimator:

$$\begin{aligned} S(t|x) &= \frac{1}{3}\sum_{\alpha=1}^{3} f_\alpha(x_\alpha) \\ f_\alpha(x_\alpha) &= \frac{1}{n}\sum_{j=1}^{n} 1(T_j > t)\,\frac{K\left(\frac{x_{\alpha j}-x_\alpha}{h}\right)}{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_{\alpha i}-x_\alpha}{h}\right)} \end{aligned}$$

for each t in $\mathcal{T}$, and where $x = [x_{ind}\ x_w\ y]$, $h = n^{-1/5}$, and $K(.)$ is the Epanechnikov kernel. For each observed variable x_α, $\alpha = 1, 2, 3$, represents the mean of that variable.

4.3.2 Summary of Results

The results are presented in Appendix B.5. It is found that men and women react differently to the costs of unemployment conditional on their education level and marital status. Conditional on education, transition rates for men are mostly driven by differences in weight functions. For women, the differences are driven by both weight functions and net wealth effects, see Figures 8 and 9. Conditional on marital status, the differences in hazard rates for men seem to be driven by differences in net wealth. For women, differences are mainly driven by weight functions, see Figure 11.

Table 5 presents the estimated rate of accumulation, ρ. The interpretation of this parameter is that during a time interval of length Δt, the expected number of losses are $\rho\Delta t$. It can be seen that high education types have the highest rate of accumulation, while low education types have the lowest rates.

duration, $t \leq 2$, respectively.

Conditional on education, Figure 8 shows that individuals with a college degree are more sensitive to earlier losses than those with a lower education level. Both men[21] and women with a college degree put more weight on earlier losses. Women with a high school degree show a decreasing sunk cost effect since they put more weight on more current losses. Higher net wealth decreases the probability of exiting unemployment for all categories except for women with a high school degree[22].

Conditional on being married, men put more emphasis on sunk costs than women[23]. Women who are not married emphasize earlier losses more than women who are married, while the opposite happens for men. The probability of exiting unemployment is decreasing in net wealth, more so for those who are not married than for those who are. Men who are married exit unemployment faster than those who are not, while for women the situation is reversed, see Figure 12.

The number of children decreases the exit probability for both men and women with a high school degree, while it increases it for those with a college degree. Conditional on education level, age has a negative effect on transition rates for all categories, with a nonlinear effect for both men and women with a high school degree. Conditional on marital status, age has a negative effect on transition rates, with a nonlinear effect for men who are not married and for married women. Education has a negative effect on exit probabilities for men who are married, while having a positive effect for all the other categories. Some of these results are corroborated by the MPH. The biggest differences between the DH and the MPH estimators are related to the effects of age for men with a high school degree and for men who are not married. The two models obtain different nonlinear relationships of transition rates with respect to age, although the turning points at which transition rates change sign with respect to age are outside the range of ages in the data. The rate parameter for the gamma distribution for the MPH is very large in comparison to that in the DH, possibly because the MPH rate parameter captures time effects which are more reasonably modeled by the DH.

Regarding survival functions (see Figures 13 and 14), the DH and the MPH produce different results depending on how flexible the MPH baseline hazard is allowed to be. When the baseline hazard is estimated by a spline function, the MPH, the DH, and the nonparametric estimator produce almost identical survival functions, for both men and women[24]. When the baseline is the generalized Weibull function, the MPH produces survival probabilities that are greater (especially in the tails) than those of the DH or of the nonparametric estimator. The MPH may overestimate the probability of survival since it treats unobserved heterogeneity as being fully correlated over the spell duration. The MPH fits the survival probability by fitting the average of the duration outcomes; the DH fits the survival probability by taking into account that those with longer durations may have a smaller realization of the stochastic process.

[21] Bootstrap results for estimators for men with a college degree can be found in Appendix B.5.

[22] This result is in line with the MPH estimator for the effect of net wealth on transition rates for women with a high school degree.

[23] Married women seem to have a turning point in the weight function at approximately 60 weeks of unemployment, although it should be noted that this could be a small sample effect, as there are few observations with such a long duration of unemployment.

[24] For the DH function for women who are not married (figure 14), the DH survival function has a kink. This is due to there being only four observations with a duration greater than 75 weeks.

4.3.3 Interpretation of the Results

Differences in weight functions can be regarded as being generated by differences in expectations, aspirations, costs of adjustment, and social norms. Outside options and work mobility can also affect perceptions of losses as well as the rates of loss accumulation.

Regarding expectations, high education types have invested in their human capital as insurance against negative labor market shocks. As such, they may perceive unemployment as a greater shock to their self-worth[25]. Additionally, reaction to losses may be stronger when the level of material benefits falls behind expectations rather than when the level of net loss is the greatest[26]. High education types may also be the ones with different career aspirations and unemployment may constitute a greater set-back in terms of their aspirations.

Social norms may affect the way individuals react to losses and the way they adapt to new situations. It may be more acceptable for women to stay at home than for men (e.g. the breadwinner syndrome) and women may have different career expectations. For example, women with a high school degree may not experience as many losses as other categories partially because they may not be as career oriented as women with a college degree, while in terms of social comparisons, they may not experience the stigma associated with staying at home that men may.

Perceived outside options, work mobility, or being associated to sectors with high turn-over rates may also influence the weight put on losses as well as the speed with which losses accumulate. Having more options available in terms of possible employers or knowing that the duration of employment in a future job is short may make one less sensitive to losses that happen earlier.

The rate at which costs accumulate can also influence the way individuals weigh their losses. When losses accumulate at a faster rate, individuals may not have enough time to adapt, resulting in more emphasis on earlier losses[27].

4.4 Counterfactuals

When losses accumulate at a slower rate, individuals may not perceive losses incurred earlier as pressing, resulting in an increasing weight function. When the rate of accumulation is higher, individuals may have less time to adjust and earlier losses may be more important, resulting in a decreasing weight function. In order to test this hypothesis, the following two counterfactuals are performed.

First, both the weight function and the function of observed covariates are estimated for women with

[25] Fact which is corroborated by both anecdotal evidence and psychological studies (see (Clark & Oswald 1994)).

[26] For example, in political science, Davies (1962) and (T.R. Gurr 1970) note that the likelihood of (political) violence is greater not under conditions of greatest suffering, but instead when the level of material benefits or rate of improvement falls behind expectations.

[27] This is in like with the framework of Chetty and Szeidl (2010). In their model, those who abandon or update consumption commitments more quickly or frequently when facing large shocks do not exhibit reference dependence. Then it could be the case that high education types update their consumption commitments faster than lower education types.

a high school degree using the estimate for the rate of accumulation for women with a college degree[28]. As can be seen in figure 15, when women with a high school degree experience losses at a higher rate of accumulation, they tend to weigh losses similarly to women with a college degree. The function of net wealth is still increasing but flatter.

Second, the net wealth function and the behavioral function were estimated for women with a college degree using the rate for men with a college degree. When women have a higher rate of accumulation, women tend to put more weight on earlier losses, resulting in a steeper weight function. Their probability of exiting unemployment also starts decreasing more in initial net wealth, see figure 16.

5 Conclusion

The paper presented, identified, and estimated a new class of duration models in which unobserved heterogeneity is time varying. The analysis is restricted to positive Lévy processes for which the distribution of the increments has finite moments. The identification of the parameters of interest varies from the nonparametric to the parametric, depending on how the effects of time can be differentiated. The stochastic integral modeling unobserved heterogeneity exhibits long-memory: All changes in heterogeneity are accumulated and the past is never forgotten. Such a specification might exaggerate the effects of unobservables when the past is gradually forgotten. It would be useful to extend the framework to allow for processes with short-memory, such as the moving average. One possible formulation of such a hazard function is

$$h\left(t|x, \{Z(u)\}_0^t\right) = \phi(x) \int_0^t f(t-u)\, dZ(u)$$

However, note that the weight function cannot be identified unless the distribution of the stochastic process is entirely parametrized. The function of observed covariates can still be identified by the arguments presented in the paper.

It would also be interesting to explore the possibility of extending the framework to semi-martingales, i.e. to allow unobserved heterogeneity to be modeled by positive transformations of general Lévy processes. This model would be less restrictive in that it would not require positive duration dependence at the individual level. When considering semi-martingales, the formulation of the survival function given in this paper will not hold exactly since the stochastic process resulting from the positive transformation applied to unobserved heterogeneity process does not have independent increments anymore. It is conjectured the identification strategies employed in this chapter may still apply after an appropriate formulation of the survival function, which needs to include a term for the quadratic variation of the process.

The estimation procedure proposed is sieve maximum likelihood, which provides a practical approach for the joint semiparametric estimation of infinite and finite dimensional parameters. The functions are

[28] The rate of accumulation for women with a college degree is about 15 times greater than that for women with a high school degree

approximated by positive transformations of polynomial splines of the second order and the position of the knots for the two splines is pre-specified. The estimators are shown to be consistent and their performance is shown through Monte Carlo studies.

The paper also introduced a first time hitting model in which unemployment duration is defined as the first time stochastically increasing losses due to unemployment reach a threshold level. The threshold models a self-imposed limit on the amount of losses individuals are willing to undertake during the spell of unemployment. The model is applied to NLSY79 data on 1240 first spells of unemployment. Transition rates out of unemployment are estimated under the assumption that unobserved losses are realizations from a gamma process, implying losses accumulate gradually, in small jumps. The application shows that men and women have different perceptions of losses and different loss accumulation rates, conditional on their education level and marital status.

A Appendix section

A.1 Lévy processes

Lévy processes are stochastic processes whose sample paths are right-continuous with left limits at every point t, and which have independent and stationary increments. The initial condition for Lévy processes is that $Z(0) = 0$. Following (J. Bertoin 1996), the formal definition of a Lévy process is:

Definition 8 (Lévy Process) *Let P be a probability measure on $(\Omega, \mathcal{F})$. We say that $\{Z(u)\}_0^t$ is a Lévy process for $(\Omega, \mathcal{F}, P)$ if for every $s, t \geq 0$, the increment $Z(s+t) - Z(t)$ is independent of the process $\{Z(v)\}_0^t$ and has the same law as $\{Z(s)\}_0^t$.*

There are two types of Lévy processes: infinite activity and finite activity. Infinite activity processes have an infinite number of jumps in a finite time interval, while finite activity processes have a finite number of jumps in a finite time interval. The number of jumps is countable for both finite and infinite activity processes. Non-negative Lévy processes, also known as subordinators, take values in $\mathbb{R}_+$, which implies that their sample paths are increasing. Figure 1 illustrates the differences in the sample paths of these two types of subordinators.

A Lévy process is uniquely characterized by its Lévy measure, $\mathcal{L}(dx)$, which is a measure on $\mathbb{R}^d - \{0\}$ such that $\int \left(1 \wedge |x|^2\right) \mathcal{L}(dx) < \infty$. The Lévy measure is indexed by a finite dimensional parameter $\theta \in \mathbb{R}^d$, where d represents the dimension of the parameter θ. Usually when $d = 2$, θ represents the scale and the rate of the jumps. For a subordinator, the Lévy measure can be interpreted as the distribution of jump sizes.

Example 9 *Let z be a random variable with the gamma distribution with shape parameter ρ and scale parameter ν, i.e. $z \sim Ga(\rho, \nu)$.*

Let $\{Z(u)\}_0^t$ be a stationary gamma process. Then $\{Z(t)\}_{t>0} \sim Ga(\rho t, \nu)$. The Lévy measure is the gamma measure, with shape parameter ρt, and scale parameter ν. Then $\theta = (\rho t, \nu)$, with $d = 2$.

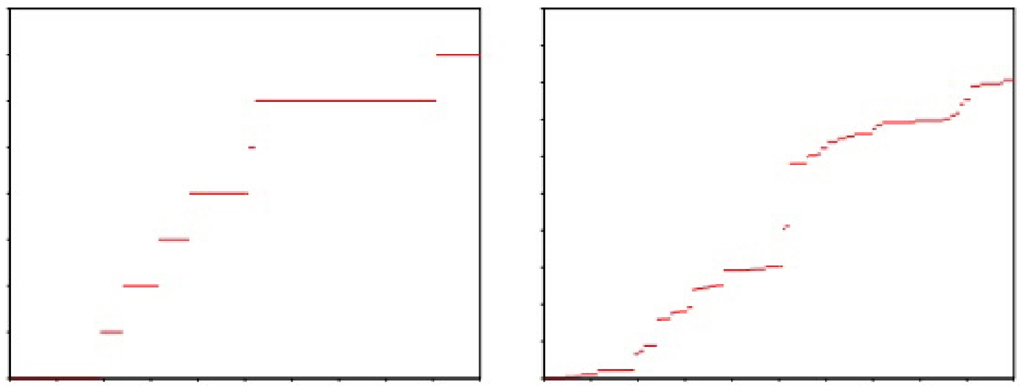


Figure 1: Sample paths of the compound Poisson process (left) and of the gamma process (right)

Subordinators are characterized uniquely by their Laplace exponent. The definition of the Laplace exponent is:

Definition 10 (Laplace Transform) *The Laplace transform*[29] *of the subordinator* $\{Z(u)\}_0^t$ *is defined as*

$$E\left(\exp\left[-\lambda Z(t)\right]\right) = \exp\left[-t\Psi(\lambda)\right], \ \lambda \geq 0$$

where $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ *is called the Laplace exponent.*

The following property[30] of Laplace exponents is key for the derivations in this paper.

Property 1 A function $\Psi(\lambda)$ is the Laplace exponent of a subordinator if, for $\lambda \geq 0$:

(i) $\Psi(\lambda)$ is infinitely differentiable with respect to $\lambda \geq 0$;

(ii) $\Psi(0) = 0$; and

(iii) $(-1)^n \frac{\partial^n}{\partial \lambda^n} \Psi(\lambda) \leq 0$ for every n.

Definition 11 (Cumulants) *The* j^{th} *cumulant of the subordinator* $\{Z(t)\}_{t \geq 0}$ *is defined as*

$$\left.\frac{d}{d\lambda^j} \Psi_j(\lambda)\right|_{\lambda=0} = k_j \tag{21}$$

Cumulants are the coefficients in the Taylor expansion about the origin of the log of the moment generating function. As such, there is a one-to-one relationship between cumulants and moments. For example, the first cumulant of the process is the mean of the distribution, while the second cumulant is the variance.

[29] The Laplace transform for a subordinator is the moment generating function of the subordinator.

[30] See (A. Gnedin & J. Pitman 2008).

This section is finished off with a few examples of subordinators and their Laplace exponents.

Example 12 (Gamma Process) *Let $\{Z(t)\}_{t\geq 0} \sim Ga(\rho t, \nu)$. The Gamma process has an infinity of very small jumps in any time interval. As such, it is commonly used to model processes that take place gradually in time, such as erosion and wear-and-tear. The Laplace exponent of $\{Z(t)\}_{t\geq 0}$ and its first two cumulants are, respectively:*

$$\begin{aligned}
\Psi(\lambda) &= \log\left(1+\frac{\lambda}{v}\right)^{\rho} \\
k_1 &= \frac{\rho}{\nu} \\
k_2 &= -\frac{\rho}{\nu^2}
\end{aligned} \tag{22}$$

If Z is a gamma process with scale parameter ν, then νZ is a gamma process with scale parameter 1. The law of Z is the same as that of νZ, which means that the distribution of the gamma process is determined up to the scale parameter ν (see (Tsilevich, Vershik & Yor 2001)).

Example 13 (Standard Compound Poisson Process) *The Poisson process has a finite number of jumps in any time interval. It plays an important role in risk analysis and it is used to model events that occur randomly in time, such as arrivals at a queue, shocks to a market, accidents, and natural disasters. For example, suppose events can happen at any time. If event arrivals are independent of one another and past arrivals do not influence future arrivals, then κt events are expected in an interval of length t, that is the number of events follows a Poisson process with rate κ.*

The jumps are distributed as $Ga(\rho, \nu)$. The standard compound Poisson process is the sum of the jumps up to time t. Then the Laplace exponent and the first three cumulants are given by, respectively:

$$\begin{aligned}
\Psi(\lambda) &= \kappa\left(1-\left(\frac{\nu}{v+\lambda}\right)^{\rho}\right) \\
k_1 &= \rho\frac{\kappa}{\nu} \\
k_2 &= -\rho(1+\rho)\frac{\kappa}{\nu^2} \\
k_3 &= \rho(1+\rho)(2+\rho)\frac{\kappa}{\nu^3}
\end{aligned} \tag{23}$$

Example 14 (Compound Poisson Process) *The compound Poisson process is a standard compound Poisson process where the jumps have a random distribution. Let the jumps be denoted by the random variable J with a random distribution with Laplace transform $L(s) = Ee^{-sJ}$. The Laplace exponent of the compound Poisson subordinator is given by*

$$\Psi(\lambda) = \kappa(1-L(s))$$

$\Psi(\lambda)$ is bounded for all $\lambda \geq 0$ if and only if the process is the compound Poisson process.

Example 15 (Stable Subordinator) *Let $\{J_i\}_1^n$ be mutually independent random variables with a common distribution D and let $S_n = \sum_{i=1}^n J_i$. Then D is stable if for each n there exist constants $c_n > 0$ and γ_n such that S_n has the same distribution as $c_n J + \gamma_n$ and D is not concentrated at the origin. For example, the normal distribution centered at zero expectation is strictly stable, i.e. $\gamma_n = 0$, with $c_n = \sqrt{n}$. The stable subordinator is a Lévy process with increments that have the strict stable distribution. The Laplace exponent has the following form:*

$$\Psi(\lambda) = \nu\lambda^{\rho}$$

where $\nu > 0$ is called the intensity parameter, $\rho \in (0, 1)$ is the exponent. The first two cumulants do not exist as the Laplace exponent is not differentiable at 0.

A.2 Survival Function: The Exponential Formula and the Laplace Exponent

A.2.1 The Exponential Formula

The hazard function (1) is defined in terms of a stochastic process. In order to apply the usual exponential formula for the survival function, $\{Z(u)\}_0^t$ needs to satisfy certain conditions guaranteeing that $P\left(T \leq t | x, \{Z(u)\}_0^t\right)$ is a martingale. If one of the conditions below is satisfied then the exponential formula for the survival function can be applied.

Condition 16 *T given $\left(x, \{Z(u)\}_0^t\right)$ is a random variable with an absolutely continuous distribution function (with respect to the Lebesgue measure).*

Remark 2 *This condition excludes jumps of the conditional survival function induced by changes in the information set, that is, changes in x and in the filtration of $\{Z(u)\}_0^t$. It allows us to work with density functions.*

Condition 17 *(i) $P\left(T \leq t | x, \{Z(u)\}_0^t\right)$ is predictable with respect to $\left(x, \{Z(u)\}_0^t\right)$ and*

(ii) For all $t > 0$, $P\left(T \leq t | x, \{Z(u)\}_0^t\right) = P\left(T \leq t | x, \{Z(u)\}_0^\infty\right)$ a.s.

Remark 3 *This condition holds if the processes generating x and the realizations of $\{Z(u)\}_0^t$ are exogenous. The condition rules out the possibility that future values of the stochastic process, not known at time t, affect the probability of exit at time t. It also excludes contemporaneous feedback between $Z(t)$ and duration. This sort of feedback is not considered since we want to exclude the case that knowledge of $Z(t)$ and its future realizations may tell us whether the event $\{T \leq t\}$ happened or not. If this feedback was not excluded, the probability $P\left(T \leq t | x, \{Z(u)\}_0^t\right)$ would be either 0 or 1. For a proper analysis of duration, $P\left(T \leq t | x, \{Z(u)\}_0^t\right)$ needs to be a value between 0 and 1.*

When the observed covariates vary with time, processes $\{X(u)\}_0^t$ and $\{Z(u)\}_0^t$ need to be independent. As long as the two processes are independent and one of the two conditions above holds, the exponential formula for the survival function can be applied.

A.2.2 The Laplace Exponent

Below, we present the derivation of the survival function (8). Let $F(u,t,x) = \int_u^t \phi(x) f(u)\, ds$ be square integrable with respect to the distribution of $\{Z(u)\}_0^t$. Using that $\{Z(u)\}_0^t$ has independent increments and letting $0 = u_{n,0} < u_{n,1} < \ldots < u_{n,n} = t$, $n = 1, 2, \ldots$ and a fixed $u_{n,j}^* \in [u_{n,j-1}, u_{n,j}]$, $j = 1, 2, \ldots, n$, obtains in mean square limit:

$$S(t|x) = E_Z \exp\left[-\int_0^t F(u,t,x)\, dZ(u)\right]$$

$$= E_Z \exp\left[-\lim_{n\to\infty}\sum_{j=1}^n F\left(u_{n,j}^*, t, x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right] \tag{24}$$

$$= E_Z \lim_{n\to\infty} \exp\left[-\sum_{j=1}^n F\left(u_{n,j}^*, t, x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right] \tag{25}$$

$$= \lim_{n\to\infty} E_Z \exp\left[-\sum_{j=1}^n F\left(u_{n,j}^*, t, x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right] \tag{26}$$

$$= \lim_{n\to\infty} E_Z \prod_{j=1}^n \exp\left(-F\left(u_{n,j}^*, t, x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right) \tag{27}$$

$$= \lim_{n\to\infty} \prod_{j=1}^n E_Z\left[\exp\left(-F\left(u_{n,j}^*, t, x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right)\right] \tag{28}$$

$$= \lim_{n\to\infty} \prod_{j=1}^n \exp\left(-\left(u_{n,j} - u_{n,j-1}\right)\Psi\left(F\left(u_{n,j}^*, t, x\right)\right)\right) \tag{29}$$

$$= \lim_{n\to\infty} \exp\left[-\sum_{j=1}^n \left(u_{n,j} - u_{n,j-1}\right)\Psi\left(F\left(u_{n,j}^*, t, x\right)\right)\right] \tag{30}$$

$$= \exp\left(-\int_0^t \Psi\left(\int_u^t f(u,x)\, ds\right)\right) du \tag{31}$$

$$= \exp\left(-\int_0^t \Psi\left(f(u,x)(t-u)\right)\right) du \tag{32}$$

(25) holds since $\exp(.)$ is a continuous function, so that:

$$\exp\left(-\lim_{n\to\infty}\sum_{j=1}^{n}G_j\right)=\lim_{n\to\infty}\exp\left(-\sum_{j=1}^{n}G_j\right)$$

(26) follows by the Bounded Convergence Theorem since:

$$\left|\exp\left(-\sum_{j=1}^{n}F\left(u_{n,j}^{*},t,x\right)\left(Z\left(u_{n,j}\right)-Z\left(u_{n,j-1}\right)\right)\right)\right|\leq 1$$

(28) follows by the independence of the increments, while (29) follows by the definition of the Laplace exponent for Levy processes. Since the process has independent increments it holds that:

$$\begin{aligned}E_Z\exp\left[-u\left(Z\left(t\right)-Z\left(s\right)\right)\right] &= E_Z\exp\left[-uZ\left(t-s\right)\right]\\ &= \exp\left[-\left(t-s\right)\Psi\left(u\right)\right]\end{aligned}$$

which in our problem obtains (29).

The calculation is finished off by switching back to integral notation in (32).

A.3 Proof of Lemma 1

Let $H(t,x)$ be defined as in (10). As explained in the main text, (10) is a well defined Volterra integral equation of the first kind since $\lim_{t\to 0}H(t,x)=0$, which follows by $\lim_{t\to 0}S_0(t|x)=1$. By Assumption ID1:

$$\begin{aligned}\frac{\partial^2}{\partial t^2}H\left(t,x\right) &= H_{tt}\left(t,x\right)\\ &= \phi\left(x\right)f\left(t\right)\Psi_1\left(0\right)+\int_0^t\phi^2\left(x\right)f^2\left(u\right)\Psi_{11}\left(\phi\left(x\right)f\left(u\right)\left(t-u\right)\right)du \qquad (33)\end{aligned}$$

where $\Psi_1(0)=k_1$ is the mean of the distribution of $\{Z(u)\}_0^t$ as defined in Appendix A.1. By assumption ID2, k_1 exists and is finite. Solving (33) for $f(t)$ obtains a nonlinear Volterra equation of the second kind with unknown kernel:

$$f\left(t\right)=\frac{1}{k_1\phi\left(x\right)}\left[H_{tt}\left(t,x\right)-\phi^2\left(x\right)\int_0^t f^2\left(u\right)\Psi_{11}\left(\phi\left(x\right)f\left(u\right)\left(t-u\right)\right)du\right]$$

A.4 Proof of Theorem 2

Consider (11). Let $x = 0$ and $t \downarrow 0$. Solving the resulting expression for k_1 obtains:

$$k_1 = \lim_{t \to 0} \frac{1}{f(t,0)} H_{tt}(t,0) = \lim_{t \to 0} H_{tt}(t,0)$$

A.5 Proof of Theorem 3

Consider (11). Let $t \downarrow 0$. By Theorem 2:

$$\phi(x) = \frac{\lim_{t \to 0} H_{tt}(t,x)}{\lim_{t \to 0} H_{tt}(t,0)}$$

A.6 Example 1 and 2

A.7 The Gamma Process

Let the gamma process have rate ρt and scale ν with $0 < \underline{\nu} \leq \nu \leq \overline{\nu} < \infty$ and $0 < \underline{\rho} \leq \rho \leq \overline{\rho} < \infty$. The first moment is $k_1 = \frac{\rho}{\nu}$. Assumption ID2 is satisfied with $k_1 \in \left[\frac{\underline{\rho}}{\overline{\nu}}, \frac{\overline{\rho}}{\underline{\nu}}\right]$.Let the weight function $f(x,t)$ be such that $0 < f(x,t) \leq M < \infty$. The Laplace exponent is $\Psi(\lambda, k_1) = \rho \log\left(1 + \frac{\lambda}{\nu}\right)$, so that assumption ID6 is verified with

$$\left|\frac{\partial^3}{\partial \lambda^3} \Psi(\lambda, k_1)\right| \leq \frac{2\overline{\rho}}{\underline{\nu}^3} = B$$

A.8 The Compound Poisson Process

Let the process have scale parameter ν, rate parameter ρt, and expected number of jumps κ with $0 < \underline{\kappa} \leq \kappa \leq \overline{\kappa} < \infty$, $0 < \underline{\nu} \leq \nu \leq \overline{\nu} < \infty$ and $0 < \underline{\rho} \leq \rho \leq \overline{\rho} < \infty$. The first moment is $k_1 = \frac{\kappa\rho}{\nu}$ so that Assumption ID2 is satisfied with $k_1 \in \left[\frac{\underline{\kappa\rho}}{\overline{\nu}}, \frac{\overline{\kappa\rho}}{\underline{\nu}}\right]$. Let the weight function $f(x,t)$ be such that $0 < f(x,t) \leq M < \infty$. The Laplace exponent is $\Psi(\lambda, k_1) = \kappa\left(1 - \frac{\nu}{\nu+\lambda}\right)^{\rho}$, so that assumption ID6 is satisfied with

$$\left|\frac{\partial^3}{\partial \lambda^3} \Psi(\lambda, k_1)\right| \leq \frac{\overline{\kappa\rho}\,(\overline{\rho}+1)\,(\overline{\rho}+2)}{\underline{\nu}^3} \equiv B$$

A.9 Proof of Theorem 4

First, it is shown the solution exists in C_w^0. Then it is shown the solution lives in C_w^s by applying an induction argument on the smoothness parameter, s. The existence and uniqueness of the solution is shown via the Banach Fixed Point Theorem, for which we prove the operator is an inclusion and a contraction. In what follows, we use the result of Lemma 2 that $k_1 = \lim_{t \to 0} H_{tt}\left(t, 0, k_1^0\right)$.

The operator satisfies the inclusion $TC_w^0 \subset C_w^0$. The operator $(Tf)(t)$ is a continuous map from $\mathbb{R}_+$ into $\mathbb{R}_+$ for any $f(t) \in C_w^0$. By ID1, $\frac{H_{tt}(t)}{k_1} \in C_w^0$. Setting $x = 0$, we have additionally:

$$\frac{1}{k_1}\int_0^t f^2(u)\,\Psi_{11}(f(u)(t-u), k_1)\,du \in C^w$$

To show this, let $v = t - u$ and consider:

$$\begin{aligned}
& \left|\left|\frac{1}{k_1}\int_0^t f^2(u)\,\Psi_{11}(f(u)(t-u), k_1)\,du\right|\right|_{w,\infty} \\
\leq\; & M^2\frac{1}{k_1}\sup_{x,t}\frac{1}{w(t)}\int_0^t\left|\frac{\partial}{\partial\lambda}\Psi_{11}(\lambda v, k_1)\right||f(u)|\,du && (34)\\
=\; & M^2\frac{1}{k_1}\sup_{x,t}\left[\frac{1}{w(t)}\int_0^t w(u)\left|\frac{\partial}{\partial\lambda}\Psi_{11}(\lambda v, k_1)\right|\left|\frac{f(u)}{w(u)}\right|du\right] \\
\leq\; & \frac{1}{k_1}\sup_{x,t}\left[M^2B\frac{1}{w(t)}\int_0^t w(u)\,du\right]||f||_{\infty,w} && (35)\\
\leq\; & \frac{1}{3}\frac{\delta}{k_1}\,||f||_{\infty,w} < \infty && (36)
\end{aligned}$$

where (34) follows by assumption ID4, (35) follows by assumption ID5, (36) follows by assumption ID6 and by the definition of the weight function $w(t)$.

The second part of the Banach Fixed Point Theorem requires us to show the operator is a contraction. For $f, g \in C_w^0$ such that $f \neq g$, the following obtains:

$$\begin{aligned}
& ||(Tf)(t) - (Tg)(t)||_{\infty,w} \\
=\; & \left|w(t)^{-1}\left[(Tf)(t) - (Tg)(t)\right]\right| \\
\leq\; & \frac{1}{k_1}\sup_{t,x}\left[\frac{3M^2B}{w(t)}\int_0^t w(u)\left|\frac{1}{w(u)}(f(u) - g(u))\right|du\right] && (37)\\
\leq\; & \frac{1}{k_1}\sup_{t,x}\left[\frac{3BM^2}{w(t)}\int_0^t w(u)\,du\right]||f - g||_{\infty,w} && (38)\\
=\; & \frac{\delta}{k_1}\,||f - g||_{\infty,w} && (39)
\end{aligned}$$

where (37) follows by the calculation below, (38) follows by assumption ID3, and (39) follows by the way the weight function is defined. First, we present the calculation for (37) and then we present the calculation of the weight function such that (39) holds.

Consider first

$$
\begin{aligned}
& \left|f^2(u)\,\Psi_{11}(f(u)v,k_1) - g^2(u)\,\Psi_{11}(g(u)v,k_1)\right| \\
\leq\ & \left|f^2(u)\,\Psi_{11}(f(u)v,k_1) - g^2(u)\,\Psi_{11}(f(u)v,k_1)\right| \\
& + \left|g^2(u)\,\Psi_{11}(f(u)v,k_1) - g^2(u)\,\Psi_{11}(g(u)v,k_1)\right| \\
\leq\ & |f(u) - g(u)|\,|f(u) + g(u)|\,|\Psi_{11}(f(u)v,k_1)| \\
& + \left|g^2(u)\right|\,|\Psi_{11}(f(u)v,k_1) - \Psi_{11}(g(u)v,k_1)| \\
\leq\ & 3M^2B\,|f(u) - g(u)|
\end{aligned}
$$

where the last inequality follows by assumption ID4 and assumption ID5.

Note that (37) holds whenever the stronger inequality (38) holds. Then, in order to finish off the contraction mapping proof, an appropriate weight function $w(t)$ needs to be defined such that

$$
3BM^2\frac{1}{w(t)}\int_0^t w(u)\,du = \delta \tag{40}
$$

Once the weight function is formulated according to (40), (39) holds. The solution to (40) is given by the solution to the following differential equation:

$$
\frac{w(t)}{w'(t)} = \frac{3BM^2}{\delta}
$$

which is

$$
w(t) = \exp\left(\frac{3BM^2}{\delta}t\right)
$$

Then, given the weight function has the form above, inequality (39) is satisfied.

Therefore, the solution exists and is an element of C_w^0. Since C_w^0 is a complete Banach space, the solution is unique.

To show the weight function exists in C_w^s, we apply an induction argument on the smoothness parameter, s. The argument is presented below:

For $s = 0$, it was shown that $f(t) \in C_w^0$.

For $s \geq 1$, let the inductive hypothesis be that $f(t) \in C_w^{s-1}$ where $f(t)$ is defined by (11). By Property A.1(i), and by the inductive hypothesis:

$$
f^2(u)\,\Psi_{11}(f(u)v,k_1) \in C_w^{s-1}
$$

Then, by the Fundamental Theorem of Calculus:

$$\int_0^t f^2(u)\,\Psi_{11}(f(u)\,v, k_1)\,du \in C_w^s$$

Additionally, by $ID6(i)$, $H_{tt}\left(t, 0, k_1^0\right) \in C_w^s$, so that $f(t) \in C_w^s$.

A.10 Proof of Theorem 5

Let there be two structures equivalent up to the covariate function, $\phi(x)$. The two structures obtain the same survival function so that

$$\int_0^t \left[\Psi(\phi_1(x)\,f(u)\,(t-u), k_1) - \Psi(\phi_2(x)\,f(u)\,(t-u), k_1)\right] du \tag{41}$$

$$= (\phi_1(x) - \phi_2(x)) \int_0^t f(u)\,(t-u)\,\Psi_1\left(\overline{\phi}(x)\,f(u)\,(t-u), k_1\right) du = 0 \ a.e. \tag{42}$$

By Property A.1(i), (42) is obtained by a mean value expansion of (41), where $\overline{\phi}$ is a mean value between ϕ_1 and ϕ_2. Since $\Psi(\lambda, k_1)$ is an increasing function in λ, (42) holds if and only if, for all x :

$$\phi_1(x) = \phi_2(x)$$

A.11 Proof of Theorem 6

First note the following notation, which will be used only for this proof. For $i \in \{1, 2, ..., s\}$ define:

$$\begin{aligned} H_{(i)} &= \lim_{t\to 0} \frac{\partial^i}{\partial t^i} H(t, 0) \\ f_{(i)} &= \lim_{t\to 0} \frac{\partial^i}{\partial t^i} f(t) \end{aligned}$$

Fix $x = 0$. Consider (11) and evaluate it in the limit as $t \to 0$. Solving the resulting expression for k_1 obtains

$$k_1 = H_{(2)} \tag{43}$$

Differentiating (11) with respect to t, evaluating the resulting expression in the limit as $t \to 0$, and using condition (B) obtains:

$$k_2 = H_{(3)} - k_1 f_{(1)} \tag{44}$$

Differentiating (11) twice with respect to t and evaluating the resulting expression in the limit as $t \to 0$ obtains

$$k_3 = H_{(4)} - k_1 f_{(2)} - 2k_2 f_{(1)} \tag{45}$$

Since by (B) $\left\{\frac{\partial^i}{\partial t^i}H(t,0)\right\}_{i\geq 1}$ exist and are well defined for all $t>0$, the process of differentiation and evaluation in the limit can be continued to obtain:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & ... & 0 \\ f_{(1)} & 1 & 0 & 0 & ... & 0 \\ f_{(2)} & 2f_{(1)} & 1 & 0 & ... & 0 \\ f_{(3)} & 3f_{(2)} & 2f_{(1)} & 1 & ... & 0 \\ ... & ... & ... & ... & ... & 0 \\ f_{(s-1)} & (s-1)\,f_{(s-2)} & (s-2)\,f_{(s-3)} & (s-3)\,f_{(s-4)} & ... & 1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ ... \\ k_s \end{bmatrix} = \begin{bmatrix} H_{(2)} \\ H_{(3)} \\ H_{(4)} \\ H_{(5)} \\ ... \\ H_{(s+1)} \end{bmatrix} \tag{46}$$

where $s\to\infty$. The system above has a unique solution since the determinant of the matrix of coefficients equals unity. Then all $s\to\infty$ cumulants of the process are identified.

Remark 4 An Infinity of Solutions*: Consider system* (46) *above. For both the weight function and the cumulants unknown, the system has an infinity of solutions. Thus, it is not possible to identify jointly the weight function and the distribution of the process via the method proposed in this paper.*

A.12 Identification with Time Varying Covariates

When observed covariates are time-varying, it is assumed as in the standard MPH literature (see (Honoré 1993), (Heckman & Taber 1994)) that $x(t)$ are jump variables. That is, they are realizations of stochastic processes with continuous sample paths. When time varying observed covariates enter the hazard function multiplicatively, the effects of time coming in through the observed covariates cannot be separated from those entering through the unobservables without imposing stronger conditions on the unobserved stochastic process. Thus, in order to identify the function of covariates, it is assumed the stochastic process is known entirely.

In order to apply the exponential formula for the survival function, the two processes are assumed to satisfy the following conditions:

Assumption P (i) The processes $\{x(u)\}_0^t$ and $\{Z(u)\}_0^t$ are independent; (ii) There is no contemporaneous feedback between $x(t)$ and $Z(t)$; and (iii) Future values of the two processes do not affect duration.

Under these conditions and under a normalization assumption on the covariate function, assumption ID8' below, the covariate function $\phi(x(t))$ is parametrically identified. The proof follows that of Theorem 5 in (Heckman & Taber 1994).

Assumption V (i) There are two different values for $x(t)$ at time t, $x_1(t)\neq x_2(t)$, such that these two different realizations at time t have the same sample paths up to t^-:

$$\{x_1(u)\}_0^{t^-} = \{x_2(u)\}_0^{t^-}$$

(ii) $\phi\left(x\left(t^{-}\right)\right)=1$

Theorem 18 *Assume the distribution of the stochastic process is entirely known. Under Property A.1(i), assumptions P and V, the covariate function* $\phi\left(x\left(t\right)\right)$ *is identified.*

Proof. Define

$$\int_{u}^{t}\phi\left(x\left(s\right)\right)ds=\Phi\left(t,u\right)$$

As before, the survival function is written as

$$\begin{aligned} S\left(t\middle|\left\{x\left(u\right)\right\}_{0}^{t}\right) &= E_{Z}\exp\left[-\int_{0}^{t}\left[\phi\left(x\left(s\right)\right)\int_{0}^{s}dZ\left(u\right)\right]ds\right] \\ &= E_{Z}\exp\left[-\int_{0}^{t}\Phi\left(t,u\right)dZ\left(u\right)\right] \\ &= \exp\left[-\int_{0}^{t}\Psi\left(\Phi\left(t,u\right)\right)du\right] \end{aligned}$$

Then

$$\begin{aligned} &\frac{\partial}{\partial t}S\left(t\middle|\left\{x\left(u\right)\right\}_{0}^{t}\right) \\ =\ &-\left[\phi\left(x\left(t\right)\right)\int_{0}^{t}\Psi_{1}\left(\Phi\left(t,u\right)\right)du\right]\exp\left[-\int_{0}^{t}\Psi\left(\Phi\left(t,u\right)\right)du\right] \end{aligned} \tag{47}$$

Evaluating (47) at the same t for two different values $x_{1}\left(t\right)$ and $x_{2}\left(t\right)$ that have the same sample path up to t^{-}, obtains:

$$\frac{\frac{\partial}{\partial t}S\left(t\middle|\left\{x_{1}\left(u\right)\right\}_{0}^{t}\right)}{\frac{\partial}{\partial t}S\left(t\middle|\left\{x_{2}\left(u\right)\right\}_{0}^{t}\right)}=\frac{\phi\left(x_{1}\left(t\right)\right)}{\phi\left(x_{2}\left(t\right)\right)}$$

Using assumption V(ii), $\phi\left(x\left(t\right)\right)$ is identified on the support of $x\left(t\right)$ for all $t\in\mathbb{R}_{+}$. ∎

A.13 The Time Deformed Framework 3

When heterogeneity is time deformed, the hazard function is modeled as 3.

A.13.1 Example: Deterioration and Repair Costs

Suppose there is a machine (or plant) that fails with a probability governed by a hazard function given by (3)[31]. While the machine is operating, there is a wear-and-tear process, $\left\{Z\left(u\right)\right\}_{0}^{t}$, undermining its lifetime[32].

[31] For a similar framework see (K. Ryu 1993).

[32] In the engineering literature, it is usually assumed the wear-and-tear process is a gamma process. See (M.D. Pandey, X.X. Yuan & J. M. Van Noortwijk 2009).

The wear-and-tear process is assumed to start once the machine starts working. The process is also assumed to be regenerative, i.e. once the machine is repaired or replaced, the process starts anew.

Rather than taking place in calendar time, the deterioration process happens in operation time. That is, letting $y(t)$ represent a continuous time measure of usage, $g(y(t))$ models the rate of usage of the machine. Then deterioration happens as a function of the rate of usage, i.e. $\{Z(g(y(u)))\}_0^t$.

Let x be some time invariant observables describing the machine, such as size, manufacturer, or geographical location. If the machine fails before being shut down for preventive repair work, it incurs a cost of being repaired, $c_F(x,t)$. The cost changes in time and it depends on x. If the machine is shut down for preventive repair work before it fails, it incurs a repair or replacement cost of $c_P(x,t)$, where $c_P(x,t) < c_F(x,t)$ for all x and t. The manager then needs to decide when to shut down the machine in order to minimize the expected cost of repair. The discounted expected cost of shut down at time t given x and the manager's subjective discount rate, r, is given by:

$$C(t|x) = e^{-rt}\left[P(t \le T|x)\, c_P(t,x) + P(t > T|x)\, c_F(t,x)\right] \tag{48}$$

where $P(t > T|x)$ represents the probability of failure. The first term on the right hand side of (48) represents expected cost in case the machine does not fail, while the second term represents expected cost in the case the machine fails before being shut down for preventive repair work. In this example, the hazard function can be used to estimate the discounted expected cost. By estimating $g(y(t))$, the manager can draw conclusions about how the timing of usage affects the lifetime of the machine. Likewise, $\phi(x)$ can be used to make inferences about the observables that affect the probability of failure.

A.13.2 Identification

Let the hazard function be described by (3). First, when the distribution of $\{Z(u)\}_0^t$ is unknown, it is shown that both the mean of the distribution of $\{Z(u)\}_0^t$ and $\phi(x)$ are identified as $t \to 0$. Then, assuming the distribution of $\{Z(u)\}_0^t$ to be known up to the mean, the time deformation function, $g(y(t))$, is identified up to an additive constant.

Let the following assumptions hold:

Assumption ID9 $H(t,x,y(t)) = -\log S(t|x,y(t))$ is twice differentiable in t for all x.

Assumption ID10 (i) $y^{-1}(.)$ exists; (ii) $g(y(0))$ is well-defined; and (iii) $g^{-1}(.)$ exists.

Assumption ID11 (i) $0 < g(y(t)) \le M_2 < \infty$; (ii) $|\Psi_{11}(\lambda, k_1)| \le B$, $\lambda > 0$ where B is a positive constant.

Assumption ID12 Define $q(y(t)) = \frac{\partial}{\partial t} g(y(t))$ and assume that $\lim_{t\to 0} q(y(t)) = 1$.

Remark 5 *Assumption ID10(iii) requires $g(.)$ to be a one-to-one mapping between calendar time and the data based time scale, while assumption ID12 is a normalization assumption on the first derivative of the time deformation function.*

Lemma 4 *Let the distribution of the subordinator* $\{Z(u)\}_0^t$ *be unknown. Under assumptions ID2, ID9, ID10, and ID12, the mean of the distribution of the process is identified and equal to*

$$k_1 = \lim_{t \to 0} \frac{\partial^2}{\partial t^2} H(t, 0, y(t)) \tag{49}$$

Proof. The identification strategy begins with the survival function associated to (3). Consider the following change of variables:

$$\eta = g(y(u)) \tag{50}$$

Using (50) obtains the following two expressions:

$$\begin{aligned} d\eta &= dg(y(u)) = g'(y(u))\, y'(u)\, du && (51) \\ y(u) &= g^{-1}(\eta) \Rightarrow u = y^{-1}\left(g^{-1}(\eta)\right) && (52) \end{aligned}$$

Additionally, let $\frac{\partial}{\partial u} g(y(u)) = g\prime(y(u))\, y'(u)$. The survival function is given by:

$$\begin{aligned} S(t|x, y(t)) &= E_Z \exp\left[-\int_0^t h\left(s|x, y(s), \{Z(u)\}_0^s\right) ds\right] \\ &= E_Z \exp\left[-\int_0^t \left[\int_0^s \phi(x)\, dZ(g(y(u)))\right] ds\right] \\ &= E_Z \exp\left[-\int_0^t \left[\int_u^t \phi(x)\, ds\right] dZ(g(y(u)))\right] \\ &= E_Z \exp\left[-\int_0^t \phi(x)(t-u)\, dZ(g(y(u)))\right] \\ &= E_Z \exp\left[-\int_{g(y(0))}^{g(y(t))} \phi(x)\left(t - y^{-1}\left(g^{-1}(\eta)\right)\right) dZ(\eta)\right] && (53) \\ &= \exp\left[-\int_{g(y(0))}^{g(y(t))} \Psi\left(\phi(x)\left(t - y^{-1}\left(g^{-1}(\eta)\right)\right)\right) d\eta\right] && (54) \\ &= \exp\left[-\int_0^t \Psi(\phi(x)(t-u))\, g\prime(y(u))\, y'(u)\, du\right] && (55) \end{aligned}$$

where (53) follows from (50), (51), and (52). Equality (54) follows from the independence and stationarity of the increments of the stochastic process, as well as from the definition of the Laplace exponent (the arguments are similar to those made previously, for the time homogeneous framework). Finally equality (55) follows by switching back to the original notation using (50).

Let

$$H(t, x, y(t)) = -\log S(t|x, y(t)) = \int_0^t \Psi(\phi(x)(t-u))\, g\prime(y(u))\, y'(u)\, du \tag{56}$$

Let $x = 0$, differentiate (56) twice with respect to t, and evaluate the resulting expression in the limit as $t \downarrow 0$ to obtain (49). ■

Consider now the identification of $\phi(x)$.

Theorem 19 *Let the distribution of the subordinator $\{Z(u)\}_0^t$ be unknown and let ID2, ID9, and ID12 hold. The function $\phi(x)$ is identified and equal to*

$$\phi(x) = \frac{\lim_{t\to 0} \frac{\partial^2}{\partial t^2} H(t, x, y(t))}{\lim_{t\to 0} \frac{\partial^2}{\partial t^2} H(t, 0, y(t))} \tag{57}$$

Proof. Differentiate (56) twice with respect to t and solve (58) for $\phi(x)$. Evaluate the resulting expression as $t \downarrow 0$ to obtain (57). ■

Lastly, consider the identification of the time-deformation function $g(y(t))$.

Theorem 20 *Let the distribution of the subordinator $\{Z(u)\}_0^t$ be known up to its mean, k_1, and assume ID2, ID5, ID6, ID8, and ID9 to ID11 hold. Then the function $g(y(t)) \in C_w^{s_2}(\mathbb{R}_+)$ is identified up to an additive constant C. The weight function $w(t)$ on $C_w^{s_2}(\mathbb{R}_+)$ is given by*

$$w(t) = \exp\left(\frac{2BM_2}{\delta} t\right)$$

Let $\int q(y(t))\, dt = g(y(t)) + C$. The function $q(y(t))$ is given by the successive approximation method, $n \geq 1$:

$$\begin{aligned} q_0(y(t)) &= \frac{\frac{\partial^2}{\partial t^2} H\left(t, 0, y(t), k_1^0\right)}{\lim_{t\to 0} \frac{\partial^2}{\partial t^2} H\left(t, 0, y(t), k_1^0\right)} \\ q_n(y(t)) &= \frac{\frac{\partial^2}{\partial t^2} H\left(t, 0, y(t), k_1^0\right)}{\lim_{t\to 0} \frac{\partial^2}{\partial t^2} H\left(t, 0, y(t), k_1^0\right)} \\ &\quad - \frac{1}{\lim_{t\to 0} \frac{\partial^2}{\partial t^2} H\left(t, 0, y(t), k_1^0\right)} \int_0^t q_{n-1}^2(y(u))\, \Psi_{11}((t-u), k_1)\, du \end{aligned}$$

such that $\lim_{n\to\infty} q_n(y(t)) = q(y(t))$.

Proof. Define the following function:

$$q(y(u)) = \frac{d}{du} g(y(u)) = g\prime(y(u))\, y'(u)$$

and note that once the distribution of the process is parametrized up to its mean, the true survival function becomes a function of the true mean, i.e.

$$H\left(t,x,y\left(t\right),k_1^0\right)=-\log S\left(t|x,y\left(t\right),k_1^0\right)=\int_0^t \Psi\left(\phi\left(x\right)\left(t-u\right),k_1\right)q\left(y\left(u\right)\right)du$$

Differentiating $H\left(t,x,y\left(t\right),k_1^0\right)$ twice with respect to t obtains the following integral equation of the second kind:

$$\begin{aligned}\frac{\partial^2}{\partial t^2}H\left(t,x,y\left(t\right),k_1^0\right) &= \frac{\partial^2}{\partial t^2}H\left(t,x,y\left(t\right),k_1^0\right)=k_1\phi\left(x\right)q\left(y\left(t\right)\right) \\ &\quad+\int_0^t \phi^2\left(x\right)q^2\left(y\left(u\right)\right)\Psi_{11}\left(\phi\left(x\right)\left(t-u\right),k_1\right)du\end{aligned} \tag{58}$$

By ID9, expression (58) becomes

$$\frac{\partial^2}{\partial t^2}H\left(t,0,y\left(t\right),k_1^0\right)=k_1 q\left(y\left(t\right)\right)+\int_0^t q^2\left(y\left(u\right)\right)\Psi_{11}\left(\left(t-u\right),k_1\right)du$$

which is an integral equation of the second kind in $q\left(y\left(t\right)\right)$:

$$\begin{aligned}q\left(y\left(t\right)\right) &= \frac{1}{\lim_{t\to 0}\frac{\partial^2}{\partial t^2}H\left(t,y\left(t\right),0,k_1^0\right)} \\ &\quad\times\left[\frac{\partial^2}{\partial t^2}H\left(t,0,y\left(t\right),k_1^0\right)-\int_0^t q^2\left(y\left(u\right)\right)\Psi_{11}\left(\left(t-u\right),k_1\right)du\right]\end{aligned}$$

This integral equation can be shown to have a unique solution, $q\left(y\left(t\right)\right)$, by the Banach Fixed Point Theorem. The identification strategy is as that employed for Theorem 4. Then the function $g\left(y\left(t\right)\right)$ is identified up to an additive constant since

$$\int q\left(y\left(u\right)\right)du=g\left(y\left(u\right)\right)+C$$

■

As before, for each of the two processes, gamma and compound Poisson, the rate parameter is identified by imposing normalization assumptions on the other elements entering the nonlinear expression for k_1. For the gamma process with $\nu=1$, and for the compound Poisson process with $(\kappa,\nu)=(1,1)$, the rate of the process obtains:

$$\rho=\lim_{t\to 0}\frac{\partial^2}{\partial t^2}H\left(t,0,y\left(t\right),k_1^0\right)$$

A.14 Proof of Theorem 7

To show the consistency of the estimators we verify the conditions of Lemma B.1 of (X. Chen & D. Pouzo 2008). First, we present Lemma B.1 adapted to our model. Then we verify its conditions.

Lemma 5 (B.1) *Let $\widehat{\alpha}_n = \left(\widehat{\phi}_n, \widehat{f}_n, \widehat{\rho}_n\right)$ be such that $\widehat{Q}_n(\widehat{\alpha}_n) \geq \sup_{\alpha\in\mathcal{A}_n} \widehat{Q}_n(\alpha) - O_p(\eta_n)$ with $\eta_n = o_p(1)$. Suppose the following conditions hold:*

B.1.1 (i) $Q(\alpha_0) < \infty$;

(ii) $\liminf_{n\to\infty} \sup_{\alpha\in\mathcal{A}_n:||\alpha-\alpha_0||\geq\varepsilon} Q(\alpha) < Q(\alpha_0)$ uniformly in $\varepsilon > 0$.

B.1.2 (i) $\mathcal{A} \subseteq \mathbf{A}$ and $(\mathbf{A},\ ||.||)$ is a metric space;

(ii) $\mathcal{A}_n \subseteq \mathcal{A}_{n+1} \subseteq \ldots \subseteq \mathcal{A}$ for all $n \geq 1$, and there exists a sequence $\Pi_n\alpha_0 \in \mathcal{A}_n$ such that $||\Pi_n\alpha_0 - \alpha_0|| \to 0$ as $n \to \infty$.

B.1.3 (i) $\widehat{Q}_n(\alpha)$ is a measurable function of the data $\{x_i, t_i\}_{i=1}^n$ for all $\alpha \in \mathcal{A}_n$;

(ii) $\widehat{\alpha}_n$ is well defined and measurable.

B.1.4 (i) Let $c(m(n)) = \sup_{\alpha\in\mathcal{A}_n} \left|\widehat{Q}_n(\alpha) - Q(\alpha)\right| = o_p(1)$;

(ii) Uniformly over $\varepsilon > 0$

$$\max\left\{c(m(n)),\ \eta_n,\ \left|Q(\Pi_n\alpha_0) - Q(\alpha_0)\right|\right\} = o(1)$$

Then $d(\widehat{\alpha}_n, \alpha_0) = o_p(1)$.

Note that since there is no penalty term, $\overline{Q}_n(.) = \overline{Q}(.) = Q(.)$ in the original Lemma B1.

Let us now check the conditions of Lemma $B.1$ above.

Condition $B.1.1$(i) is satisfied by assumptions $C1$ and $C2$. In order for the criterion function $Q(\alpha_0) = E_{t,x}\log p(t|x,\alpha_0) < \infty$ and in anticipation of the information inequality, we show that

$$E_{t,x}p(t|x,\alpha_0) < \infty$$

where $p(t|x,\alpha_0) > 0$. The joint probability distribution of T and X is denoted as $P(t,x)$, while the marginal densities of $X|T$ and of T are denoted as $\mu_t(x)$ and $\pi(t)$, respectively. Then

$$\begin{aligned}
E_{t,x}p(t|x,\alpha_0) &= \int_{\mathcal{X}\times\mathcal{T}} p(s|w,\alpha_0)\,dP(s,w) &(59)\\
&= \int_{\mathcal{T}}\left[\int_{\mathcal{X}} p(s|w,\alpha_0)\,\mu_t(w)\,dw\right]\pi(s)\,ds &(60)\\
&\leq M_\phi M_f M_1 \int_{\mathcal{T}}\left[\int_{\mathcal{X}} \mu_t(w)\,dw\right]\pi(s)\,ds &(61)\\
&= M_\phi M_f M_1 \int_{\mathcal{T}} \mu(s)\,ds < \infty &(62)
\end{aligned}$$

where (61) follows since

$$\begin{aligned} p\left(t|x,\alpha_0\right) &= S\left(t|x;\phi_0,f_0,\rho_0\right)\int_0^t \phi_0\left(x\right)f_0\left(u\right)\Psi_1\left(\phi_0\left(x\right)f_0\left(u\right)\left(t-u\right),\rho_0\right)du \\ &\le \int_0^t \phi_0\left(x\right)f_0\left(u\right)\Psi_1\left(\phi_0\left(x\right)f_0\left(u\right)\left(t-u\right),\rho_0\right)du \\ &\le M_\phi M_f M_1 \end{aligned}$$

where we used that $\sup_x \phi_0\left(x\right) \equiv M_\phi$, $\sup_t f\left(t\right) \equiv M_f$, and $\sup_{\lambda,\rho}\Psi_1\left(\lambda,\rho\right) \equiv M_1$. Equality (62) follows since $\int_{\mathcal{X}} \mu_t\left(w\right)dw = 1$ a.s. in w. Since $E_{t,x}p\left(t|x,\alpha_0\right) < \infty$, the stronger condition that $E_{t,x}\log p\left(t|x,\alpha_0\right) < \infty$ is satisfied. Thus $Q\left(\alpha_0\right) < \infty$ so that Condition $B.1.1\left(i\right)$ is satisfied.

Condition $B.1.1$(ii) is implied by assumptions $C1$ and $C2$. Since α_0 is identified and $E_{t,x}p\left(t|x,\alpha_0\right) < \infty$, by the information inequality, $Q\left(\alpha\right) - Q\left(\alpha_0\right) < 0$ for $\alpha \in \mathcal{A}_n$ with $\alpha \neq \alpha_0$.

Define:

$$\delta\left(m\left(n\right),\varepsilon\right) \equiv \sup_{\alpha\in\mathcal{A}_n:||\alpha-\alpha_0||_\infty\ge\varepsilon} Q\left(\alpha\right) - Q\left(\alpha_0\right)$$

Since $\mathcal{A}_n$ is compact, there exists a $\alpha_n^* \in \mathcal{A}_n$ with $||\alpha_n^*-\alpha_0||_\infty \ge \varepsilon > 0$ such that

$$\alpha_n^* = \arg\max_{\alpha\in\mathcal{A}_n:||\alpha_n^*-\alpha_0||_\infty\ge\varepsilon} Q\left(\alpha\right)$$

Then, for some constant $C > 0$,

$$\begin{aligned} \delta\left(m\left(n\right),\varepsilon\right) &= Q\left(\alpha_n^*\right) - Q\left(\alpha_0\right) \\ &= Q\left(\alpha_n^*\right) - Q\left(\Pi_n\alpha_0\right) + Q\left(\Pi_n\alpha_0\right) - Q\left(\alpha_0\right) \\ &\ge C\left|\left|\alpha_n^* - \Pi_n\alpha_0\right|\right|_\infty^2 + o\left(1\right) \end{aligned}$$

Suppose $Q\left(\alpha_n^*\right) - Q\left(\alpha_0\right) \to 0$, then $||\alpha_n^*-\Pi_n\alpha_0||_\infty^2 \to 0$. However, since

$$||\alpha_n^*-\alpha_0||_\infty^2 \le ||\alpha_n^*-\Pi_n\alpha_0||_\infty^2 + ||\Pi_n\alpha_0-\alpha_0||_\infty^2$$

then $||\alpha_n^*-\alpha_0||_\infty^2 \to 0$, which is a contradiction to $||\alpha_n^*-\alpha_0||_\infty^2 \ge \varepsilon > 0$. Therefore

$$\lim\inf_{n\to\infty} \delta\left(m\left(n\right),\varepsilon\right) > 0$$

Condition $B.1.2$ is implied by the way the parameter and the sieve spaces are defined in $(17a)-(17b)$ and $(18a)-(18b)$.

Condition $B.1.3$ is implied by assumptions C1, C2, and C3. In order to check B.1.3 we apply Remark B.1(1)(a) in (Chen & Pouzo 2008). First note that by construction, $\mathcal{A}_n$ is a compact subset of $\mathcal{A}$ for each n

under the norm defined in (19). Before showing the continuity of the criterion function in the consistency norm, let $\alpha = (\gamma, \rho)$ and define the following terms:

$$\begin{aligned}
\gamma(u, x) &= \phi(x) f(u) \\
\Upsilon(\gamma, t, \rho) &= \int_0^t \gamma(x, u) \Psi_1(\gamma(t-u), \rho) \, du \\
\Gamma_1(\gamma(t-u), \rho) &= f(u)(1 - (t-u)\Upsilon(\gamma, t, \rho)) \Psi_1(\gamma(t-u), \rho) \\
&\quad + \phi(x) f^2(u)(t-u) \Psi_{11}(\gamma(t-u), \rho) \\
\Gamma_2(\gamma(t-u), \rho) &= \phi(x)(1 - (t-u)\Upsilon(\gamma, t, \rho)) \Psi_1(\gamma(t-u), \rho) \\
&\quad + \phi^2(x) f(u)(t-u) \Psi_{11}(\gamma(t-u), \rho) \\
\Gamma_3(\gamma(t-u), \rho) &= \phi(x) f(u) \Psi_{12}(\gamma(t-u), \rho) \\
&\quad - \Upsilon(\gamma, t, \rho) \Psi_2(\gamma(t-u), \rho)
\end{aligned}$$

By assumption C2 and by letting $M_f \equiv \sup_t f(t)$ and $M_\phi \equiv \sup_x \phi(x)$, we have that

$$\begin{aligned}
\sup \Gamma_1(\gamma(t-u), \rho) &\leq M_f M_1 \\
\sup \Gamma_2(\gamma(t-u), \rho) &\leq M_\phi M_1 \\
\sup \Gamma_3(\gamma(t-u), \rho) &\leq M_\phi M_f M_{12}
\end{aligned}$$

and by assumption C1(i) and Property 1(iii), we have that for all α and x and almost all t :

$$\Upsilon(\gamma, t, \rho) \geq \xi \neq 0$$

By a mean value expansion of $\widehat{Q}_n(\alpha_1)$ about $\alpha = (\phi, f, \rho)$, with $\widetilde{\alpha}$ the mean value between α_1,$\alpha \in \mathcal{A}$ obtains:

$$\left| \widehat{Q}_n(\alpha_1) - \widehat{Q}_n(\alpha) \right| \tag{63}$$

$$\leq \frac{1}{n} \sum_i \left| \frac{1}{\Upsilon(\gamma_i, t_i, \widetilde{\rho})} \right| \left[\begin{array}{l} |(\phi_1 - \phi)(x_i)| \int_0^{t_i} |\Gamma_1(\widetilde{\gamma}_i(t_i - u), \widetilde{\rho})| \, du \\ + \left(\sup\limits_{0 < u \leq t} |\Gamma_2(\widetilde{\gamma}_i(t_i - u), \widetilde{\rho})| \right) \\ \times \left(\int_0^{t_i} |(f_1 - f)(u)| \, du \right) \\ + |\rho_1 - \rho| \int_0^{t_i} |\Gamma_3(\widetilde{\gamma}_i(t_i - u), \widetilde{\rho})| \, du \end{array} \right] \tag{64}$$

$$\leq \left[\frac{1}{\xi} \left(|(M_f + c_1 M_\phi) M_1| + |M_\phi M_f M_{12}| \right) \right] ||\alpha_1 - \alpha||_\infty$$

Condition $B.1.4$(i) is implied by assumptions C1 through C3. To show the uniform convergence of the criterion function over the sieve space, we have to show that:

$$\sup_{\alpha \in \mathcal{A}_{m(n)}} \left| \widehat{Q}_n (\alpha) - Q(\alpha) \right| = o_p(1)$$

which holds if the class of functions indexing the criterion is Glivenko-Cantelli. That is, we need to show that the class of functions (65) is Glivenko-Cantelli

$$\mathcal{L} = \{ l(t|x, \alpha) = \log p(t|x, \alpha) : \alpha \in \mathcal{A}_n \} \tag{65}$$

By Theorem 2.4.1 of van der Vaart and Wellner (1998) if the bracketing number $N_{[]}(\varepsilon, \mathcal{L}, L_1)$ is finite for all $\varepsilon > 0$, then $\mathcal{L}$ is Glivenko-Cantelli. We proceed now to calculate the bracketing number of the class $\mathcal{L}$.

Define

$$\begin{aligned} v &= t - u \\ \gamma_j^L &= \phi_j^L(x) f_j^L(u) \\ \gamma_j^U &= \phi_j^U(x) f_j^U(u) \end{aligned}$$

where $\gamma_j^L < \gamma_j^U$ for some $j = 1, ..., m(n)$ and $i = 1, ..., k$, where it is known that the minimum value of k is of order $O(1/\varepsilon)$, $\varepsilon > 0$. For $\rho^L \leq \rho \leq \rho^U$ such that $\left| \rho_i^U - \rho_i^L \right| \leq \varepsilon$, $i = 1, ..., \frac{c}{\varepsilon}$, define:

$$\begin{aligned} l_{ij}^U(x, t, \gamma, \rho) &= \log \int_0^t \gamma_j^U \Psi_1\left(\gamma_j^U v, \rho_i^U\right) du - \int_0^t \Psi\left(\gamma_j^L v, \rho_i^L\right) du \\ l_{ij}^L(x, t, \gamma, \rho) &= \log \int_0^t \gamma_j^L \Psi_1\left(\gamma_j^L v, \rho_i^L\right) du - \int_0^t \Psi\left(\gamma_j^U v, \rho_i^U\right) du \end{aligned}$$

By Property 1(iii) and by assumption C3, $\Psi(\lambda, \rho)$ is increasing in both λ and ρ, so for each $\alpha \in \mathcal{A}_n$ and for some $j = 1, ..., m(n)$ and $i = 1, ..., k$:

$$l_{ij}^L(x, t, \gamma, \rho) \leq l(x, t, \gamma, \rho) \leq l_{ij}^U(x, t, \gamma, \rho)$$

Furthermore, letting $\overline{\gamma}_j$ and $\overline{\rho}_i$ be mean values between $\left(\gamma_j^L, \gamma_j^U\right)$ and $\left(\rho_i^L, \rho_i^U\right)$ respectively, by assumptions C1 and C3 obtains:

$$\begin{aligned}
& \left|\int_0^t \Psi\left(\gamma_j^U v, \rho_i^U\right) du - \int_0^t \Psi\left(\gamma_j^L v, \rho_i^L\right) du\right| \qquad (66)\\
= & \left|\int_0^t v\Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\left(\gamma_j^U - \gamma_j^L\right) du + \left(\rho_i^U - \rho_i^L\right)\int_0^t \Psi_2\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du\right| \\
\leq & \; t \sup_{0<u\leq t}\left|v\Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right|\left[\int_0^t \left|\gamma_j^U - \gamma_j^L\right| du\right] \\
& + \left|\rho_i^U - \rho_i^L\right| \int_0^t \left|\Psi_2\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right| du
\end{aligned}$$

By a mean value expansion and by assumptions C1 and C3 obtains

$$\begin{aligned}
& \left|\log\int_0^t \gamma_j^U \Psi_1\left(\gamma_j^U v, \rho_i^U\right) du - \log\int_0^t \gamma_j^L \Psi_1\left(\gamma_j^L v, \rho_i^L\right) du\right| \\
\leq & \left|\frac{\int_0^t \left[\Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) + \overline{\gamma}_j v\Psi_{11}\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right]\left(\gamma_j^U - \gamma_j^L\right) du}{\int_0^t \overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}\right| \\
& + \left|\frac{\int_0^t \overline{\gamma}_j \Psi_{12}\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}{\int_0^t \overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}\right|\left|\rho_i^U - \rho_i^L\right| \\
\leq & \; \frac{t\sup_{0<u\leq t}\left|\Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) + \overline{\gamma}_j v\Psi_{11}\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right|}{\int_0^t \left|\overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right| du}\left[\int_0^t \left|\gamma_j^U - \gamma_j^L\right| du\right] \qquad (67)\\
& + \left|\frac{\int_0^t \overline{\gamma}_j \Psi_{12}\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}{\int_0^t \overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}\right|\left|\rho_i^U - \rho_i^L\right|
\end{aligned}$$

Combining (66) and (67) obtains

$$\begin{aligned}
& \left|l_{ij}^U(t,\gamma,\rho) - l_{ij}^L(t,\gamma,\rho)\right| \\
\leq & \; t\left[\frac{\sup_{0<u\leq t}\left|\Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) + \overline{\gamma}_j v\Psi_{11}\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right|}{\int_0^t \left|\overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right| du} + \sup_{0<u\leq t}\left|v\Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right)\right|\right] \\
& \times\int_0^t \left|\gamma_j^U - \gamma_j^L\right| du \\
& + \left|\int_0^t \Psi_2\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du + \frac{\int_0^t \overline{\gamma}_j \Psi_{12}\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}{\int_0^t \overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{\rho}_i\right) du}\right|\left|\rho_i^U - \rho_i^L\right|
\end{aligned}$$

Let

$$C = \sqrt{c_2}\left(\frac{M_1}{\xi} + M_1\right) + \frac{M_\phi M_f \left|M_{12}\right|}{\xi} + M_2$$

for all x and almost all t.

By using a result of (Shen & Wong 1994) (page 597) and by using a bracketing entropy preservation result of Kosorok (2008, Lemma 9.25) we show below that

$$\left|\left|l_{ij}^{U}\left(t,\gamma,\rho\right)-l_{ij}^{L}\left(t,\gamma,\rho\right)\right|\right|_{\infty}\leq C\varepsilon$$

First, notice that $\left|\rho_i^U-\rho_i^L\right|\leq\varepsilon/2$ holds as ρ is a finite dimensional parameter and the covering number of Θ is of order $O\left(\frac{1}{\varepsilon}\right)$. Then we show that $\int_0^t\left|\gamma_j^U-\gamma_j^L\right|du\leq\varepsilon/2$ holds. According to a result on page 597 of (Shen & Wong 1994), the bracketing entropy of Φ_n is bounded by

$$\log N_{[]}\left(\tfrac{\varepsilon}{2M_\phi},\Phi_n,||.||_\infty\right)\leq C^{'}m_n\log\left(\tfrac{2M_\phi}{\varepsilon}\right)$$

where the envelope of the class of functions indexing Φ_n is M_ϕ and where we used that if $\mathcal{F}$ is a class of functions with envelope equal to 1, then $M\mathcal{F}$, where M is a constant, has $N_{[]}\left(\varepsilon M,\mathcal{F},||.||\right)=N\left(\frac{\varepsilon}{M},\mathcal{F},||.||\right)$. Also, $\mathcal{F}_n^{int}$, the space of functions indexed by $\int_0^t f_n\left(u\right)du$ is a finite dimensional linear space with envelope $\int_0^t f_n\left(u\right)du\leq M_f$. Applying the same result in (Shen & Wong 1994), we have that the bracketing entropy of $\mathcal{F}_n^{int}$ is bounded by

$$\log N_{[]}\left(\tfrac{\varepsilon}{2M_f},\mathcal{F}_n^{int},||.||_\infty\right)\leq C^{''}m\left(n\right)\log\left(\frac{2M_f}{\varepsilon}\right)$$

By bracketing entropy preservation results[33], since both $\frac{\phi_n(x)}{M_\phi}$ and $\frac{1}{M_f}\int_0^t f_n\left(u\right)du$ are uniformly bounded by 1, letting $K=\max\left(C^{'},C^{''}\right)$ and defining the class of functions indexed by $\phi\left(x\right)\int_0^t f_n\left(u\right)du$ as Δ, we have that the class Δ is bounded by

$$\log N_{[]}\left(\varepsilon,\Delta,||.||_\infty\right)\leq Km_n\log\left(\frac{4M_\phi M_f}{\varepsilon}\right)$$

which means there exists a set of functions $\left\{\phi_j^Lf_j^L,\ \phi_j^Uf_j^U\right\}_{j=1}^{(4M_\phi M_f/\varepsilon)^{Km_n}}$ such that the following two expressions hold for some $j=1,...,\left(\frac{4M_\phi M_f}{\varepsilon}\right)^{Km_n}$

$$\phi_j^Lf_j^L\leq\Lambda\leq\phi_j^Uf_j^U$$
$$\left|\left|\phi_j^Lf_j^L-\phi_j^Uf_j^U\right|\right|_\infty\leq\varepsilon/2$$

[33] Let $\mathcal{F}$ and $\mathcal{G}$ be classes of measurable functions. Then for any probability measure P and any $1\leq r\leq\infty$, provided $f\in\mathcal{F}:|f|\leq L$ and $g\in\mathcal{G}:|g|\leq K$

$$N_{[]}\left(\varepsilon,\mathcal{F}\cdot\mathcal{G},L_r\left(P\right)\right)\leq N_{[]}\left(\frac{\varepsilon}{2L},\mathcal{F},L_r\left(P\right)\right)N_{[]}\left(\frac{\varepsilon}{2K},\mathcal{G},L_r\left(P\right)\right)\tag{68}$$

Then, the class of functions $\mathcal{L}$ is bounded by

$$\begin{aligned}\log N_{[]}\left(\varepsilon,\mathcal{L},||.||_{\infty}\right) &\leq \log N_{[]}\left(\varepsilon,\Delta,||.||_{\infty}\right)+\log N_{[]}\left(\varepsilon,\Theta,||.||_{E}\right)\\ &= Km_n\log\left(\frac{4M_{\phi}M_f}{\varepsilon}\right)+\log\left(\frac{2}{\varepsilon}\right)\end{aligned}$$

so that the class $\mathcal{L}$ is Glivenko-Cantelli. Moreover, $\mathcal{L}$ is Donsker. Then we can find $c^{\widehat{Q}}\left(m_n\right)$ explicitly by calculating the integral below

$$\int_0^1\sqrt{Km_n\log\left(\frac{4M_{\phi}M_f}{\varepsilon}\right)+\log\left(\frac{2}{\varepsilon}\right)}d\varepsilon$$

which obtains a result of order $O\left(\sqrt{1+m_n}\right)$. Therefore:

$$c^{\widehat{Q}}\left(m_n\right)=\left(\frac{1+m_n}{n}\right)^{1/2}$$

(ii) The second part of condition B.1.4 states that

$$c^{\widehat{Q}}\left(m_n\right) = o\left(1\right) \tag{69a}$$

$$\left|Q\left(\Pi_n\alpha_0\right)-Q\left(\alpha_0\right)\right| = o\left(1\right) \tag{69b}$$

$$\eta_n = o\left(1\right) \tag{69c}$$

$(69a)$ holds since d_{θ} is fixed and by construction $m_n\to\infty$ at a rate slower than $n.(69b)$ is satisfied by the continuity of $Q\left(\alpha\right)$ and by $\Pi_n\alpha_0\to\alpha_0$ from Condition B.1.2(ii). $(69c)$ holds with η_n small enough by the uniform convergence of the criterion function.

B Monte Carlo

B.1 DH DGP Estimated by DH

This subsection presents simulation exercise 1. For the Monte Carlo simulations, $\{X_i\}_{i=1}^{n=1000}$ is generated from the uniform distribution on $[0,1]$. The duration T_i associated to each X_i is calculated by solving (15) for $\{T_i\}_{i=1}^{n=1000}$. For estimation, the stochastic process has $\nu=1$. Three simulation studies are presented, where the true functions and the true ρ are summarized in table 1 below:

Table 1: Simulation Studies

True Parameters	Study 1	Study 2	Study 3
$\phi(x)$	$\exp(2x-3x^2)$	$\exp(2x-3x^2)$	$1+\sqrt{x}-x^3$
$f(t)$	$1-t+\frac{2}{3}t^3$	$1-t+\frac{2}{3}t^3$	$1-t+\frac{2}{3}t^3$
ρ	1	2	1

For all cases, the functions are approximated by polynomial splines of the second degree:

$$\begin{aligned}\log\phi_n(x) &= \sum_{j=0}^{2} a_j x^j + b_1 \max\{x-q_1^x,0\}^2,\ a_0=0\\ \log f_n(t) &= \sum_{j=0}^{2} c_j x^j + \sum_{j=1}^{3} d_j \max\left\{t-q_j^t,0\right\}^2,\ c_0=0\end{aligned}$$

where $q_1^x = 0.5$ quantile for x and the $q_1^t = 0.2$, $q_2^t = 0.5$, and $q_3^t = 0.8$ are the quantiles for t. The simulation results for the two functions of interest are included in figures 2, 3, and 4, where the green lines are the functions from each simulation, the blue lines are the averages over the simulations, and the red lines are the true function. The average of the rate parameters are shown in table 2 below.

Table 2: Rate Estimators

	True ρ	Estimated ρ	Standard Deviation	Simulations
Study 1	1	1.09	0.156	300
Study 2	2	2.16	0.24	300
Study 3	1	1.157	0.239	250

B.2 DH DGP Estimated by MPH

For simulation exercise 2, the data were generated as in study 1 presented above.

The estimating model is the MPH model with gamma heterogeneity, with scale parameter 1 and the rate to be estimated. There are two different studies: in the first one both functions ϕ and λ are estimated by second degree polynomial splines:

$$\begin{aligned}\log\phi_n(x) &= \sum_{j=0}^{2} a_j x^j + b_1 \max\{x-q_1^x,0\}^2,\ a_0=0\\ \log\lambda_n(t) &= \sum_{j=0}^{2} c_j x^j + \sum_{j=1}^{3} d_j \max\left\{t-q_j^t,0\right\}^2,\ c_0=0\end{aligned}$$

where $q_1^x = 0.5$ quantile for x and the $q_1^t = 0.2,\ q_2^t = 0.5$, and $q_3^t = 0.8$ are the quantiles for t.

For the second study, both functions are parametrized according to forms that are used in practice:

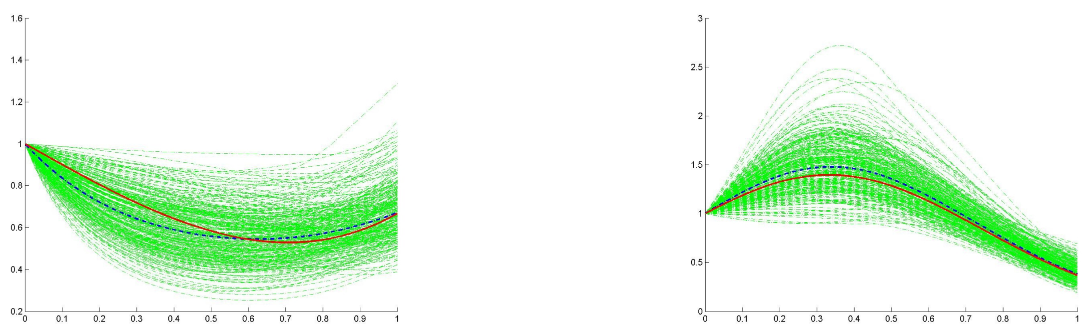
$$\begin{aligned} \phi(x) &= \exp\left(ax + bx^2\right) \\ \lambda(t) &= \alpha_1\alpha_2\left(\alpha_3 + t\right)^{\alpha_2 - 1},\ \alpha_i > 0 \end{aligned} \tag{70}$$

where the baseline hazard is the generalized Weibull function.

The results of the simulations are included in figures 5 and 6 where, as before, the green lines are the functions from each simulation, the blue lines are the averages over the simulations, and the red lines are the true function.

B.3 MPH DGP Estimated by DH

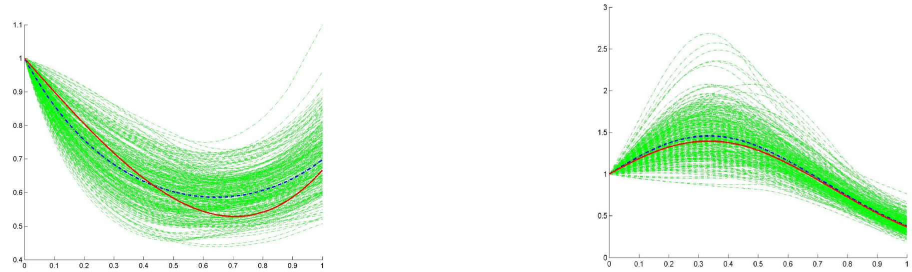
For the third simulation exercise (see figure 7), the data were generated by the MPH with gamma unobserved heterogeneity with scale and shape parameters equal to 1. The baseline hazard is the Weibull function, $\alpha_1\alpha_2 t^{\alpha_2-1}$, with $\alpha_1 = 1,\ \alpha_2 = 2$, and $\phi(x) = \exp\left(ax + bx^2\right)$, with $a = 2$, $b = -3$. $\{X_i\}_{i=1}^{n=1000}$ is generated from the uniform distribution on $[0,1]$. The duration T_i associated to each X_i is calculated by solving the MPH for $\{T_i\}_{i=1}^{n=1000}$. The DH is fitted as in study 1.



(a) Function f(t)

(b) Function phi(x)

Figure 2: Study 1 Simulations



(a) Function f(t)

(b) Function phi(x)

Figure 3: Study 2 Simulations

(a) Function f(t)

(b) Function phi(x)

Figure 4: Study 3 Simulations

(a) Estimate of phi(x)

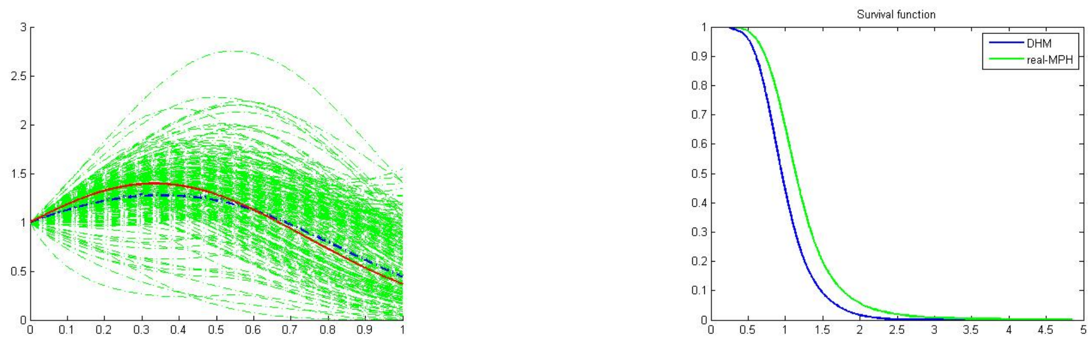
(b) Estimate of survival function

Figure 5: Spline MPH Estimators for True GDP with Stochastic Heterogeneity

(a) Estimate of phi(x)

(b) Estimate of survival function

Figure 6: Parametric MPH Estimators for True GDP with Stochastic Heterogeneity



(a) Estimate of phi(x): True (red), Average of estimators (blue), Estimators (green)

(b) Estimate of survival function: Average estimator (blue)

Figure 7: True GDP is MPH with Gamma Heterogeneity

B.4 Data Description and Summary Statistics

The empirical analysis uses data from the NLSY79, on the first spell of unemployment for both men and women. The NLSY79 is a national probability sample of 12,686 men (6,403) and women (6,283) who were between 14 and 22 years old when they were first surveyed in 1979. Information on demographic characteristics and labor force participation of the individuals has been collected annually from 1979 to 1993, and biennially from 1994 to the present.

Using the NLSY79 data set, it is possible to construct complete work histories since information for the non-interview period will be recovered in the next interview period. That is, for an individual last interviewed in 1980, and not interviewed until 1982, the information between $1980 - 1982$ will be recovered in the 1982 interview. The individual's labor force history is constructed by having the individual fill in the weeks between reported start and end dates for different activities with the appropriate labor status code. The codes range from: no information provided, unemployed, out of the labor force, associated with an employer but not currently working, being in active military service, and employed[34]. Note that a job held for any day of the week is considered as the job for the whole week.

The analysis sample is composed of 1240 individuals, out of which 616 are men. The original sample is restricted to completed first unemployment spells ranging from 1979 to 2002. Completed spells are defined as transitions out of employment to unemployment and then back to employment. Unemployed workers are those who did not work at all during the survey week but have searched for a job in the four weeks prior to the survey and during the survey week. The gaps during the unemployment spell, when individuals are

[34] It is noted in Appendix 18 of the NLSY documentation files that the quality or completeness of the work history files are not compromised by the possible inconsistencies generated by having the respondent fill in the gaps between start and end dates of jobs.

unemployed for part of the spell and out of the labor force for the other part of the spell, are problematic. Such situations are not considered in the analysis, as they do not respect the definition of a completed spell. Additionally, situations for which no information was provided for part of the time that individuals were unemployed are also not considered. Out-of-labor force and no-information provided categories are easy to drop as the NLSY has a special code for such situations.

The duration of a spell is the difference in weeks between the start and the end of the spell. Considering completed spells eliminates censoring issues as well as issues concerning missing data, occurrence dependence, and lagged dependence[35].

The application uses net wealth at the beginning of the unemployment spell. Net wealth is calculated as the sum of the market value of residential property, farm or business property, vehicles, money assets (savings), other assets each worth more than \$500, IRS, tax deferred plans, CDs, loans, stocks, bonds, mutual funds, and investment trust to be received, minus the amount of mortgage debt and back taxes, other debt on property, debt on farm/business property, amount of money owned on vehicles, and other debts over \$500.

The sample is divided first in two groups based on education level: low education and high education. Low-education are those who either have less than twelve years of schooling, where we do not differentiate between high-school drop-outs and successful high-school graduates. The high-education group is formed of those who have more than twelve years of schooling, where again, we do not differentiate between those who have and those who have not successfully graduated college. The sample is also divided by marital status. Individuals who are not married can be either single, separated, divorced, or widowed. For a table containing descriptive statistics for the data, see Tables 3 and 4.

[35] Occurrence (lagged) dependence takes place when the probability of leaving unemployment depends on the number (duration) of previous unemployment spells.

Table 3: Descriptive Statistics by Education Level

			College		
Men	Total 223				
		min	max	mean	std
	Duration (weeks)	1	116	13.06	15.23
	Net wealth (dollars)	-31,401	499,000	27,715	68,912
	Age	20	41	26.87	4.42
	Children	0	5	0.37	0.85
			High School		
	Total 391				
		min	max	mean	std
	Duration (weeks)	1	128	16.23	18.75
	Net wealth (dollars)	-147,000	193,700	11,912	29,682
	Age	20	41	27.26	4.27
	Children	0	7	0.73	1.1
			College		
Women	Total 335				
		min	max	mean	std
	Duration (weeks)	1	100	12.83	17.44
	Net wealth (dollars)	-249,000	487,067	20,717	49,897
	Age	20	43	27.75	4.47
	Children	0	5	0.7	1.05
			High School		
	Total 281				
		min	max	mean	std
	Duration (weeks)	1	142	15.88	20.53
	Net wealth (dollars)	-86,400	310,680	11,734	11,894
	Age	20	42	27.16	4.80
	Children	0	6	1.26	1.24

Table 4: Descriptive Statistics by Marital Status

		Married			
Men	Total 209				
		min	max	mean	std
	Duration (weeks)	1	116	13.66	17.05
	Net wealth (dollars)	-31,401	499,000	36,883	73,739
	Age	20	41	28.73	4.19
	Children	0	5	1.42	1.14
		Not Married			
	Total 415				
		min	max	mean	std
	Duration (weeks)	1	128	15.5	17.61
	Net wealth (dollars)	-147,000	221,500	8,209	23,898
	Age	20	41	26.3	4.17
	Children	0	7	0.18	0.65
		Married			
Women	Total 236				
		min	max	mean	std
	Duration (weeks)	1	100	13.71	18.04
	Net wealth (dollars)	-86,400	487,067	30,282	60,595
	Age	20	42	27.92	4.6
	Children	0	5	1.4	1.2
		Not Married			
	Total 380				
		min	max	mean	std
	Duration (weeks)	1	142	15.1	20.06
	Net wealth (dollars)	-33,000	219,075	8,134	25,562
	Age	20	43	26.32	4.53
	Children	0	6	0.67	1.05

B.5 Application Results

Table 5: DH Estimators

	Men		Women	
	College	High School	College	High School
rate	1.577	0.807	0.864	0.056
age	-0.1489	-0.1268	-0.0645	-0.005
age2	-0.009	0.0006	-0.003	0.0005
kids	0.0003	0.3703	-0.0103	-0.0006
marital	-0.009	0.2439	0.0361	0.00014
	Married	Not Married	Married	Not Married
rate	1.853	0.875	0.566	1.655
age	-0.1557	-0.0996	-0.1319	-0.1998
age2	-0.0062	0.0037	0.0164	-0.0172
kids	0.1721	0.00997	-0.4636	-0.2461
education	-0.0844	0.0062	0.9198	0.2915

Table 6: MPH Estimators

	Men		Women	
	College	High School	College	High School
rate	8665.835	0.4394	118.0114	77.0892
age	-0.3871	0.0391	-0.0211	-0.0036
age2	-0.0805	-0.0088	-0.005	-0.0126
kids	0.0048	0.0395	-0.043	-0.0744
marital	-0.0144	0.2348	0.1384	-0.0842
wealth	0.1166	0.3767	-0.5682	0.4498
	Married	Not Married	Married	Not Married
rate	138.5528	0.2553	108.8406	48.6275
age	-0.0492	0.0671	-0.0271	-0.0001
age2	-0.0056	-0.0026	-0.002	-0.0039
kids	0.0448	-0.0338	-0.0009	-0.0694
education	-0.0232	0.2925	0.1681	0.0891
wealth	0.2176	-2.9242	-0.0672	-0.6894

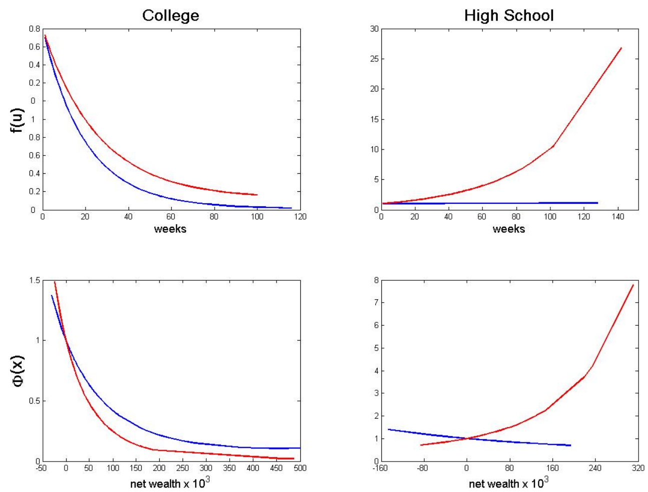


Figure 8: Estimators by Education for Men (blue) and Women (red)

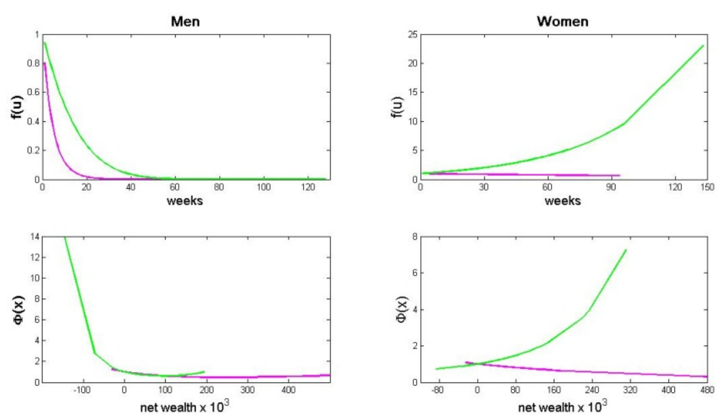


Figure 9: Estimators for College (green) and Highschool (magenta)

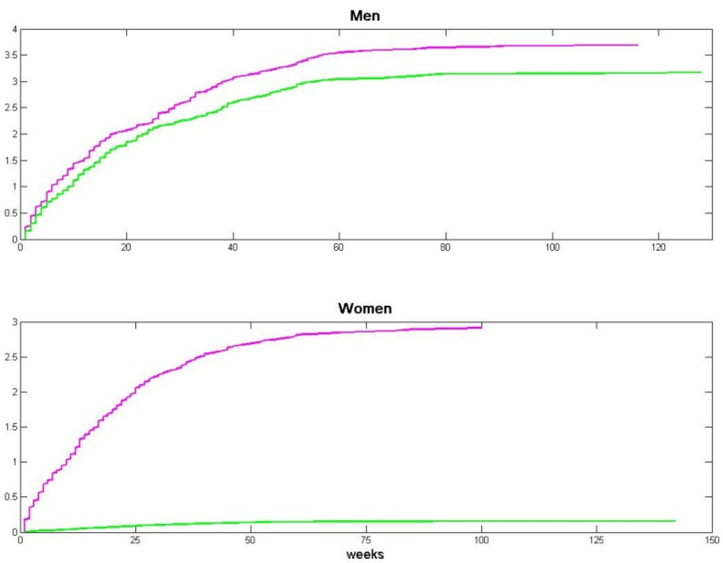


Figure 10: Hazards for College (green) and Highschool (magenta)

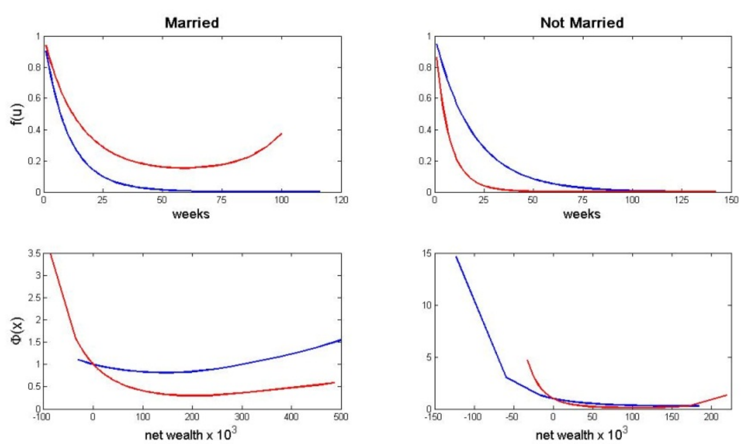


Figure 11: Estimators by Marital Status for Men (blue) and Women (red)

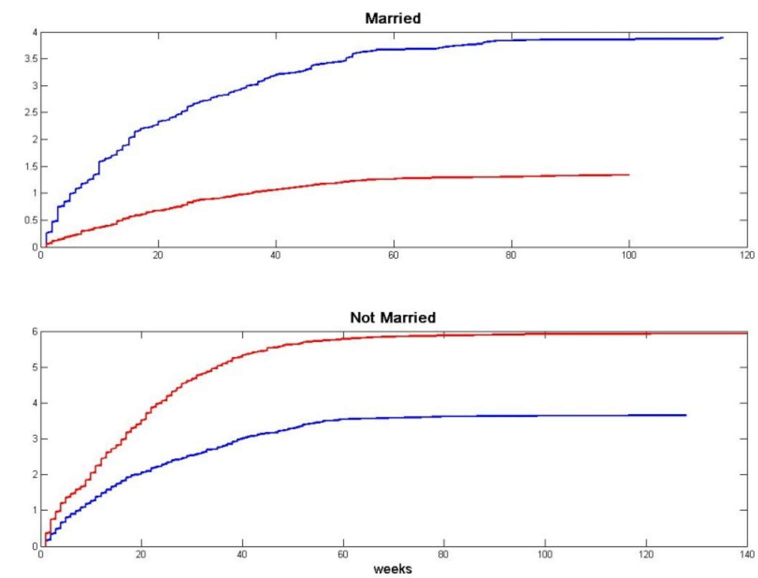
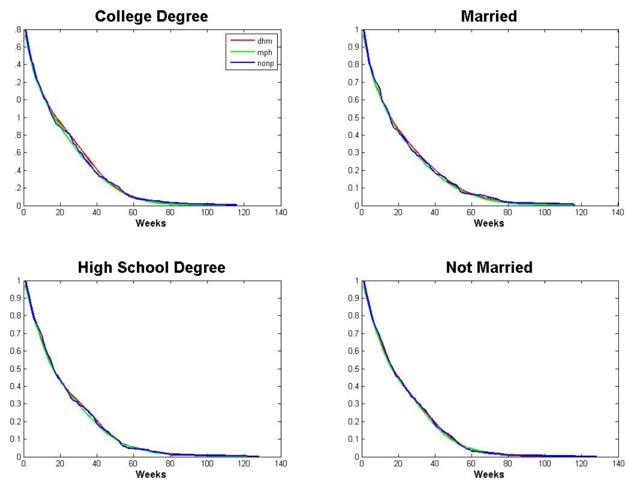
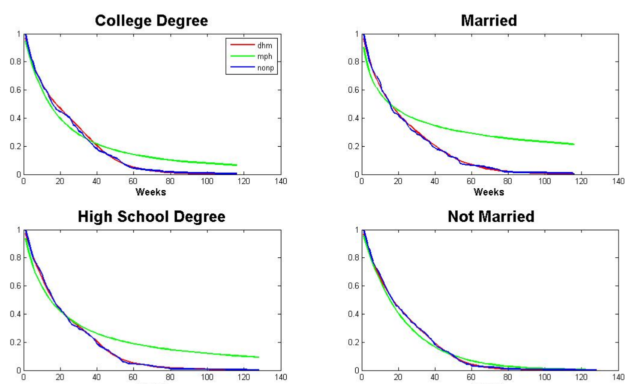


Figure 12: Hazard Functions by Marital Status for Men (blue) and Women (red)

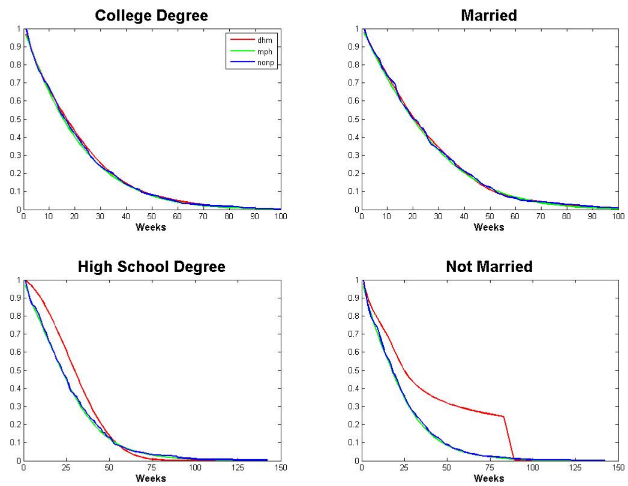


(a) Case S Survival Functions

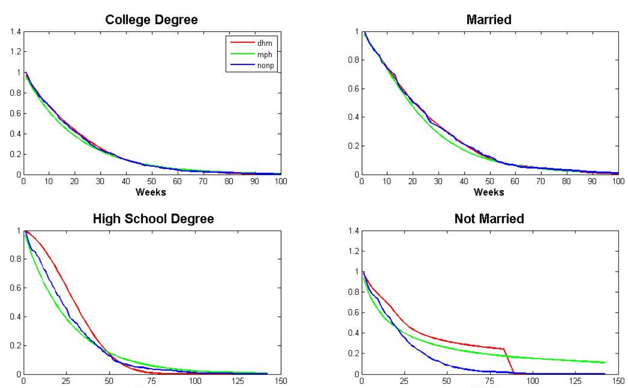


(b) Case P Survival Functions

Figure 13: Survival Function for Men: DH (red), MPH (green), Nonparametric (blue)



(a) Case S Survival Functions



(b) Case P Survival Functions

Figure 14: Survival Function for Women: DH (red), MPH (green), Nonparametric (blue)

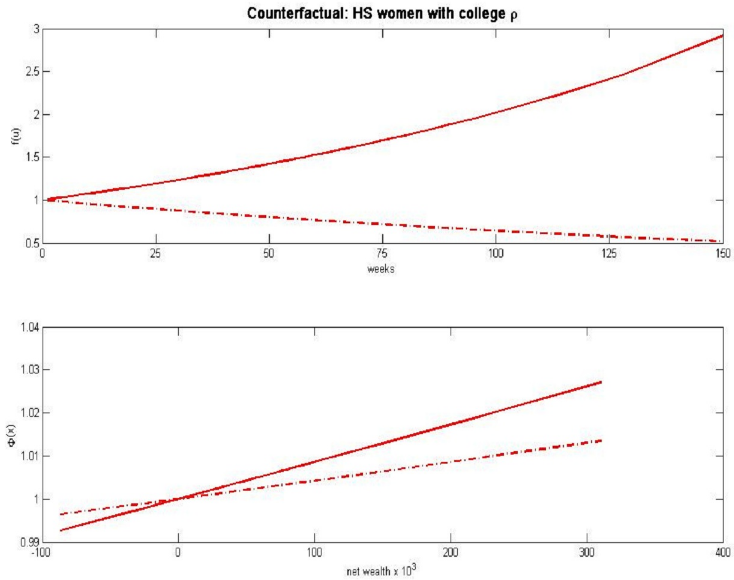


Figure 15: Counterfactual (dotted line) vs original estimators (continuous line)

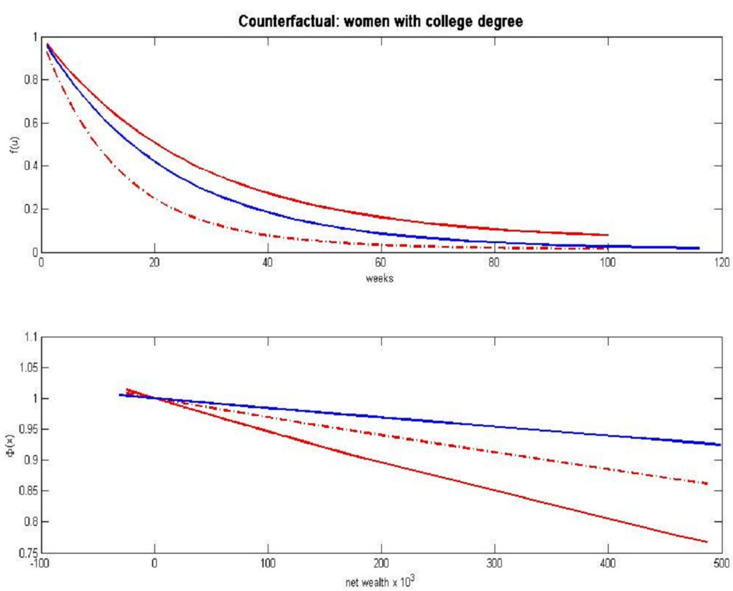


Figure 16: Women's counterfactual (dotted red) vs men's (blue) and women's original estimators (red continuous)

References

Aalen, O., and N.L. Hjört. 2002. "Frailty Models that Yield Proportional Hazards." *Statistics and Probability Letters*, 58: 335–342.

Aalen, O., H.K. Gjessing, and N.L. Hjört. 2003. "Frailty Models Based on Lévy Processes." *Advances in Applied Probability*, 35: 532–550.

Abbring, J.H. 2011. "Mixed Hitting-Time Models." *Econometrica.* forthcoming.

Ai, C., and X. Chen. 2003. "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions." *Econometrics*, 71: 1795–1844.

Arkes, H., and C. Blumer. 1985. "The Psychology of Sunk Cost." *Organizational Behavior and Human Decision Process*, 35: 124–140.

Bertoin, J. 1996. *Lévy Processes.* Cambridge University Press.

Brinch, C.N. 2007. "Nonparametric Identification of the Mixed Hazards Model with Time-Varying Covariates." *Econometric Theory*, 23: 349–354.

Brockner, J., and J.Z. Rubin. 1985. *Entrapment in escalating conflicts: A social psychological analysis.* New York: Springer-Verlag.

Brown, M., C. Flinn, and A. Schotter. 2011. "Real-Time Search in the Laboratory and the Market." *American Economic Review*, 2: 948–974.

Chen, X. 2007. "Large Sample Sieve Estimation of Semi-Nonparametric Models." In *Handbook of Econometrics.* Vol. 6 of *Handbook of Econometrics*, , ed. J.J. Heckman and E.E. Leamer, Chapter 76, 5550–5632. Elsevier.

Chen, X., and D. Pouzo. 2008. "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals." Cowles Foundation for Research in Economics, Yale University Cowles Foundation Discussion Papers 1650R.

Clark, A. E., and A.J. Oswald. 1994. "Unhappiness and Unemployment." *Economic Journal*, 104(424): 648–659.

Crowder, M. 2001. *Classical Competing Risks.* Chapman and Hall/CRC.

Dabrowska, D.M. 2005. "Estimation in a class of semiparametric transformation models."

de Boor, C. 1978. *A Practical Guide to Splines.* Springer-Verlag, New York.

Dragulescu, A., and V.M. Yakovenko. 2001. "Evidence for the exponential distribution of income in the U.S.A." *The European physical journal B*, 20(4): 585–589.

Elbers, C., and G. Ridder. 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies*, 49(3): 403–409.

Faber, H.S. 2008. "Reference-dependent preferences and labor supply: The case of New York city taxi drivers." *American Economic Review*, 98(2): 1069–1082.

Feller, W. 1970. *An Introduction to Probability Theory and Its Applications.* Vol. II. 2 ed., Wiley, New York.

Gallant, A.R., and D.W. Nychka. 1987. "Semi-nonparametric Maximum Likelihood Estimation." *Econometrica*, 55(2): 363–90.

Geman, S., and C. Hwang. 1982. "Nonparametric Maximum Likelihood Estimation by the Method of Sieves." *Annals of Statistics*, 10: 401–414.

Gnedin, A., and J. Pitman. 2008. "Moments of Convex Distribution Functions and Completely Alternating Sequences." In *Probability and Statistics: Essays in Honor of David A. Freedman.* , ed. D. Nolan and T. Speed, 30–41. Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Gourville, J.T., and D. Soman. 1998. "Payment depreciation: The behavioral effects of temporally separating payments from consumption." *Journal of consumer research*, 25: 160–174.

Gruber, J. 2001. "The wealth of the unemployed." *Industrial and labor relations review*, 55(1): 79–94.

Gurr, T.R. 1970. *Why Men Rebel.* Princeton University Press, Princeton.

Heath, C. 1995. "Escalation and de-escalation of committment in response to sunk costs: The role of budgeting in mental accounting." *Organizational Behavior and Human decision processes*, 62(1): 38–54.

Heckman, J.J., and B. Singer. 1984. "Econometric duration analysis." *Journal of Econometrics*, 24(1-2): 63–132.

Heckman, J.J., and C.R. Taber. 1994. "Econometric Mixture Models and More General Models for Unobservables in Duration Analysis." National Bureau of Economic Research, Inc NBER Technical Working Papers 0157.

Honoré, B.E. 1993. "Identification Results for Duration Models with Multiple Spells." *Review of Economic Studies*, 60(1): 241–46.

Huang, J. 1996. "Efficient Estimation for the Cox Model with Interval Censoring." *Annals of Statistics*, 24: 540–568.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263–91.

Kebir, Y. 1991. "On Hazard Rate Processes." *Naval Research Logistics*, 6(38): 865–876.

Khan, S., and E. Tamer. 2010. "Irregular Identification, Support Conditions, and Inverse Weight Estimation." *Econometrica*, 78(6): 2021–2042.

Kogut, C.A. 1990. "Consumer search behavior and sunk costs." *Journal of Economic Behavior & Organization*, 14(3): 381–392.

Lee, M-L.T., and G.A. Whitmore. 2004. "First Hitting Time Models for Lifetime Data." In *Advances in Survival Analysis.* , ed. C.R. Rao and N. Balakrishnan, 227–235. Birkhäuser, Boston.

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

Lopez-Ruiz, R., J.L. Lopez, and X. Calbet. 2011. "Exponential wealth distribution: A new approach from functional iteration theory." *arXiv:1103.1501v1.*

McCall, B.P. 1993. "Identifying State Dependence in Duration Models with Time-Varying Regressions." Minnesota - Industrial Relations Center Papers 93-06.

McKean, K. 1985. "Decisions, Decisions, Decisions." *Discover*, 22–31.

Pandey, M.D., X.X. Yuan, and J. M. Van Noortwijk. 2009. "The Influence of Temporal Uncertainty of Deterioration on Life-Cycle Management of Structures." *Structure and Infrastructure Engineering*, 5(2): 145–156.

Ryu, K. 1993. "Structural Duration Analysis of Management Data." *Journal of Econometrics*, 57(1-3): 91–115.

Schumaker, L.L. 1981. *Spline Functions, Basic Theory.* John Wiley & Sons, New York.

Shefrin, H., and M. Statement. 1985. "The disposition to sell winner too early and ride losers too long: The theory and evidence." *Journal of Finance*, 40: 777–790.

Shen, X. 1995. "On Methods of Sieves and Penalization." *The Annals of Statistics*, 17(6): 2555–2591.

Shen, X., and W.H. Wong. 1994. "Convergence Rate of Sieve Estimates." *The Annals of Statistics*, 22(2): 580–615.

Shmer, R. 2008. "The probability of finding a job." *American economic review*, 98(2): 268–273.

Singpurwalla, N.D. 1995. "Survival in Dynamic Environments." *Statistical Science*, 10(1): 86–103.

Singpurwalla, N.D. 2006. "The Hazard Potential: Introduction and Overview." *Journal of the American Statistical Association*, 101(476): 1705–1717.

Soman, D. 2001. "The mental accounting of sunk time costs: Why time is not like money." *Journal of Behavioral decision making*, 14: 169–185.

Stock, J.H. 1988. "Estimating Continuous-Time Processes Subject to Time Deformation: An Application to Postwar U.S. GNP." *Journal of the American Statistical Association*, 83(401): 77–85.

Stone, C.J., M.H. Hansen, C. Kooperberg, and Y.K. Truong. 1997. "Polynomial Splines and their Tensor Products in Extended Linear Modeling." *The Annals of Statistics*, 25(4): 1371–1425.

Thaler, R. 1980. "Toward a positive theory of consumer choice." *Journal of Economic Behavior & Organization*, 1(1): 39–60.

Thaler, R., and E. Johnson. 1990. "Gambling with the house money and trying to break even: The effects of prior outcomes in risky choices." *Management Science*, 36(6): 643–660.

Tsilevich, N., A. Vershik, and M. Yor. 2001. "An Infinite-Dimensional Analogue of the Lebesgue Measure and Distinguished Properties of the Gamma Process." *Journal of Functional Analysis*, 185: 274–296.

Van den Berg, G.J. 2001. "Duration models: specification, identification and multiple durations." In *Handbook of Econometrics*. Vol. 5 of *Handbook of Econometrics*, , ed. J.J. Heckman and E.E. Leamer, Chapter 55, 3381–3460. Elsevier.

Wellner, J.A., and Y. Zhang. 2007. "Two Likelihood-Based Semiparametric Estimation Methods for Panel Count Data with Covariates." *The Annals of Statistics*, 35(5): 2106–2142.