Botosaru, I. (2020). Nonparametric Analysis of a Duration Model with Stochastic Unobserved Heterogeneity. *Journal of Econometrics*, 217(1), 112-139. <https://doi.org/10.1016/j.jeconom.2019.06.006>

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.jeconom.2019.06.006](https://doi.org/10.1016/j.jeconom.2019.06.006)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://www.sciencedirect.com/science/article/pii/S030440761930243X> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Nonparametric Analysis of a Duration Model with Stochastic Unobserved Heterogeneity*

Irene Botosaru[†]

April 20, 2019

Abstract

This paper develops nonparametric identification and estimation results for a single-spell hazard model, where the unobserved heterogeneity is specified as a Lévy subordinator. The identification approach solves a nonlinear Volterra integral equation of the first kind with an unknown kernel function defined on a non-compact support. Both the kernel of the integral operator, which models the distribution of the unobserved heterogeneity, and the functions that enter it nonlinearly are identified given regularity conditions and minimal variation in the observed covariates. The paper proposes a shape-constrained nonparametric two-step sieve minimum distance estimator. The second step estimates the kernel of the integral operator, exploiting a monotonicity property. Rates of convergence are derived and Monte Carlo experiments show the finite sample performance of the estimator.

JEL classification: C14, C41

Keywords: duration model, Lévy process, nonlinear Volterra integral equation of the first kind, shape restricted two-step nonparametric estimation.

*I would like to thank the editor, associate editor, and two referees for their comments and suggestions, which have greatly improved the paper. I thank Jaap Aabring, Donald Andrews, Donald Brown, Xiaohong Chen, Kirill Evdokimov, Jean-Pierre Florens, David Freeman, James Heckman, Bo Honoré, Anke Kessler, Yuichi Kitamura, Simon Lee, Chris Muris, Whitney Newey, Krishna Pendakur, Franco Peracchi, Peter Phillips, Christoph Rothe, Sami Stouli, Francis Vella, Frank Windmeijer, and Simon Woodcock for their comments and suggestions. I thank participants at various seminars and conferences for questions and comments. Financial support from Cowles Foundation via the Carl Arvid Anderson Prize and from the SSHRC through IDG 430-2015-00073 is gratefully acknowledged.

[†]University of Bristol, Department of Economics, The Priory Road Complex, 12 Priory Road, Bristol BS8 1TU, United Kingdom. Email: irene.botosaru@bristol.ac.uk.

1 Introduction

Duration analysis is used to model transitions across states as well as the duration of a spell in a given state.¹ Standard duration models incorporate both observed and unobserved heterogeneity. Up to now, identification of these models has been shown only for the case where the unobserved heterogeneity is fixed over time. In this paper, I consider a single-spell duration model that features time-varying unobserved heterogeneity. In particular, the model allows for unobserved persistent individual shocks that accumulate during the spell, in a data context where at most one transition is observed for each individual. I first show the nonparametric identification of the model by establishing sufficient conditions for the unique solvability of a nonlinear Volterra integral equation of the first kind with unknown kernel function. I then propose a shape-constrained nonparametric two-step sieve minimum distance estimator, derive its rate of convergence, and show its finite sample properties.

The model considered belongs the class of models proposed in biostatistics by Gjessing, Aalen, and Hjort (2003). This class expands the field of duration modeling to contexts where unobserved heterogeneity ought to be time-varying, and, in particular, where it ought to accumulate during the spell. Examples in economics include: job-specific skills that accumulate during the spell of employment affecting the probability of being promoted; signals on an employee's productivity that accumulate during the employment spell affecting the probability of being laid-off; the private value of an innovation that accumulates during the development phase affecting the probability of patenting; occupational injuries and stress that accumulate during the duration of a job affecting the probability of early retirement. In all of these cases, time-invariant unobserved heterogeneity - the standard approach - is overly restrictive, and imposition of this restriction on duration modeling could lead the researcher to false conclusions.

The model considered is as follows. All random variables and stochastic processes appearing in this paper are defined on a filtered probability space, $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$, satisfying the usual conditions. Let T be a non-negative random variable modeling the length of time of an individual spell, let X denote exogenous observed random variables, and let $\{Z(t)\}_{t \geq 0} \equiv \{Z(t, \omega) : t \geq 0, \omega \in \Omega\}$ be an

¹In economics, duration models have been applied to study the length of wars, strikes, unemployment, marriage, as well as the likelihood of bankruptcy or technological breakthrough, see e.g. van den Berg (2001) for a complete list of applications.

exogenous unobserved stochastic process modeling individual time-varying unobserved heterogeneity. The process $\{Z(t)\}_{t \geq 0}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$, where

$$\mathcal{F}_t = \sigma(Z(u), X, 0 \leq u \leq t), \text{ for any } t \geq 0, \quad (1)$$

and conditional on $X = x \in \mathcal{X} \subseteq \mathbb{R}^q$, $q \geq 1$, an individual's spell duration is characterized by the following hazard function:

$$h(t|\mathcal{F}_t) \equiv \phi(x) \int_0^t f(u) dZ(u), \text{ for any } t \geq 0, \quad (M)$$

where $\phi : \mathcal{X} \rightarrow (0, \infty)$ is a function of observed covariates, $f : [0, \infty) \rightarrow (0, \infty)$ is a non-random function, and $\{Z(t)\}_{t \geq 0}$ satisfies the following assumption:

Assumption Z. The stochastic process $\{Z(t)\}_{t \geq 0}$ is a Lévy subordinator.

Lévy subordinators are stochastic processes with increments that are non-negative, stationary, independent, and of finite variation, see Appendix A for the formal definition and Figure 1 for an arbitrary sample path of a subordinator. Since the hazard function is non-negative for any $t \geq 0$, Assumption Z is a sufficient but not necessary assumption for $h(t|\mathcal{F}_t) \geq 0$ for any t .² Under Assumption Z, the distribution of $\{Z(t)\}_{t \geq 0}$ is characterized by the distribution of $Z(1)$, which denotes the random variable $Z(t)$ at $t = 1$, i.e.

$$Z(1) \equiv \{Z(t=1, \omega), \omega \in \Omega\}.$$

Let F_Z denote this distribution function.

The main contribution of this paper is establishing sufficient conditions for the nonparametric identification of the functions (ϕ, f, F_Z) , where the arguments of the functions have been suppressed for convenience. The identification strategy involves solving a *nonlinear* Volterra integral equation of the first kind with an *unknown* kernel function that is defined on a non-compact support. The kernel

²Given specification (M), the hazard function is non-negative whenever the sum of all negative increments is smaller than the sum of all positive increments. Imposing such an assumption, though, seems much more difficult to justify in applied work than Assumption Z. Hence, I do not pursue this approach here.

of the integral operator, which models F_Z , and the functions (ϕ, f) that enter the kernel, are identified given regularity assumptions and minimal variation in the observed covariates. The nonparametric identification results in this paper extend those in the integral equation literature to a class of models not covered by existing identification results. In particular, both the nonlinearity of the Volterra integral operator and the fact that the kernel function is unknown and non-compactly supported are new features when compared to those in the recent econometrics literature on integral operators, see Literature Review.

The paper then proposes a shape-constrained nonparametric two-step sieve minimum distance estimator, where both the first-step and the second-step involve infinite-dimensional parameters. The second step estimator estimates three functions. One of the functions, the kernel of the integral operator, is monotone on its entire domain. The second-step estimator exploits this shape restriction by approximating the kernel function with a Bernstein polynomial and imposing specific restrictions on the sieve coefficients. The paper derives the asymptotic properties of the second-step estimators and shows their performance in finite samples.

Interpretation of the Model. Interpreting the increments of the stochastic process as shocks to individual unobserved heterogeneity, Assumption Z implies that the shocks are non-negative, have random magnitude, accumulate during the spell, and happen at random times during the spell. In addition, Assumption Z implies that there is persistence in levels but that *changes* in levels corresponding to non-overlapping time periods are statistically independent, e.g., for $0 \leq t_1 < t_2 < t_3$, $Z(t_1) \leq Z(t_2) \leq Z(t_3)$ but knowing $Z(t_2) - Z(t_1)$ does not help predict $Z(t_3) - Z(t_2)$. The Gamma process and the compound Poisson process are examples of stochastic processes that satisfies this assumption. The latter process has been used in economics to model the evolution over the development phase of the value of an innovation, see, e.g., Hopenhayn and Squintani (2011).

Function f changes the way that shocks to the unobserved heterogeneity affect the likelihood of exit depending on when they happen during the spell. For example, given a sample path of $\{Z(t)\}_{t \geq 0}$, a decreasing f increases the hazard function early on in the spell, thus increasing the probability of exiting earlier (a steeper survival function). On the other hand, an increasing f causes shocks that occur later on in the spell to increase the likelihood of exiting later, thus increasing the probability

The Effect of $f(t)$ on the Hazard Function and on the Survival Function

Figure 1: The figure plots a path of a gamma subordinator (top left panel), three different functions $f(t)$ (top right panel), and the effect of each $f(t)$ on the hazard function (bottom left panel) and on the survival function (bottom right panel). The underlying process $Z(t)$ is the gamma process with scale and shape parameters equal to 1. The three different functions $f(t)$ are: constant, $f(t) = 1$ (continuous line), decreasing, $f(t) = \exp(-t)$ (dotted line), and increasing, $f(t) = 1 + \arctan(t)$ (dashed line). The bottom left panel plots the sample paths of the resulting hazards, given the sample path of $Z(t)$ in the top left panel and the functions $f(t)$ in the top right panel. This panel shows that a decreasing $f(t)$ increases the hazard early on in the spell, while an increasing $f(t)$ increases the hazard later on in the spell. The bottom right panel shows the survival functions associated to the hazard functions in the bottom left panel. The steepest survival function corresponds to a decreasing $f(t)$, which suggests an increased probability of exit early on in the spell. The less steep survival function corresponds to an increasing $f(t)$ suggesting an increase in the probability of exit later on in the spell.

of exiting later (a less steep survival function).³ Figure (1) illustrates the effect of the shape of f on both the hazard and the survival functions.

To summarize, the model given by the hazard function in (M) and Assumption Z implies that there are non-negative shocks to unobserved heterogeneity that accumulate during the spell, and that the timing of the shocks matters, i.e. depending on the shape of f , a shock that takes place earlier on in the spell may have a different effect on the probability of exit than a shock that takes place later on in the spell. To demonstrate the relevance of this model, consider the reduced-form example

³Mathematically, f transforms the stationary process $\{Z(t)\}_{t \geq 0}$ into a non-stationary process, $\{h(t|\mathcal{F}_t)\}_{t \geq 0}$. The increments of $\{Z(t)\}_{t \geq 0}$ are identically distributed, but those of the process $\{h(t|\mathcal{F}_t)\}_{t \geq 0}$ are not. For example, $\{Z(t)\}_{t \geq 0}$ is a homogeneous process, with a constant expected number of jumps per unit time, while $\{h(t|\mathcal{F}_t)\}_{t \geq 0}$ is a nonhomogeneous process, with a time-varying expected number of jumps per unit time.

below.⁴

Example. Suppose that the job-specific skills of employed individuals evolve during the employment spell according to an exogenous Lévy subordinator, $\{Z(t)\}_{t>0}$. Under this assumption, skills accumulate with tenure such that the stock of skills at some time $\tau + 1$ is at least as large as that at time τ , but such that knowing how much skills changed between periods τ and $\tau + 1$ does not help predict how much skills change between periods $\tau + 1$ and $\tau + 2$. Additionally, having job-specific skills accumulate for a longer time increases the likelihood of promotion – support for this assumption can be found in, e.g., Waldman (2013). Treating the transition hazard to promotion as a primitive and specifying it as in (M) implies that shocks to job-specific skills affect directly the risk of being promoted. The shocks are unobserved to the econometrician.

Assumption Z is similar to that usually made in the literature modeling human capital accumulation, where the process of accumulation is modeled as an AR(1) process with i.i.d. shocks, see, e.g., Khan and Lange (2014). The main differences between modeling human capital as an AR(1) process versus a Lévy subordinator are as follows. First, under the AR(1) specification, skills change continuously, while under the subordinator specification, they change in jumps. If agents learn when faced with tasks that arrive at random times, e.g., according to a Poisson distribution, then it seems reasonable to assume that the job-specific skill evolves in jumps. Second, the subordinator assumption produces considerable dispersion of skills at any time period due to the stochastic nature of the evolution of skills; this is not the case under the AR(1) specification. Third, *while employed*, the subordinator assumption restricts job-specific skills to be non-negative. That is, while on the job, shocks to job-specific skills either *improve* the level of the skills or leave the skills unchanged, but shocks never “harm” or decrease the level of job-specific skills. Under the AR(1) assumption, shocks to skills can have a negative impact and skills at time τ can be less than skills at $\tau - 1$. The AR(1) assumption is usually made in this literature since the scope is to model *lifetime* human capital, which includes negative shocks such as layoffs. The subordinator assumption would be better suited to model human

⁴Recent works by Singpurwalla (2006) and Liu (2016) suggest that it may be possible to apply the identification arguments of this paper to optimal stopping time models driven by Lévy subordinators. Rigorously developing the connection between durations generated by the reduced-form model of this paper and those generated by optimal stopping problems is left for future work, as it requires the development of new results from the stochastic processes literature concerned with the absolute continuity of Volterra processes driven by Lévy subordinators.

capital while employed at a specific firm, rather than lifetime human capital.

Model (M) can then be used to analyze how the probability of promotion varies with observed covariates X given that skills accumulate with tenure. The shape of the function f may be informative about the type of incentive schemes provided by the firm or industry. If f is decreasing, skills accumulated earlier on result in a higher likelihood of promotion, so the scheme may be categorized as “fast track,”⁵ while if f is increasing, it is skills accumulated later on that increase the probability of promotion and the scheme may be categorized as “late selection.” Fast track jobs provide incentive schemes that encourage the accumulation of human capital early on in the spell of employment. These are jobs in industries where the production process and human capital are complementary, see, e.g., Clemens (2012) and Ishida et al. (2012). On the other hand, late selection jobs reward skills that are accumulated later on in the tenure. These would be careers in “seniority oriented” firms where “differentiation within each cohort is delayed,” see Ariga et al. (1999).

Organization. The rest of the paper is organized as follows. Section 2 reviews the relevant duration model and integral equation literatures. Section 3 presents the identification results. After formulating the survival function in terms of the Laplace transform of F_Z in Section 3.1, the structure (ϕ, f, F_Z) is identified in Section 3.2. Section 4 introduces the estimators, and Section 4.1 presents their asymptotic properties. Section 5 shows the small sample performance of the proposed estimators via a Monte Carlo study. Section 6 concludes. The Appendix contains all proofs. Section A introduces Lévy subordinators and their properties that are used in the main text and in the proofs.

Notation. In what follows, “ $\equiv$ ” defines a term or introduces a notation. For any interval $(a, b) \in \mathbb{R}$, $a \lesssim b$ means $a < Cb$, where C is a generic positive and finite constant. For any function g defined on a set $\mathcal{D}$ denote $\|g\|_\infty \equiv \sup_{\chi \in \mathcal{D}} |g(\chi)|$. Denote by $C^m(\mathcal{D})$ be the space of m -times continuously differentiable real-valued functions on $\mathcal{D}$. Given a d -tuple of non-negative integers, $\alpha = (\alpha_1, \dots, \alpha_d)$, set $|\alpha| = \sum_{i=1}^d \alpha_i$, and $D^\alpha \equiv \frac{\partial^{|\alpha|}}{\partial \xi_1^{\alpha_1} \dots \partial \xi_d^{\alpha_d}}$. Let $m \in \mathbb{N}$, $\kappa \in (0, 1]$, $p \equiv m + \kappa$, and $\mathcal{D} \subseteq \mathbb{R}^d$. A function $g : \mathcal{D} \rightarrow \mathbb{R}$ is p -smooth if $g \in C^m(\mathcal{D})$ and, for some generic finite constant $b > 0$, it holds that $\sup_{|\alpha| \leq m} \|D^\alpha g\|_\infty \leq b$, and $\sup_{|\gamma|=m} \sup_{\xi_1, \xi_2 \in \mathcal{D}, \xi_1 \neq \xi_2} \frac{|D^\gamma g(\xi_1) - D^\gamma g(\xi_2)|}{\|\xi_1 - \xi_2\|_e^\kappa} \leq b$, where $\|\cdot\|_e$ is the $\mathbb{R}^d$ Euclidean norm. Define

⁵This is in line with Clemens (2012), who finds that for fast-track jobs the speed of job-specific skill accumulation is higher and it happens earlier on in the spell of employment, which in turn translates into higher rates of promotion.

the Hölder norm $\|g\|_{C^p} \equiv \sup_{|\alpha| \leq m} \|D^\alpha g\|_\infty + \sup_{|\gamma|=m} \sup_{\xi_1, \xi_2 \in \mathcal{D}, \xi_1 \neq \xi_2} \frac{|D^m g(\xi_1) - D^m g(\xi_2)|}{\|\xi_1 - \xi_2\|_e^\kappa}$ and denote by $(C_b^p(\mathcal{D}), \|\cdot\|_{C^p})$ the space $C_b^p(\mathcal{D})$ endowed with the norm $\|g\|_{C^p}$. For $1 \leq p \leq \infty$, let $L_p(\mathcal{D})$ denote the space of measurable functions $g : \mathcal{D} \rightarrow \mathbb{R}$ such that $\|g\|_{L_p} = \left(\int_{\mathcal{D}} |g(u)|^p \mu(u) du \right)^{1/p} < \infty$, where $\mu : \mathcal{D} \rightarrow \mathbb{R}_+$ is a continuous nonnegative weight function, and $\|\cdot\|_{L_p}$ is the L_p -norm.

2 Literature Review

The standard duration model applied in economics is the mixed proportional hazard model (MPH), which specifies the hazard function as

$$h(t|x, z) = \phi(x) \lambda(t) z$$

where $\lambda(t)$ is a positive deterministic function of time and z is non-negative a random variable denoting unobserved heterogeneity. The MPH has been extensively studied and applied, see, e.g., van den Berg (2001). Early nonparametric identification results have been provided by Elbers and Ridder (1982) and Heckman and Singer (1984). These identification results hinge on the multiplicative separability of t and z , and as such they cannot be used to study the identification of the model considered in this paper.

Despite its modeling flexibility, the MPH does not allow for non-proportional hazards or for time-varying unobserved heterogeneity, which restricts its applicability. Two alternative classes of duration models have been proposed in the (bio)statistics literature to address these issues, see, e.g., Aalen et al. (2008). One is the class of hitting-time duration models, whose identification has been studied by Abbring (2012), and the other is the class of duration models with hazard functions driven by stochastic processes, which is the subject of this paper.

The model analyzed by Abbring (2012) - the mixed hitting-time (MHT) model, specifies durations as the first time a spectrally negative Lévy process hits a threshold from above:

$$T_{MHT} = \inf \{t > 0 : Y(t) \leq \phi(X)z\}, \quad (2)$$

where $\{Y(t)\}_{t>0}$ is the Lévy process and z is a time-invariant random variable. Since the threshold is multiplicative in X and z , Abbring (2012) is able to use insights from the identifiability of the MPH in order to obtain the nonparametric identification of the MHT model, i.e. of ϕ and of the distributions of z and $\{Y(t)\}_{t\geq 0}$.

Although Abbring (2012) does not provide this interpretation, if the process $\{Y(t)\}_{t>0}$ in (2) were interpreted as unobserved heterogeneity, the MHT could be seen as relaxing the time-invariance assumption of the unobserved heterogeneity. Then the MHT and the model of this paper are implicitly imposing different assumptions on the unobserved heterogeneity: the former assumes heterogeneity is a spectrally negative Lévy process (a Lévy process with non-positive jumps perturbed by a diffusion), while the latter assumes that it is a Lévy subordinator (a pure-jump non-decreasing Lévy process). As such the two specifications apply to different phenomena. For example, the MHT framework cannot be used if unobserved heterogeneity stands for the accumulation of skills, signals, or the effects of aging.

In a subsequent paper, Liu (2016) proposes a hitting-time model based on Lévy subordinators, where durations are specified as:

$$T_L = \inf \{t > 0 : \exp(X'\beta) Y(\Lambda(t)) \geq z\},$$

where $\{Y(t)\}_{t>0}$ is a Lévy subordinator, Λ is a deterministic function, β is a finite dimensional parameter, and $z \sim \exp(1)$. He shows the identification of (β, Λ) and of the distribution of $\{Y(t)\}_{t>0}$, and he proposes a sieve maximum partial likelihood estimator for β . The model in Liu (2016) and that of this paper are non-nested, since, for example, the former always obtains proportional population (observed) hazards, while the latter does not.

The class of duration models in which unobserved heterogeneity is specified as a subordinator has been first proposed by Kebir (1991), while similar ideas have been discussed by Singpurwalla (1995), Aalen and Hjort (2002), Gjessing, Aalen, and Hjort (2003), Singpurwalla (2006), and references therein. The papers in this literature provide detailed *descriptive implications* of duration models driven by Lévy subordinators, but they do *not* study the identification of the model parameters.

Furthermore, the estimators proposed in this literature are fully parametric, with the distribution of the unobserved heterogeneity parametrized up to a finite dimensional parameter. On the other hand, this paper considers both the *nonparametric* identification and estimation of a class of duration models where the unobserved heterogeneity evolves over time as a Lévy subordinator.

In deriving sufficient conditions for the nonparametric identification of the model, the paper also contributes to the recent econometrics literature on the solvability of integral equations. Recent work in econometrics focuses on linear integral operators with either known or unknown kernel functions, e.g., Hu, Schennach, and Shiu (2017) and Hu and Shiu (2017), respectively. When the integral operator is linear, unique solvability of the linear integral equation reduces to the injectivity of the linear operator. For example, Hu and Shiu (2017) provide sufficient conditions for the injectivity of a linear Fredholm operator with unknown kernel, while Hu, Schennach, and Shiu (2017) analyze the injectivity of a linear Volterra operator with known kernel that is compactly supported. In contrast, this paper derives sufficient conditions for the solvability⁶ of a *nonlinear* Volterra integral equation, where the kernel function is *unknown* and non-compactly supported.⁷ Results from the linear-operator literature cannot be used in this paper due to the nonlinearity and nonparametric nature of the kernel function.

3 Identification

This section derives sufficient conditions for the nonparametric identification of (ϕ, f, F_Z) . A key element in the identification strategy is the survival function associated to (M) . I first derive the closed-form expression for the survival function. Then I show the nonparametric identification of the three functions ϕ , f , and F_Z .

⁶The concept of injectivity is not used for nonlinear operators.

⁷To the best of my knowledge, Hu and Shiu (2017) is the only other paper in the econometrics literature that considers an integral equation with unknown kernel. In their case, the equation is a linear Fredholm integral equation and the kernel function is a conditional density function.

3.1 The Survival Function

Let $P(T \leq t | \mathcal{F}_t)$ denote the conditional probability function of T whose hazard function is given by (M).

Assumption S1. (i) For all $t \geq 0$, $P(T \leq t | \mathcal{F}_t) = P(T \leq t | \mathcal{F}_\infty)$ almost surely; (ii) $P(T \leq t | \mathcal{F}_t)$ is predictable with respect to $\{\mathcal{F}_t\}_{t \geq 0}$.

Assumption S1(i) rules out the possibility that the unobserved heterogeneity at time t predicts the future survival probability. In particular, the event $\{T \leq t\}$ and the σ -algebra, $\mathcal{F}_\infty$, are independent given $\mathcal{F}_t$. Letting $\mathcal{F}_{t-} \equiv \sigma(Z(s), X, 0 \leq s < t)$, Assumption S1(ii) implies that $P(T \leq t | \mathcal{F}_t)$ is $\mathcal{F}_{t-}$ measurable for each t . Importantly, Assumption S1 implies that the survival function is absolutely continuous with respect to the Lebesgue measure, see Yashin and Arjas (1988), so that the survival function has the conventional exponential representation.⁸ That is, under S1, the survival function associated to the hazard function in (M) is given by:

$$S(t|x) \equiv E_Z \left[\exp \left(- \int_0^t h(s|x, \{Z_s\}_{s \geq 0}) ds \right) \right], \text{ for any } t \geq 0, x \in \mathcal{X},$$

where the expectation is taken with respect to F_Z .

Assumption S2. $f : [0, \infty) \rightarrow (0, \infty)$ is continuous on $[0, \infty)$.

Given Assumption Z, $\{Z(t)\}_{t \geq 0}$ is of bounded variation. Then Assumption S2 guarantees that the stochastic integral appearing in (M) is well defined for any $t \geq 0$, see, e.g., Rajput and Rosinski (1989). In particular, Assumption S2 guarantees that f is locally integrable on $[0, \infty)$, i.e. integrable on $[0, t]$ for any $t < \infty$.

Lemma 1 below expresses the survival function associated to (M) in terms of the Laplace exponent of $\{Z(t)\}_{t \geq 0}$. The Laplace exponent, $\Psi : [0, \infty) \rightarrow [0, \infty)$, is defined as:

$$E \left[e^{-\chi Z(t)} \right] = e^{-t\Psi(\chi)}, \text{ for all } t, \chi \geq 0. \quad (3)$$

⁸For a derivation of the survival function as an exponential function of the hazard function, see, e.g., Barlow and Proschan (1965), Kalbfleisch and Prentice (2002).

By the uniqueness theorem for Laplace transforms, Ψ determines F_Z uniquely; for a list of properties of the Laplace exponent see Property 1 in the Appendix.

Lemma 1 *Given assumptions Z, S1, and S2, for each $t \geq 0$ and $x \in \mathcal{X}$, the conditional survival function associated to (M) is given by*

$$S(t|x) = \exp \left[- \int_0^t \Psi(\phi(x) f(u)(t-u)) du \right], \quad (4)$$

where Ψ is the Laplace exponent of $\{Z(t)\}_{t \geq 0}$ defined in (3).

Proof. The proof of Lemma 1 is in Appendix B.1. ■

The expression for the survival function in Lemma 1 is similar to that of Theorem 1 in Gjessing, Aalen, and Hjort (2003), except that there the authors do not control for covariates, and they have a different specification of the stochastic process modeling unobserved heterogeneity.

3.2 Identification of the Model

Denoting by $S_0(t|x)$ the population conditional survival function appearing on the left hand side of (4), the identification of (ϕ, f, F_Z) is based on solving the integral equation below for the triple of functions (ϕ, f, Ψ) :

$$H(t, x) \equiv -\log S_0(t|x) = \int_0^t \Psi(\phi(x) f(u)(t-u)) du, \quad t \geq 0. \quad (5)$$

Equation (5) is a nonlinear Volterra integral equation of the first kind, where the kernel function, Ψ , is *unknown*.

For any $k \in \mathbb{N}$ and for all $\chi \geq 0$, define the following:

$$\begin{aligned} \Psi^{(k)}(\chi) &\equiv \frac{d^k}{d\chi^k} \Psi(\chi), \\ f^{(k)}(\chi) &\equiv \frac{d^k}{d\chi^k} f(\chi), \\ \frac{\partial^k}{\partial t^k} H(0, x) &\equiv \frac{\partial^k}{\partial t^k} H(t, x) \Big|_{t=0}, \text{ for any } x \in \mathcal{X}. \end{aligned}$$

To show the identification of (ϕ, f, Ψ) entering (5) consider the following assumptions.

Assumption ID0. (i) For each $x \in \mathcal{X}$, $H(\cdot, x) \in C^\infty((0, \infty))$; (ii) For all $\kappa \in \mathbb{N}$, there exist integrable functions $g_\kappa : \mathcal{X} \rightarrow \mathbb{R}_+$ such that $\left| \frac{\partial^\kappa}{\partial t^\kappa} \Psi(\phi(x) f(u)(t-u)) \right| \leq g_\kappa(x)$ for all $0 \leq u \leq t < \infty$ and almost every $x \in \mathcal{X}$, (iii) $f \in C^\infty((0, \infty))$.

As the proof of identification is based on differentiating (5) repeatedly with respect to t , Assumption ID0 is a regularity assumption that allows the application of the Dominated Convergence Theorem to each resulting derivative. Assumption ID0(ii) is a high-level assumption; strong but sufficient assumptions for ID0(ii) are, e.g., ϕ is integrable on $\mathcal{X}$, f is bounded on $(0, \infty)$, and $|\Psi^{(\kappa)}(\chi)| < \infty$ for all $\chi \geq 0$ and $\kappa \in \mathbb{N}$. Assumption ID0(iii) strengthens Assumption S2 by requiring f to be infinitely many times differentiable on $(0, \infty)$.

Assumption ID1. For all $\kappa \in \mathbb{N}$, (i) $\frac{\partial^\kappa}{\partial t^\kappa} H(0, x)$ exists for each $x \in \mathcal{X}$; (ii) $\Psi^{(\kappa)}(0)$ exists and $\Psi^{(1)}(0) \neq 0$; (iii) $f^{(\kappa)}(0)$ exists.

Assumption ID1 allows for the resulting derivatives of (5) with respect to t to be evaluated at $t = 0$. Assumption ID1(ii) requires the existence of $\{\Psi^{(\kappa)}(0)\}_{\kappa \geq 1}$, which implies that all moments of $Z(1)$ exist.⁹ This assumption excludes subordinators with fat tails such as the stable process. Additionally, ID1(ii) requires that the mean of $Z(1)$ be different from zero. Since $Z(1)$ is a non-negative random variable, this assumption implies that $Z(1)$ is not a degenerate random variable at zero.

Assumption ID2. $\phi : \mathcal{X} \rightarrow (0, \infty)$ is such that (i) $\phi(x_0) = 1$ for some arbitrary $x_0 \in \mathcal{X}$; (ii) $\phi(x_0) \neq \phi(x_1)$ for some $x_0 \neq x_1 \in \mathcal{X}$.

Assumption ID2(i) imposes a normalization assumption at x_0 , while ID2(ii) requires ϕ to be non-constant, with the variation required being minimal, see also, e.g., Elbers and Ridder (1982), Kortram et al. (1995), and Abbring (2012).

⁹The sequence $\{\Psi^{(\kappa)}(0)\}_{\kappa \geq 1}$ represents the sequence of cumulants of $Z(1)$, and there is a one-to-one relation between cumulants and moments, e.g., Lukács (1955). For example, $E(Z(1)) = \Psi^{(1)}(0)$, i.e. the mean of $Z(1)$ equals the first cumulant of $Z(1)$. More generally, the k^{th} moment of $Z(1)$ can be obtained from the cumulants by the following formula involving the incomplete Bell polynomial, $B_{k,j}(v_1, v_2, \dots, v_{k-j-1})$:

$$E\left((Z(1))^k\right) = \sum_{m=0}^k B_{k,k-m}\left(\Psi^{(1)}(0), \Psi^{(2)}(0), \dots, \Psi^{(m+1)}(0)\right).$$

Assumption ID3. $f : [0, \infty) \rightarrow (0, \infty)$ is such that $f(0) = 1$.

Assumption ID3 imposes a normalization assumption on f . Without this assumption, the identification results hold up to $f(0)$.

For any integer $k \geq 4$, let:

$$D_k(x) \equiv \frac{\partial^k}{\partial t^k} H(0, x) - \sum_{j=2}^{k-2} \phi^j(x) \Psi^{(j)}(0) \left(\frac{d^{k-1-j}}{dt^{k-1-j}} f^j(t) \Big|_{t=0} \right), \quad x \in \mathcal{X}. \quad (6)$$

Theorem 1 *Let assumptions Z, S1, and S2 hold such that the survival function of T given $X = x$ is given by (4). Additionally, let assumptions ID0 to ID3 hold. Then for some $x_0 \neq x_1 \in \mathcal{X}$, the function $\phi : \mathcal{X} \rightarrow (0, \infty)$ and the sequences $\{f^{(\kappa)}(0)\}_{\kappa \geq 1}$ and $\{\Psi^{(\kappa)}(0)\}_{\kappa \geq 1}$ are identified and given by:*

$$\phi(x) = \frac{\frac{\partial^2}{\partial t^2} H(0, x)}{\frac{\partial^2}{\partial t^2} H(0, x_0)}, \quad x \in \mathcal{X}, \quad (7)$$

$$\Psi^{(1)}(0) = \frac{\partial^2}{\partial t^2} H(0, x_0), \quad (8)$$

$$f^{(1)}(0) = \frac{\phi^2(x_1) \frac{\partial^3}{\partial t^3} H(0, x_0) - \frac{\partial^3}{\partial t^3} H(0, x_1)}{\phi(x_1) [\phi(x_1) - 1] \Psi^{(1)}(0)}, \quad (9)$$

$$f^{(\kappa)}(0) = \frac{\phi^{\kappa+1}(x_1) D_{\kappa+2}(x_0) - D_{\kappa+2}(x_1)}{\phi(x_1) [\phi^\kappa(x_1) - 1] \Psi^{(1)}(0)}, \quad \kappa \geq 2, \quad (10)$$

$$\Psi^{(2)}(0) = \frac{\frac{\partial^3}{\partial t^3} H(0, x_1) - \phi(x_1) \frac{\partial^3}{\partial t^3} H(0, x_0)}{\phi(x_1) [\phi(x_1) - 1]}, \quad (11)$$

$$\Psi^{(\kappa)}(0) = \frac{D_{\kappa+1}(x_1) - \phi(x_1) D_{\kappa+1}(x_0)}{\phi(x_1) [\phi^{\kappa-1}(x_1) - 1]}, \quad \kappa \geq 3. \quad (12)$$

Proof. The proof of Theorem 1 is in Appendix B.2. ■

Theorem 1 identifies $\{\Psi^{(\kappa)}(0)\}_{\kappa \geq 1}$, which is the sequence of cumulants of $Z(1)$. To identify Ψ itself the following properties for functions on $\mathbb{R}$ are used: (1) Ψ is a Laplace exponent, so it is a Bernstein function; (2) the first derivative of a Bernstein function is completely monotone; (3) completely monotone functions are real analytic; and (4) real analytic functions are uniquely determined by their values on a non-empty open set, see, e.g., Abbring and van den Berg (2003). To be able to use this sequence of arguments on functions defined on $[0, \infty)$ (rather than $\mathbb{R}$), the assumption below is

imposed.

Assumption ID4. (i) $\Psi^{(1)}(\chi)$ is right-continuous at $\chi = 0$; (ii) The moment generating function of $Z(1)$ exists.

Since $\Psi(\chi)$ is a Laplace exponent, it is a Bernstein function on $[0, \infty)$. Hence, its first derivative $\Psi^{(1)}(\chi)$ is completely monotone on $(0, \infty)$, e.g., Widder (2011). Requiring ID4(i) guarantees that $\Psi^{(1)}(\chi)$ is completely monotone on $[0, \infty)$ instead. Assumptions ID4 and ID1 guarantee that $\Psi^{(1)}(\chi)$ is real analytic at $\chi = 0$,¹⁰ see Theorem 15 in Koumandos (2014).¹¹

Remark 1 *Within the context of identification of duration models, Assumption ID4(ii) may seem strong as it requires existence of all moments rather than just the mean (Elbers and Ridder (1982), Kortram et al. (1995)), or a tail condition on the density of the unobserved heterogeneity (Heckman and Singer (1984)), or regular variation of the distribution of unobserved heterogeneity (Abbring and Ridder (2015), Abbring (2012)).¹² However, this should not be surprising given that the model of this paper is much more flexible in regards to how the unobserved heterogeneity is allowed to change over time. As opposed to standard duration models, where the unobserved heterogeneity stays constant and equal to its realization at $t = 0$ for the entire duration of the spell, in this model, individual unobserved heterogeneity is allowed to change during the spell of interest.*

Nonetheless, the assumptions above can be somewhat weakened at the expense of identifying the distribution of the unobserved heterogeneity. For example, assuming existence of $\kappa = 1$ moments and $f(0) = 1$, identifies the function ϕ and the mean of $Z(1)$. Further assuming existence of $\kappa = 1, 2, \dots, M$ moments, minimal variation in X , and $f \in C^{\kappa-1}([0, \infty))$, identifies all M moments of $Z(1)$. Finally, assuming existence of all moments, i.e. $\kappa = 1, 2, \dots$, minimal variation in X , and $f \in C^\infty([0, \infty))$, identifies all moments of $Z(1)$.

¹⁰Related assumptions on real analyticity of the moment generating function are made in the literature on random coefficients, see e.g. Fox et al. (2012).

¹¹Note that the assumptions in Koumandos (2014) are expressed in terms of moments of $Z(1)$, e.g. letting m_k be the k^{th} moment of $Z(1)$, the assumption is stated as $\limsup_{k \rightarrow \infty} \left[\frac{1}{k} (m_k)^{1/k} \right] < \infty$. By Corollary 2 of Lin (2017), this assumption is equivalent to requiring that the moment generating function of $Z(1)$ exists.

¹²The identification strategy in these papers takes advantage of the special functional form of the survival function as the Laplace transform of the unobserved heterogeneity, which allows one to use different properties related to real analyticity of the Laplace transform.

Corollary 1 *Under the assumptions of Theorem 1 and Assumption ID4, the Laplace exponent Ψ is identified and given by:*

$$\Psi(\chi) = \sum_{j=1}^{\infty} \frac{\chi^j}{j!} \Psi^{(j)}(0), \quad \chi \geq 0. \quad (13)$$

Proof. Since $\Psi : [0, \infty) \rightarrow [0, \infty)$ is a Bernstein function, and given ID4(i), $\Psi^{(1)}(\chi)$ is a completely monotone function on $[0, \infty)$. Assumptions ID1(ii) and ID4(ii) verify the conditions in Theorem 15 of Koumandos (2014), so that $\Psi^{(1)}(\chi)$ is real analytic at $\chi = 0$. Hence:

$$\Psi^{(1)}(\chi) = \sum_{j=0}^{\infty} \frac{\chi^j}{j!} \Psi^{(j+1)}(0),$$

where $\{\Psi^{(j)}(0)\}_{j \geq 1}$ are given in the statement of Theorem 1. Therefore, Ψ is identified and given by:

$$\Psi(\chi) = \int \Psi^{(1)}(v) dv = \sum_{j=0}^{\infty} \frac{\chi^{j+1}}{(j+1)!} \Psi^{(j+1)}(0) = \sum_{j=1}^{\infty} \frac{\chi^j}{j!} \Psi^{(j)}(0),$$

where $\Psi(0) = 0$ was used. ■

Since Theorem 1 identifies the sequence of derivatives $\{f^{(\kappa)}(0)\}_{\kappa \geq 1}$, imposing the (sufficient) assumption below obtains identification of f .

Assumption ID5. $f(t)$ is real analytic at $t = 0$.

The class of real analytic functions includes, e.g., polynomials, exponential functions, power functions, logit-type functions, and trigonometric functions.

Corollary 2 *Let the assumptions for Theorem 1 and Corollary 1 hold, and let Assumption ID5 hold. Then f is identified and given by*

$$f(t) = 1 + \sum_{j=1}^{\infty} \frac{t^j}{j!} f^{(j)}(0), \quad t \geq 0, \quad (14)$$

where $\{f^{(j)}(0)\}_{j \geq 1}$ are defined in (9) and (10).

Proof. Theorem 1 identifies the derivatives of all orders of f evaluated at zero. Then Assumption ID5 obtains (14). ■

Remark 2 *Assumption ID5 can be relaxed at the expense of further restrictions on (ϕ, f, Ψ) , such as boundedness of $\left(\phi, f, \frac{\partial^2}{\partial t^2} H(t, x)\right)$ and a uniform Lipschitz condition on $\Psi^{(1)}(.)$. To see this, notice that once (ϕ, Ψ) have been identified, differentiating (5) twice with respect to t , obtains the following equation:*

$$f(t) = \frac{1}{\phi(x) \Psi^{(1)}(0)} \left[\frac{\partial^2}{\partial t^2} H(t, x) - \phi^2(x) \int_0^t f^2(u) \Psi^{(1)}(\phi(x) f(u)(t-u)) du \right]. \quad (15)$$

This equation is a Volterra integral equation of the second kind for f where the kernel function $\Psi^{(1)}$ is known by Corollary 1. Existence and uniqueness of f can be shown via a Fixed Point argument. For example, since T takes values in an unbounded interval, it is possible to show that the operator associated with (15) is a contraction in a weighted Banach space by verifying the conditions of Theorem 2.4 or Theorem 3.3 in Rejto and Taboada (1992). On the other hand, assuming that T takes values in a compact set, it is possible to show identification of f by verifying the conditions of Theorem 2.1 in Hu, Schennach, and Shiu (2017).

4 Estimation

Let $W = (X, T)' \in \mathcal{W} = \mathcal{X} \times \mathcal{T}$, $\mathcal{X} \subset \mathbb{R}^q$ and $\mathcal{T} \subset [0, \infty)$, be a random vector whose distribution satisfies the hazard model described in the sections above. Denote by $\theta_0 = \Psi_0(\phi_0, f_0) = (\phi_0, f_0, \Psi_0)$ the true value of the infinite-dimensional parameters of interest.

Although the identification argument for θ_0 is constructive, the estimator proposed is not the analog estimator due to the following two issues. First, the analog estimator would involve estimation of the k^{th} derivative of the log conditional survival function evaluated at a boundary point, i.e. estimation of $\frac{\partial^k}{\partial t^k} H(0, x)$. This would lead to issues related to boundary effects (such as inconsistency). Second, estimation of the Laplace exponent would involve estimation of a growing number of moments, $\{\Psi^{(\kappa)}(0)\}_{\kappa \rightarrow \infty}$, which would require some sort of trimming in order for the estimation procedure to be feasible. Addressing these two issues would be of independent interest and is beyond the scope of this paper.

Instead, I propose a two-step estimation procedure for θ_0 , starting from equation (5). That is,

letting $\Theta \equiv \Phi \times \mathcal{F} \times \ominus$ denote the parameter space (to be defined later), define:

$$\theta_0 = \arg \max_{\theta \in \Theta} \left[-E \left(-\log S_0(T|X) - \int_0^T \Psi_0(\phi_0(X) f_0(u) (T-u)) du \right)^2 \right],$$

where $S_0(T|X)$ is the true conditional survival function.

The estimation method proceeds as follows. Let $\hat{S}_n(t|x)$ denote the first-step estimator of the conditional survival function $S_0(t|x)$. This step does not require a specific procedure, however, certain high-level assumptions (below) should be satisfied. Given $\hat{S}_n(t|x)$, I then estimate θ_0 via sieve-minimum distance (MD) under a monotonicity restriction on an otherwise nonparametric estimator of the Laplace exponent, Ψ_0 . This restriction is theoretically driven since Ψ_0 is the Laplace exponent of a subordinator and so it is monotone on its domain of definition. Note that this shape restriction is not used for identification, however, imposing it in finite samples improves considerably the performance of the estimator, see also Horowitz and Lee (2017).

The estimator of θ_0 is then defined as:

$$\hat{\theta}_n \equiv \arg \max_{\theta \in \Theta_n} -\frac{1}{n} \sum_{i=1}^n \left[-\log \hat{S}_n(T_i|X_i) - \int_0^{T_i} \Psi(\phi(X_i) f(u) (T_i-u)) du \right]^2, \quad (16)$$

where $\Theta_n \equiv \Phi_n \times \mathcal{F}_n \times \ominus_n$ is a finite-dimensional linear space such that $\Theta_n \subset \Theta_{n+1} \subset \Theta$ with the set Θ_∞ complete in Θ , and where each sieve space, Φ_n and $\mathcal{F}_n$, is a linear span of finitely many polynomial basis functions, while $\ominus_n$ is a linear span of finitely many Bernstein polynomial basis functions. That is,

$$\Phi_n = \left\{ \phi : \phi(x) = g \left(\sum_{j=0}^{J_\phi(n)} c_{\phi,j} B_{\phi,j}(x) \right), \phi(0) = 1 \right\}, \quad (17)$$

$$\mathcal{F}_n = \left\{ f : f(t) = \tilde{g} \left(\sum_{j=0}^{J_f(n)} c_{f,j} B_{f,j}(t) \right), f(0) = 1 \right\}, \quad (18)$$

$$\ominus_n = \left\{ \Psi : \Psi(\chi) = \sum_{j=1}^{J_\Psi(n)} c_{\Psi,j} B_{\Psi,j}(\chi), A_{J_\Psi(n)} C_{J_\Psi(n)} \geq 0 \right\}, \quad (19)$$

where $g, \tilde{g}$ are known positive transformation functions, e.g., $g(\chi) = \exp(\chi)$ (exponential) or $g(\chi) =$

χ^2 (quadratic); $B_{\phi,j}$ and $B_{f,j}$ are polynomial basis functions while $B_{\Psi,j}$ are Bernstein polynomial basis functions; $c_{l,j}$, $l = \phi, f, \Psi$, $j \in \mathbb{N}$ are coefficients, and $C_{J_{\Psi}(n)} = (c_{\Psi,1}, \dots, c_{\Psi,J_{\Psi}(n)})'$ is a vector of $J_{\Psi}(n)$ coefficients;¹³ $A_{J_{\Psi}(n)}$ is the $(J_{\Psi}(n) - 1) \times J_{\Psi}(n)$ full row rank matrix defined below; and each $J_{\phi}(n)$, $J_f(n)$, $J_{\Psi}(n) \rightarrow \infty$ as $n \rightarrow \infty$.

The space Θ_n is a constrained Bernstein polynomial sieve space,¹⁴ where the Bernstein polynomial basis functions are defined as:

$$B_{\Psi,j}(\chi) = \frac{J_{\Psi}(n)!}{j!(J_{\Psi}(n) - j)!} \chi^j (1 - \chi)^{J_{\Psi}(n) - j}, \quad j = 1, \dots, J_{\Psi}(n),$$

and the elements of $A_{J_{\Psi}(n)}$ are chosen such that the monotonicity restriction on Ψ holds, i.e.

$$A_{J_{\Psi}(n)} = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ & \ddots & & & \\ \dots & 0 & -1 & 2 & -1 \end{bmatrix}.$$

For details on different type of shape restrictions, see, e.g., Wang and Ghosh (2012).¹⁵

For the implementation of the estimation method, $J(n) \equiv (J_{\phi}(n), J_f(n), J_{\Psi}(n))$ has to be determined. I suggest considering different values for these three unknowns and then choosing the combination that minimizes a V -fold cross-validation criterion function as in Wang and Ghosh (2012), i.e.

$$CV(J(n)) = \frac{1}{V} \sum_{v=1}^V \sum_{i \in I_{-v}} - \left[-\log \hat{S}_n(T_i | X_i) - \int_0^{T_i} \tilde{\Psi}_v(\tilde{\phi}_v(X_i) \tilde{f}_v(u) (T_i - u)) du \right]^2,$$

where $(\tilde{\Psi}_v, \tilde{\phi}_v, \tilde{f}_v)$ is computed by solving (16) with the v^{th} data set consisting of $\lfloor n \frac{V-1}{V} \rfloor$ observations and where I_{-v} denotes the excluded group containing the remaining observations. The value of $J(n)$

¹³Notice that the restriction $\Psi(0) = 0$ is already imposed by setting $c_{\Psi,0} = 0$.

¹⁴For advantages of using Bernstein polynomials when estimating shape restricted functions, see, e.g., Lorentz (1986), Zhou et al. (2017).

¹⁵Notice that since $\Psi(0) = 0$, the matrix $A_{J_{\Psi}(n)}$ has only $J_{\Psi}(n)$ rather than $J_{\Psi}(n) + 1$ columns, as it is usually specified for the coefficient matrix associated with Bernstein polynomials.

is selected as

$$\hat{J}(n) = \arg \min_{J(n) \in [2, J_{\max}]} CV(J(n)),$$

where $2 < J_{\max} < \lfloor n^{\frac{V-1}{V}} - 1 \rfloor$.

4.1 Asymptotic Theory

Let $H_0(W) \equiv -\log S_0(T|X)$, and let $I(\theta, W) : \Theta \times \mathcal{W} \rightarrow \mathbb{R}_+$ and $m_{H,W}(\theta) : \Theta \times \mathcal{H} \times \mathcal{W} \rightarrow \mathbb{R}$ be defined as, respectively:

$$\begin{aligned} I(\theta, W) &\equiv \int_0^T \Psi(\phi(X) f(u)(T-u)) du, \\ m_{\theta,H}(W) &\equiv -[H(W) - I(\theta, W)]^2. \end{aligned}$$

The population criterion function is then defined as:

$$M(\theta, H) \equiv E[m_{\theta,H}(W)],$$

and

$$\theta_0 = \arg \max_{\Theta} M(\theta, H_0).$$

Letting $\hat{H}_n(W)$ be the first step estimator of $H_0(W)$, the sample criterion function is defined as

$$\begin{aligned} M_n(\theta, \hat{H}_n) &\equiv \frac{1}{n} \sum_{i=1}^n m_{\theta, \hat{H}_n}(W_i) \\ &= \frac{1}{n} \sum_{i=1}^n -[\hat{H}_n(W_i) - I(\theta, W_i)]^2, \end{aligned}$$

while the estimator $\hat{\theta}_n$ is defined as

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta_n} M_n(\theta, \hat{H}_n).$$

The norm under which both consistency and the rate of convergence of $\hat{\theta}_n = (\hat{\phi}_n, \hat{f}_n, \hat{\Psi}_n)$ are

derived is defined as:

$$d\left(\widehat{\theta}_n, \theta_0\right) = \left\|\widehat{\phi}_n - \phi_0\right\|_{\infty} + \left\|\widehat{f}_n - f_0\right\|_{\infty} + \left\|\widehat{\Psi}_n - \Psi_0\right\|_{\Theta},$$

where

$$\left\|\widehat{\Psi}_n - \Psi_0\right\|_{\Theta} = \sup_{\phi, f \in \Phi \times \mathcal{F}} \left\|\widehat{\Psi}_n(\phi, f, \cdot) - \Psi_0(\phi, f, \cdot)\right\|_{\infty}.$$

The functions H_0 , ϕ , f , and Ψ are assumed to belong, respectively, to the following spaces:

$$\mathcal{H} = \mathcal{C}^{\infty}(\mathcal{W}) \cap L_2(\mathcal{W}), \quad (20)$$

$$\Phi = \left\{\phi \in \left(C_{b_{\phi}}^{p_{\phi}}(\mathcal{X}), \|\cdot\|_{C^{p_{\phi}}}\right) : \phi(0) = 1, p_{\phi} > \frac{1}{2}\right\}, \quad (21)$$

$$\mathcal{F} = \left\{f \in \left(C_{b_f}^{p_f}(\mathcal{T}), \|\cdot\|_{C^{p_f}}\right) : f(0) = 1, p_f > \frac{1}{2}\right\}, \quad (22)$$

$$\Theta = \left\{\Psi \in \left(C_{b_{\Psi}}^{p_{\Psi}}([0, 1]), \|\cdot\|_{C^{p_{\Psi}}}\right) : p_{\Psi} > \max\{p_{\phi}, p_f\}\right\}, \quad (23)$$

where p_{Ψ} can be arbitrarily large.¹⁶ In the analysis that follows, I assume that b_{ϕ} , b_f , and b_{Ψ} are known. This assumption can be weakened at the expense of redefining the sieve spaces, see, e.g., Shen and Wong (1994). For sieve approximation, let $\phi_n \in \Phi_n$, $f_n \in \mathcal{F}_n$, and $\Psi_n \in \Theta_n$ be defined by (17), (18), and (19), respectively. Further, let $\Pi_n \theta_0$ be the projection of θ_0 onto the sieve space Θ_n , i.e.

$$\Pi_n \theta_0 = \arg \max_{\theta \in \Theta_n} M\left(\theta, \widehat{H}_n\right) = \arg \max_{\theta \in \Theta_n} -E\left(\widehat{H}_n(W) - I(\theta, W)\right)^2.$$

Assumption C1

- (i) The data $(X_i, T_i)_{i=1}^n$ are i.i.d. with joint distribution function satisfying (4);
- (ii) $\mathcal{T}$ and $\mathcal{X}$ are compact with $\mathcal{T} = [0, 1] = \mathcal{X}$;
- (iii) The density of X is bounded and bounded away from zero on $\mathcal{X}$, and the conditional density of T given $X = x$ is bounded away from zero for every $x \in \mathcal{X}$.

¹⁶By the properties of Laplace exponents, $\Psi \in C^{\infty}(\mathbb{R}_+)$. However, since the rate depends on $\min(p_{\phi}, p_f, p_{\Psi})$, for an arbitrarily large value of p_{Ψ} such that $p_{\Psi} > \max\{p_{\phi}, p_f\}$, it is only $\min(p_{\phi}, p_f)$ that affects the rate of convergence.

Assumption C1(i) is usually made in the duration literature, while Assumption C1(ii) is usually made in nonparametric analysis, and, for most applications, it is not too restrictive.¹⁷ Assumption C1(iii) guarantees that $\log S(W)$ is defined on its domain.

Assumption C2

- (i) $\left\| \widehat{H}_n - H_0 \right\|_{\infty} = o_p(1)$;
- (ii) $\left\| \widehat{H}_n - H_0 \right\|_{L_2} = O_p(\varepsilon_n)$, where $\{\varepsilon_n\}_{n \geq 1}$ is a non-stochastic sequence such that $r_n \varepsilon_n \leq 1$ for each n .

Assumption C2 imposes restrictions on the first step estimator: C2(i) requires uniform consistency of $\widehat{H}_n$, while C2(ii) requires an L_2 -rate of convergence of r_n .

Assumption C3

- (i) $\phi \in \Phi$, $f \in \mathcal{F}$, $\Psi \in \Theta$ for ϕ, f, Ψ in a neighborhood of ϕ_0, f_0, Ψ_0 in the norm $d(.,.)$;
- (ii) Θ_n and Θ are convex and closed in $d(.,.)$, and $\Theta_n \subset \Theta_{n+1} \subset \Theta$;
- (iii) For every $\theta \in \Theta$ there exists a $\Pi_n \theta_0 \in \Theta_n$ with $d(\Pi_n \theta_0, \theta_0) = o(1)$ as $J_l(n) \rightarrow \infty$, $\frac{J_l(n)}{n} \rightarrow 0$ as $n \rightarrow \infty$ for $l = \phi, f, \Psi$.

Assumption C3(i) imposes smoothness restrictions on the unknown functions. The choice of Hölder spaces is common in the literature on series estimation, see, e.g., Chen (2007). For technical reasons, the smoothness of ϕ and f is restricted to be smaller than that of the Laplace exponent. Additionally, Assumption C3(i) implies that the drift of the subordinator is finite, which excludes linear processes, i.e. processes whose sample paths are linear functions of time. Finally, Assumption C3(i) guarantees that the parameter space Θ is compact in $d(.,.)$, see, e.g., Theorem 1.34 in Adams and Fournier (2003). Compactness of the parameter space under the norm $d(.,.)$ is important for the proof of consistency, see Gallant and Nychka (1987). This assumption could be relaxed by introducing a

¹⁷The assumption can be relaxed to allow $\mathcal{T}$ and $\mathcal{X}$ to be unbounded at the expense of defining the nonparametric components on Banach spaces with weighted norms, with the asymptotic theory derived in terms of the weighted norms. I do not pursue this approach here.

penalty function as in Chen and Pouzo (2012), but this approach is not pursued here. Assumption C3(ii) requires the sieve space Θ_n to grow dense in Θ ; this assumption is satisfied by polynomial sieves, see, e.g., Newey (1997) and Chen (2007). Finally, Assumption C3(iii) restricts the complexity of the sieve space; this assumption is satisfied by polynomial sieves for which the approximation rate is of order $O(J_l(n)^{-p_l})$, $l = \phi, f, \Psi$.

Theorem 2 (Consistency) *Let $\hat{\theta}_n$ be the second-stage sieve estimation defined by (16). Let assumptions C1 to C3 and the assumptions of Theorem 1 hold. Then $d(\hat{\theta}_n, \theta_0) \rightarrow^p 0$.*

Proof. The consistency of $\hat{\theta}_n$ is shown in Appendix C.1. ■

The local behavior of $M(\theta, H)$ at the point (θ_0, H_0) is central to the analysis of the rate of convergence of the sieve estimator $\hat{\theta}_n$. This local behavior is described by the concept of pathwise derivative. Given the special form of the criterion function $M(\theta, H)$, it is sufficient to impose pathwise differentiability on $I(\theta, W)$ only. For any $\theta \in \Theta$ and almost all $W \in \mathcal{W}$, $I(\theta, W)$ is pathwise differentiable at θ in the direction $[\theta^* - \theta]$ if $\{\theta + \tau(\theta^* - \theta) : \tau = 0\} \subset \Theta$ and $\lim_{\tau \rightarrow 0} \frac{1}{\tau} [I(\theta + \tau(\theta^* - \theta), W) - I(\theta, W)]$ exists. The pathwise derivative at θ in the direction $[\theta^* - \theta]$ is given by:

$$\frac{\partial}{\partial \theta} I(\theta, W)(\theta^* - \theta) \equiv \left. \frac{\partial}{\partial \tau} I(\theta + \tau(\theta^* - \theta), W) \right|_{\tau=0}.$$

Assumption C4

The pathwise derivative of $I(\theta_0, \cdot)$ exists in all directions $[\theta - \theta_0]$.

Theorem 3 (Rate of convergence) *Let $\hat{\theta}_n$ be the second-stage sieve estimator defined by (16). If the assumptions of Theorem 2 and Assumption C4 hold, then*

$$d(\hat{\theta}_n, \theta_0) = O_p \left(\max \left\{ d(\hat{\theta}_n, \Pi_n \theta_0), d(\theta_0, \Pi_n \theta_0) \right\} \right),$$

where $r_n \leq \max \left\{ d(\hat{\theta}_n, \Pi_n \theta_0), d(\theta_0, \Pi_n \theta_0) \right\}$ and where

$$\begin{aligned} d(\hat{\theta}_n, \Pi_n \theta_0) &= O \left(\sqrt{\frac{\max \{J_\phi(n), J_f(n), J_\Psi(n)\}}{n}} \right), \\ d(\theta_0, \Pi_n \theta_0) &= O \left(\max \{J_\phi(n)^{-p_\phi}, J_f(n)^{-p_f}, J_\Psi(n)^{-p_\Psi}\} \right), \end{aligned}$$

with $p_\Psi > \max \{p_\phi, p_f\}$.

Proof. The rate of convergence of $\hat{\theta}_n$ is established in Appendix C.2. ■

Remark 3 *When all functions in the second-step are parametrized up to finite dimensional parameters, consistency and asymptotic normality of the estimators of the finite-dimensional parameters follows from standard arguments such as those in Newey and McFadden (1994). When all functions in the second step but the Laplace exponent are parametrized up to finite dimensional parameters, the second step in the M-estimation involves both infinite- and finite-dimensional parameters, where the finite-dimensional parameters enter as arguments of the infinite-dimensional parameter. Establishing rates of convergence for the finite-dimensional parameters in this case would require extending the analysis in either Hahn, Liao, and Ridder (2018), to allow for a kernel estimator in the first step and a bundled parameter in the second-step, or Ding and Nan (2011) to the case of two-step estimators. These extensions are beyond the scope of this paper.*

5 Monte Carlo Simulations

This section shows the finite sample properties of the estimators $(\hat{\phi}_n, \hat{f}_n, \hat{\Psi}_n)$ via Monte Carlo simulations. I investigate the finite sample performance of the estimators for different sample sizes and different specifications of (ϕ_0, f_0, Ψ_0) .

The descriptions in this section are necessarily concise since there are no estimators in the literature with which to compare the finite-sample performance of the proposed estimator. That is, in the (bio)statistics literature that has considered similar duration models driven by Lévy subordinators, Ψ_0 is always parametrically specified and there are no functions that depend on the covariates such as ϕ_0 , see, e.g., Aalen et al. (2008).

Data Generating Process. For the generation of the observed data, let S denote the survival probability, and let $S \sim U(0, 1)$ and $X \sim \text{Bernoulli}(0.5)$.¹⁸ For a given simulation, using the

¹⁸A binary covariate provides the minimal variation sufficient for identification, see assumption ID2(ii).

generated pairs $(X_i, S_i)_{i=1}^n$, I calculate the duration T_i as the solution to the equation:

$$S_i - \exp \left[- \int_0^{T_i} \Psi_0 (\phi_0 (X_i) f_0 (u) (T_i - u)) du \right] = 0,$$

where (ϕ_0, f_0, Ψ_0) take different functional forms, see below.

The process $\{Z(t)\}_{t \geq 0}$ is a Poisson process, so that the Laplace exponent Ψ_0 is given by:

$$\Psi_0(\chi) = 10(1 - e^{-\chi}), \quad \chi \geq 0. \quad (24)$$

The functional form specifications of (ϕ_0, f_0) are given in the table below, where the transformation functions $(g, \tilde{g})$ are defined in (17), (18), respectively:

Design	$\phi_0(x)$	$f_0(t)$	g	$\tilde{g}$
1	$1 + x^2 - \frac{1}{10}x^3$	$1 - \frac{1}{4}t^2 + 2t^3$	exponential	exponential
2S	$1 + x^2 - \frac{1}{10}x^3$	$\exp(\gamma_0 t), \gamma_0 = 1$	exponential	exponential
2N	$1 + x^2 - \frac{1}{10}x^3$	$\exp(t)$	exponential	identity
3S	$\exp(\beta_0 x), \beta_0 = 1$	$1 - \frac{1}{4}t^2 + 2t^3$	exponential	exponential
3N	$\exp(x)$	$1 - \frac{1}{4}t^2 + 2t^3$	identity	exponential
4S	$(1 + \beta_0 x)^2, \beta_0 = 1$	$(1 + \gamma_0 t)^2, \gamma_0 = 1$	quadratic	quadratic

(25)

Designs 1, 2N, 3N consider nonparametric estimators for (ϕ_0, f_0) , as described in Section 4. Designs 2S, 3S, and 4S consider semiparametric estimators for f_0, ϕ_0 , and (ϕ_0, f_0) , respectively where the functions ϕ_0 and f_0 are parametrized up to the finite dimensional parameters β_0 and γ_0 , respectively. Across all designs, the estimator for Ψ_0 is nonparametric and restricted to be monotonic as described in Section 4. There are 1000 simulations per design.

When the functions (ϕ_0, f_0) are not parametrized, as in designs 1, 2N, and 3N, I report the integrated mean square error (IMSE) of the estimators averaged over the simulations, $j = 1, \dots, 1000$. The reported IMSE for an estimator $\hat{\eta}$ is defined as:

$$IMSE(\hat{\eta}) = \frac{1}{1000} \sum_{j=1}^{1000} \left[\frac{1}{n} \sum_{i=1}^n (\eta_0(Y_i) - \hat{\eta}_j(Y_i))^2 \right],$$

where $(\eta_0, \hat{\eta})$ are the true function and its estimator, respectively, n is the sample size, and $\{Y_i\}_{i=1}^n$ is the generated data. When the functions (ϕ_0, f_0) are parametrized up to (β_0, γ_0) , as in designs 2S, 3S, and 4S, I report the bias, standard deviation, and the root mean square error (RMSE) of the estimators averaged over the simulations. The observed data is $(X_i, T_i)_{i=1}^n$ for each simulation, and for each semiparametric semiparametric simulation (2S, 3S, 4S), the sample size $n \in \{1000, 2000, 4000\}$, while for each nonparametric simulation (1, 2N, 3N), the sample size $n \in \{500, 2000\}$.

Estimators. Conditional on $X = x$, $\hat{S}_n(t|x)$ is the Kaplan-Meier estimator and $\hat{H}_n(t|x) = -\log \hat{S}_n(t|x)$. The second-step sieve estimator is as explained in Section 4, with the polynomial basis functions for (ϕ, f) being Chebyshev bases. The degrees of the three polynomial sieves are chosen via 7-fold cross-validation as in Wang and Ghosh (2012).

Results. Table 1 and Figure 2 show results for Design 1. The functions (ϕ_0, f_0) are estimated nonparametrically by Chebyshev polynomials with $J_\phi(n) = 1$ and $J_f(n) = 2$, respectively. Table 1 shows a decreasing IMSE for both estimators as the sample size increases. Figure 2 shows the estimator for Ψ_0 approximated by a restricted Bernstein polynomial with $J_\Psi(n) = 3$. The sample size is $n = 500$ for the top graph and $n = 2000$ for the bottom graph. This figure shows that the shape of the Laplace exponent is approximated very well, with wider interquantile range for the smaller sample size as expected.

Table 2 shows results for Design 2S and for Design 2N. For Design 2S, ϕ_0 is estimated nonparametrically by a Chebyshev polynomial of degree $J_\phi(n) = 1$ with g as in (25), while f_0 is parametrized up to $\gamma_0 = 1$. Both the average IMSE for $\hat{\phi}$ and the RMSE for $\hat{\gamma}$ are decreasing in n . For Design 2N, ϕ_0 is estimated nonparametrically by a Chebyshev polynomial of degree $J_\phi(n) = 1$ and with g as in (25), while f_0 is estimated nonparametrically by a Chebyshev polynomial of degree $J_f(n) = 3$, with the first coefficient normalized to unity (as per Assumption ID3(i)), and with $\tilde{g}$ as in (25). The IMSE of $\hat{\phi}$ and that of $\hat{f}$ decrease with the sample size. Figure 3 shows the estimator for the Laplace exponent of the Poisson process for Design 2N. The Laplace exponent is approximated by a Bernstein polynomial with $J_\Psi(n) = 3$. The figure shows results for $n = 2000$ only, since the graph for $n = 500$ is similar with the interquantile range slightly wider.

Table 3 shows results for Design 3S and for Design 3N. For Design 3S, ϕ_0 is parametrized up to the finite dimensional parameter $\beta_0 = 1$, while f_0 is estimated nonparametrically by a Chebyshev polynomial of degree $J_f(n) = 2$ with $\tilde{g}$ as in (25). Both the RMSE for $\hat{\beta}$ and the IMSE for $\hat{f}$ decrease with the sample size. Design 3N estimates ϕ_0 nonparametrically by a Chebyshev polynomial of degree 1, with the first coefficient normalized to unity (as per Assumption ID2(i)) and with g as in (25), while f_0 is estimated nonparametrically by a Chebyshev polynomial of degree $J_f(n) = 2$ with $\tilde{g}$ as in (25). The IMSE of $\hat{\phi}$ and that of $\hat{f}$ decrease with the sample size. Figure 4 shows the estimator for the Laplace exponent for Design 3N, approximated by a Bernstein polynomial with $J_\Psi(n) = 3$. The sample size for this graph is $n = 2000$.

Table 4 and its associated figure, Figure 5, show results for the case where ϕ_0 and f_0 are parametrized up to β_0 and γ_0 , respectively, but the Laplace exponent is estimated nonparametrically as described in Section 4. Table 4 considers two different specification for the distribution of the unobserved heterogeneity. The process $\{Z(t)\}_{t \geq 0}$ is a Gamma subordinator (top), with Laplace exponent given by:

$$\Psi_0(\chi) = 30 \log(1 + \chi), \quad \chi \geq 0, \quad (26)$$

and a Poisson process (bottom) with Laplace exponent given in (24). The table shows that the bias for the estimators of the finite dimensional parameters is negligible and decreasing with the sample size, while the RMSE is driven by the standard deviation. Although the paper does not provide rates of convergence for the estimators of the finite-dimensional parameters for this case, the estimators for (β_0, γ_0) appear to have $\sqrt{n}$ -rates of convergence. As mentioned in Remark 3, providing theoretical results for this case is left for future work. The properties of the estimators for the finite-dimensional parameters are robust to the specification of the distribution of $\{Z(t)\}_{t \geq 0}$. Figure 5 shows the estimated Laplace exponent, $\hat{\Psi}_n$, when the true subordinator is a Gamma process (top) and a Poisson process (bottom). For both the Gamma and the Poisson process, the degree of the Bernstein polynomial $J_\Psi(n) = 3$. The figure shows results for $n = 4000$ only; the results for $n \in \{1000, 2000\}$ are similar, with a slightly wider interquantile range that becomes narrower as n increases. As the graphs suggest, the shape of the Laplace exponent is recovered extremely well.

Table 1: Simulation Results for Design 1

	Estimator for ϕ_0	Estimator for f_0
	IMSE	IMSE
Poisson process		
$n = 500$	0.0918	0.0583
$n = 2000$	0.0097	0.0303

NOTE: The estimators for ψ_0 and f_0 are as described in Section 4 with $J_\phi(n) = 1$ and $J_f(n) = 2$, respectively.

Table 2: Simulation Results for Design 2

(a) Simulation Results for Design 2S

	Estimator for ϕ_0	Estimator for $\gamma_0 = 1$		
	IMSE	Bias	SD	RMSE
Poisson process				
$n = 1000$	0.0292	0.0520	0.2020	0.2084
$n = 2000$	0.0089	-0.0011	0.1208	0.1208
$n = 4000$	0.0026	-0.0082	0.0656	0.0661

NOTE: The estimator for ϕ_0 is as described in Section 4 with $J_\phi(n) = 1$. The function f_0 is parametrized up to the finite dimensional parameter γ_0 .

(b) Simulation Results for Design 2N

	Estimator for ϕ_0	Estimator for f_0
	IMSE	IMSE
Poisson process		
$n = 500$	0.0365	0.2618
$n = 2000$	0.0087	0.0783

NOTE: The estimators for ψ_0 and f_0 are as described in Section 4 with $J_\phi(n) = 1$ and $J_f(n) = 3$, respectively.

Design 1: Simulation results for Laplace exponent

Design 1: Simulation results for Laplace exponent

Figure 2: Simulation results for the Laplace exponent of a Poisson process associated to Design 1. The top graph corresponds to $n = 500$, while the bottom corresponds to $n = 2000$. For both, $J_{\Psi}(n) = 3$ computed as described in Section 4.

Figure 3: Simulation results for the Laplace exponent of a Poisson process associated to Design 2N. The number of basis functions for the Bernstein polynomial is $J_{\Psi}(n) = 3$ computed as described in Section 4. Sample size is $n = 2000$.

Table 3: Simulation Results for Design 3

(a) Simulation Results for Design 3S

	Estimator for $\beta_0 = 1$			Estimator for f_0
	Bias	SD	RMSE	IMSE
Poisson process				
$n = 1000$	-0.0271	0.1303	0.1330	0.0833
$n = 2000$	-0.0118	0.0685	0.0695	0.0252
$n = 4000$	-0.0056	0.0326	0.0331	0.0076

NOTE: The function ϕ_0 is parametrized up to β_0 , while the estimator for f_0 is as described in Section 4 with $J_f(n) = 2$.

(b) Simulation Results for Design 3N

	Estimator for ϕ_0	Estimator for f_0
	IMSE	IMSE
Poisson process		
$n = 500$	0.0915	0.1138
$n = 2000$	0.0096	0.0599

NOTE: The estimators for ψ_0 and f_0 are as described in Section 4 with $J_\phi(n) = 1$ and $J_f(n) = 2$, respectively.

Figure 4: Simulation results for the Laplace exponent of a Poisson process associated to Design 3N. The number of basis functions for the Bernstein polynomial is $J_{\Psi}(n) = 3$ computed as described in Section 4. Sample size is $n = 2000$.

Table 4: Simulation Results for Design 4S

	Estimator for $\beta_0 = 1$			Estimator for $\gamma_0 = 1$		
	Bias	SD	RMSE	Bias	SD	RMSE
Gamma process						
$n = 1000$	0.0711	0.1497	0.1656	0.0396	0.2256	0.2290
$n = 2000$	0.0285	0.0787	0.0834	-0.0229	0.1345	0.1363
$n = 4000$	0.0126	0.0346	0.0375	-0.0291	0.0735	0.0791
Poisson process						
$n = 1000$	0.0088	0.2207	0.2207	-0.0045	0.1308	0.1307
$n = 2000$	0.0011	0.0583	0.0582	-0.0084	0.0707	0.0712
$n = 4000$	-0.0100	0.0387	0.0405	-0.0027	0.0424	0.0429

NOTE: The function ϕ_0 and f_0 are parametrized up to finite dimensional parameters β_0 and γ_0 , respectively, while Ψ_0 is estimated as described in Section 4.

6 Conclusion

This paper establishes conditions for the nonparametric identification of a duration model where the unobserved heterogeneity is specified as a Lévy subordinator. The model belongs to the class of models proposed in biostatistics by Gjessing, Aalen, and Hjort (2003). A shape-restricted two-step sieve minimum distance estimator is proposed and the rate of convergence of the nonparametric estimator is derived. Numerical studies show that the estimator has very good finite-sample performance.

The model considered in this paper represents an initial step towards allowing for a more flexible heterogeneity structure in duration analysis. I discuss below extensions that may significantly expand the applicability of duration models with time-varying unobserved heterogeneity to dynamic environments. These extensions are beyond the scope of this paper.

Censoring. When data is right censored and the censoring is non-informative, the identification

Design 4S: Simulation results for Gamma Laplace exponent

Design 4S: Simulation results for Poisson Laplace exponent

Figure 5: Simulation results for the Laplace exponent of a Gamma process (top) and a Poisson process (bottom) for Design 4S. For both, $J_{\Psi}(n) = 3$ computed as described in Section 4. Sample size is $n = 4000$.

strategy presented in this paper remains unchanged. The estimation procedure changes in that the first stage estimator of the survival function must account for censoring.

Duration Dependence. Assumption Z serves two purposes: (i) it renders a non-negative hazard function, and (ii) it facilitates the derivation of a closed-form expression for the survival function. However, the assumption is not innocuous. This assumption implies positive duration dependence at the individual level. That is, the likelihood of exiting a state increases with the duration of the state. There may be different ways of relaxing this assumption. One way would be to model the unobserved heterogeneity as a spectrally positive Lévy process instead of a Lévy subordinator. This assumption changes the functional form of the survival function, which may result in a different identification method.

Initial Conditions. Another restriction of modeling unobserved heterogeneity as a Lévy process is that $Z(0) = 0$ with probability one. This implies that at $t = 0$, the individual hazard is zero with probability one. However, the process generating the unobserved heterogeneity starts as soon as individuals enter the spell, so that at $t + \varepsilon$, where $\varepsilon > 0$ is arbitrarily small, each individual has a different realization of the hazard function. Although there is no unobserved heterogeneity at the beginning of the spell, unobserved heterogeneity kicks in as soon as individuals enter the spell. One possible way of introducing heterogeneity at the beginning of the spell would be to assume that initial heterogeneity is due to observable heterogeneity y , and specify the hazard as, e.g. $y\beta + \phi(x) \int_0^t f(u) dZ(u)$. As long as y is observable, β is identified from the observations with $t \rightarrow 0$, while the identification of the other parameters would remain unchanged conditional on y .

Time Deformed Unobserved Heterogeneity. Modeling dynamics of unobserved heterogeneity in this set-up is more flexible than in the MPH. For example, it is possible to express $Z(t)$ as a function of time varying observed covariates, i.e. $\{Z(g(x(t)))\}_{t>0}$ where $x(t)$ are time-varying covariates and $g(\cdot)$ is an unknown function of $x(t)$. Stochastic processes of this form are known as time-deformations since the time scale of the process changes from calendar time, t , to $g(x(t))$. Stock (1988) mentions situations when it may be more realistic to model certain phenomena as evolving in operational or economic time rather than in calendar time. In a different paper, I discuss the interpretability and

the identification of g when the hazard function is specified as

$$\phi(y) \int_0^t dZ(g(x(u))) du,$$

where y are time-invariant observed covariates and $x(t)$ are time-varying observed covariates.

Multiple Spells. An interesting extension would be to data where more than one transition is observed for each individual. This would, for example, allow the observed covariates to be correlated with past realizations of the unobserved heterogeneity, which would significantly expand the applicability of duration models. Important considerations here would be whether the distribution of the unobserved heterogeneity is the same across spells, as well as whether there is serial dependence among spells.

7 Appendix

A Subordinators

This section contains definitions and properties of Lévy subordinators, which are referred to in the main text and in the proofs. The main references for this section are Applebaum (2009) and Bertoin (1996).

Definition 1 (Lévy process) *A stochastic process $\{Z(t)\}_{t \geq 0}$ is a Lévy process with respect to a filtration $\mathcal{F}_t$ if for all $0 \leq u \leq t < \infty$, the following properties hold:*

P(i) $P(Z(t) - Z(u) \in A) = P(Z(t - u) - Z(0) \in A)$ for all Borel sets A .

P(ii) For $0 \leq t_1 \leq t_2 \leq \dots \leq t_n < \infty$, the random variables $Z(t_1) - Z(0), \dots, Z(t_n) - Z(t_{n-1})$ are independent.

P(iii) $P(Z(0) = 0) = 1$.

P(iv) For all $a > 0$ and all $u \geq 0$, $\lim_{t \rightarrow u} P(|Z(t) - Z(u)| > a) = 0$.

By P(i), the increments of the process are identically distributed, while by P(ii) they are independent. By P(iii), the paths of the process start at the origin. This is a normalization that can

be relaxed to $P(Z(0) = z) = 1$, where z is an element of the support of the random variable $Z(0)$. Finally, by P(iv) the paths of the process are stochastically continuous.

Definition 2 (Subordinator) *A subordinator, $\{Z(t)\}_{t \geq 0}$, is a one-dimensional Lévy process that is nondecreasing with probability 1.*

Definition 3 (Laplace exponent) *The function $\Psi : [0, \infty) \rightarrow [0, \infty)$ is the Laplace exponent of a subordinator $\{Z(t)\}_{t \geq 0}$ given by:*

$$\exp(-t\Psi(\chi)) = E[e^{-\chi Z(t)}] \text{ for all } \chi \in [0, \infty), \quad (27)$$

where the expectation is taken with respect to the randomness in the subordinator. The Laplace exponent is non-negative, non-decreasing, and concave.

Laplace exponents have the following property, see Gneden and Pitman (2008):

Property 1 A function $\Psi(\chi)$, $\chi \in [0, \infty)$, is the Laplace exponent of a subordinator if:

- (i) $\Psi(\chi)$ is infinitely differentiable with respect to $\chi \in [0, \infty)$;
- (ii) $\Psi(0) = 0$;
- (iii) $(-1)^n \frac{\partial^n}{\partial \chi^n} \Psi(\chi) \leq 0$ for all $n \in \mathbb{N}$ and $\chi \in [0, \infty)$.

Property 1(iii) shows that the Laplace exponent is a Bernstein function on $[0, \infty)$. By definition, the first derivative of a Bernstein function is completely monotone on $(0, \infty)$ and, hence, real analytic on $(0, \infty)$.

Bernstein functions have the following Lévy-Khinchine representation:

Lévy-Khinchine formula For a subordinator $\{Z(t)\}_{t \geq 0}$, there exists a Lévy measure μ such that

$$\mu(-\infty, 0) = 0 \text{ and } \int_0^\infty (z \wedge 1) \mu(dz) < \infty.$$

For $0 \leq d < \infty$ (the drift of the process), the Laplace exponent is:

$$\Psi(\chi) = d\chi + \int_0^\infty (1 - e^{-\chi z}) \mu(dz).$$

Example 1 (Gamma Process) Let Z be a gamma random variable with shape parameter $\rho > 0$ and scale parameter $s > 0$, i.e. $Z \sim \text{Ga}(\rho, s)$. Then the gamma process $\{Z(t)\}_{t \geq 0}$ is distributed as:

$$\{Z(t)\}_{t \geq 0} \sim \text{Ga}(\rho t, s), \quad t \geq 0,$$

and its Laplace exponent is given by:

$$\Psi(\chi) = \log \left(1 + \frac{\chi}{s} \right)^\rho.$$

Finally, for $f(t)$ a non-random function as in (M), Dykstra and Laud (1981) showed that:

$$\int_0^t f(u) dZ(u) \sim \text{Ga}(\rho t, s f(t)),$$

which implies that the increments of the resulting process (the stochastic integral) are not identically distributed.

B Identification Proofs

B.1 Proof of Lemma 1

This section presents the derivation of the survival function in (4).

Define

$$F(u, t, x) \equiv \phi(x) \int_u^t f(u) ds = \phi(x) f(u) (t - u), \quad (28)$$

and let $\{u_{n,j}\}_{j=1}^n$, $n = 1, 2, \dots$, be a sequence of partitions of $[0, t]$ such that $0 = u_{n,0} < u_{n,1} < \dots < u_{n,n} = t$, $\lim_{n \rightarrow \infty} \sup_{1 \leq j \leq n} |u_{n,j} - u_{n,j-1}| = 0$, and $u_{n,j}^* \in [u_{n,j-1}, u_{n,j}]$, $1 \leq j \leq n$.

Consider the survival function associated with the hazard function in (M) :

$$S(t|X = x) = E_Z \left[\exp \left(- \int_0^t h(s|x, \{Z_s\}_{s \geq 0}) ds \right) \right] \quad (29)$$

$$= E_Z \left[\exp \left(- \int_0^t F(u, t, x) dZ(u) \right) \right] \quad (30)$$

$$= E_Z \left[\exp \left(- \lim_{n \rightarrow \infty} \sum_{j=1}^n F(u_{n,j}^*, t, x) (Z(u_{n,j}) - Z(u_{n,j-1})) \right) \right] \quad (31)$$

$$= E_Z \left[\lim_{n \rightarrow \infty} \exp \left(- \sum_{j=1}^n F(u_{n,j}^*, t, x) (Z(u_{n,j}) - Z(u_{n,j-1})) \right) \right] \quad (32)$$

$$= \lim_{n \rightarrow \infty} E_Z \left[\exp \left(- \sum_{j=1}^n F(u_{n,j}^*, t, x) (Z(u_{n,j}) - Z(u_{n,j-1})) \right) \right] \quad (33)$$

$$= \lim_{n \rightarrow \infty} E_Z \left[\prod_{j=1}^n \exp \left(- F(u_{n,j}^*, t, x) (Z(u_{n,j}) - Z(u_{n,j-1})) \right) \right] \quad (34)$$

$$= \lim_{n \rightarrow \infty} \prod_{j=1}^n E_Z \left[\exp \left(- F(u_{n,j}^*, t, x) (Z(u_{n,j}) - Z(u_{n,j-1})) \right) \right] \quad (35)$$

$$= \lim_{n \rightarrow \infty} \prod_{j=1}^n \exp \left(- (u_{n,j} - u_{n,j-1}) \Psi(F(u_{n,j}^*, t, x)) \right) \quad (36)$$

$$= \lim_{n \rightarrow \infty} \left[\exp \left(- \sum_{j=1}^n (u_{n,j} - u_{n,j-1}) \Psi(F(u_{n,j}^*, t, x)) \right) \right] \quad (37)$$

$$= \exp \left(- \int_0^t \Psi \left(\int_u^t \phi(x) f(u) ds \right) du \right) \quad (38)$$

$$= \exp \left(- \int_0^t \Psi(\phi(x) f(u) (t - u)) du \right) \quad (39)$$

where (29) follows by Assumption S1 and the expectation is taken with respect to the distribution of the process $\{Z(t)\}_{t \geq 0}$; (30) follows by a change in the order of integration and using (28), i.e.

$$\begin{aligned} \int_0^t h(s|x, \{Z_s\}_{s \geq 0}) ds &= \int_0^t \left(\int_0^s \phi(x) f(u) dZ(u) du \right) ds \\ &= \int_0^t \left(\int_u^t \phi(x) f(u) ds \right) dZ(u) \\ &= \int_0^t F(u, t, x) dZ(u); \end{aligned}$$

(31) follows by Assumption Z, which implies that the process is of finite variation, and by Assumption

S2; together, these two assumptions allow us to write the stochastic integral in (30) as a Riemann sum; (32) holds since $\exp(\cdot)$ is a continuous function, so that:

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

(33) follows by the Bounded Convergence Theorem since

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

(35) follows by the independence of the increments of $\{Z(t)\}_{t \geq 0}$; (36) follows by the definition of the Laplace exponent and by Assumption Z, in particular, by the independence of the increments of $\{Z(t)\}_{t \geq 0}$, which allows to write that:

$$E_Z[\exp(-u(Z(t) - Z(s)))] = E_Z[\exp(-uZ(t-s))] = \exp[-(t-s)\Psi(u)];$$

and, finally, (38) follows by continuity of $\exp(\cdot)$ and by switching back to integral notation, which is possible given the local integrability of the Laplace exponent. This last property follows since the Laplace exponent is continuous on its domain.

B.2 Proof of Theorem 1

Repeated differentiation of (5) at $t = 0$ obtains the following formula for at least two different values of $x \in \mathcal{X}$:

$$\begin{aligned} \frac{\partial^\kappa}{\partial t^\kappa} H(0, x) &= \left(\frac{\partial^\kappa}{\partial t^\kappa} \int_0^t \Psi(\phi(x) f(u)(t-u)) du \right) \Big|_{t=0} \\ &= \sum_{j=1}^{\kappa-1} [\phi(x)]^j \times [\Psi^{(j)}(0)] \times \left(\frac{d^{\kappa-j-1}}{dt^{\kappa-j-1}} [f(t)]^j \right) \Big|_{t=0}, \quad \kappa = 2, 3, \dots \end{aligned} \quad (40)$$

Note that it is possible to differentiate (5) with respect to t by Assumption ID0, which allows interchanging differentiation and integration on the right hand-side of (5); further, Assumption ID1

guarantees the existence of $\left\{ \frac{\partial^\kappa}{\partial t^\kappa} H(0, x), f^{(\kappa)}(0), \Psi^{(\kappa)}(0) \right\}_{\kappa \geq 1}$. Finally, the term $\Psi(0)$ drops out since $\Psi(0) = 0$ by the properties of the Laplace exponent.

Expression (40) will be used to show the identification of the function $\phi(\cdot)$ and of the sequence

$$\left\{ f^{(\kappa-1)}(0), \Psi^{(\kappa-1)}(0) \right\}_{\kappa \geq 2}.$$

The identification strategy proceeds sequentially, with the function $\phi(\cdot)$ and the mean, or the first cumulant, $\Psi^{(1)}(0)$, being identified by setting $\kappa = 2$ in (40). Given identification of $\{\phi(\cdot), \Psi^{(1)}(0)\}$, the derivative $f^{(1)}(0)$ and the second cumulant, $\Psi^{(2)}(0)$, are identified by setting $\kappa = 3$ in (40). Given identification of $\{\phi(\cdot), f^{(1)}(0), \Psi^{(1)}(0), \Psi^{(2)}(0)\}$, the derivative $f^{(2)}(0)$ and the second cumulant, $\Psi^{(3)}(0)$, are identified by setting $\kappa = 4$ in (40), and so on. I show this sequential argument below.

Let $\kappa = 2$ in (40). Then, given Assumption ID3,

$$\frac{\partial^2}{\partial t^2} H(0, x) = \phi(x) \Psi^{(1)}(0). \quad (41)$$

Letting $x = x_0$ in (41) and using Assumption ID2(i), obtains

$$\frac{\partial^2}{\partial t^2} H(0, x_0) = \Psi^{(1)}(0), \quad (42)$$

which is expression (8) in Theorem 1. Letting x vary in (41) and using Assumption ID2(ii), obtains

$$\phi(x) = \frac{\frac{\partial^2}{\partial t^2} H(0, x)}{\frac{\partial^2}{\partial t^2} H(0, x_0)}, \quad (43)$$

where (42) was used. Expression (43) is expression (7) in the statement of Theorem 1.

Now, let $\kappa = 3$ in (40). Given Assumption ID3,

$$\begin{aligned} \frac{\partial^3}{\partial t^3} H(0, x) &= \sum_{j=1}^2 \phi^j(x) \Psi^{(j)}(0) \left(\frac{d^{2-j}}{dt^{2-j}} f^j(t) \right) \Big|_{t=0} \\ &= \phi(x) \Psi^{(1)}(0) f^{(1)}(0) + \phi^2(x) \Psi^{(2)}(0). \end{aligned} \quad (44)$$

Consider now (44) evaluated at two different values $x_0 \neq x_1 \in \mathcal{X}$. Assumptions ID2 and ID3 obtain the following system of equations for $(f^{(1)}(0), \Psi^{(2)}(0))$:

$$\begin{bmatrix} \frac{\partial^3}{\partial t^3} H(0, x_0) \\ \frac{\partial^3}{\partial t^3} H(0, x_1) \end{bmatrix} = \begin{bmatrix} \Psi^{(1)}(0) & 1 \\ \phi(x_1) \Psi^{(1)}(0) & \phi^2(x_1) \end{bmatrix} \begin{bmatrix} f^{(1)}(0) \\ \Psi^{(2)}(0) \end{bmatrix} \quad (45)$$

where $(\phi(x_1), \Psi^{(1)}(0))$ have been identified in the $\kappa = 2$ step above. By Assumption ID2, the determinant of the coefficient matrix is different from zero, i.e.

$$\phi(x_1) \Psi^{(1)}(0) (\phi(x_1) - 1) \neq 0$$

Then the system in (45) can be solved uniquely for the parameters of interest to obtain:

$$\begin{aligned} \Psi^{(2)}(0) &= \frac{-\phi(x_1) \frac{\partial^3}{\partial t^3} H(0, x_0) + \frac{\partial^3}{\partial t^3} H(0, x_1)}{\phi(x_1) [\phi(x_1) - 1]}, \\ f^{(1)}(0) &= \frac{\phi^2(x_1) \frac{\partial^3}{\partial t^3} H(0, x_0) - \frac{\partial^3}{\partial t^3} H(0, x_1)}{\phi(x_1) [\phi(x_1) - 1] \Psi^{(1)}(0)}, \end{aligned}$$

which are expressions (11) and (9), respectively, in the statement of Theorem 1.

Letting $\kappa = 4$ in (40) and given Assumption ID3 obtains:

$$\begin{aligned} \frac{\partial^4}{\partial t^4} H(0, x) &= \sum_{j=1}^3 \phi^j(x) \Psi^{(j)}(0) \left(\frac{d^{3-j}}{dt^{3-j}} f^j(t) \right) \Big|_{t=0} \\ &= \phi(x) \Psi^{(1)}(0) f^{(2)}(0) + 2\phi^2(x) \Psi^{(2)}(0) f^{(1)}(0) + \phi^3(x) \Psi^{(3)}(0). \end{aligned}$$

Evaluating the expression above for two different $x_0 \neq x_1 \in \mathcal{X}$, and given assumptions ID2 and ID3, obtains the following system of equations for $(f^{(2)}(0), \Psi^{(3)}(0))$:

$$\begin{bmatrix} D_4(x_0) \\ D_4(x_1) \end{bmatrix} = \begin{bmatrix} \Psi^{(1)}(0) & 1 \\ \phi(x_1) \Psi^{(1)}(0) & \phi^3(x_1) \end{bmatrix} \begin{bmatrix} f^{(2)}(0) \\ \Psi^{(3)}(0) \end{bmatrix}, \quad (46)$$

where, using expression (6) with $\kappa = 4$,

$$\begin{aligned} D_4(x) &= \frac{\partial^4}{\partial t^4} H(0, x) - \sum_{k=2}^2 \phi^k(x_l) \Psi^{(k)}(0) \left(\frac{d^{3-k}}{dt^{3-k}} f^k(t) \right) \Big|_{t=0} \\ &= \frac{\partial^4}{\partial t^4} H(0, x) - 2\phi^2(x) \Psi^{(2)}(0) f^{(1)}(0). \end{aligned}$$

Notice that $(f^{(1)}(0), \Psi^{(2)}(0))$ have been identified in the $\kappa = 3$ step, while $\phi(\cdot)$ has been identified in the $\kappa = 2$ step. Since the determinant of the coefficient matrix in (46) is different from zero by Assumption ID2, the system in (46) can be solved uniquely for the parameters of interest:

$$\begin{aligned} \Psi^{(3)}(0) &= \frac{D_4(x_1) - \phi(x_1) D_4(x_0)}{\phi(x_1) [\phi^2(x_1) - 1]}, \\ f^{(2)}(0) &= \frac{\phi^3(x_1) D_4(x_0) - D_4(x_1)}{\phi(x_1) [\phi^2(x_1) - 1] \Psi^{(1)}(0)}. \end{aligned}$$

Letting $\kappa = M \geq 5$ in (40) obtains

$$\begin{aligned} \frac{\partial^M}{\partial t^M} H(0, x) &= \sum_{j=1}^{M-1} \phi^j(x) \Psi^{(j)}(0) \left(\frac{d^{M-1-j}}{dt^{M-1-j}} f^j(t) \right) \Big|_{t=0} \\ &= \phi(x) \Psi^{(1)}(0) \left(\frac{d^{M-2}}{dt^{M-2}} f(t) \right) \Big|_{t=0} \\ &\quad + \phi^{M-1}(x) \Psi^{(M-1)}(0) \left(\frac{d^{M-(M-1)-1}}{dt^{M-(M-1)-1}} f^{M-1}(t) \right) \Big|_{t=0} \\ &\quad + \sum_{j=2}^{M-2} \phi^j(x) \Psi^{(j)}(0) \left(\frac{d^{M-1-j}}{dt^{M-1-j}} f^j(t) \right) \Big|_{t=0} \\ &= \phi(x) \Psi^{(1)}(0) f^{(M-2)}(0) + \phi^{M-1}(x) \Psi^{(M-1)}(0) \\ &\quad + \sum_{j=2}^{M-2} \phi^j(x) \Psi^{(j)}(0) \left(\frac{d^{M-1-j}}{dt^{M-1-j}} f^j(t) \right) \Big|_{t=0}, \end{aligned} \tag{47}$$

where Assumption ID3 was used in the second equality sign above, and the first and the last terms containing the unknowns, $f^{(M-2)}(0)$ and $\Psi^{(M-1)}(0)$, were taken out of the summation. Evaluating (47)

at two different $x_0 \neq x_1 \in \mathcal{X}$, obtains the following system of equations for $(f^{(M-2)}(0), \Psi^{(M-1)}(0))$:

$$\begin{bmatrix} D_M(x_0) \\ D_M(x_1) \end{bmatrix} = \begin{bmatrix} \Psi^{(1)}(0) & 1 \\ \phi(x_1) \Psi^{(1)}(0) & \phi^{M-1}(x_1) \end{bmatrix} \begin{bmatrix} f^{(M-2)}(0) \\ \Psi^{(M-1)}(0) \end{bmatrix} \quad (48)$$

where $D_{M+1}(x)$ is as defined in (6) with $\{\phi(x_1), \Psi^{(j)}(0), f^{(j-1)}(0)\}_{j=2}^{M-2}$ having been identified previously. By Assumption ID2 the determinant of the coefficient matrix in (48) is different from zero, so the system in (48) can be solved uniquely for the parameters of interest to obtain:

$$\begin{aligned} \Psi^{(M-1)}(0) &= \frac{D_M(x_1) - \phi(x_1) D_M(x_0)}{\phi(x_1) [\phi^{M-2}(x_1) - 1]}, \\ f^{(M-2)}(0) &= \frac{\phi^{M-1}(x_1) D_M(x_0) - D_M(x_1)}{\phi(x_1) [\phi^{M-2}(x_1) - 1] \Psi^{(1)}(0)}. \end{aligned}$$

Since ϕ and $\{\Psi^{(j)}(0), f^{(j-1)}(0)\}_{j=1}^M$ are identified for all positive integers M , it follows that they are identified in the limit as $M \rightarrow \infty$. Hence, the infinite sequence $\{\Psi^{(j)}(0), f^{(j-1)}(0)\}_{j \geq 2}$ is identified.

C Estimation proofs

C.1 Proof of Theorem 2

To show that $d(\widehat{\theta}_n, \theta_0) \xrightarrow{P} 0$, I check the conditions of Theorem 1 of Kristensen (2009). These conditions are:

Condition 1 For all $\varepsilon > 0$

$$\sup_{\theta \in \Theta: d(\theta, \theta_0) > \varepsilon} M(\theta, H_0) < M(\theta_0, H_0)$$

Condition 2 For all $\delta > 0$

$$\sup_{\theta \in \Theta: d(H, H_0) < \delta} |M_n(\theta, H) - M(\theta, H_0)| \xrightarrow{P} 0$$

Condition 3 *Let $H, H_* \in \mathcal{H}$. As $H \rightarrow H_*$*

$$\sup_{\theta \in \Theta} |M(\theta, H) - M(\theta, H_*)| \rightarrow 0$$

Condition 1 holds since, for any $\theta \neq \theta_0$,

$$\begin{aligned} M(\theta, H_0) &= -E[H_0(W) - I(\theta_0, W) + I(\theta_0, W) - I(\theta, W)]^2 \\ &= -E[I(\theta_0, W) - I(\theta, W)]^2 \\ &< M(\theta_0, H_0) = 0. \end{aligned}$$

Condition 2 is satisfied if the class of functions below is P-Glivenko Cantelli:

$$\mathcal{M} = \{m_{\theta, H}(\cdot) : \theta \in \Theta, H \in \mathcal{H}\}.$$

Verifying this condition requires showing that the bracketing number $N_{[]}(\varepsilon, \mathcal{M}, L_1)$ is finite for every $\varepsilon > 0$, see Theorem 2.4.1 of van der Vaart and Wellner (1996). To do so, I first show that $m_{\theta, H}$ is Lipschitz continuous in θ and H , and then I apply Theorem 2.7.11 in van der Vaart and Wellner (1996).

First, note that the bounds below exist by assumptions C1(ii), (21), (22), and (23) :

$$b_I \equiv \sup_{\theta, W} I(\theta, W) < \infty \text{ and } b_H \equiv \sup_{W \in \mathcal{W}} H(W) < \infty.$$

Second, letting $H_1, H_2 \in \mathcal{H}$ and $\theta_1, \theta_2 \in \Theta$, $m_{\theta, H}$ can be shown to be Lipschitz continuous in θ and H for every $W \in \mathcal{W}$. To show this, consider first:

$$\begin{aligned}
& |m_{\theta_1, H_1}(W) - m_{\theta_2, H_2}(W)| \\
&= |(H_1(W) - I(\theta_1, W))^2 - (H_2(W) - I(\theta_2, W))^2| \\
&\leq |H_1^2(W) - H_2^2(W)| + 2|H_1(W)I(\theta_1, W) - H_2(W)I(\theta_2, W)| + |I^2(\theta_1, W) - I^2(\theta_2, W)| \\
&\leq |H_1(W) + H_2(W)||H_1(W) - H_2(W)| \\
&\quad + 2|H_1(W)I(\theta_1, W) - H_1(W)I(\theta_2, W) + H_1(W)I(\theta_2, W) - H_2(W)I(\theta_2, W)| \\
&\quad + |I(\theta_1, W) + I(\theta_2, W)||I(\theta_1, W) - I(\theta_2, W)| \\
&\leq 2(b_H + b_I) \left[\|H_1 - H_2\|_\infty + \sup_W |I(\theta_1, W) - I(\theta_2, W)| \right]
\end{aligned}$$

Further, letting

$$\gamma_j(X, u) \equiv \phi_j(X) f_j(u), \quad j = 1, 2,$$

it is possible to show that the function $I(\theta, \cdot)$ is Lipschitz continuous in θ :

$$\begin{aligned}
& \sup_W |I(\theta_1, W) - I(\theta_2, W)| \\
&= \sup_W \left| \int_0^T [\Psi_1(\gamma_1(X, u)(T-u)) - \Psi_2(\gamma_2(X, u)(T-u))] du \right| \\
&\leq \sup_W \int_0^T |\Psi_1(\gamma_1(X, u)(T-u)) - \Psi_1(\gamma_2(X, u)(T-u))| du \tag{49}
\end{aligned}$$

$$+ \sup_W \int_0^T |\Psi_1(\gamma_2(X, u)(T-u)) - \Psi_2(\gamma_2(X, u)(T-u))| du \tag{50}$$

$$\lesssim \max(b_\phi, b_f, 1) [\|\phi_1 - \phi_2\|_\infty + \|f_1 - f_2\|_\infty + \|\Psi_1 - \Psi_2\|_\ominus] \tag{51}$$

Expression (51) follows by the arguments below. Using the Lévy-Khintchine formula, which shows that Ψ is Lipschitz continuous,¹⁹ the term in (49) can be bounded by:

$$\begin{aligned}
& \sup_W \int_0^T |\Psi_1(\gamma_1(X, u)(T-u)) - \Psi_1(\gamma_2(X, u)(T-u))| du \\
& \lesssim \sup_{W, u} T^2 |\gamma_1(X, u) - \gamma_2(X, u)| \\
& \lesssim \sup_T T^2 [b_\phi \|f_1 - f_2\|_\infty + b_f \|\phi_1 - \phi_2\|_\infty] \\
& \lesssim \max(b_\phi, b_f) [\|f_1 - f_2\|_\infty + \|\phi_1 - \phi_2\|_\infty]
\end{aligned}$$

while the term in (50) is bounded by:

$$\sup_W \int_0^T |\Psi_1(\gamma_2(X, u)(T-u)) - \Psi_2(\gamma_2(X, u)(T-u))| du \leq \|\Psi_1 - \Psi_2\|_\Theta$$

Let $K \equiv 2(b_H + b_I) \max(1, b_\phi, b_f)$. Then by Theorem 2.7.11 in van der Vaart and Wellner (1996) and by the inequality between the bracketing number and the covering number of a class of functions, we get that:

$$N_{[]}(\varepsilon, \mathcal{M}, L_1) \leq N_{[]} \left(\frac{\varepsilon}{4K}, \mathcal{H}, \|\cdot\|_\infty \right) N_{[]} \left(\frac{\varepsilon}{4K}, \mathcal{F}, \|\cdot\|_\infty \right) N_{[]} \left(\frac{\varepsilon}{4K}, \Phi, \|\cdot\|_\infty \right) N_{[]} \left(\frac{\varepsilon}{4K}, \Theta, \|\cdot\|_\Theta \right)$$

Since $\mathcal{H}$ is the class of monotone functions (by construction), which is continuous by Assumption ID1 and bounded by Assumption C1, its ε -bracketing entropy is of order $\frac{1}{\varepsilon}$, i.e.

$$\log N_{[]}(\varepsilon, \mathcal{H}, \|\cdot\|_\infty) \lesssim \frac{C}{\varepsilon}$$

¹⁹By the Lévy-Khintchine formula, it can be shown that the Laplace exponent is Lipschitz continuous in its argument:

$$\begin{aligned}
|\Psi(\chi_1) - \Psi(\chi_2)| & \leq |d(\chi_1 - \chi_2)| + \left| \int_0^\infty (e^{-\chi_2 u} - e^{-\chi_1 u}) \mu(du) \right| \\
& \leq d|\chi_1 - \chi_2| + \int_0^\infty | -ue^{-\tilde{\chi}u} | |\chi_1 - \chi_2| \mu(du) \\
& \leq \tilde{d}|\chi_1 - \chi_2|,
\end{aligned}$$

where $\tilde{\chi} \in [\chi_1, \chi_2]$, $0 < \tilde{d} \neq d$ is a generic finite constant, and $\int_0^\infty | -ue^{-\tilde{\chi}u} | \mu(du) < \int_0^\infty u \mu(du) < \infty$ by the properties of the Lévy measure for subordinators.

where C is some constant that is not a function of ε . By assumptions C1, (21), (22), and (23), the ε -bracketing entropies of $\mathcal{F}$ and Φ are given by:

$$\begin{aligned}\log N_{[]}(\varepsilon, \Phi, \|\cdot\|_\infty) &\lesssim \left(\frac{b_\phi}{\varepsilon}\right)^{1/p_\phi} \\ \log N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) &\lesssim \left(\frac{b_f}{\varepsilon}\right)^{1/p_f}\end{aligned}$$

see, e.g., van der Vaart (1994). Since Ψ is Lipschitz continuous, by Theorem 2.7.11 of van der Vaart and Wellner (1996):

$$\begin{aligned}\log N_{[]}(\varepsilon, \Psi, \|\cdot\|_\Theta) &\leq \log N\left(\frac{\varepsilon}{4C}, \Phi, \|\cdot\|_\infty\right) + \log N\left(\frac{\varepsilon}{4C}, \mathcal{F}, \|\cdot\|_\infty\right) \\ &\lesssim \left(\frac{b_\phi}{\varepsilon}\right)^{1/p_\phi} + \left(\frac{b_f}{\varepsilon}\right)^{1/p_f}\end{aligned}$$

where C is a constant that is not a function of ε .

Putting everything together and by the relation between covering and bracketing numbers:

$$\log N_{[]}(\varepsilon, \mathcal{M}, \|\cdot\|_\infty) \lesssim \frac{1}{\varepsilon} + \left(\frac{b_f}{\varepsilon}\right)^{1/p_f} + \left(\frac{b_\phi}{\varepsilon}\right)^{1/p_\phi}$$

which implies that the class $\mathcal{M}$ is Glivenko Cantelli.

Finally, letting $H_* \in \mathcal{H}$, Condition 3 follows by:

$$\begin{aligned}
& \sup_{\theta \in \Theta} |M(\theta, H) - M(\theta, H_*)| \\
& \leq \sup_{\theta \in \Theta} E |m_{\theta, H}(W) - m_{\theta, H_*}(W)| \\
& = \sup_{\theta \in \Theta} E |(H(W) - I(\theta, W))^2 - (H_*(W) - I(\theta, W))^2| \\
& \leq \sup_{\theta \in \Theta} E |H^2(W) - H_*^2(W) - 2I(\theta, W)(H(W) - H_*(W))| \\
& \leq \sup_{\theta \in \Theta} E |H(W) - H_*(W)| |H(W) + H_*(W) - 2I(\theta, W)| \\
& \leq \|H - H_*\|_\infty \sup_{\theta \in \Theta} E |H(W) + H_*(W) - 2I(\theta, W)| \\
& \leq \|H - H_*\|_\infty \left[E |H(W) + H_*(W)| + 2 \sup_{\theta \in \Theta} E |I(\theta, W)| \right] \\
& \leq 2 \|H - H_*\|_\infty [b_H + b_I] \\
& \rightarrow 0 \text{ as } H \rightarrow H_*
\end{aligned}$$

C.2 Proof of Theorem 3

To establish the rate of convergence of $\widehat{\theta}_n$, I check the conditions of Lemma 2 of Chen, Lee, and Sung (2014). These conditions are included below.

Let ε_n be non-stochastic positive real-sequence and let $C > 0$ be a constant. Define the ball:

$$\mathcal{H}_n(C) \equiv \{H \in \mathcal{H}_n : \|H - H_0\|_{L_2} \leq C\varepsilon_n\}$$

Condition 4 For $C > 0$ and n large enough

$$\sup_{H \in \mathcal{H}_n(C)} |M(H, \theta_0) - M(H_0, \theta_0)| \leq (C\varepsilon_n)^2$$

Condition 5 There is a sequence of non-stochastic functions $e_n : \mathcal{H}_n(C) \times \Theta_n \rightarrow \mathbb{R}$ such that for all

sufficiently small $\delta > 0$ and for all $(H, \theta) \in \mathcal{H}_n(C) \times \Theta_n$ such that $d(\theta, \theta_0) \leq \delta$ it holds that:

$$\begin{aligned} M(H, \theta) - M(H_0, \theta_0) + e_n(H, \theta) &\lesssim -d^2(\theta, \theta_0) + \|H - H_0\|_{L_2}^2 \\ \sup_{d(\theta, \theta_0) \leq \delta, (H, \theta) \in \mathcal{H}_n \times \Theta_n} |e_n(H, \theta)| &\lesssim C\delta\varepsilon_n \end{aligned}$$

Condition 6 Define

$$\begin{aligned} \widetilde{M}_n(H, \theta) &\equiv [M_n(H, \theta) - M(H, \theta)] - [M_n(H, \theta_0) - M(H, \theta_0)] \\ &= \frac{1}{\sqrt{n}} G_n(m_{\theta, H} - m_{\theta_0, H}), \quad G_n = \sqrt{n}(P_n - P) \end{aligned}$$

Then

$$E \left[\sup_{d(\theta, \theta_0) \leq \delta, (H, \theta) \in \mathcal{H}_n(C) \times \Theta_n} \left| \widetilde{M}_n(H, \theta) \right| \right] \lesssim \frac{\varphi_n(\delta)}{\sqrt{n}}$$

where $\varphi_n(\delta)$ is a sequence of functions on $(0, \infty)$ such that $\frac{\varphi_n(\delta)}{\delta^\rho}$ is decreasing for some $\rho < 2$.

Condition 7 $\left\| \widehat{H}_n - H_0 \right\|_{L_2}^2 = O_p(\varepsilon_n)$ and $d(\widehat{\theta}_n, \theta_0) = o_p(1)$

Condition 8 There exists a non-stochastic positive real sequence $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ such that, for every n ,

$$\varphi_n(\delta_n) \leq \sqrt{n}\delta_n^2$$

I now check these conditions.

To check Condition 4 consider the difference:

$$\begin{aligned} \sup_{H \in \mathcal{H}_n(C)} |M(H, \theta_0) - M(H_0, \theta_0)| &\leq \sup_{H \in \mathcal{H}_n(C)} E |[H(W) - H_0(W)]^2| \\ &\leq (C\varepsilon_n)^2 \end{aligned}$$

which holds by construction.

To check Condition 5 consider the following argument. Since $M(H_0, \theta_0) = 0$, we have that:

$$\begin{aligned}
M(H, \theta) - M(H_0, \theta_0) &= -E[H(W) - I(\theta, W)]^2 \\
&= -E[H(W) - H_0(W) + I(\theta_0, W) - I(\theta, W)]^2 \\
&= -E[H(W) - H_0(W)]^2 - E[I(\theta_0, W) - I(\theta, W)]^2 \\
&\quad - 2E[H(W) - H_0(W)][I(\theta_0, W) - I(\theta, W)]
\end{aligned}$$

Letting

$$e_n(\theta, H) = 2E[H(W) - H_0(W)][I(\theta_0, W) - I(\theta, W)],$$

obtains

$$M(H, \theta) - M(H_0, \theta_0) + e_n(\theta, H) = -E[H(W) - H_0(W)]^2 - E[I(\theta_0, W) - I(\theta, W)]^2.$$

Define now:

$$\begin{aligned}
\Delta_\phi &\equiv \phi - \phi_0 \\
\Delta_f &\equiv f - f_0 \\
\Delta_\Psi &\equiv \Psi - \Psi_0 \\
\gamma_0(W, u) &\equiv \phi_0(X) f_0(u) (T - u) \\
\xi_0(W, u) &\equiv (T - u) \Psi_0^{(1)}(\gamma_0(W, u))
\end{aligned}$$

Notice that given Assumption C1, $|\xi_0(W, u)| \leq b_\xi < \infty$. This is used below.

Consider the expansion:

$$\begin{aligned}
& E(I(\theta, W) - I(\theta_0, W))^2 \\
&= E\left(\frac{\partial}{\partial \theta} I(\theta_0, W)(\theta - \theta_0)\right)^2 \\
&= E\left(\frac{\partial}{\partial \tau} I(\theta_0 + \tau(\theta - \theta_0), W)\Big|_{\tau=0}\right)^2 \\
&= E\left(\frac{\partial}{\partial \tau} \int_0^T [\Psi_0((\phi_0 + \tau \Delta_\phi)(f_0 + \tau \Delta_f)(T - u)) + \tau \Delta_\Psi] du \Big|_{\tau=0}\right)^2 \\
&= E\left(\int_0^T f_0(u) \xi_0(W, u) \Delta_\phi(X) du + \int_0^T \phi_0(X) \xi_0(W, u) \Delta_f(u) du + \int_0^T \Delta_\Psi(\gamma_0(W, u)) du\right)^2 \\
&\leq E\left(\int_0^T f_0(u) \xi_0(W, u) \Delta_\phi(X) du\right)^2 + E\left(\int_0^T \phi_0(X) \xi_0(W, u) \Delta_f(u) du\right)^2 \\
&\quad + E\left(\int_0^T \Delta_\Psi(\gamma_0(W, u)) du\right)^2 \\
&\leq \sup_{u, W} (f_0(u) \xi_0(W, u))^2 E\left(\int_0^T \Delta_\phi(X) du\right)^2 + \sup_{u, W} (\phi_0(X) \xi_0(W, u))^2 E\left(\int_0^T \Delta_f(u) du\right)^2 \\
&\quad + E\left|\int_0^T \Delta_\Psi(\gamma_0(W, u)) du\right|^2 \\
&\leq K [\|\phi - \phi_0\|_\infty^2 + \|f - f_0\|_\infty^2 + \|\Psi - \Psi_0\|_\ominus^2]
\end{aligned}$$

where

$$K = T^2 \max(b_f^2 b_\xi^2, b_\phi^2 b_\xi^2, 1)$$

The inequalities above were derived using:

$$\begin{aligned}
E\left(\int_0^T \Delta_\phi(X) du\right)^2 &\leq E\left(\int_0^T |\phi(X) - \phi_0(X)| du\right)^2 \leq \|\phi - \phi_0\|_\infty^2 E(T^2) \\
E\left(\int_0^T \Delta_f(u) du\right)^2 &\leq E\left(\int_0^T |f(u) - f_0(u)| du\right)^2 \leq \|f - f_0\|_\infty^2 E(T^2) \\
E\left(\int_0^T \Delta_\Psi(\gamma_0(W, u)) du\right)^2 &\leq E\left(\int_0^T |\Psi(\gamma_0(W, u)) - \Psi_0(\gamma_0(W, u))| du\right)^2 \leq \|\Psi - \Psi_0\|_\ominus^2 E(T^2)
\end{aligned}$$

Then

$$\begin{aligned}
\sup_{\|H-H_0\|_{L_2} \leq C\varepsilon_n, \|\theta-\theta_0\|_\infty \leq \delta} |e_n(H, \theta)| &\leq E |H(W) - H_0(W)| |I(\theta_0, W) - I(\theta, W)| \\
&\leq \sup_W |I(\theta_0, W) - I(\theta, W)| (E |H(W) - H_0(W)|^2)^{1/2} \\
&\lesssim \|H - H_0\|_{L_2} (\|\phi - \phi_0\|_\infty + \|f - f_0\|_\infty + \|\Psi - \Psi_0\|_\infty) \\
&\lesssim C\delta\varepsilon_n
\end{aligned}$$

The condition is now verified.

Condition 6 is a stochastic equicontinuity condition. To verify this condition, I derive an upper bound for the bracketing entropy number of the following class of functions:

$$\mathcal{M}_{n,\delta,\varepsilon} = \{m_{\theta,H} - m_{\theta_0,H} : H \in \mathcal{H}_n(C), \theta \in \Theta_n, \|H - H_0\|_{L_2} \leq \varepsilon_n, \|\theta - \theta_0\|_\infty \leq \delta\}$$

After which I apply Corollary 19.35 in van der Vaart (2000).

First, given that $I(\theta, \cdot)$ is Lipschitz continuous in θ , it is possible to show that the envelope function of $\mathcal{M}_{n,\delta,\varepsilon}$ is bounded in δ :

$$\begin{aligned}
\sup_{(H,\theta) \in \mathcal{H}_n(C) \times \Theta_n} \|m_{\theta,H} - m_{\theta_0,H}\|_\infty &= \sup_{(H,\theta) \in \mathcal{H}_n(C) \times \Theta_n} \sup_W E |(H(W) - I(\theta_0, W))^2 - (H(W) - I(\theta, W))^2| \\
&\lesssim \sup_{(H,\theta) \in \mathcal{H}_n(C) \times \Theta_n} \sup_W E |H(W)| |I(\theta, W) - I(\theta_0, W)| \\
&\quad + \sup_{(H,\theta) \in \mathcal{H}_n(C) \times \Theta_n} \sup_W E |I(\theta, W) - I(\theta_0, W)| |I(\theta, W) + I(\theta_0, W)| \\
&\lesssim \delta
\end{aligned} \tag{52}$$

Second, the bracketing entropy of $\mathcal{M}_{n,\delta,\varepsilon}$ can be bounded by

$$\log N_{[]}(\epsilon, \mathcal{M}_{n,\delta,\varepsilon}, \|\cdot\|_\infty) \lesssim \log \left(1 + \frac{C}{\epsilon} \right) \max \{J_\phi(n), J_f(n), J_\Psi(n)\} \tag{53}$$

where C is a constant that depends on b_ϕ , b_f , and b_Ψ (but not on ϵ). This is shown as follows. Each of the sieve spaces, Φ_n , $\mathcal{F}_n$, and Θ_n , is a linear combination of finite basis functions that are bounded

from above by b_ϕ , b_f , and b_Ψ , respectively. Then by Corollary 2.6 in van de Geer (2000):

$$\begin{aligned}\log N(\epsilon, \Phi_n, \|\cdot\|_\infty) &\lesssim J_\phi(n) \log\left(1 + \frac{4b_\phi}{\epsilon}\right) \\ \log N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) &\lesssim J_f(n) \log\left(1 + \frac{4b_f}{\epsilon}\right) \\ \log N(\epsilon, \Theta_n, \|\cdot\|_\Theta) &\lesssim J_\Psi(n) \log\left(1 + \frac{4b_\Psi}{\epsilon}\right)\end{aligned}$$

Since $\mathcal{M}_{n,\delta,\varepsilon}$ is a space of functions satisfying

$$\|m_{\theta,H} - m_{\theta_0,H}\|_\infty \lesssim \|\phi - \phi_0\|_\infty + \|f - f_0\|_\infty + \|\Psi - \Psi_0\|_\Theta$$

Theorem 2.7.11 in van der Vaart and Wellner (1996) obtains that:

$$\log N_{[]}(\epsilon, \mathcal{M}_{n,\delta,\varepsilon}, \|\cdot\|_\infty) \leq \log N\left(\frac{\epsilon}{3}, \mathcal{F}_n, \|\cdot\|_\infty\right) + \log N\left(\frac{\epsilon}{3}, \Phi_n, \|\cdot\|_\infty\right) + \log N\left(\frac{\epsilon}{3}, \Theta_n, \|\cdot\|_\Theta\right)$$

Combining the results on entropy bounds above obtains the result in (53).

Let M_δ be the envelope function of $\mathcal{M}_{n,\delta,\varepsilon}$ where it was shown in (52) that $M_\delta \leq \delta$. Let the bracketing integral entropy be defined as:

$$J_{[]} (M_\delta, \mathcal{M}_{n,\delta,\varepsilon}, \|\cdot\|_\infty) = \int_0^{M_\delta} \sqrt{\log N_{[]}(\epsilon, \mathcal{M}_{n,\delta,\varepsilon}, \|\cdot\|_\infty)} d\epsilon$$

Letting $J(n) \equiv \max\{J_\phi(n), J_f(n), J_\Psi(n)\}$, the bracketing entropy integral of $\mathcal{M}_{n,\delta,\varepsilon}$ is bounded by

$$\begin{aligned}J_{[]} (M_\delta, \mathcal{M}_{n,\delta,\varepsilon}, \|\cdot\|_\infty) &\leq \sqrt{J(n)} \int_0^\delta \sqrt{\log\left(1 + \frac{C}{\epsilon}\right)} d\epsilon \\ &\lesssim \delta \sqrt{J(n)}\end{aligned}$$

Then by Corollary 19.35 in van der Vaart (2000), we can take

$$\varphi_n(\delta) = \delta \sqrt{J(n)}$$

Condition 7 holds by Assumption C2 and Theorem 2.

To check Condition 8 notice that since $\varphi_n(\delta) = \delta\sqrt{J(n)}$, the rate of convergence δ_n is found by solving

$$\delta_n\sqrt{J(n)} \leq \sqrt{n}\delta_n^2$$

which obtains that $\delta_n = O\left(\sqrt{\frac{J(n)}{n}}\right)$.

References

- Aalen, O.O., Borgan, O., and Gjessing, H.K. (2008) *Survival and Event History Analysis: A Process Point of View*. Springer.
- Aalen, O. and Hjört, N.L. (2002) Frailty Models that Yield Proportional Hazards. *Statistics and Probability Letters*, 58: 335-342.
- Abbring, J.H. (2012) Mixed Hitting-Time Models. *Econometrica*, 80: 783-819.
- Abbring, J. and van den Berg, G. (2003) The Identifiability of the Mixed Proportional Hazards Competing Risks Model. *J. R. Statist. Soc. B*, 65, Part 3: 701-710.
- Abbring, J.H. and G. Ridder. Regular Variation and the Identification of the Generalized Accelerated Failure-Time Models. *Econometric Theory*, 31: 1229-1248.
- Adams, R. A. and J. J. Fournier (2003) *Sobolev spaces*. Vol. 140, Academic Press, 2nd ed.
- Applebaum, D. (2009) *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2nd ed.
- Ariga, K., Ohkusa, Y., and Brunello, G. (1999) Fast Track: Is It in the Genes? The Promotion Policy of a Large Japanese Firm. *Journal of Economic Behavior and Organization*, 38(1): 385-402.
- Barlow, R. E. and F. Proschan (1965). *Mathematical Theory of Reliability*. J. Wiley & Sons. Reprinted (1996) SIAM, Philadelphia, PA.
- Bertoin, J. (1996) *Lévy Processes*. Cambridge University Press.

- Chen, X. (2007) Large Sample Sieve Estimation of Semi-Nonparametric Models. In Heckman, J.J. and Leamer, E.E. (eds), *Handbook of Econometrics*, Vol. 6, Ch. 76, 5550-5632.
- Chen, X. and D. Pouzo (2012) Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals. *Econometrica*, 80: 277-321.
- Chen, L., Lee, S. and Sung, M.J. (2014) Maximum Score Estimation with Nonparametrically Generated Regressors. *The Econometrics Journal*, 17(3): 271-300.
- Çinlar, E. (1979) On Increasing Continuous Processes. *Stochastic Processes and Their Applications*, 9: 147-154.
- Clemens, A. (2012) Position-Specific Promotion Rates and the “Fast Track” Effect. In Polachek, S.W. and Tatsiramos, K., *Research in Labour Economics*, Vol. 36., 77-108.
- Ding, Y. and Nan, B. (2011) A Sieve M-Theorem for Bundled Parameters in Semiparametric Models, with Application to the Efficient Estimation in a Linear Model for Censored Data. *The Annals of Statistics*, 39(6): 3032-3061.
- Dykstra, R., and Laud, P. (1981) Bayesian Nonparametric Approach toward Reliability. *Annals of Statistics*, 9: 356–367.
- Elbers, C. and Ridder, G. (1982) True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. *Review of Economic Studies*, 49(3): 403-409.
- Fox, J., Kim, K., Ryan, S.P., and Bajari, P. (2012) The Random Coefficients Logit Model Is Identified. *Journal of Econometrics*, 166: 204-212.
- Gallant, A.R. and Nychka, D.W. (1987) Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, 55(2): 363-390.
- Gjessing, H.K., Aalen, O.O., and Hjort, N.L. (2003) Frailty Models Based on Lévy Processes. *Advances in Applied Probability*, 35: 532-550.

- Gnedin, A. and Pitman, J. (2008) Moments of Convex Distribution Functions and Completely Alternating Sequences. In Nolan, D. and Speed, T. (eds), *Probability and Statistics: Essays in Honor of David A. Freedman*, Vol. 2, 30-41.
- Gripenberg, G., Londen, S.O., and Staffans, O. (1990). *Volterra Integral and Functional Equations*. Cambridge University Press.
- Hahn, J., Liao, Z., and Ridder, G. (2018) Nonparametric Two-Step Sieve M Estimation and Inference. *Econometric Theory*, doi:10.1017/S0266466618000014.
- Heckman, J.J. and Singer, B. (1984) Econometric Duration Analysis. *Journal of Econometrics*, 24(1-2): 63-132.
- Hopenhayn, H.A. and Squintani, F. (2011) Preemption Games with Private Information. *Review of Economic Studies*, 78(2): 667-692.
- Horowitz, J.L. and S. Lee (2017) Nonparametric Estimation and Inference Under Shape Restrictions. *Journal of Econometrics*, 201(1): 108-126.
- Hu, Y. and Shiu, J-L. (2017) Nonparametric Identification Using Instrumental Variables: Sufficient Conditions for Completeness. *Econometric Theory*, doi:10.1017/S0266466617000251.
- Hu, Y., Schennach, S., and Shiu, J-L. (2017) Injectivity of a Class of Integral Operators with Compactly Supported Kernels. *Journal of Econometrics*, 200: 48-58.
- Ishida, H., Su, K. H., and Spilerman, S. (2002) Models of Career Advancement in Organizations. *European Sociological Review*, 18(2): 179-198.
- Khan, L. and Lange, F. (2014) Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures. *Review of Economic Studies*, 81(4): 1575-1613.
- Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Inc., New York, 2nd edition.
- Kebir, Y. (1991) On Hazard Rate Processes. *Naval Research Logistics*, 6(38): 865-876.

- Kortram, R.A., A.C.M. Rooij, A.J. Lenstra, and G. Ridder (1995) Constructive Identification of the Mixed Proportional Hazards Model. *Statistica Neerlandica*, 49(3): 269-281.
- Koumandos, S. (2014) On Completely Monotonic and Related Functions. In Pardalos, P.M. and Rassias, T.M. (eds), *Mathematics Without Boundaries: Surveys in Pure Mathematics*. Springer.
- Kristensen, D. (2009) Semiparametric Modelling and Estimation: A Selective Overview. CREATES Research Paper No. 2009-44.
- Lin, G.D. (2017). Recent Developments on the Moment Problem. *Journal of Statistical Distributions and Applications*, 4:5, DOI 10.1186/s40488-017-0059-2.
- Liu, R. (2016) A Single-Index Cox Model Driven by Lévy Subordinators. Working paper.
- Lorentz, G.G. (1986). *Bernstein Polynomials*. AMS Chelsea Publishing.
- Lukács, E. (1955) Applications of Faà di Bruno's Formula in Mathematical Statistics. *American Mathematical Monthly*, 62: 340-48.
- Newey, W.K. (1997) Convergence Rates and Asymptotic Normality for Series Estimators. *Journal of Econometrics*, 79: 147-168.
- Newey, W.K. and D. McFadden. Large Sample Estimation and Hypothesis Testing. In *Handbook of Econometrics*. Vol. 4. Edited by Daniel McFadden and Robert Engle. Amsterdam, The Netherlands: Elsevier, North-Holland, 1999, chapter 36, pp. 2113-2245.
- Rajput, B.S. and J. Rosinski (1989) Spectral Representations of Infinitely Divisible Processes. *Probability Theory Related Fields*, 82(3): 451-448.
- Rejto, R. and M. Taboada (1992) Unique Solvability of Nonlinear Volterra Equations in Weighted Spaces. *Journal of Mathematical Analysis and Applications*, 162(2): 368-381.
- Shen, X. and Wong, W.H. (1994) Convergence Rate of Sieve Estimates. *The Annals of Statistics*, 22(2):580-615.
- 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.
- 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.
- van de Geer, S. (2000) *Empirical Processes in M-Estimation*. Cambridge University Press.
- Van den Berg, G.J. (2001) Duration Models: Specification, Identification, and Multiple Durations. In J.J. Heckman and E. Leamer (editors), *Handbook of Econometrics*, Volume V, North Holland, Amsterdam.
- van der Vaart, A. (1994) Weak Convergence of Smoothed Empirical Processes. *Scandinavian Journal of Statistics*, 21(4): 501-504.
- van der Vaart, A.W. (2000) *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.
- van der Vaart, A.W., and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes: with Applications to Statistics*. New York: Springer-Verlag.
- van der Vaart, A., and Wellner J. A. (2011) A Local Maximal Inequality under Uniform Entropy. *The Electronic Journal of Statistics*, 5: 192-203.
- Waldman, M. (2013) Theory and Evidence in Internal Labor Markets. In Gibbons, R. and Roberts, J. (editors), *The Handbook of Organizational Economics*, 520-574.
- Wang, J. and Ghosh, S.K. (2012) Shape Restricted Nonparametric Regression with Bernstein Polynomials. *Computational Statistics and Data Analysis*, 56: 2729-2741.
- Widder, D.V. (2011) *The Laplace Transform*. Dover Publications.
- Yashin, A., and Arjas, E. (1988) A Note On Random Intensities and Conditional Survival Functions. *Journal of Applied Probability*, 25: 630 - 635.

Zhou, Q., Hu, T., and Sun, J. (2017) A Sieve Semiparametric Maximum Likelihood Approach for Regression Analysis of Bivariate Interval-Censored Failure Time Data. *Journal of the American Statistical Association*, 112(518): 664-672. DOI: 10.1080/01621459.2016.1158113.