Quantitative Economics 11 (2020), 535–577 1759-7331/20200535

A competing risks model with time-varying heterogeneity and simultaneous failure

Ruixuan Liu
Department of Economics, Emory University

This paper proposes a new bivariate competing risks model in which both durations are the first passage times of dependent Lévy subordinators with exponential thresholds and multiplicative covariates effects. Our specification extends the mixed proportional hazards model, as it allows for the time-varying heterogeneity represented by the unobservable Lévy processes and it generates the simultaneous termination of both durations with positive probability. We obtain nonparametric identification of all model primitives given competing risks data. A flexible semiparametric estimation procedure is provided and illustrated through the analysis of a real dataset.

Keywords. Duration analysis, competing risks, first passage times, nonparametric identification.

JEL classification. C14, C34, C41.

1. Introduction

Threshold-crossing models for analyzing durations or time-to-events have recently attracted considerable interest. Instead of directly specifying the hazard rate function, the basic setup in this paper is to view the duration as the first passage time of a latent stochastic process crossing a random threshold (de Paula (2009), Abbring (2012)). The main advantages of such a threshold-crossing model can be summarized as follows. First, it offers a transparent view of the underlying failure/survival mechanism, as the duration of interest typically represents the termination of certain process reaching a critical level (Lancaster (1972), Aalen, Borgan, and Gjessing (2008)). Second, it captures the essential feature of optimal stopping time problems in which the generic solutions are characterized exactly by such threshold-crossing rules (Heckman and Navarro (2007), Honoré and de Paula (2010), Abbring (2012)). Last but not least, the model generalizes the mixed proportional hazards model (Lancaster (1979)), as the latter can be seen as a special case with particular assumptions about the latent process and threshold (Honoré and de Paula (2018)). Along with a variety of promising aspects, it remains challenging to model multiple durations defined via threshold-crossing rules involving continuous-time latent stochastic processes.

Ruixuan Liu: ruixuan.liu@emory.edu
I am deeply indebted to my thesis advisor Yanqin Fan for her continuing guidance and support. I am grateful to Jaap Abbring, Irene Botosaru, Emmanuel Guerre, Roger Klein, Qi Li, Zhongjian Lin, Essie Maasoumi, Elie Tamer, and Jon Wellner for many thoughtful discussions. I want to thank three anonymous referees for their insightful comments that have greatly improved the paper. The usual disclaimer applies.

 https://doi.org/10.3982/QE1159

In this paper, we construct a bivariate threshold-crossing durations model in which both durations T_1 and T_2 are the first passage times driven by bivariate latent stochastic processes. Conditional on observable covariates, the model contains both static and time-varying heterogeneity terms. The static heterogeneity ν is a positive random variable that changes the time deformation, which turns out to be equivalent to the frailty term (van den Berg (2001)) in the mixed proportional hazards model. Considering the time-varying heterogeneity represented by the unobservable stochastic processes, throughout this paper we focus on Lévy subordinators $\mathbf{L}(t) \equiv (L_1(t), L_2(t))$, that is, continuous-time processes with stationary, independent, and nonnegative increments (Sato (2013)). The Lévy subordinator belongs to a subclass of the general Lévy process and its nonnegative increment or, equivalently, its nondecreasing sample path encapsulates the notion that the mechanism leading to the failure/exit is irreversible and always accumulates in one direction. Meanwhile, the family of Lévy subordinators nests commonly used (compound) Poisson, gamma, and stable processes as special cases. Furthermore, the dependence of two marginal Lévy subordinators $L_1(t)$ and $L_2(t)$ is completely characterized by the Lévy copula (Kallsen and Tankov (2006)).

We show that the conditional joint survivor function of duration variables admits a closed-form expression when the thresholds are independent draws from the unit exponential distribution and the covariates enter the model in a multiplicative way. The specification could be seen as a natural variant of the bivariate mixed proportional hazards model in the sense that it is boils down to the proportional hazards model with a shared frailty (Oakes (1989)) in the absence of Lévy processes. By explicitly incorporating the latent stochastic processes, our model embeds time-varying heterogeneity and generates the simultaneous termination of both durations, because two dependent Lévy processes can simultaneously jump over the thresholds. Interestingly, the conditional survivor function of the two durations extends the bivariate exponential distribution from Marshall and Olkin (1967) and the conditional copula function falls into the general framework laid down by Marshall and Olkin (1988). Compared with the original and existing generalizations of the Marshall–Olkin-type model, durations in our model are defined as the first passage times of latent processes and the concurrent termination of both durations is not necessarily due to some independent and common shock, unlike the constructions in Marshall and Olkin (1967) and An, Christensen, and Gupta (2004). Thus, we name the present model the extended Marshall–Olkin model.[1]

The empirical content of the extended Marshall–Olkin model is studied with competing risks data where only the minimum duration or failure time and the cause of failure are observed. For the two durations (T_1, T_2) of interest, economists or econometricians only observe $V = \min(T_1, T_2)$, with the indicator D recording the failure type. The application of competing risks models in economics goes back to Flinn and Heckman (1982) and the delicate identifiability issues of such models continue to attract attention from researchers, including Heckman and Honoré (1989, 1990), Abbring and van den Berg (2003), Honoré and Lleras-Muney (2006), Khan and Tamer (2009), Lee and Lewbel (2013), and Fan and Liu (2018), to name just a few. Our work sheds new light on the

[1] Honoré and de Paula (2010) showed that the Marshall–Olkin model is observationally equivalent to a stopping game model with additive interaction effects; see Section 4.2 in Honoré and de Paula (2010).

modeling framework concerning competing risks in the sense that the aforementioned works typically regard one individual as having different ways to exit a given state; for example, an unemployed person can leave unemployment by finding a job or going out of the labor force (Flinn and Heckman (1982)). These applications do not involve simultaneous failures at all. In this paper, however, the competing risks arise as two different agents being exposed to the risk of exiting in one state; they can either do it at the same time or not.[2] The main theoretical contribution here is combining insights from the literature on the identification of duration models with Lévy processes theory to obtain the sampling information and nonparametric identification results for this new competing risks model. Because the conditional survivor copula function of two durations varies with covariates and it has a singular component, existing results do not apply. Instead, we make use of the analytical properties of Lévy subordinators and the special Marshall–Olkin structure to achieve nonparametric identification. When the covariates effects are parameterized, as in the standard Cox regression, we present a straightforward semiparametric estimation procedure following the identification strategy closely and we offer an empirical illustration using data from Honoré and de Paula (2018).

The construction of our model was motivated by empirical studies on the joint retirement decisions of married couples with both spouses working at time zero. It is well documented for various datasets that a significant portion of married couples choose to retire at the same time (An, Christensen, and Gupta (2004), Honoré and de Paula (2018)). We reexamine the joint retirement problem of married couples using data drawn from eight waves of the Health and Retirement Study. Here, the duration variables of interest (T_1, T_2) are the retirement dates of the wife and husband. The duration V represents the first entry into retirement for the corresponding member of the household, and $D = 1, 2$, or 3 depending on whether the wife retires first, the husband first, or they retire simultaneously. Our methodology offers a continuous-time model in which wives and husbands are timing their retirement dates subject to health shocks that arrive randomly. One special case of our setup is employing a compound Poisson process in which the arrival rate is governed by $\nu\lambda_0(t)$ and the magnitudes of shocks are represented by $(L_1(t), L_2(t))$. More generally, we allow $(L_1(t), L_2(t))$ to be Lévy subordinators without any parametric assumption. Various observable household characteristics such as wealth and health expenditures can be included to account for the covariates effects $\phi_1(x)$ and $\phi_2(x)$ for wives and husbands, respectively. In particular, we assume that $\phi_k(x)$ acts multiplicatively on the corresponding shock process $L_k(t)$, which implies that an increase (decrease) of $\phi_k(x)$ amplifies (dampens) the magnitude of shocks and leads to earlier (later) exit for individual k, with $k = 1$ or 2. An important feature of the model is the provision for correlated health shocks among the couple and, as a by-produce, the explicit allowance for simultaneous retirement. This property is also expected to be useful in other economic applications, including patent racing (Reinganum (1981, 1982), Choi (1991)), technology adoption (Jensen (1982), Farzin, Huisman, and Kort (1998)), and the smoking cessation of married couples (Abbring and Yu (2015)), to name a few.

[2] I want to thank a knowledgeable referee for this clarifying point.

Compared with the voluminous work on the mixed proportional hazards models (van den Berg (2001)), the literature on structural threshold-crossing models driven by unobservable stochastic processes is relatively scarce. An ancestor in econometrics is the strike model by Lancaster (1972), in which a Brownian motion is employed to represent the level of disagreement. Abbring (2012) obtained the nonparametric identification results of more general threshold-crossing duration models, in which the duration is the first passage time of a spectral negative Lévy process crossing a heterogeneous threshold. Our paper complements Abbring (2012) and the existing literature in several ways. The increasing sample path of Lévy subordinators is more suitable for representing the accumulation of knowledge in patent-race game (Reinganum (1981)), the improving profitability in the technology adoption process (Farzin, Huisman, and Kort (1998)), the underlying aging process in joint retirement decisions (Coile (2004)), and the degradation of health condition due to smoking (Abbring and Yu (2015)), as opposed to the spectrally negative Lévy process in Abbring (2012), which fluctuates up and down. Another notable feature of our model is that its empirical content is studied by examining the conditional (sub)survivor functions directly, whereas the structural parameters in Abbring (2012) are expressed through the conditional Laplace transform; see equation (6) on page 797 of Abbring (2012). As such, it holds great convenience in facilitating the estimation of our model when the durations are subject to censoring. To extend the univariate duration model driven by a single Lévy subordinator (Gjessing, Aalen, and Hjort (2003), Botosaru (2016)), we need to consider the dependence structure of Lévy subordinators and how it translates into the dependence of structural durations. For this purpose, we rely on the Lévy copula (Cont and Tankov (2004), Kallsen and Tankov (2006)) in constructing our model and we show this dependence measure can be identified based on competing risks data only, which is an advantage, as (V, D, X) corresponds to the coarsened data of (T_1, T_2, X) where both durations are observed.

Apart from the univariate stopping time model of Abbring (2012), we would like to mention de Paula (2009) and Alvarez, Borovickova, and Shimer (2016), who model multiple durations as first passage times. A simultaneous duration model is proposed by de Paula (2009) with multiple decision makers to exit a given state. The problem of determining the existence and uniqueness of equilibrium stopping strategies is analyzed in de Paula (2009), with state variables evolving as general diffusion processes. Interestingly, de Paula (2009) showed that the social interaction effect in this stopping game is sufficient and necessary for the simultaneous failure. Alvarez, Borovickova, and Shimer (2016) proposed an empirical search model where the duration of each worker's unemployment spell is generated when a Brownian motion hits a barrier. In the absence of heterogeneity, unemployment duration follows an inverse Gaussian distribution (Lancaster (1972)). Alvarez, Borovickova, and Shimer (2016) allowed for arbitrary heterogeneity across workers in the parameters of this inverse Gaussian distribution and prove that the distribution of these parameters is identified based on multiple spells per worker. The identification results in de Paula (2009) and Alvarez, Borovickova, and Shimer (2016) are not for competing risks data. The only existing competing risks model that involves a latent continuous-time process is found in Ryu (1993). Ryu (1993) considered two competing forces with one representing the natural failure mechanism caused by a compound Poisson process and the other one from economic agent's endogenous optimization behavior. There is no formal identification result for such a complicated structural model.

The rest of this paper is organized as follows. In Section 2, we construct the extended Marshall–Olkin model for bivariate durations and discuss its probabilistic features. Section 3 presents its empirical content under competing risks and suggests a straightforward semiparametric estimation procedure when the covariates effects are parameterized. We consider a further generalization in Section 4 that incorporates additional observable time-varying covariates. The empirical application in Section 5 illustrates how the semiparametric method can be used to obtain meaningful estimates of behavioral effects with minimal parametric restrictions on the primitives of the model. We conclude in Section 6. Appendix A collects the proofs of main theorems and Appendix B contains some auxiliary results and more technical proofs.

2. The extended Marshall–Olkin model

Because both duration variables are modeled as the first passage times of Lévy subordinators, it is necessary to introduce the definition and characterization of a bivariate Lévy subordinator $\mathbf{L}(t) = (L_1(t), L_2(t))$ with marginal subordinators denoted by $L_1(t)$ and $L_2(t)$, respectively (Sato (2013)).

Definition 2.1. A bivariate Lévy subordinator $\mathbf{L}(t)$ is a right-continuous stochastic process with left limits such that for every t and $r \geq 0$, the increment $\mathbf{L}(t+r) - \mathbf{L}(t)$ is independent of $\{\mathbf{L}(s); 0 \leq s \leq t\}$ and has the same distribution as $\mathbf{L}(r)$. Furthermore, almost surely it has nondecreasing sample path, that is, one has $\mathbf{L}(t) \geq \mathbf{L}(s)$ for $t \geq s$.

The class of Lévy subordinators is flexible enough to incorporate a variety of interesting parametric precesses and it also admits an elegant characterization according to the well-known Lévy–Khintchine representation. For $z_1, z_2 \in \mathcal{R}^+$, the joint Laplace transform can be expressed as

$$E\{\exp[-z_1 L_1(t) - z_2 L_2(t)]\} = \exp[-t\Phi_{12}(z_1, z_2)], \tag{2.1}$$

where $\Phi_{12}(\cdot,\cdot)$ is known as the joint Lévy–Laplace or Lévy exponent function (Cont and Tankov (2004)). The bivariate function $\Phi_{12}(\cdot,\cdot)$ is further determined by the underlying jump measure or Lévy measure $\Pi_{12}(\cdot,\cdot)$ via the following integral transform:

$$\Phi_{12}(z_1, z_2) = \iint_{\mathcal{R}_+^2} [1 - e^{-y_1 z_1 - y_2 z_2}]\, d\Pi_{12}(dy_1, dy_2). \tag{2.2}$$

When it comes to each marginal Lévy subordinator $L_k(t)$, we have

$$E\{\exp[-z_k L_k(t)]\} = \exp[-t\Phi_k(z_k)], \tag{2.3}$$

where the marginal Lévy exponent $\Phi_k(\cdot)$ is expressible as

$$\Phi_k(z_k) = \int_{\mathcal{R}_+} [1 - e^{-z_k y}]\Pi_k(dy), \tag{2.4}$$

with the marginal Lévy measure $\Pi_k(\cdot)$ for $k=1,2$. Also, the Lévy density function is denoted by $\pi_k(\cdot)$ when $\Pi_k(\cdot)$ is absolutely continuous with respect to the Lebesgue measure for $k=1,2$.

The dependence of two Lévy subordinators is determined by the Lévy copula; see Cont and Tankov (2004) and Kallsen and Tankov (2006). Note that the Lévy copula is not a standard copula (Nelsen (2006)) that acts on distribution or survivor functions; rather, it operates on Lévy measures which determine the jump intensity. Nevertheless, it does share certain similarities with the standard copula function such as linking the joint and marginal Lévy measures. For most Lévy subordinators, the jump intensity diverges to infinity as the jump size approaches zero, that is, $\lim_{u\to 0}\pi_k(u)=\infty$ for $k=1,2$. Hence, it is the tail integral or *survivor functional version* of a Lévy measure that is more tractable. The Lévy copula $\mathcal{C}_L(\cdot,\cdot)$ is defined as the unique mapping that couples two marginal Lévy tail integrals $\overline{\Pi}_k(y)=\int_y^\infty \pi_k(y)\,dy$ to generate the joint tail integral:

$$\overline{\Pi}_{12}(u_1,u_2)=\mathcal{C}_L\big(\overline{\Pi}_1(u_1),\overline{\Pi}_2(u_2)\big), \tag{2.5}$$

where $\overline{\Pi}_{12}(u_1,u_2)$ is the two-dimensional tail integral (see Appendix B for the precise definition of the two-dimensional tail integral). The role of the Lévy copula is more transparent in pinning down the joint Laplace exponent function given two marginal Lévy densities π_1 and π_2, according to the following identity from Cont and Tankov (2004):

$$\Phi_{12}(z_1,z_2)=\iint\left[1-e^{-z_1y_1-z_2y_2}\right]\frac{\partial^2}{\partial u_1\,\partial u_2}\mathcal{C}_L(\cdot,\cdot)\Big|_{\substack{u_1=\overline{\Pi}_1(y_1)\\ u_2=\overline{\Pi}_2(y_2)}}\pi_1(y_1)\pi_2(y_2)\,dy_1\,dy_2. \tag{2.6}$$

In sum, a pair of Lévy subordinators is entirely determined by triple *time-invariant* measures, including two marginal Lévy measures (Π_1,Π_2) and the Lévy copula $\mathcal{C}_L$. This time-invariant feature forms the basis of the identifiability of latent processes by only observing the corresponding first passage times.

2.1 *The model setup*

The duration often records the end of an event, in which the underlying process leading to this event may not be directly observable. Our econometric model consists of two agents where each individual k makes an optimal timing decision, which generates the duration T_k to transit out of the current state with $k=1,2$. Let the nonnegative pay-off or utility flow of switching to the alternative state be $L_k(t)$ at time t and let the lump-sum cost be ϵ_k associated with this transition for $k=1,2$. Both T_1 and T_2 are first passage times of the underlying stochastic processes $(L_1(t),L_2(t))$, as motivated by the generic solutions in optimal stopping time problems (Honoré and de Paula (2010), Abbring (2012)). We also introduce the observed covariates X and an unobserved heterogeneity ν. Formally, each duration T_k is defined as the first passage time of a latent process $L_k(t)$ crossing a random threshold ϵ_k:

$$T_k\equiv\inf\big\{t_k:\phi_k(X)L_k\big(\nu\Lambda_0(t_k)\big)\geq\epsilon_k\big\}\quad\text{for } k=1,2, \tag{2.7}$$

where the nonnegative functions $\phi_1(\cdot)$, $\phi_2(\cdot)$ represent the effects of covariates and the monotone time deformation $\Lambda_0(t)$ reveals how time unfolds (Stock (1988)). Considering the way that covariates influence the underlying failure/survival mechanism, an increase (decrease) of $\phi_k(x)$ amplifies (dampens) the jumping magnitude of $L_k(\cdot)$, which induces earlier (later) exiting, *ceteris paribus*. The static heterogeneity ν changes the magnitude of $\Lambda_0(t)$ and is introduced in the same spirit as Feller (1943), who first used the mixed Poisson process to demonstrate the distinction between genuine and spurious duration dependence (Elbers and Ridder (1982), Heckman and Singer (1984a)). For numerous economic models, it is routine to employ a compound Poisson process (which is indeed a Lévy subordinator) to model certain monotone pay-off process like $L_k(t) = \sum_{i=1}^{N(t)} \xi_{i,k}$, where the arrival of *shocks* is governed by a common homogeneous Poisson process $N(t)$ and (possibly) correlated shocks $(\xi_{i,1}, \xi_{i,2})$ are assumed to be nonnegative for $k = 1, 2$. Our specification offers flexibility regarding the speed or arrival rate by a random deformation $\nu\Lambda_0(t)$ so that one has the composition $(L_1(\nu\Lambda_0(t)), L_2(\nu\Lambda_0(t)))$. Note that we restrict the speed $\nu\Lambda_0(t)$ for which the shocks arrive to be the same for both individuals, whereas the magnitude $(\xi_{i,1}, \xi_{i,2})$ can be different, though still correlated.

Considering the rich structure involving various modeling components, it is desirable to strive for a generality that encompasses commonly encountered situations, while making necessary assumptions for tractability. We make the following model assumptions throughout the paper.

Assumption 2.1. Random variables $X, \nu, \mathbf{L}(\mathbf{t})$, and (ϵ_1, ϵ_2) are mutually independent.

Assumption 2.2. The static heterogeneity ν is nonnegative with its distribution function $F_\nu(\cdot)$ and Laplace transform $\psi(\cdot)$.

Assumption 2.3. The stochastic process $\mathbf{L}(\mathbf{t}) = (L_1(t), L_2(t))$ is a bivariate Lévy subordinator.

Assumption 2.4. The thresholds ϵ_1 and ϵ_2 are two independent random variables drawn from the unit exponential distribution.

Assumption 2.5. The support of covariates X is $\mathcal{X}$, which is an open set in $\mathcal{R}^d$. The function $\phi_k(x)$ is nonnegative and measurable for $k = 1, 2$.

Assumption 2.6. The deformation function $\Lambda_0(t)$ is continuously differentiable with a nonnegative derivative function $\lambda_0(t)$.

To elaborate on the different roles played by the static heterogeneity ν and the time-varying heterogeneity $L(t)$, we consider a univariate compound Poisson process[3] as $L(t) = \sum_{i=1}^{N(t)} \xi_i$, in which the arrival of shocks is governed by a standard Poisson process $N(t)$ and individual shocks $\{\xi_i\}$ are assumed to be nonnegative. In the composition of $L(\nu\Lambda_0(t))$, $\nu\Lambda_0(t)$ determines the speed or arrival rate of the Poisson process

[3]We refer readers to Marshall and Shaked (1979) for the multivariate compound Poisson process.

and $\xi_1, \xi_2, \dots$ represent the jump sizes. It is clear that the heterogeneity regarding the speed or arrival rate is determined by ν at time 0, whereas the sequence of shocks $\xi_1, \xi_2, \dots, \xi_{N(t)}$ is generated as time unfolds, demonstrating that the latter source of heterogeneity indeed varies over time.

The independence and distributional assumption on thresholds may seem stringent at first, but it is motivated by recasting the mixed proportional hazards model into the current framework where the duration variable is viewed as certain first passage time. For a generic duration T modeled by the mixed proportional hazards model with covariates X and heterogeneity ν, it is straightforward to see that T can also be characterized by

$$T \equiv \inf\{t : \phi(X)\nu\Lambda_0(t) \geq \epsilon\}, \tag{2.8}$$

where the threshold ϵ is drawn from the unit exponential distribution and the underlying trend $\Lambda_0(t)$ is the cumulative baseline hazard function (Ridder (1990)). In other words, given that the cumulative hazard function is deterministic and the randomness comes from the draw that is unobserved by the econometrician, the duration outcome (transition out of a given state) is observed only when this deterministic function hits the threshold.[4] Regarding the bivariate mixed proportional hazards model, it models each marginal duration separately while requiring the independence assumption on exponential thresholds and the dependence is generated from static heterogeneity (Honoré and de Paula (2018)). This type of construction rules out simultaneous failure from the very beginning, that is, $P\{T_1 = T_2|x\} = 0$ for any x. The target of Ridder (1990) is to relax the exponential mixture structure,[5] extending it to the generalized accelerated failure time (GAFT) models. Here, we focus on exponential thresholds and explore another path by incorporating time-varying heterogeneity. Such an extension is desirable for empirical applications because given the realization of ν, one only has a deterministic trend $\Lambda_0(t)$ approaching the exponential threshold from below in the mixed proportional hazards model. The individual heterogeneity ν does not evolve over time along the entire duration.

The importance of allowing for time-varying heterogeneity and simultaneous failure is further exemplified via the following discussion of motivating applications and the interpretation of Lévy subordinators for use. Obviously, our model is applicable to the traditional reliability study in which the equipment is subject to random shocks that can knock out one or more components at the same time; see Ryu (1993) or Chapter 16 of Crowder (2012).

Example 2.1. In a sequence of research, Reinganum (1981, 1982) studied the patent-race game in which the durations of interest are the times to develop new products for firms engaging in Research and Development (R&D) activities. The time-varying heterogeneity is of central importance here, because the R&D progress is inherently stochastic.

[4] I want to thank one knowledgeable referee for this point.

[5] The exponential mixture structure has been utilized by Heckman and Singer (1984b) for deriving nonparametric maximum likelihood estimates, by Horowitz (1999) for obtaining deconvolution of the frailty term, and by Ridder and Woutersen (2003) for characterizing the nonsingular information bound.

The increasing sample path of Lévy subordinators signifies the accumulation of knowledge or research efforts in R&D. Our specification restricts that the speed or arrival rate $\nu\lambda_0(t)$ for which the innovation becomes available to be the same for both firms (Choi (1991)), whereas the extent to which each firm utilizes the innovation can be different, though still correlated, that is, $L_1(t)$ and $L_2(t)$ are dependent. In these theoretical R&D models (Reinganum (1981, 1982), Choi (1991)), it is routine to employ a Poisson process to model the innovation process for both firms, which is indeed a Lévy subordinator. The covariates X that act multiplicatively on the latent processes can be various measures of firm size and research investment.

Example 2.2. The preceding example concerns the supply (creation) of new technology; however, our model is also applicable to the demand (adoption) of new technology among firms. Once a new technology becomes available, firms consider the optimal timing to adopt it. In this context, the latent process encodes the expected profitability (or cost savings) of adoption (Jensen (1982)). The increasing sample path of Lévy subordinators $(L_1(t), L_2(t))$ reveals that the new technology is indeed more efficient than the current state of the art and the uncertainty regarding its profitability is reduced over time. Here, $\nu\lambda_0(t)$ represents the speed of the reduction of uncertainty, whereas the Lévy subordinator $(L_1(t), L_2(t))$ encodes the magnitude or degree of profit gains (Farzin, Huisman, and Kort (1998)). The covariates X include the consumers' response to the new product and the market structure.

Example 2.3. A significant portion of married couples choose to retire at the same time, as documented across different datasets (An, Christensen, and Gupta (2004), Honoré and de Paula (2018)). Regarding these joint retirement decisions, the underlying stochastic processes $(L_1(t), L_2(t))$ characterize the degradation of health conditions in the elderly. The identical time deformation $\nu\lambda_0(t)$ determines the arrival rate of health shocks and the jumping behavior of Lévy subordinators captures sudden health shocks such as heart attacks or new cancer diagnoses (An, Christensen, and Gupta (2004), Coile (2004)). The covariates X include health expenditures and family income or wealth information.

Example 2.4. Our model is also inspired by the stopping-time model in Abbring and Yu (2015), which determines couples' decisions to quit smoking. In this context, the duration variables are the time to stop smoking for wives and husbands and the covariates X collect the family or demographic characteristics of the couple. The latent Lévy processes $(L_1(t), L_2(t))$ represent the degradation of health conditions due to the smoking behavior and the corresponding jumps mimic shocks due to the diagnosis of lung cancer. Our model is also applicable to a related study by Drepper and Effraimidis (2015), where they examine the first time use of tabacoo/drug for siblings within a family.

2.2 *Dependence properties*

We present probabilistic features of our model in this subsection. We begin with some definitions that lead up to the statement of Theorem 2.1. First, define

$$\Psi_1(x) = \Phi_{12}\big(\phi_1(x), \phi_2(x)\big) - \Phi_2\big(\phi_2(x)\big), \tag{2.9}$$

$$\Psi_2(x)=\Phi_{12}\big(\phi_1(x),\phi_2(x)\big)-\Phi_1\big(\phi_1(x)\big), \tag{2.10}$$

$$\Psi_3(x)=\Phi_1\big(\phi_1(x)\big)+\Phi_2\big(\phi_2(x)\big)-\Phi_{12}\big(\phi_1(x),\phi_2(x)\big). \tag{2.11}$$

We prove in Lemma B.1 of Appendix B that $\Psi_j(x)\geq 0$ for all x and $j=1,2$, and 3. Moreover, let

$$\omega_1(x)=\frac{\Psi_3(x)}{\Psi_1(x)+\Psi_3(x)} \quad \text{and} \quad \omega_2(x)=\frac{\Psi_3(x)}{\Psi_2(x)+\Psi_3(x)}. \tag{2.12}$$

The following function belongs to the Marshall–Olkin copula family:

$$K(u_1,u_2|x)=u_1u_2\min\big(u_1^{-\omega_1(x)},u_2^{-\omega_2(x)}\big); \tag{2.13}$$

see Section 3.1.1 in Nelsen (2006). Let the conditional joint survivor function be

$$S_{T_1,T_2}(t_1,t_2|x)\equiv P\{T_1>t_1,T_2>t_2|X=x\}, \tag{2.14}$$

with its marginal $S_{T_k}(t|x)\equiv P\{T_k>t|X=x\}$ for $k=1,2$. The lower-case x stands for the realization of the random variable X. Also, denote the conditional survivor copula function (Nelsen (2006)) by $C(u_1,u_2|x)$. Recall that the Laplace transform of the static heterogeneity ν is $\psi(z)=\int e^{-zu}\,dF_\nu(u)$. Its derivative is denoted by $\psi'(z)$ accordingly. The explicit joint survivor function is stated by the following theorem.

Theorem 2.1. *If Assumptions 2.1 to 2.6 hold, then the conditional joint survivor function of (T_1,T_2) is*

$$S_{T_1,T_2}(t_1,t_2|x)=\psi\big\{\Lambda_0(t_1)\Psi_1(x)+\Lambda_0(t_2)\Psi_2(x)+\Lambda_0(t_1\vee t_2)\Psi_3(x)\big\}, \tag{2.15}$$

with the conditional marginal survivor function equal to

$$S_{T_k}(t|x)=\psi\big\{\Lambda_0(t)\Phi_k\big(\phi_k(x)\big)\big\} \quad \text{for } k=1,2. \tag{2.16}$$

The conditional survivor copula function is

$$C(u_1,u_2|x)=\psi\big\{-\log K\big(e^{-\psi^{-1}(u_1)},e^{-\psi^{-1}(u_2)}|x\big)\big\}. \tag{2.17}$$

Remark 2.1. The static and time-varying unobserved terms play different roles in determining the conditional joint survivor function. Specifically, the characteristics from the Lévy subordinators are directly absorbed on the covariate effect. This is reminiscent of McFadden (1974) for binary choice models where different distributions of the random error cause changes to the covariate effect.[6] Meanwhile, the Laplace transform of the static heterogeneity term ν becomes the outer link function coupling the covariate effect and time deformation $\Lambda_0(t)$. If we define the conditional cumulative hazard function as $\Theta_k(t|x)\equiv -\log S_{T_k}(t|x)$, it becomes

$$\Theta_k(t|x)=-\log\big[\psi\big\{\Lambda_0(t)\Phi_k\big(\phi_k(x)\big)\big\}\big],$$

[6] I am grateful to one referee for pointing out this connection.

for $k = 1, 2$. Now, it is obvious that the time deformation $\Lambda_0(t)$ is equivalent to the cumulative baseline hazard function and ν plays the same role as the frailty term in the mixed proportional hazards model (van den Berg (2001)).

The hazard rate plays a fundamental role in econometric duration analysis as both a theoretical and descriptive tool. It has interpretive content in the sense that the duration dependence can be encapsulated by the shape of the hazard rate function (Elbers and Ridder (1982), Heckman and Singer (1984a)). To illustrate the different roles of structural components, we plot conditional hazard rate functions in Figure 1:

$$\theta(t|x) = -\frac{\partial S(t|x)/\partial t}{S(t|x)}, \tag{2.18}$$

where $S(t|x)$ is the conditional survivor function, as in (2.16) based on various specifications of the time deformation, static heterogeneity, and Lévy processes. We include a scalar regressor with $\phi(x) = \exp(x)$ and depict $\theta(t|x)$ for $x = 0.5$, 1, and 2. In the baseline specification, we parameterize $\Lambda_0(t) = t^2$ and let ν be a gamma random variable with the variance parameter equal to 1, for example, $\psi(u) = (1+u)^{-1}$. The Lévy exponent function is taken to be $\Phi(z) = z^{2/3}/2$, which corresponds to a stable process with a power parameter equal to 2/3 and a scale parameter equal to 1/2. The resulting conditional hazard rates shown in the top left panel all exhibit an inverse U-shape in the baseline setup. Then we examine $\theta(t|x)$, which differs only in one of the three components (L, Λ_0, ψ) in the remaining panels. The roles of $\Lambda_0(t)$ and ν are similar to their roles in standard mixed proportional hazards models (van den Berg (2001)). The bottom left panel plots the conditional hazard rates for a different choice of $\Lambda_0(t) = t$, which causes the hazard shape to decrease. The bottom right panel displays hazard rates for a different gamma frailty ν with the variance parameter equal to 2, which makes the hazard rates flatter. The top right panel demonstrates the role of time-varying heterogeneity represented by the Lévy processes. In particular, we switch to a gamma process with the Lévy exponent function $\Phi(z) = 2\log((2+z)/2)$. Consistent with (2.16), it is clear in the plot that the Lévy process mainly changes how the covariate acts on the conditional hazard rates by mapping $\phi(x)$ to $\Phi \circ \phi(x)$.

The dependence structure of our model falls into the framework of Marshall and Olkin (1988) who comprehensively studied multivariate distributions generated from power mixtures. Let K be a copula function that is *min-infinitely divisible* in the sense of Joe (2014),[7] then

$$S_{\nu,K}(\cdot,\cdot) = \int_0^\infty K^u(\cdot,\cdot)\, dF_\nu(u) \tag{2.19}$$

is a well-defined multivariate survivor function with its survivor copula equal to

$$C_{\psi,K}(u_1, u_2) = \psi\bigl\{-\log K\bigl(e^{-\psi^{-1}(u_1)}, e^{-\psi^{-1}(u_2)}\bigr)\bigr\}. \tag{2.20}$$

[7] The power transform $K^u(\cdot,\cdot)$ should be a well-defined copula function itself for any $u > 0$, which is satisfied by (2.13) and more generally by any min-infinitely divisible copula; see Section 3.5 of Joe (2014).

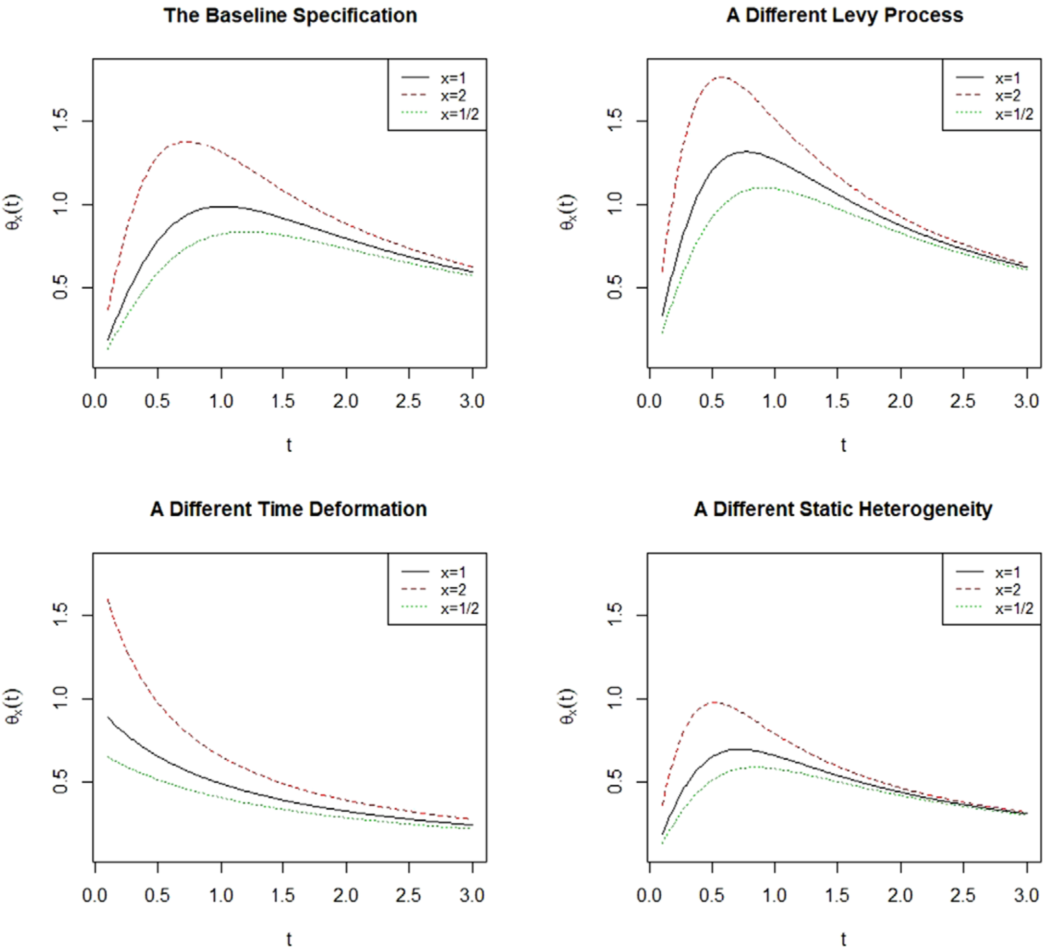

FIGURE 1. Plots of the conditional hazard rate function $\theta_x(t) \equiv \theta(t|x)$ for different specifications of $(L(t), \Lambda_0(t), \nu)$. The top left panel depicts the baseline specification, where $\Phi(z) = z^{2/3}/2$, $\Lambda_0(t) = t^2$, and $\psi(u) = (1+u)^{-1}$. The top right panel shows the conditional hazard rate functions for a changed Lévy process with $\Phi(z) = 2\log((2+z)/2)$. The bottom left panel shows the conditional hazard rates for a different time deformation $\Lambda_0(t) = t$. The bottom left panel shows the conditional hazard rates for another ν with $\psi(u) = (1+2u)^{-1/2}$.

In the extended Marshall–Olkin model, the conditional survivor copula is of the same form except that $K(u, v|x)$ is defined by (2.13) and it varies with the covariate value x. Next, we discuss two important subclasses of our model, including the shared-frailty models (Oakes (1989), van den Berg (2001)) and the multivariate shock models (Marshall and Olkin (1967), Marshall and Shaked (1979)). This illustrates that the extended Marshall–Olkin model offers a unifying framework for traditionally nonoverlapping approaches to model multiple durations.

REMARK 2.2. If two Lévy subordinators are independent, then $\Phi_{12}(u_1, u_2) = \Phi_1(u_1) + \Phi_2(u_2)$ (see the independent Lévy copula in Appendix B). This in turn leads to $\Psi_3(\cdot) = 0$

and

$$S_{T_1,T_2}(t_1, t_2|x) = \psi\{\Lambda_0(t_1)\Psi_1(x) + \Lambda_0(t_2)\Psi_2(x)\},$$

with $\Psi_k(x) = \Phi_k(\phi_k(x))$ for $k = 1, 2$. Therefore, the model is basically identical to the shared frailty model in Oakes (1989) and Honoré (1993). The conditional survivor copula function of two durations is Archimedean; we refer readers to Fan and Liu (2018) for a systematic treatment.

REMARK 2.3. In the absence of the static heterogeneity term, our method contains a new construction of the celebrated Marshall–Olkin model. The original model is generated by three independent exponential random variables: one leading to the simultaneous failure and two causing individual failures. Maintaining the presence of a common shock, An, Christensen, and Gupta (2004) generalized the Marshall–Olkin model with covariates and frailty terms as follows. The two durations are generated by $T_1 = \min(T_1^*, T_3^*)$ and $T_2 = \min(T_2^*, T_3^*)$, where all the latent failure times T_1^*, T_2^*, and T_3^* are assumed to obey the mixed proportional hazards structure. Clearly, the common failure source T_3^* is the one that produces the simultaneous failure among T_1 and T_2. Honoré and de Paula (2010, 2018) criticized this type of generalization due to the lack of structural interpretation. It is important to note that our modeling strategy has circumvented the shortcomings from two aspects. First, the durations in our model are defined via the structural threshold-crossing rules rather than set by the reduced-form hazard functions for T_1^*, T_2^*, and T_3^*. Second, the simultaneous failure in our model is generated by the dependent jumping behavior of bivariate Lévy subordinators, not necessarily through some common component T_3^* that terminates both durations.

REMARK 2.4. A convenient parametric family is the Lévy–Clayton copula with a single parameter $\alpha > 0$:

$$\mathcal{C}_{L,\alpha}(u_1, u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha}\right)^{-1/\alpha}. \tag{2.21}$$

One nice property related to the Lévy–Clayton copula is that the joint Lévy–Laplace exponent function can be further simplified as

$$\Phi_{12}(z_1, z_2) = \Phi_1(z_1) + \Phi_2(z_2) - \kappa(z_1, z_2; \alpha),$$

where

$$\kappa(z_1, z_2; \alpha) = z_1 z_2 \iint_{\mathcal{R}_+^2} e^{-y_1 z_1 - y_2 z_2} \mathcal{C}_{L,\alpha}\left(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\right) dy_1\, dy_2; \tag{2.22}$$

see Proposition 1 in Epifani and Lijoi (2010). Note that $\kappa(z_1, z_2; \alpha)$ summarizes the entire dependence structure when the Lévy–Clayton copula is adopted. Given the representation (2.22), we have the following expressions for functions in the conditional joint survivor function:

$$\begin{aligned} \Psi_1(x) &= \Phi_1\left(\phi_1(x)\right) - \kappa\left(\phi_1(x), \phi_2(x); \alpha\right), \\ \Psi_2(x) &= \Phi_2\left(\phi_2(x)\right) - \kappa\left(\phi_1(x), \phi_2(x); \alpha\right), \\ \Psi_3(x) &= \kappa\left(\phi_1(x), \phi_2(x); \alpha\right). \end{aligned} \tag{2.23}$$

Remark 2.5. If one starts with a model with different time deformation functions for first passage times:

$$T_k \equiv \{t : \phi_k(X) L_k(\nu \Lambda_k(t)) > \epsilon_k\}, \quad \text{for } k = 1, 2, \tag{2.24}$$

then we could obtain

$$S_{T_1,T_2}(t_1, t_2|x) = \psi\{\Lambda_1(t_1)\Psi_1(x) + \Lambda_2(t_2)\Psi_2(x) + [\Lambda_1(t_1) \vee \Lambda_2(t_2)]\Psi_3(x)\}, \tag{2.25}$$

along the same lines as the proof of Theorem 2.1. Note that the singularity of this type of model implies that there is positive probability such that $\Lambda_1(T_1) = \Lambda_2(T_2)$. However, in applications, the simultaneous failure, that is, $P\{T_1 = T_2\} > 0$, is mainly of interest (Honoré and de Paula (2018)), which forces the time deformation to be the same. In Section 4, we present a generalization that can incorporate different baseline hazards and time-varying covariates while maintaining the simultaneous stopping feature.

3. Identification and estimation with competing risks

In a bivariate competing risks model, two types of durations may occur on a subject, but only the one occurring first is observed together with its occurrence time; the other duration is censored. Referring to the conditional joint survivor functions, we encounter the dependent censoring problem in our extended Marshall–Olkin model, which is known to raise delicate concerns with identifiability (Tsiatis (1975)). Interested readers are referred to Heckman and Taber (1994) and Crowder (2012) for comprehensive reviews. Extra difficulties arise because the conditional survival copula function changes with x, in contrast with the existing literature that assumes an invariant copula, such as in Heckman and Honoré (1989), Abbring and van den Berg (2003), and Lee and Lewbel (2013). The conditional survival copula reveals that the current model does not fit into the framework in Fan and Liu (2018), since the conditional survivor copula is neither Archimedean, nor does it admit a copula density function for dependent Lévy subordinators. A separate analysis is required for our model.

3.1 *Sampling information*

Instead of having full access to both durations (T_1, T_2), the observational setting here comprises independent and identically distributed (*i.i.d.*) copies of the minimum $V = \min(T_1, T_2)$, the indicator of failure type (Honoré and de Paula (2010))

$$D = \begin{cases} 1 & \text{if } T_1 < T_2, \\ 2 & \text{if } T_1 > T_2, \\ 3 & \text{if } T_1 = T_2, \end{cases}$$

and a vector of covariates X. The sampling information is summarized by the conditional subsurvivor functions

$$S_{V,D=j}(t|x) \equiv P\{V > t, D = j | X = x\},$$

or by their corresponding conditional subdensity functions (Tsiatis (1975)):

$$f_{V,D=j}(t|x) \equiv -\partial S_{V,D=j}(t|x)/\partial t,$$

for $j = 1, 2$, and 3. Adding up all conditional subsurvivor functions leads to the conditional survivor function of the minimum V, meaning $S_V(t|X = x) = \sum_{j=1}^{3} S_{V,D=j}(t|x)$. We also write its conditional distribution function as $F_V(t|X = x)$. Given the explicit conditional joint survivor function, it is straightforward to compute the conditional sub-density functions. The following theorem generalizes Basu and Ghosh (1978), who present the subdensity functions of the original Marshall–Olkin distribution.

Theorem 3.1. *If Assumptions 2.1 to 2.6 hold, then the conditional survival function of V is*

$$S_V(t|x) = \psi\{\Lambda_0(t)\Phi_{12}(\phi_1(x), \phi_2(x))\}, \tag{3.1}$$

whereas the conditional subdensity functions are

$$f_{V,D=j}(t|x) = -\psi'\{\Lambda_0(t)\Phi_{12}(\phi_1(x), \phi_2(x))\}\lambda_0(t)\Psi_j(x), \tag{3.2}$$

for $j = 1, 2$, and 3.

Remark 3.1. Referring to the motivating economics applications, the patent-race game is indeed a competing risks model (Reinganum (1981, 1982), Choi (1991)), as the firm that produces the invention first wins the game and is awarded a patent. However, in the technology adoption or joint retirement problems, both durations could be observed for some pairs (but not all due to the additional censoring problem). For the latter case, the competing risks data corresponds to a coarsening of the bivariate durations data. It is an advantage that all model primitives can be identified from the competing risks data. If the bivariate durations data were available, then one has direct access to the conditional joint survivor functions of both durations, which opens up the possibility of introducing dependent thresholds or distinct static heterogeneity terms (ν_1, ν_2) in defining the bivariate durations; see the model from Begun and Yashin (2018) and Proposition 1 therein.

3.2 *Nonparametric identifiability*

We assume that an infinite sample from the distribution of (V, D, X) is available so that we can recover the subdensity functions $f_{V,D=j}(\cdot|X)$ for $j = 1, 2$, and 3, conditional on observable covariates. Now, we consider assumptions made on model primitives. Because the trajectory of latent processes is unobserved, some normalization assumptions cannot be avoided, even when dealing with the univariate duration model. We show the potential nonidentifiability related to stable processes in Appendix B. A complete characterization of observational equivalent pairs like in Ridder (1990) or Abbring (2012) is not yet known, as there is no analogous result for multivariate functions such as the joint Lévy exponent or copula function, so it will not be attempted here.

Assumption 3.1. There exist points t_0 and t_1 such that $\Lambda_0(t_0)/\lambda_0(t_0) = 1$ and $\Lambda_0(t_1) = 1$. Moreover, the range of $\Lambda_0(\cdot)$ contains a nonempty open set.

Assumption 3.2. There exists a known constant r_1 such that $\psi(1) = r_1$.

Assumption 3.3. The marginal Lévy measure $\Pi_k(u)$ is absolutely continuous, with its density function denoted by $\pi_k(u)$ for $k = 1, 2$. There exist some known constants $r_{2,k}$ such that $\Phi_k(1) = r_{2,k}$ for $k = 1, 2$.

Assumption 3.4. The distribution of d-dimensional covariate X is absolutely continuous with respect to Lebesgue measure for $d \geq 2$. We partition it into two parts $X = (X_1, X_{-1})$ and assume that the effect of covariate $\phi_k(x)$ is continuously differentiable and multiplicatively separable in the following sense:

$$\phi_k(x) = \phi_{k,1}(x_1)\phi_{k,-1}(x_{-1}) \quad \text{for } k = 1, 2. \tag{3.3}$$

For given points x_1^0, x_1^1, the normalization conditions hold as

$$\phi_{k,1}(x_1^0) = \phi'_{k,1}(x_1^0) \quad \text{and} \quad \phi_{k,1}(x_1^1) = 1, \tag{3.4}$$

where $\phi'_{k,1}(\cdot)$ is the derivative of $\phi_{k,1}(\cdot)$. Moreover, the support of $(\phi_1(x), \phi_2(x))$ contains an open rectangle in $\mathcal{R}_+^2$ for $x \in \mathcal{X}$.

Assumption 3.1 imposes the necessary location and scale normalization, following Assumption (I*2) in Jacho-Chavez, Lewbel, and Linton (2010). This offers some flexibility regarding the chosen normalization point, because the condition imposed on the time deformation does not restrict the baseline hazard density at time 0 (Ridder and Woutersen (2003)). The variation restriction in (A3) on the covariate effect is close to minimal in view of Abbring and van den Berg (2003). In the same spirit as Abbring (2012), many of the high-level assumptions in identifying GAFT models are automatically satisfied for threshold-crossing models driven by Lévy processes, due to their neat analytical characterization.

Theorem 3.2. *Suppose Assumptions 2.1 to 2.6 and Assumptions 3.1 to 3.4 hold, then the covariates effects (ϕ_1, ϕ_2), the time deformation function Λ_0, the Laplace transform ψ, the marginal Lévy measures (Π_1, Π_2), and the Lévy copula $\mathcal{C}_L$ are identified given the competing risks data (V, D, X).*

We offer a heuristic discussion of our identification strategy here. It is insightful to take a detour to examine the structure of the marginal distribution of each duration and its implication for identification. Recall that (2.16) gives

$$S_{T_k}(t|x) = \psi\{\Lambda_0(t)\Phi_k(\phi_k(x))\}, \quad \text{for } k = 1, 2.$$

Hence, the static heterogeneity ν and time-varying heterogeneity $L_k(t)$ affect $S_{T_k}(t|x)$ in different ways; that is, the Lévy–Laplace exponent function $\Phi_k(\cdot)$ directly transforms the covariate effect $\phi_k(\cdot)$, whereas $\psi(\cdot)$ serves as the unknown link function coupling the multiplicative components $\Lambda_0(t)$ and $\Phi_k(\phi_k(x))$. If $S_{T_k}(t|x)$ were available, one could separately identify $\psi(\cdot)$, $\Lambda_0(\cdot)$, and $\Phi_k \circ \phi_k(\cdot)$ following the approaches in Ekeland, Heckman, and Nesheim (2004), Jacho-Chavez, Lewbel, and Linton (2010), and Chiappori, Komunjer, and Kristensen (2015). Then one could disentangle $\Phi_k(\cdot)$ and $\phi_k(\cdot)$ based on the variation of covariates and additional structure on $\phi_k(\cdot)$, as stated in our Assumption 3.4. Obviously, the marginal survivor functions are not available in the competing risks data and one also must account for the dependence produced by the bivariate Lévy subordinators. Fortunately, the conditional survivor function of the observed minimum V in (3.1) exhibits a similar structure such that the aforementioned analysis could be adapted to identify all components, including ψ, Λ_0, and $\Phi_{12}(\phi_1(x), \phi_2(x))$, utilizing the results of Jacho-Chavez, Lewbel, and Linton (2010). Thereafter, we make use of the conditional subdensity functions in (3.2) and the algebraic relationships in defining $\Psi_1(x)$ and $\Psi_2(x)$ in Section 2 to identify $\Phi_k(\phi_k(x))$ for $k = 1, 2$. Thus, the multiplicative structure of covariate effects in Assumption 3.4 identifies the marginal Lévy measure Π_k and individual covariate effect ϕ_k for $k = 1, 2$. Finally, the Lévy copula is identified invoking the corresponding Sklar theorem in Kallsen and Tankov (2006).

Remark 3.2. An alternative path that one may take in formulating the multiple durations model is to incorporate dependent thresholds. When we allow arbitrary dependence between two exponential thresholds, the conditional marginal survivor function of individual duration still has a closed form. Following the partial identification methods in Honoré and Lleras-Muney (2006) and Khan and Tamer (2009), one could conduct analysis by bounding the conditional marginal survivor function via inequalities (1) and (2) from Bedford and Meilijson (1997), which improve the worst-case bounds from Peterson (1976) in the presence of the simultaneous failure.

3.3 *Semiparametric estimation*

We propose a flexible semiparametric procedure to obtain consistent estimates of the model primitives with minimal parametric restrictions. The exact regularity conditions and large sample properties of the estimates are deferred to future work, considering the amount of details required. The estimation of our model is based on a random sample $\{(V_i, D_i, X_i) : i = 1, \dots, n\}$ of (V, D, X). We adopt the commonly used parameterization by setting $\phi_1(x) = \exp(x^\top \gamma_1)$ and $\phi_2(x) = \exp(x^\top \gamma_2)$, where γ_1 and γ_2 are the regression coefficients as in the standard Cox regression. This specification is consistent with our Assumption 3.4. For rotational simplicity, we denote $\Omega_{12}(x) = \Phi_{12}(\phi_1(x), \phi_2(x))$.

One important observation regarding the regression coefficient γ_j is that it can be viewed as the linear coefficient in the single-index model (Ichimura (1993)) because of

$$\Phi_1\big(\phi_1(x)\big) = \Phi_1\big(\exp\big(x^\top \gamma_1\big)\big) \quad \text{and} \quad \Phi_2\big(\phi_2(x)\big) = \Phi_2\big(\exp\big(x^\top \gamma_2\big)\big), \tag{3.5}$$

where the Lévy–Laplace exponent functions become the unknown link functions. Therefore, γ_j is proportional to the following partial derivative (Stoker (1986), Powell, Stock, and Stoker (1989)):

$$\gamma_j \propto \frac{\partial}{\partial x}\Phi_j\big(\phi_j(x)\big) = \frac{\partial}{\partial x}\big[\Omega_{12}(x) - \Psi_k(x)\big], \quad \text{for } j, k = 1, 2 \text{ and } j \neq k; \tag{3.6}$$

given the relationships between $\Psi_1(x)$, $\Psi_2(x)$, and $\Psi_3(x)$ in Section 2.2.

Our semiparametric estimation procedure can be roughly divided into three steps. First, we explore the structure associated with the conditional distribution function of V to obtain consistent estimates of $\widehat{\Lambda}_0(\cdot)$, $\widehat{\psi}(\cdot)$, and $\widehat{\Omega}_{12}(\cdot)$. Second, we use certain weighted average derivatives to estimate the regression coefficients γ_1 and γ_2. Then the marginal Lévy exponent functions Φ_1 and Φ_2 can also be estimated as link functions in the associated single-index models, for example, via (3.5). In the last step, we can construct a minimum distance type estimator for the Lévy copula function or the finite dimensional parameter of a parameterized Lévy copula function. The details are as follows.

First, referring to the identities in our Theorem 3.1, we use the kernel-type method to consistently estimate the conditional distribution function of V and its conditional (sub)density functions, denoted by $\widehat{F}_V(t|x)$, $\widehat{f}_V(t|x)$, and $\widehat{f}_{V,D=j}(t|x)$, for $j = 1, 2$, and 3. We denote $\widehat{r}_i \equiv \widehat{r}(X_i, V_i) = \widehat{F}_V(V_i|X_i)$ and $\widehat{s}_i \equiv \widehat{s}(X_i, V_i) = \widehat{f}_V(V_i|X_i)$ for $i = 1, \dots, n$. Then we obtain a consistent estimator of $\bar{q}(z, t_0) \equiv \mathbb{E}[s(X, V)|r(X, V) = z, V = t_0]$ by running a nonparametric regression of $\widehat{s}_i$ on $\widehat{r}_i$ and V_i for those $\widehat{r}_i$ and V_i in the local neighborhood of (z, t_0). Denote the resulting estimate by $\widehat{\bar{q}}(z, t_0)$. Thereafter, given the multiplicative structure of $F_V(t|x)$, one can resort to Steps (3*)–(5*) on page 395 of Jacho-Chavez, Lewbel, and Linton (2010) to consistently estimate the functions $\widehat{\Lambda}_0(t)$, $\widehat{\psi}(u)$, and $\widehat{\Omega}_{12}(x)$. Specifically, for a constant r_1, define an estimate of $M(x, t) \equiv \Omega_{12}(x)\Lambda_0(t)$ by

$$\widehat{M}(x, t) \equiv \exp\left(\int_{r_1}^{\widehat{r}(x,t)} \frac{dz}{\widehat{\bar{q}}(z, t_0)}\right). \tag{3.7}$$

Then $\Omega_{12}(x)$ and $\Lambda_0(t)$ are estimated (up to a scale factor) by the marginal integration:

$$\widehat{\eta}_{P_1}(x) = \int \widehat{M}(x, t)\, dP_1(t), \qquad \widehat{\eta}_{P_2}(t) = \int \widehat{M}(x, t)\, dP_2(x), \tag{3.8}$$

for some normalizing measures P_1 and P_2. Define the normalizing constant

$$\tilde{c} \equiv (1/2)\left[\int \widehat{\eta}_{P_1}(x)\, dP_2(x) + \int \widehat{\eta}_{P_2}(t)\, dP_1(t)\right],$$

and scaled estimates $\widehat{\Omega}_{12}(x) = \widehat{\eta}_{P_1}(x)/\tilde{c}$, $\widehat{\Lambda}_0(t) = \widehat{\eta}_{P_2}(t)/\tilde{c}$, and $\tilde{M}(X_i, V_i) = \widehat{\Omega}_{12}(X_i) \times \widehat{\Lambda}_0(V_i)\tilde{c}$. The link function $\psi(\cdot)$ and its derivative $\psi'(\cdot)$ can be estimated by the intercept term and slope coefficient, respectively, in a local quadratic regression (Fan and Gijbels (1996)) of $(1 - \widehat{r}_i)$ on the regressor $\tilde{M}(X_i, V_i)$. We denote the resulting estimates by $\widehat{\psi}(\cdot)$ and $\widehat{\psi}'(\cdot)$.

Second, we explore (3.2) and borrow the key insight from Horowitz (1996) to obtain

$$\widehat{\Psi}_j(x) = -\int \frac{\widehat{f}_{V,D=j}(t|x) w_T(t)}{\widehat{\lambda}_0(t)\widehat{\psi}'\big\{\widehat{\Lambda}_0(t)\widehat{\Omega}_{12}(x)\big\} \vee \rho}\, dt, \quad j = 1, 2, \tag{3.9}$$

for a weight function $w_T(t)$ along the time dimension. Also, we introduce a trimming parameter ρ to overcome the instability of the denominator $\widehat{\lambda}_0(t)\widehat{\psi}'\{\widehat{\Lambda}_0(t)\widehat{\Omega}_{12}(x)\}$. In (3.9), the derivative of deformation function $\lambda_0(t)$ is estimated by the following kernel-smoothed estimate:

$$\widehat{\lambda}_0(t) = \frac{1}{h_\lambda}\int K\big((t-u)/h_\lambda\big)\,d\widehat{\Lambda}_0(u), \tag{3.10}$$

with the corresponding kernel function $K(\cdot)$ and bandwidth h_λ. Regarding the regression coefficients, we estimate the weighted average derivatives (Powell, Stock, and Stoker (1989)):

$$\widehat{\gamma}_j^* = \int \frac{\partial}{\partial x}\big[\widehat{\Omega}_{12}(x) - \widehat{\Psi}_k(x)\big]w_X(x)\,dx, \quad j,k = 1,2 \text{ and } j \neq k, \tag{3.11}$$

with some weight function $w_X(x)$ operating on the support of X. These estimates deliver the estimated regression coefficients upon scale normalization, that is, $\widehat{\gamma}_j = \widehat{\gamma}_j^*/|\widehat{\gamma}_j^*|$ for $j = 1, 2$, if we impose the normalization that $|\gamma_j| = 1$ for both $j = 1$ and 2. Then we denote $\widehat{\phi}_j(x) \equiv \exp(x^\top\widehat{\gamma}_j)$ for $j = 1, 2$. The consistency results of the aforementioned estimates can be shown by adapting Theorem 1 of Horowitz (1996).

With the estimated $\widehat{\Omega}_{12}(x)$, $\widehat{\Psi}_1(x)$, and $\widehat{\Psi}_2(x)$ at hand, we can obtain estimated $\widehat{\Phi}_j(\cdot)$ by collecting the intercept term in the local linear regression (Fan and Gijbels (1996)) of $(\widehat{\Omega}_{12}(X_i) - \widehat{\Psi}_k(X_i))$ on $\exp(X_i^\top\widehat{\gamma}_j)$ for $j, k \in \{1, 2\}$ and $j \neq k$. Finally, the Lévy copula function $\mathcal{C}_L$ can be estimated using the sieve minimum distance estimator (see Section 2.2.4 of Chen (2007)):

$$\min_{\mathcal{C}_L\in\mathcal{S}}\big\|\widehat{\Omega}_{12}(x) - \Phi_{12}\big(\widehat{\phi_1}(x), \widehat{\phi}_2(x)\big)\big\|_\Phi, \tag{3.12}$$

where $\mathcal{S}$ is an appropriate sieve space and $\|\cdot\|_\Phi$ denotes the suitable norm for the joint Lévy–Laplace exponent function. In particular, Wood (2003) advocateed thin plane splines for approximating smooth bivariate functions. Alternatively, one can also parameterize the Lévy copula function, such as the Lévy–Clayton copula in Remark 2.4, and conduct a parametric estimation in the last step.

4. Generalizations

In the benchmark case of our extended Marshall–Olkin model studied in Section 2, the time deformation function $\Lambda_0(t)$ is equivalent to a common cumulative baseline hazard function for both durations T_1 and T_2. Because a baseline hazard function encodes genuine duration dependence (Elbers and Ridder (1982)), it is desirable to consider a more flexible specification that allows different baseline hazards. Moreover, one often encounters time-varying covariates in practice. In this section, we propose a more general model that accommodates these features and study the empirical content with competing risks data.

In defining structural durations, we add idiosyncratic trend functions $\Lambda_k(t|W)$ driven by time-varying covariates $W(t)$ in the following way:

$$T_k \equiv \inf\big\{t_k : \phi_k(X)L_k\big(\nu\Lambda_0(t_k)\big) + \Lambda_k(t_k|W) \geq \epsilon_k\big\} \quad \text{for } k = 1, 2, \tag{4.1}$$

with

$$\Lambda_k(t|W)=\int_0^t \beta_k^\top(s)W(s)\,ds, \tag{4.2}$$

where $\beta_k(t)$ stands for the time-varying covariate effect in the sense of Aalen (1980) for $k=1,2$. The additive-multiplicative specification resembles the symmetric entry model of Bresnahan and Reiss (1991). Adopting a static threshold-crossing model, Bresnahan and Reiss (1991) specified the potential benefit of entry as a multiplicative function in terms of the variable profit and market size, whereas the cost function is decomposed additively into observable and unobservable parts.

Considering the technology development/adoption examples in Section 2, $W(t)$ represents various characteristics that do not directly intervene with the latent innovation process, such as the firm's financial capacities, control of various costs of borrowing capital, hiring labor, and use of input. All these factors are allowed to vary with time as well. Likewise, for the retirement decisions, $W(t)$ collects individual job characteristics for husbands and wives, such as the skill level of the job, union status, etc. Regarding the smoking cessation, the time-varying covariates, which do not directly interact with the latent health process, may include the external changes of cigarette prices and tobacco control policies.

Before we present more details of the model, it is useful to introduce some shorthand notation to facilitate the remaining discussion. We define

$$\Gamma(t,x)\equiv\psi\big\{\Lambda_0(t)\Phi_{12}\big(\phi_1(x),\phi_2(x)\big)\big\}, \tag{4.3}$$

and

$$\gamma(t,x)\equiv-\psi'\big\{\Lambda_0(t)\Phi_{12}\big(\phi_1(x),\phi_2(x)\big)\big\}\lambda_0(t). \tag{4.4}$$

We assume that $\Lambda_1(\cdot|W)$ and $\Lambda_2(\cdot|W)$ are absolutely continuous with the nonnegative derivatives denoted by $\lambda_1(t|W)=\beta_1^\top(t)W(t)$ and $\lambda_2(t|W)=\beta_2^\top(t)W(t)$, respectively. In our context, the time-varying covariates $W(t)$ are assumed to be predictable or weakly exogenous processes (Ridder and Tunali (1999)) and independent of Lévy subordinators $\mathbf{L}(t)=(L_1(t),L_2(t))$. In this context, the conditional survivor function $P\{T_i>t|X,\bar{W}(t)\}$ becomes a stochastic process and $\bar{W}(t)$ collects the history of time-varying covariates $\{W(s):0\le s\le t\}$ (Ridder and Tunali (1999)). The lower-case $w(t)$ and $\bar{w}(t)$ stand for the realization of the processes $W(t)$ and $\bar{W}(t)$.

The next two theorems generalize Theorems 2.1 and 3.1. The proofs are straightforward modifications based on Theorems 2.1 and 3.1; thus, they are deferred to Appendix B.

Theorem 4.1. *For the generalized version defined by* (4.1), *the joint survivor function is*

$$\begin{aligned}&S_{T_1,T_2}\big(t_1,t_2|x,\bar{w}(t_1\vee t_2)\big)\\&\quad=e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}\psi\big\{\Lambda_0(t_1)\Psi_1(x)+\Lambda_0(t_2)\Psi_2(x)+\Lambda_0(t_1\vee t_2)\Psi_3(x)\big\}.\end{aligned} \tag{4.5}$$

The conditional marginal survivor function is

$$S_{T_k}\big(t|x,\bar{w}(t)\big)=e^{-\Lambda_k(t|w)}\psi\big\{\Lambda_0(t)\Phi_k\big(\phi_k(x)\big)\big\},\quad \textit{for } k=1,2. \tag{4.6}$$

THEOREM 4.2. *For the generalized version defined by* (4.1), *the conditional survivor function for V is*

$$S_V\big(t|x,\bar{w}(t)\big)=e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}\Gamma(t,x). \tag{4.7}$$

And the conditional subdensity functions are

$$\begin{aligned} f_{V,D=1}\big(t|x,\bar{w}(t)\big)&=\lambda_1(t|w)e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}\Gamma(t,x)\\ &\quad+e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}\gamma(t,x)\Psi_1(x), \end{aligned} \tag{4.8}$$

$$\begin{aligned} f_{V,D=2}\big(t|x,\bar{w}(t)\big)&=\lambda_2(t|w)e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}\Gamma(t,x)\\ &\quad+e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}\gamma(t,x)\Psi_2(x), \end{aligned} \tag{4.9}$$

$$f_{V,D=3}\big(t|x,\bar{w}(t)\big)=e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}\gamma(t,x)\Psi_3(x). \tag{4.10}$$

Referring to the conditional cumulative hazard function $\Theta_k(t|x)$ for $k=1,2$, we get

$$\Theta_k\big(t|x,\bar{w}(t)\big)=\Lambda_k(t|w)-\log\big[\psi\big(\Lambda_0(t)\Phi_k\big(\phi_k(x)\big)\big)\big], \tag{4.11}$$

which simplifies to

$$\Theta_k\big(t|x,\bar{w}(t)\big)=\Lambda_k(t|w)+\Lambda_0(t)\Phi_k\big(\phi_k(x)\big), \tag{4.12}$$

without the time-invariant heterogeneity ν. Thus, the idiosyncratic trend function $\Lambda_k(t|W)$ measures the excess risk on top of $\Lambda_0(t)$ for $k=1,2$. A further simplification occurs when both marginal Lévy subordinators are stable processes with a power parameter b_0 and a unit scale parameter (see Section B.2 in Appendix B) and covariates effects are parameterized as in Section 3.3. In this case, the conditional hazard rate function is

$$\theta_k\big(t|x,\bar{w}(t)\big)\equiv\frac{\partial}{\partial t}\Theta_k\big(t|x,\bar{w}(t)\big)=w(t)^\top\beta_k(t)+\lambda_0(t)\exp\big(b_0x^\top\gamma_k\big),\quad \text{for } k=1,2, \tag{4.13}$$

which coincides with the specification of Martinussen and Scheike (2002); see their equation (1) on page 283. Interestingly, this generalized version of our model induces the additive-multiplicative hazards model of Martinussen and Scheike (2002), which encompasses both the Cox proportional hazards model and the Aalen additive hazards model.

We need an extra assumption that states the full rank condition on the time-varying covariates $W(t)$ and allows us to resort to the identification-at-limit strategy.

ASSUMPTION 4.1. (i) There exists a limit point x_0 in the support of X such that

$$\lim_{x\to x_0}\phi_1(x)=\lim_{x\to x_0}\phi_2(x)=0.$$

(ii) For any t, the matrix $E[W(t)W^\top(t)|X=x_0]$ is of full rank.[8]

[8] I am grateful to an anonymous referee who suggested the rigorous formulation.

The introduction of time-varying covariates $W(t)$ functions as the exclusion restriction in the sense that the observable process $W(t)$ does not directly interact with the latent Lévy processes. Another type of exclusion restrictions for competing risks models is utilized by Heckman and Honoré (1990), in which the instrument exists and affects only one type of latent duration; see Section 2.3 in Heckman and Honoré (1990). Our purpose is somewhat different, as the existence of $W(t)$ offers flexible conditional (on observables) marginal hazard functions. The following theorem asserts that we can point identify all structural components with the help of Assumption 4.1. Intuitively speaking, if we let the covariates effect from static covariates X converge to zero and condition on the time-varying covariates $W(t)$, the model can be seen as the independent competing risks model from Berman (1963) where both idiosyncratic trends $\Lambda_1(t|W)$ and $\Lambda_2(t|W)$ can be explicitly solved. Next, the time-varying covariate effects $\beta_k(t)$ are identified utilizing the rank condition in Assumption 4.1(ii) for $k = 1, 2$. Then the identification of other structural components are shown by analogous arguments used to prove Theorem 3.2. The detailed proof is relegated to Appendix B.

Theorem 4.3. *Suppose Assumptions 2.1 to 2.6, Assumptions 3.1 to 3.4, and 4.1 hold, then all structural components of the model defined by (4.1) are identified given the competing risks data (V, D, X, W).*

5. An empirical application

To illustrate the model (2.7), we apply the proposed semiparametric estimation procedure to the joint retirement problem of married couples, using the data in Honoré and de Paula (2018) drawn from eight waves of the Health and Retirement Study (every 2 years from 1992 to 2006). Here, the duration variables of interest (T_1, T_2) are the retirement dates of the wife and husband.[9] In terms of the measurement of individual durations, we follow Honoré and de Paula (2018), who measure time in terms of *family age*, which is set to zero when the older partner in the couple reaches age 60; the duration of the other spouse is then recorded by examining the age difference. The original retirement information is recorded at a monthly frequency and we take the logarithmic transformation to avoid a long right tail in the distribution. The competing risks data (V_i, D_i) for the ith household is understood as follows for $i = 1, \dots, n$. The duration V represents the first entry into retirement for the corresponding member of the household, and $D = 1, 2$, or 3 depending on whether the wife retires first, the husband first, or they retire simultaneously. In this example, the latent stochastic processes $(L_1(t), L_2(t))$ characterize the aging processes for the elderly. These processes are assumed to be irreversible and accumulate incrementally in one direction. Also, the jumping behavior of Lévy subordinators captures sudden health shocks such as heart attacks or new cancer diagnoses (An, Christensen, and Gupta (2004), Coile (2004)). Moreover, the common

[9] We adopt the retirement classification suggested by the Rand Corporation, as in Honoré and de Paula (2018), which classifies a respondent as retired if she/he is not working and not looking for work or there is any mention of retirement in employment status or in the answers to questions that ask the respondent whether he or she considers him or herself to be retired.

time deformation $\nu\lambda_0(t)$ governs the speed of arrival rate of health shocks. Two continuous covariates[10] $X = (X_1, X_2)^\top$ include (1) the total health expenditure of the household in the previous 12 months for the first two waves and the previous 2 years for the subsequent years (inflation adjusted using the CPI to Jan/2000 dollars); (2) the financial wealth of the family (inflation adjusted using the CPI to Jan/2000 dollars). For these two covariates, we use the transformation $\text{sgn}(x)\sqrt{x}$, in the spirit of a logarithmic transformation of positive variables.[11] Referring to the effects of covariates, we adopt the following parameterization: $\phi_1(x) = \exp(x_1\gamma_{1,1} + x_2\gamma_{1,2})$ and $\phi_2(x) = \exp(x_1\gamma_{2,1} + x_2\gamma_{2,2})$. We exclude observations with missing values for health expenditure or the minimum duration V being censored, which leaves us with 821 households in total. Further details of the data set are given in Honoré and de Paula (2018) and not repeated here. In principle, the discrete regressors can also be included because the key estimation steps of Jacho-Chavez, Lewbel, and Linton (2010) or the weighted average derivatives for the covariate effects can be extended to allow for discrete regressors; see Section 6 in Jacho-Chavez, Lewbel, and Linton (2010) and Horowitz and Hardle (1996). Note that the data in Honoré and de Paula (2018) contains a wealth of discrete covariates such as socioeconomic, demographic, and self-reported health information. The empirical usefulness of incorporating such discrete regressors will be pursued in future work.

We briefly discuss the distinction with respect to the specification of Honoré and de Paula (2018). First, the key difference is regarding the modeling framework, as Honoré and de Paula (2018) employ a cooperative game to model the bargaining and joint retirement decisions. In comparison, the couples act noncooperatively in our model. As a matter of fact, our model resembles Honoré and de Paula (2010) more closely, because Honoré and de Paula (2010) showed the Marshall–Olkin model is observationally equivalent to a particular version of their noncooperative stopping game model with additive interaction effects. For a general discussion of noncooperative games for within-marriage decisions, we refer readers to Lundberg and Pollak (1994). In terms of the estimation and inference of the complicated model in Honoré and de Paula (2018), a fully parametric indirect inference is adopted. One advantage of the model from Honoré and de Paula (2018) is that the baseline hazard functions are allowed to be different for wives and husbands. In comparison, our semiparametric procedure allows the common time deformation, Laplace transform of ν, and marginal Lévy exponent functions to be completely nonparametric.

Here, we fill in the details regarding the implementation of our semiparametric estimation procedure[12]. We start with the kernel-type estimation of the conditional subdis-

[10]In Appendix B, we report additional estimation results with the inclusion of a third regressor, "age difference." Given that the age difference is neither significant for the wife nor for the husband, we delegate the details to Section B.4 in Appendix B.

[11]In the computations, we also divide the transformed variable by 100 for total health expenditure and by 1000 for financial wealth to avoid overflow, as in Honoré and de Paula (2018).

[12]Liu (2020) contains the code for the implementation of our estimators.

tribution and density functions as follows:

$$\widehat{F}_{V,D=j}(t|x)=\frac{\sum_{i=1}^{n}\mathbb{I}\{V_i\le v,D=j\}\mathbf{K}\big((X_i-x)/\mathbf{h_x}\big)}{\sum_{i=1}^{n}\mathbf{K}\big((X_i-x)/\mathbf{h_x}\big)},\quad\text{and}$$

$$\widehat{f}_{V,D=j}(t|x)=\frac{\sum_{i=1}^{n}\mathbb{I}\{D=j\}K\big((V_i-t)/h_v\big)\mathbf{K}\big((X_i-x)/\mathbf{h_x}\big)}{h_v\sum_{i=1}^{n}\mathbf{K}\big((X_i-x)/\mathbf{h_x}\big)},\quad j=1,2,3,$$

where we use the quartic kernel function $K(u)=\frac{15}{16}(1-u^2)^2\mathbb{I}\{|u|\le 1\}$ and the boldfaced version $\mathbf{K}(\cdot)$ to denote the product kernel, for example, $\mathbf{K}((X_i-x)/\mathbf{h_x})=\prod_{j=1}^{2}K((X_{ij}-x_j)/h_{x_j})$. The bandwidths (h_v,h_{x_1},h_{x_2}) are chosen by the Sheather–Jones rules (Sheather and Jones (1991)) for the dependent variable and two covariates. The conditional distribution function $\widehat{F}_V(t|x)$ or the density function $\widehat{f}_V(t|x)$ are taken to be the sum of the above estimates over three categories. Then we set $\widehat{r}_i=\widehat{F}_V(V_i|X_i)$ and $\widehat{s}_i=\widehat{f}_V(V_i|X_i)$ as the initial inputs of the approach in Jacho-Chavez, Lewbel, and Linton (2010). To estimate the $\bar{q}(z,t_0)$ function, we use the k-nearest-neighbor estimator with bivariate regressors $\widehat{r}_i$ and V_i. The number of neighbors is set to be the integer part of $n^{3/5}$. The numerical integration involved in $\widehat{M}(x,t)$ is computed by the trapezoidal rule and the normalizing scalar r_1 is the sample mean of V. Regarding the marginal integration in the fourth and fifth steps of Jacho-Chavez, Lewbel, and Linton (2010), we set $P_1(t)$ and $P_2(x)$ to be $F_V(t)$ and $F_X(x)$, which are the distribution functions of V and X, respectively. In the actual estimation, we replace these marginal distributions by their empirical counterparts so that

$$\widehat{\eta}_{P_1}(x)=\frac{1}{n}\sum_{i=1}^{n}\widehat{M}(V_i,x),\qquad \widehat{\eta}_{P_2}(t)=\frac{1}{n}\sum_{i=1}^{n}\widehat{M}(t,X_i).$$

Regarding the final step of Jacho-Chavez, Lewbel, and Linton (2010), we run the local quadratic regression with the dependent variables $(1-\widehat{r}_i)_{i=1}^n$ to obtain the estimated Laplace transform $\widehat{\psi}$ and its derivative $\widehat{\psi}'$, where the latter is required for estimating the finite dimensional coefficients via (3.9). The kernel bandwidth here is selected using the method of Ruppert, Sheather, and Wand (1995). In (3.10), the bandwidth h_λ is taken to be the same as h_v. The weighting function $w_T(\cdot)$ is the uniform kernel with the same support of V and $w_X(\cdot)$ in constructing the weighted average derivatives is taken to be the standard Gaussian kernel. Finally, the trimming parameter ρ in (3.9) is set to be 0.01.

We now present our estimation results in Table 1. The effects of two continuous covariates are rather similar for wives and husbands in terms of the magnitude and sign. The reported standard errors are calculated by perturbing all the nonparametric esti-

TABLE 1. Estimated effects of covariates.

Covariates	Estimated Coeff.	S.E.
Health expenditure-wife	0.862	0.057
Financial wealth-wife	0.507	0.106
Health expenditure-husband	0.999	0.403
Financial wealth-husband	0.030	0.352

mates with the Bayesian bootstrap weights[13] (Rubin (1981)), based on 100 replications. Namely, the bootstrap multipliers are given by $M_{ni} = \varpi_i/(\sum_{i=1}^{n} \varpi_i)$ for $1 \leq i \leq n$ and we take ϖ_i to be *i.i.d.* draws from the unit exponential distribution (Rubin (1981)). For example, the conditional subdistribution function is estimated by

$$\widehat{F}^*_{V,D=j}(t|x) = \frac{\sum_{i=1}^{n} M_{ni}\mathbb{I}\{V_i \leq v, D = j\}\mathbf{K}\big((X_i - x)/\mathbf{h_x}\big)}{\sum_{i=1}^{n} M_{ni}\mathbf{K}\big((X_i - x)/\mathbf{h_x}\big)}, \quad \text{for } j = 1, 2, 3,$$

in the bootstrap sample. The same set of weights is also applied to the other nonparametric estimates in order to quantify the estimation uncertainty. The total health expenditure serves as a significantly positive predictor for both males and females as expected (Coile (2004)). The positive effect of financial wealth on both partners owes to the fact that the richer family is less concerned about the income effect due to retirement. Nonetheless, this effect is only significant for the wife. The insignificance of financial wealth for the husband is also evident in the model of Honoré and de Paula (2018).

Figure 2 depicts the estimated time deformation function (in the left panel) and the estimated Laplace transform of the static heterogeneity ν (in the right panel). The pattern of the time deformation (or the baseline cumulative hazard function) exhibits positive duration dependence overall, confirming that retirement becomes more likely as the household ages. This type of duration dependence is also found in Honoré and de Paula (2018) for their model with Weibull-type baseline hazard functions. The decreasing tendency of $\widehat{\psi}(\cdot)$ is also consistent with the shape of the Laplace transform of a positive random variable, despite some minor fluctuations in the middle range due to the unconstrained nonparametric estimation.

Regarding the features of latent Lévy processes, we first run the local linear regressions to obtain the marginal Lévy–Laplace exponent functions $\Phi_k(\cdot)$ for $k = 1, 2$, as displayed in Figure 3. The bandwidths are selected according to the rules given by Ruppert, Sheather, and Wand (1995). Clearly, $\widehat{\Phi}_1(\cdot)$ (in the left panel) is an increasing function, consistent with the shape of a Lévy–Laplace exponent function, so is $\widehat{\Phi}_2(\cdot)$ (in the right panel), except for the parts near the left or right boundaries. There is substantial difference between these two marginal Lévy–Laplace exponent functions, which translates

[13]The Bayesian bootstrap is a smooth alternative to the nonparametric bootstrap, because all observations are assigned with positive probability mass in the resampling procedure.

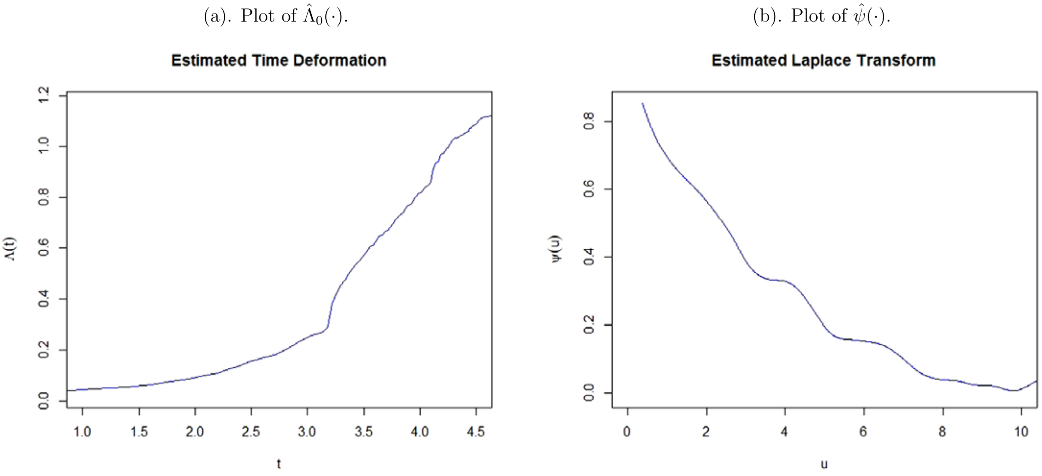

FIGURE 2. Estimated Common Time Deformation and Laplace Transform of Static Heterogeneity for the Couple.

into different aging processes of males and females. Regarding the dependence measure of the two marginal Lévy processes, one could use the thin plate spline basis (Wood (2003)) for approximating the Lévy copula function. To offer a more concise summary, we parameterize the Lévy copula function to be Lévy–Clayton with the unknown scalar α, as in Section 2.2. Hence, the functions $\Psi_j(x)$ with $j = 1, 2$, or 3 are of the forms presented in Remark 2.4. Maintaining this parametric specification, the minimum distance estimation (3.12) boils down to the nonlinear parametric version with a single unknown

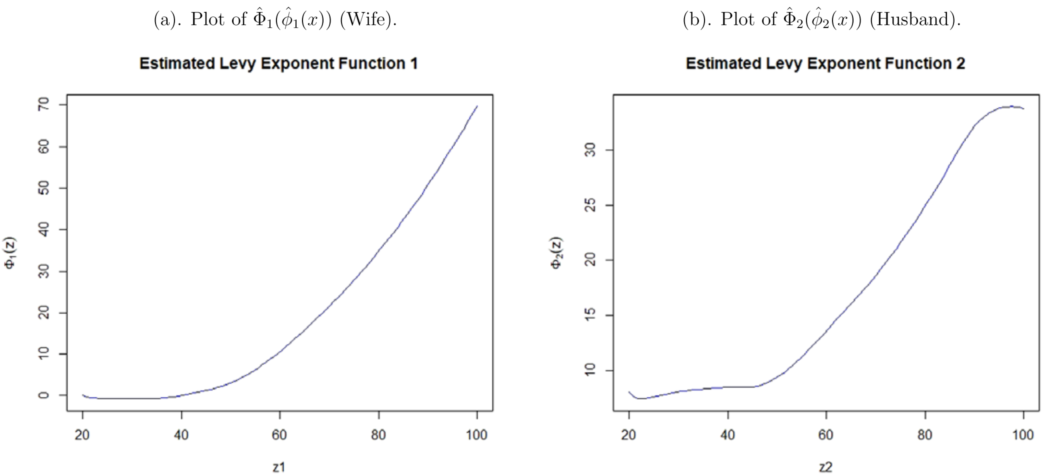

FIGURE 3. Estimated marginal Lévy exponent functions for the couple.

parameter α:

$$\widehat{\alpha} \equiv \arg\min_{\alpha} \sum_{i=1}^{n} \left(\widehat{\Omega}_{12}(X_i) - \Phi_{12}\left(\widehat{\phi_1}(X_i), \widehat{\phi}_2(X_i); \alpha\right)\right)^2, \tag{5.1}$$

where $\Phi_{12}(\cdot, \cdot; \alpha)$ signifies the dependence of the joint Lévy exponent function on α. The estimate $\widehat{\alpha}$ is found to be 0.018 (with the bootstrap standard error equal to 0.006), indicating that the dependence of health shocks to the couple is relatively weak but still significant. This is not too surprising given that around 3.7% of the observed couples retired simultaneously. The percentage of *approximate* joint retirement increases to 7.4% for couples who retire within 1 month of each other and to 10.2% for couples retiring within 2 months.

6. Conclusion

In this paper, we propose a new bivariate durations model in which the underlying durations are driven by dependent continuous-time stochastic processes, without any parametric assumption. The duration variables in our model are closely related to the decision rules suggested by optimal stopping time models, while at the same time providing a rich and tractable statistical description of the underlying failure/survival mechanism. This extended Marshall–Olkin model can be seen as the natural variant of the mixed proportional hazards model from a process point of view and it is transparent how the structural parameters can be identified from reduced-form functions with competing risks data. Because the sampling information is explicitly presented by the conditional (sub)survivor functions for our model, this opens the possibility of flexible semiparametric estimation.

Numerous extensions are possible. First, one could incorporate endogenous and time-varying covariates by explicitly modeling additional stochastic processes, as in Renault, van den Heijden, and Werker (2014). Second, it remains challenging to consider arbitrarily distributed thresholds combined with Lévy subordinators, as the characterization is more complicated and only available for the univariate case. Very little is known about the analytical properties of the first passage times associated with general multivariate stochastic processes. Last but not least, it is much more ambitious to characterize the empirical content of stochastic game models driven by latent Lévy processes, extending de Paula (2009) and Honoré and de Paula (2018).

Appendix A: Proofs of main results

In this Appendix, we prove the theorems stated in Sections 2 and 3 of the paper.

Proof of Theorem 2.1. The main appealing property of Lévy subordinators combined with exponential thresholds lies in the concrete analytical characterization in (2.1). Moreover, the monotonic sample path implies that if $T_k > t_k$, then the process

$\phi_k(x)L_k(\nu\Lambda_0(t_k))$ is still below the threshold ϵ_k for $k=1,2$. Conditional on the time-varying and static heterogeneity terms and for $t_1 \le t_2$, we have

$$\begin{aligned} &P\{T_1 > t_1, T_2 > t_2 | x, \mathbf{L}, \nu\} \\ &\quad = P\big\{\epsilon_1 > \phi_1(x)L_1\big(\nu\Lambda_0(t_1)\big), \epsilon_2 > \phi_2(x)L_2\big(\nu\Lambda_0(t_2)\big) | x, \mathbf{L}, \nu\big\} \\ &\quad = e^{-\phi_1(x)L_1(\nu\Lambda_0(t_1))-\phi_2(x)L_2(\nu\Lambda_0(t_2))} \\ &\quad = e^{-\phi_1(x)L_1(\nu\Lambda_0(t_1))-\phi_2(x)L_2(\nu\Lambda_0(t_1))} e^{-\phi_2(x)[L_2(\nu\Lambda_0(t_2))-L_2(\nu\Lambda_0(t_1))]}. \end{aligned}$$

Given that Lévy subordinators have independent increments, we obtain

$$\begin{aligned} &P\{T_1 > t_1, T_2 > t_2 | x, \nu\} \\ &\quad = E\big[e^{-\phi_1(x)L_1(\nu\Lambda_0(t_1))-\phi_2(x)L_2(\nu\Lambda_0(t_1))} | x, \nu\big] E\big[e^{-\phi_2(x)[L_2(\nu\Lambda_0(t_2))-L_2(\nu\Lambda_0(t_1))]} | x, \nu\big] \\ &\quad = e^{-\nu\Lambda_0(t_1)\Phi_{12}(\phi_1(x),\phi_2(x))} e^{-\nu(\Lambda_0(t_2)-\Lambda_0(t_1))\Phi_2(\phi_2(x))}, \end{aligned}$$

by integrating out $\mathbf{L} = (L_1, L_2)$ and using equations (2.1) and (2.3). Integration with respect to ν gives

$$P\{T_1 > t_1, T_2 > t_2 | x\} = \psi\big\{\Lambda_0(t_1)\Phi_{12}\big(\phi_1(x), \phi_2(x)\big) + \big[\Lambda_0(t_2) - \Lambda_0(t_1)\big]\Phi_2\big(\phi_2(x)\big)\big\}.$$

Over the region where $t_1 \ge t_2$, one could proceed in a similar fashion to obtain

$$P\{T_1 > t_1, T_2 > t_2 | x\} = \psi\big\{\Lambda_0(t_2)\Phi_{12}\big(\phi_1(x), \phi_2(x)\big) + \big[\Lambda_0(t_1) - \Lambda_0(t_2)\big]\Phi_1\big(\phi_1(x)\big)\big\}.$$

The desired formula of the conditional joint survivor function can be obtained by combining the two cases together. To get the conditional marginal survivor function, we simply set either t_1 or t_2 to zero.

Given the conditional joint and marginal survivor functions for both durations, it is straightforward to obtain the conditional survivor copula function as

$$\begin{aligned} &C(u_1, u_2 | x) = S_{T_1,T_2}\big(S_{T_1}^{-1}(u_1|x), S_{T_2}^{-1}(u_2|x) | x\big) \\ &\quad = \begin{cases} \psi\big\{\psi^{-1}(u_1) + \big(1-\omega_2(x)\big)\psi^{-1}(u_2)\big\}, & \text{if } \dfrac{\psi^{-1}(u_1)}{\Psi_1(x)+\Psi_3(x)} \ge \dfrac{\psi^{-1}(u_2)}{\Psi_2(x)+\Psi_3(x)} \\ \psi\big\{\big(1-\omega_1(x)\big)\psi^{-1}(u_1) + \psi^{-1}(u_2)\big\}, & \text{if } \dfrac{\psi^{-1}(u_1)}{\Psi_1(x)+\Psi_3(x)} \le \dfrac{\psi^{-1}(u_2)}{\Psi_2(x)+\Psi_3(x)}, \end{cases} \end{aligned}$$

which is of the same form as in (2.20). □

Proof of Theorem 3.1. The conditional survivor function of the minimum V is obtained by letting $t_1 = t_2 = t$ in Theorem 2.1. The simplification occurs because

$$\Psi_1(x) + \Psi_2(x) + \Psi_3(x) = \Phi_{12}\big(\phi_1(x), \phi_2(x)\big).$$

The conditional subdensity functions are computed using a modified version of Tsiatis' (1975) theorem stated in Theorem 1 of Arnold and Brockett (1983) that copes with simultaneous failure. Considering the case where $D = 1$, the conditional subdensity function is calculated by

$$f_{V,D=1}(t|x) = -\lim_{s\uparrow t} \frac{\partial}{\partial s} S_{T_1,T_2}(s, t|x). \tag{A.1}$$

Over the range where $t_1 < t_2$, the conditional joint survivor function is

$$S_{T_1,T_2}(t_1, t_2|x) = \psi\{\Psi_1(x)\Lambda_0(t_1) + [\Psi_2(x) + \Psi_3(x)]\Lambda_0(t_2)\}. \tag{A.2}$$

The claimed formula of $f_{V,D=1}(t|x)$ follows immediately from (A.2) and (A.1). A similar derivation gives $f_{V,D=2}(t|x)$. Finally, the subdensity attached to the singular part follows by subtracting the sum of $f_{V,D=1}(t|x)$ and $f_{V,D=2}(t|x)$ from the conditional density $f_V(\cdot|x)$:

$$f_{V,D=3}(t|x) = f_V(t|x) - f_{V,D=1}(t|x) - f_{V,D=2}(t|x).$$

□

Proof of Theorem 3.2. Due to the length of the proof, we outline the key steps and defer some technical details to Appendix B. We proceed by the following steps.

Step 1. First, consider the conditional survivor function of the minimum V:

$$S_V(t|x) = \psi\{\Lambda_0(t)\Phi_{12}(\phi_1(x), \phi_2(x))\},$$

which belongs to the GAFT model in Ridder (1990). We define $r(t, x) = S_V(t|x)$, $s(t, x) = \partial r(t, x)/\partial t$, and $q(t, z) = E[s(V, X)|V = t, r(V, X) = z]$. Then Corollary 2.1 in Jacho-Chavez, Lewbel, and Linton (2010) gives us

$$\Lambda_0(t)\Phi_{12}(\phi_1(x), \phi_2(x)) = \exp\left(\int_{r_1}^{r(t,x)} \frac{dz}{q(t_0, z)}\right).$$

We present the detailed verification of assumptions required by Jacho-Chavez, Lewbel, and Linton (2010) in Lemma B.3 in Appendix B. Invoking the normalization condition $\Lambda_0(t_1) = 1$, we have

$$\Phi_{12}(\phi_1(x), \phi_2(x)) = \exp\left(\int_{r_1}^{r(t_1,x)} \frac{dz}{q(t_0, z)}\right),$$

which becomes identified. The identification of $\Lambda_0(t)$ is also obvious. The outer link function ψ is the Laplace transform and it is identified given the variation of $\Lambda_0(t)$ on an open interval by Proposition 1 in Abbring and van den Berg (2003).

Step 2. From the conditional subdensity functions $f_{V,D=1}(t|x)$ and $f_{V,D=2}(t|x)$, we can identify

$$\Psi_k(x) = -\frac{f_{V,D=k}(t|x)}{\psi'(\Lambda_0(t)\Phi_{12}(\phi_1(x), \phi_2(x)))\lambda_0(t)}, \tag{A.3}$$

for any t and $k = 1, 2$. This leads to the identification of $\Phi_1(\phi_1(x))$ and $\Phi_2(\phi_2(x))$ because

$$\Phi_k(\phi_k(x)) = \Phi_{12}(\phi_1(x), \phi_2(x)) - \Psi_j(x), \quad j, k \in \{1, 2\}, k \neq j, \tag{A.4}$$

and $\Phi_{12}(\phi_1(x), \phi_2(x))$ has already been determined from Step 1. Given the multiplicative separable structure on

$$\Phi_k(\phi_k(x)) = \Phi_k\{\phi_{k,1}(x_1)\phi_{k,-1}(x_{-1})\}, \tag{A.5}$$

the same identification strategy in Step 1 can be adapted to point identify $\phi_{k,1}$ and $\phi_{k,-1}$ for $k = 1, 2$. The details are shown in Lemma B.4 of Appendix B.

Step 3. Given the identification of ϕ_k, we show that the marginal Lévy exponent functions Φ_k can be identified without the large support condition for ϕ_k for $k = 1, 2$. Recall that the marginal Lévy exponent function is

$$\Phi_k(z) = \int [1 - e^{-yz}] \Pi_k(dy), \quad \text{for } k = 1, 2,$$

which is a Bernstein function, meaning its derivative is completely monotone as in Theorem 3.2 in Schilling, Song, and Vondracek (2012). Thus, by Criterion 2 on page 417 in Feller (1966), the function $\exp(\Phi_k(z))$ is completely monotone and, therefore, real analytic. Hence, by Proposition 1 in Abbring and van den Berg (2003), we can identify $\Phi_k(z)$, as long as we get enough variation on some nonempty open sets as in Assumption 3.4. Thereafter, the marginal Lévy measures Π_k are also identified for $k = 1, 2$ by the uniqueness of Lévy–Khintchine representation in Sato (2013).

Step 4. It suffices to identify the Lévy copula function to fully pin down the characteristics of the Lévy subordinators. So far, we have already identified $\Phi_{12}(\phi_1(x), \phi_2(x))$ and $\phi_k(x)$ for $k = 1, 2$. Thus, $\Phi_{12}(\cdot, \cdot)$ is identified according to the bivariate version of Proposition 1 in Abbring and van den Berg (2003), given that it is real analytical as stated in Lemma B.2 in Appendix B. Therefore, we can identify the joint Lévy measure $\Pi_{12}(\cdot, \cdot)$. Finally, the Lévy copula function $\mathcal{C}_L$ is unique by the Sklar theorem in Kallsen and Tankov (2006) given the continuity of the tail integrals associated with two marginal Lévy measures. □

Appendix B: Auxiliary results and proofs of technical lemmas

In Appendix B, we first collect necessary notions and key theorems related to the Lévy copulas and Lévy exponent functions in Section B.1. A nonidentifiability result is stated in Section B.2. In Section B.3, we prove some auxiliary results that have been used in Appendix A and the identification results concerning the generalized model from Section 4. Section B.4 reports additional empirical results for the regression coefficients, when a third regressor, "age difference," is included.

B.1 *Lévy copula*

There are mainly two advantages of the Lévy copula compared with the standard distributional copula (Nelsen (2006)). First, the laws of bivariate Lévy subordinators are conveniently specified by their Lévy measures; only in a few cases the distribution or probability density function could be given in closed forms. Second, the distributional copula implied by the joint Lévy processes is typically time-varying, whereas the Lévy copula is time-invariant. Nevertheless, it does share certain similarities with the standard distributional copula function, including the Sklar theorem and Frechet–Hoeffding-type inequalities. We refer interested readers to Cont and Tankov (2004) and Kallsen and Tankov (2006) for the authoritative treatment.

For most Lévy processes, the jump intensity would grow or explode as the jump size converges to zero (Sato (2013)), therefore it is the tail integral or *survivor functional* version of Lévy measure that is more tractable. The two marginal Lévy tail integrals are simply defined as $\overline{\Pi}_k(y) = \int_y^\infty \pi_k(u)\, d\mu(u)$ with marginal Lévy densities as π_k for $k = 1, 2$. Before we formally introduce the two-dimensional tail integral, we need some terminology from Kallsen and Tankov (2006).

Definition B.1. A bivariate function $\mathcal{C}(\cdot, \cdot)$ is called 2-increasing if

$$\mathcal{C}(b_1, b_2) - \mathcal{C}(a_1, b_2) - \mathcal{C}(b_1, a_2) + \mathcal{C}(a_1, a_2) \geq 0,$$

for any $a_1 \leq b_1$ and $a_2 \leq b_2$.

Definition B.2. A bivariate function $\mathcal{C}(u_1, u_2)$ is said to be grounded if $\mathcal{C}(u_1, u_2) \neq +\infty$ for $(u_1, u_2) \neq (+\infty, +\infty)$.

Definition B.3 (Tail Integral). A two-dimensional tail integral is a function $\overline{\Pi} : \mathcal{R}_+^2 \to \mathcal{R}_+$ such that:

(1) $\overline{\Pi}$ is a 2-increasing function;
(2) $\overline{\Pi}$ is equal to zero if one of its arguments is equal to $+\infty$;
(3) $\overline{\Pi}$ is finite everywhere except possibly at zero.

With the tail integral in hand, we define the Lévy copula and highlight its role in connecting the joint and marginal Lévy measures.

Definition B.4. (Lévy copula) A two-dimensional Lévy copula is a 2-increasing grounded function $\mathcal{C}_L(u_1, u_2) : [0, \infty]^2 \to [0, \infty]$ with uniform margins, that is, $\mathcal{C}_L(u, \infty) = \mathcal{C}_L(\infty, u) = u$.

Theorem B.1 (Sklar theorem). *Let $\overline{\Pi}_{12}$ be a two-dimensional tail integral with margins $\overline{\Pi}_1, \overline{\Pi}_2$, then there exists a Lévy copula $\mathcal{C}_L$ such that*

$$\overline{\Pi}_{12}(u_1, u_2) = \mathcal{C}_L\big(\overline{\Pi}_1(u_1), \overline{\Pi}_2(u_2)\big). \tag{B.1}$$

If $\overline{\Pi}_1, \overline{\Pi}_2$ are continuous, then the copula function $\mathcal{C}_L$ is unique. Conversely, for a given Lévy copula $\mathcal{C}_L$ and two marginal tail integrals $\overline{\Pi}_1, \overline{\Pi}_2$, $\mathcal{C}_L(\overline{\Pi}_1(u_1), \overline{\Pi}_2(u_2))$ defines a two-dimensional tail integral.

Theorem B.2 (Frechet–Hoeffding inequalities). *For any bivariate Lévy subordinator, its Lévy copula $\mathcal{C}_L$ is bounded between two extreme cases*:

$$\mathcal{C}_{L,\perp}(u_1, u_2) \leq \mathcal{C}_L(u_1, u_2) \leq \mathcal{C}_{L,\parallel}(u_1, u_2), \tag{B.2}$$

where

$$\mathcal{C}_{L,\perp}(u_1, u_2) = u_1 \cdot I\{u_2 = \infty\} + u_2 \cdot I\{u_1 = \infty\}, \tag{B.3}$$

$$\mathcal{C}_{L,\parallel}(u_1, u_2) = \min\{u_1, u_2\}. \tag{B.4}$$

$\mathcal{C}_{L,\perp}(u_1, u_2)$ *is the independent Lévy copula and* $\mathcal{C}_{L,\parallel}(u_1, u_2)$ *stands for the completely dependent Lévy copula.*

Lemma B.1. *The following inequalities hold for any bivariate Lévy subordinator*:

$$\max\{\Phi_1(z_1), \Phi_2(z_2)\} \le \Phi_{12}(z_1, z_2) \le \Phi_1(z_1) + \Phi_2(z_2),$$

which induce $\Psi_j(\cdot) \ge 0$ *for* $j = 1, 2,$ *and* 3.

Proof. Note that $F_{z_1,z_2}(y_1, y_2) \equiv 1 - e^{-y_1 z_1 - y_2 z_2}$ is the joint distribution function of a pair of independent exponential random variables with hazard rates equal to (z_1, z_2). It is clear that

$$\begin{aligned}
&\Phi_{12}(z_1, z_2) \\
&\quad = \iint_{\mathcal{R}_+^2} \left[1 - e^{-y_1 z_1 - y_2 z_2}\right] d\mathcal{C}_L\big(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\big) \\
&\quad = -\iint_{\mathcal{R}_+^2} \mathcal{C}_L\big(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\big)\, dF_{z_1,z_2}(y_1, y_2) \\
&\quad \le -\iint_{\mathcal{R}_+^2} \mathcal{C}_{L,\perp}\big(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\big)\, dF_{z_1,z_2}(y_1, y_2) \\
&\quad = \Phi_1(z_1) + \Phi_2(z_2),
\end{aligned}$$

where integration-by-parts is used in the second equality and the first part of Frechet–Hoeffding inequalities has been applied. Similarly, we have

$$\begin{aligned}
&\Phi_{12}(z_1, z_2) \\
&\quad = -\iint_{\mathcal{R}_+^2} \mathcal{C}_L\big(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\big)\, dF_{z_1,z_2}(y_1, y_2) \\
&\quad \ge -\iint_{\mathcal{R}_+^2} \mathcal{C}_{L,\parallel}\big(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\big)\, dF_{z_1,z_2}(y_1, y_2) \\
&\quad = \iint_{\mathcal{R}_+^2} \left[1 - e^{-y_1 z_1 - y_2 z_2}\right] d\mathcal{C}_{L,\parallel}\big(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\big) \\
&\quad \ge \max\{\Phi_1(z_1), \Phi_2(z_2)\},
\end{aligned}$$

where the final inequality follows from the fact that

$$1 - e^{-y_1 z_1 - y_2 z_2} \ge \max\{1 - e^{-y_1 z_1}, 1 - e^{-y_2 z_2}\}, \tag{B.5}$$

and the integration is computed over the set $\{(y_1, y_2) : \overline{\Pi}_1(y_1) = \overline{\Pi}_2(y_2)\}$ on which the measure $\mathcal{C}_{L,\|}$ is supported. □

Here, we record a lemma stating the analytical property of the joint Lévy–Laplace exponent function, see also Begun and Yashin (2018).

Lemma B.2 (Theorem 2.1 in Brychkov, Glaeske, Prudnikov, and Tuan (1992)). *The bivariate Lévy–Laplace exponent function $\Phi_{12}(z_1, z_2)$ is an analytic function on*

$$H_b \equiv \{(z_1, z_2)|\mathbf{Re}(z_1) > -b, \mathbf{Re}(z_1) > -b\}$$

and, therefore, it is a real analytical function on $\mathbf{Re}H_b$ for any positive b. Since $\Phi_1(z) = \Phi_{12}(z, 0)$ and $\Phi_2(z) = \Phi_{12}(0, z)$, we also have $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ to be real analytical.

B.2 *A nonidentifiability result*

We state a nonidentifiability result of model primitives even when the conditional marginal distribution $S_{T_1}(t|x)$ or $S_{T_2}(t|x)$ is identified. Recall that the conditional marginal survivor function is

$$S_{T_k}(t|x) = \psi\{\Lambda_0(t)\Phi_k(\phi_k(x))\}, \quad k = 1, 2.$$

Without necessary normalization, we cannot separately identify ψ, Λ_0, Φ_k, and ϕ_k. The structure is in analog with the GAFT model, in which the observationally equivalent pairs can be constructed if the heterogeneity is drawn from the stable distribution (Lancaster (1979), Ridder (1990)). Our construction of observationally equivalent pairs makes use of the stable process.[14] The stable process is commonly parameterized by a power parameter $b \in (0, 1)$ and a scale parameter $a \in (0, 2]$, such that the Lévy exponent function is equal to $\Phi_S(u) = au^b$; see Aalen, Borgan, and Gjessing (2008).

Theorem B.3. *Let $a, \tilde{a} \in (0, 2]$ and $b, \tilde{b} \in (0, 1)$. For any given ψ, Λ_0, ϕ_k, and $\Phi_k(u) = au^b$, there exists an observationally equivalent construction consisting of $\tilde{\psi}$, $\tilde{\Lambda}_0$, $\tilde{\phi}_k$, and $\tilde{\Phi}_k(u) = \tilde{a}u^{\tilde{b}}$ for $k = 1, 2$ where*

$$\tilde{\psi}(u) = \psi(u^{1/l}), \tag{B.6}$$

$$\tilde{\phi}_k(x) = \phi_k^m(x), \tag{B.7}$$

$$\tilde{\Lambda}_0(t) = a^l \Lambda_0^l(t)/\tilde{a}, \tag{B.8}$$

for some positive numbers $l > 1$ and $m = bl/\tilde{b}$. The observational equivalence means

$$\psi\{\Lambda_0(t)\Phi_k(\phi_k(x))\} = \tilde{\psi}\{\tilde{\Lambda}_0(t)\tilde{\Phi}_k(\tilde{\phi}_k(x))\}, \tag{B.9}$$

for all t, x, and $k = 1, 2$.

[14] I would like to thank one knowledgeable referee who suggests this example.

Proof. As shown in problem 13 on page 439 in Feller (1966), for any Laplace transform $\psi(u)$ of ν and $l > 1$, the random variable $\nu^l\tau$ has a Laplace transform equal to $\psi(u^{1/l})$, where τ is an independent stable random variable with stable index equal to $1/l$. Therefore, $\tilde{\psi}(u)$ is a legitimate Laplace transform (Ridder (1990)). It is straightforward to verify both sets of model primitives lead to the same conditional marginal survivor function $S_{T_k}(t|x)$ for $k = 1, 2$. □

B.3 *Auxiliary lemmas and proofs*

We prove three auxiliary lemmas that have been used in the proof of Theorems 3.2 and 4.3. The first two lemmas verify the conditions in Jacho-Chavez, Lewbel, and Linton (2010).

Lemma B.3. *If Assumption 2.1 to Assumption 2.6 and Assumption 3.1 to 3.4 hold, then we get*

$$\Lambda_0(t)\Phi_{12}\big(\phi_1(x), \phi_2(x)\big) = \exp\left(\int_{r_1}^{r(t,x)} \frac{dz}{q(t_0, z)}\right), \tag{B.10}$$

where $r(t, x) = S_V(t|x)$ *and* $q(z, t) = E[s(V, X)|V = t, r(V, X) = z]$ *with* $s(t, x) = \partial r(t, x)/\partial t$.

Proof. We shall verify Assumption (I*) in Jacho-Chavez, Lewbel, and Linton (2010). First, the set of (I*1) is directly satisfied by assumptions. Considering the set of (I*2)(i), the outer link function ψ is a strictly monotonic and continuously differentiable function, as the Laplace tranform is know to be completely monotone, that is, it is infinitely order differentiable with derivatives of alternating signs. As for the two normalization conditions within the current context, it is true that $\Lambda_0(t_0)/\lambda_0(t_0) = 1$ and $\psi(1) = r_1$ for some constant r_1. □

Lemma B.4. *If Assumption 2.1 to Assumption 2.6 and Assumption 3.1 to 3.4 hold, then we get*

$$\phi_{k,1}(x_1)\phi_{k,-1}(x_{-1}) = \exp\left(\int_{r_{2,k}}^{\zeta_k(x_1,x_{-1})} \frac{dz}{\eta_k(x_1^0, z)}\right), \tag{B.11}$$

where $\zeta_k(x_1, x_{-1}) = \Phi_k(\phi_k(x_1, x_{-1}))$ *and* $\eta_k(x_1, z) = E[\xi_k(x_1, z)|\zeta_k(X_1, X_{-1}) = z, X_1 = x_1]$ *with* $\xi_k(x_1, x_{-1}) = \partial\zeta_k(x_1, x_{-1})/\partial x_1$.

Proof. By a similar argument as in the proof of Lemma B.3, one gets identification of $\phi_{k,1}(x_1)\phi_{k,-1}(x_{-1})$ where the outer link function is the Laplace exponent function Φ_k for $k = 1, 2$. Since the Laplace exponent function has a derivative being completely monotone, the required monotonicity and differentiability are again satisfied. □

The remaining part is about the proofs of theorems regarding the generalization in Section 4.

Proof of Theorem 4.1. Given the assumed nonnegative $\lambda_k(t|W)$ for $k=1,2$, both first passage times are driven by increasing trends crossing exponential thresholds. Thus, we can proceed in the same way as the proof of Theorem 2.1.

Conditional both on the observable covariates and heterogeneity terms, we have

$$\begin{aligned}
&P\{T_1>t_1,T_2>t_2|x,\bar{w},\mathbf{L},\nu\}\\
&\quad=P\bigl\{\epsilon_1>\Lambda_1(t_1|w)+\phi_1(x)L_1\bigl(\nu\Lambda_0(t_1)\bigr),\epsilon_2>\Lambda_2(t_2|w)+\phi_2(x)L_2\bigl(\nu\Lambda_0(t_2)\bigr)|x,\bar{w},\mathbf{L},\nu\bigr\}\\
&\quad=e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}e^{-\phi_1(x)L_1(\nu\Lambda_0(t_1))-\phi_2(x)L_2(\nu\Lambda_0(t_2))}\\
&\quad=e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}e^{-\phi_1(x)L_1(\nu\Lambda_0(t_1))-\phi_2(x)L_2(\nu\Lambda_0(t_1))}e^{-\phi_2(x)[L_2(\nu\Lambda_0(t_2))-L_2(\nu\Lambda_0(t_1)))]},
\end{aligned}$$

for $t_1\le t_2$. After integrating out the time-varying and static heterogeneity terms, one has

$$\begin{aligned}
&P\{T_1>t_1,T_2>t_2|x,\bar{w}\}\\
&\quad=e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}\psi\bigl\{\Lambda_0(t_1)\Phi_{12}\bigl(\phi_1(x),\phi_2(x)\bigr)+\bigl[\Lambda_0(t_2)-\Lambda_0(t_1)\bigr]\Phi_2\bigl(\phi_2(x)\bigr)\bigr\}.
\end{aligned}$$

Over the region where $t_1\ge t_2$, one could proceed in a similar fashion to obtain

$$\begin{aligned}
&P\{T_1>t_1,T_2>t_2|x,\bar{w}\}\\
&\quad=e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}\psi\bigl\{\Lambda_0(t_2)\Phi_{12}\bigl(\phi_1(x),\phi_2(x)\bigr)+\bigl[\Lambda_0(t_1)-\Lambda_0(t_2)\bigr]\Phi_1\bigl(\phi_1(x)\bigr)\bigr\}.
\end{aligned}$$

The desired formula of the conditional joint survivor function follows by combining two cases together. To get the conditional marginal survivor function, we simply set either t_1 or t_2 as zero. □

Proof of Theorem 4.2. The conditional survivor function of the minimum V is obtained by letting $t_1=t_2=t$ in Theorem 4.1. Considering the case where $D=1$, the conditional subdensity function is calculated by

$$f_{V,D=1}(t|x,\bar{w})=-\lim_{s\uparrow t}\frac{\partial}{\partial s}S_{T_1,T_2}(s,t|x,\bar{w}).$$

Over the range where $t_1<t_2$, the conditional joint survivor function is

$$S_{T_1,T_2}(t_1,t_2|x,\bar{w})=e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}\psi\bigl\{\Psi_1(x)\Lambda_0(t_1)+\bigl[\Psi_2(x)+\Psi_3(x)\bigr]\Lambda_0(t_2)\bigr\}.$$

The claimed identify of $f_{V,D=1}(t|x,\bar{w})$ follows immediately from preceding two formulas. A similar derivation gives $f_{V,D=2}(t|x,\bar{w})$. Moreover, the conditional density $f_V(t|x,\bar{w})$ of the minimum V can be computed directly by differentiating $-S_V(t|x,\bar{w})$:

$$\begin{aligned}
f_V(t|x,\bar{w})={}&\bigl(\lambda_1(t|w)+\lambda_2(t|w)\bigr)e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}\Gamma(t,x)\\
&+e^{-\Lambda_1(t_1|w)-\Lambda_2(t_2|w)}\gamma(t,x)\Phi_{12}\bigl(\phi_1(x),\phi_2(x)\bigr).
\end{aligned}$$

Finally, the subdensity attached to the singular part follows by subtracting the sum of $f_{V,D=1}(t|x,\bar{w})$ and $f_{V,D=2}(t|x,\bar{w})$ from $f_V(t|x,\bar{w})$:

$$f_{V,D=3}(t|x,\bar{w})=f_V(t|x,\bar{w})-f_{V,D=1}(t|x,\bar{w})-f_{V,D=2}(t|x,\bar{w}).$$ □

Lemma B.5. *If Assumption 2.1 to Assumption 2.6, Assumption 3.1 to 3.4, and Assumption 4.1 hold, then we get*

$$\lim_{x\to x_0} f_{V,D=j}\big(t|x,\bar{w}(t)\big)=\lambda_j(t|w)e^{-\Lambda_1(t|w)-\Lambda_2(t|w)},\quad j=1,2,$$

and

$$\lim_{x\to x_0} S_V\big(t|x,\bar{w}(t)\big)=e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}.$$

Proof. The results follow from analytical properties about the Laplace transformation and Lévy–Laplace exponent function. Specifically, we have

$$\lim_{u\to 0}\psi(u)=1, \tag{B.12}$$

and

$$\lim_{z\to 0}\Phi_k(z)=0,\quad k=1,2. \tag{B.13}$$

Meanwhile, for the joint Lévy–Laplace exponent, we get

$$\lim_{z_1\to 0,z_2\to 0}\Phi_{12}(z_1,z_2)=0. \tag{B.14}$$

The first fact in (B.12) is well known, since $\psi(0)=1$ and $\psi(\cdot)$ is completely monotone hence continuous at 0. The statement regarding Lévy subordinator in (B.13) follows from the preceding argument, because the Laplace exponent function is simply the negative logarithm of the Laplace transform of $L_k(t)$ evaluated at $t=1$. The conclusion in (B.14) follows from an analogous result regarding the bivariate Laplace transform. Now, it is clear that for any t,

$$\lim_{x\to x_0}\Gamma(t,x)=1, \tag{B.15}$$

and

$$\lim_{x\to x_0}\Psi_j(x)=0, \tag{B.16}$$

for $j=1,2$, and 3, which give rise to the stated results. □

Proof of Theorem 4.3. We start with an extra step to identify $\Lambda_1(t)$ and $\Lambda_2(t)$ as follows.

For any realization of the time-varying covariates $w(t)$, we have shown the following results hold:

$$\lim_{x\to x_0} f_{V,D=j}\big(t|x,\bar{w}(t)\big)=\lambda_j(t|w)e^{-\Lambda_1(t|w)-\Lambda_2(t|w)},\quad j=1,2, \tag{B.17}$$

and

$$\lim_{x\to x_0} S_V\big(t|x,\bar{w}(t)\big)=e^{-\Lambda_1(t|w)-\Lambda_2(t|w)}. \tag{B.18}$$

Thus, the derivatives of two idiosyncratic trend functions are both identified following the classical result of Berman (1963) as

$$\lambda_j(t|w) = \frac{\lim_{x\to x_0} f_{V,D=j}\big(t|x,\bar{w}(t)\big)}{\lim_{x\to x_0} S_V\big(t|x,\bar{w}(t)\big)}, \quad j=1,2. \tag{B.19}$$

Then the time-varying coefficient can be identified by the standard least squares argument:

$$\beta_k(t) = E\big[W(t)W^\top(t)\big]^{-1}E\big[W(t)\lambda_k(t|W)\big], \quad k=1,2, \tag{B.20}$$

following Aalen (1980) for additive hazards models.

Given the identification of $\Lambda_1(t|w)$ and $\Lambda_2(t|w)$ and the equality

$$S_V\big(t|x,\bar{w}(t)\big)e^{\Lambda_1(t|w)+\Lambda_2(t|w)} = \psi\big\{\Lambda_0(t)\Phi_{12}\big(\phi_1(x),\phi_2(x)\big)\big\}, \tag{B.21}$$

we can point identify $\psi(u)$, $\Lambda_0(t)$ and $\Phi_{12}(\phi_1(x),\phi_2(x))$ as in Step 1 of the proof of Theorem 3.2. The rest of the proof goes through with no essential change. Starting from the conditional subdensities, we can pin down $\Phi_k(\phi_k(x))$ for $k=1,2$. The identifiability of covariates effect and characteristics of Lévy subordinators follow the same argument in the proof of Theorem 3.2, mutatis mutandis. □

B.4 *Additional empirical results*

In this section, we report estimation results with the inclusion of an additional regressor X_{i3}, the age difference between the couples as in Honoré and de Paula (2018). Recall that individual durations are measured in terms of *family age*, which is set to zero when the older partner of the couple reaches age 60 and the age difference allows us to track the duration of the other spouse. The original age difference is measured as the husband's age minus wife's age in years and we use the standardized version to avoid the overflow and discreteness, for example, $X_3 =$ (Age Difference−Mean(Age Difference))/std(Age Difference). For the data set in Honoré and de Paula (2018), we also must exclude those observations whose age differences are recorded as *"Inf,"* which leaves us with overall 798 observations.

The implementation of our semiparametric estimation extend in a straightforward way. For example, the conditional subdistribution function is estimated by

$$\widehat{F}_{V,D=j}(t|x) = \frac{\sum_{i=1}^n \mathbb{I}\{V_i \le v, D=j\}\mathbf{K}\big((X_i - x)/\mathbf{h_x}\big)}{\sum_{i=1}^n \mathbf{K}\big((X_i - x)/\mathbf{h_x}\big)}, \quad j=1,2,3,$$

where $X_i = (X_{i1}, X_{i2}, X_{i3})^\top$ and $\mathbf{K}((X_i - x)/\mathbf{h_x}) \equiv \prod_{j=1}^3 K((X_{ij} - x_j)/h_{x_j})$. Regarding the bandwidth choices for each coordinate in the kernel estimation, we still follow the Sheather–Jones rule in estimating the conditional subdensity functions and adopt the

Table 2. Estimated effects of covariates.

Covariates	Estimated Coeff.	S.E.
Age difference-wife	−0.460	0.339
Health expenditure-wife	0.879	0.316
Financial wealth-wife	0.131	0.192
Age difference-husband	0.195	0.232
Health expenditure-husband	0.912	0.381
Financial wealth-husband	0.360	0.386

Ruppert–Sheather–Wand choice in the local quadratic regression or local linear regression. When it comes to the estimation of $\widehat{q}(\cdot,\cdot)$, we set the number of neighborhoods in the k-nearest-neighbor to be the largest integer not exceeding $n^{2/5}$.

The estimated coefficients with the bootstrap standard errors are reported in Table 2. The reported standard errors are calculated by perturbing all the nonparametric estimates with the Bayesian bootstrap weights (Rubin (1981)), based on 100 replications. The effects from the total health expenditure or the financial wealth exhibit patterns very similar to the ones reported in the main text excluding the age difference, in terms of both the sign and magnitude. As expected, the standard errors have all increased (except for the health expenditure for the husband) due to the less precise nonparametric estimates in the presence of the additional regressor. In particular, both coefficients associate with the age difference for wives and husbands are insignificant. Nonetheless, the estimated signs are consistent with the model from Honoré and de Paula (2018), so that the age difference tends to increase the retirement hazard for men and decrease it for women. This empirical result can be understood as Honoré and de Paula (2018) did, by noting that men are typically older and that "family age" is counted from the 60th year of the older partner; thus a larger age difference implies that the wife is younger at time zero and less likely to retire at any "family age" than an older woman would be.

References

Aalen, O. (1980), "A model for nonparametric regression analysis of counting processes." In *Mathematical Statistics and Probability Theory*, 1–25. [554, 571]

Aalen, O., O. Borgan, and H. Gjessing (2008), *Survival and Event History Analysis: A Process Point of View*. Springer. [535, 567]

Abbring, J. H. (2012), "Mixed hitting-time models." *Econometrica*, 80 (2), 783–819. [535, 538, 540, 549, 550]

Abbring, J. H. and G. J. van den Berg (2003), "The identifiability of the mixed proportional hazards competing risks model." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (3), 701–710. [536, 548, 550, 563, 564]

Abbring, J. H. and Y. Yu (2015), "The empirical content of synchronization games." Working paper. [537, 538, 543]

Alvarez, F., K. Borovickova, and R. Shimer (2016), "Decomposing duration dependence in a stopping time model." NBER working paper, No. 22188. [538]

An, M. Y., B. J. Christensen, and N. D. Gupta (2004), "Multivariate mixed proportional hazard modelling of the joint retirement of married couples." *Journal of Applied Econometrics*, 19 (6), 687–704. [536, 537, 543, 547, 556]

Arnold, B. and P. Brockett (1983), "Identifiability for dependent multiple decrement/competing risk models." *Scandinavian Actuarial Journal*, 2, 117–127. [562]

Basu, A. and J. Ghosh (1978), "Identifiability of the multinomial and other distributions under competing risks model." *Journal of Multivariate Analysis*, 8, 413–429. [549]

Bedford, T. and I. Meilijson (1997), "A characterization of marginal distributions of (possibly dependent) lifetime variables which right censor each other." *The Annals of Statistics*, 25, 1622–1645. [551]

Begun, A. and A. Yashin (2018), "Study of the bivariate survival data using frailty models based on Lévy processes." *AStA Advances in Statistical Analysis*, 1–31. [549, 567]

Berman, S. (1963), "Note on extreme values, competing risks and semi-Markov processes." *The Annals of Mathematical Statistics*, 34, 1104–1106. [556, 571]

Botosaru, I. (2016), "A duration model with dynamic unobserved heterogeneity." Working paper. [538]

Bresnahan, T. F. and P. C. Reiss (1991), "Entry and competition in concentrated markets." *Journal of Political Economy*, 99, 977–1009. [554]

Brychkov, Y., H. Glaeske, A. Prudnikov, and V. Tuan (1992), *Multidimensional Integral Transformations*. Gordon and Breach Science Publisher. [567]

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models." In *Handbook of Econometrics*, Vol. 6, 5549–5632, Elsevier. [553]

Chiappori, P. A., I. Komunjer, and D. Kristensen (2015), "Nonparametric identification and estimation of transformation models." *Journal of Econometrics*, 188 (1), 22–39. [551]

Choi, J. (1991), "Dynamic R&D competition under hazard rate uncertainty." *The RAND Journal of Economics*, 22, 596–610. [537, 543, 549]

Coile, C. (2004), "Health shocks and couples' labor supply decisions." NBER working paper, No. 10810. [538, 543, 556, 559]

Cont, R. and P. Tankov (2004), *Financial Modelling With Jump Processes*. Chapman & Hall. [538, 539, 540, 565]

Crowder, M. (2012), *Multivariate Survival Analysis and Competing Risks*. CRC Press. [542, 548]

de Paula, Á. (2009), "Inference in a synchronization game with social interactions." *Journal of Econometrics*, 148, 56–71. [535, 538, 561]

Drepper, B. and G. Effraimidis (2015), "Social interaction effects in duration models." Working paper. [543]

Ekeland, I., J. Heckman, and L. Nesheim (2004), "Identification and estimation of hedonic models." *Journal of Political Economy*, 112 (1), 60–109. [551]

Elbers, C. and G. Ridder (1982), "True and spurious duration dependence: The identifiability of the proportional hazard model." *The Review of Economic Studies*, 49 (3), 403–409. [541, 545, 553]

Epifani, I. and A. Lijoi (2010), "Nonparametric priors for vectors of survival functions." *Statistica Sinica*, 20, 1455–1484. [547]

Fan, J. and I. Gijbels (1996), *Local Polynomial Regression*. Chapman and Hall. [552, 553]

Fan, Y. and R. Liu (2018), "Partial identification and inference in censored quantile regression." *Journal of Econometrics*, 206, 1–38. [536, 547, 548]

Farzin, Y., K. Huisman, and P. Kort (1998), "Optimal timing of technology adoption." *Journal of Economic Dynamics and Control*, 22, 779–799. [537, 538, 543]

Feller, W. (1943), "On a general class of "contagious" distributions." *The Annals of Mathematical Statistics*, 14, 389–400. [541]

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications*, Vol. 2. John Wiley & Sons. [564, 568]

Flinn, C. and J. Heckman (1982), "New methods for analyzing structural models of labor force dynamics." *Journal of Econometrics*, 18, 115–168. [536, 537]

Gjessing, H., O. Aalen, and N. Hjort (2003), "Frailty models based on Lévy processes." *Advances in Applied Probability*, 35, 532–550. [538]

Heckman, J. J. and B. E. Honoré (1989), "The identifiability of competing risks model." *Biometrika*, 76, 325–330. [536, 548]

Heckman, J. J. and B. E. Honoré (1990), "The empirical content of the Roy model." *Econometrica: Journal of the Econometric Society*, 58, 1121–1149. [536, 556]

Heckman, J. J. and S. Navarro (2007), "Dynamic discrete choice and dynamic treatment effects." *Journal of Econometrics*, 136, 341–396. [535]

Heckman, J. J. and B. Singer (1984a), "The identifiability of the proportional hazard model." *The Review of Economic Studies*, 51 (2), 231–241. [541, 545]

Heckman, J. J. and B. Singer (1984b), "A method for minimizing the impact of distributional assumptions in econometric models for duration data." *Econometrica*, 52, 271–320. [542]

Heckman, J. J. and C. Taber (1994), "Econometric mixture models and more general models for unobservables in duration analysis." *Statistical Methods in Medical Research*, 3, 279–299. [548]

Honoré, B. E. (1993), "Identification results for duration models with multiple spells." *Review of Economic Studies*, 60, 241–246. [547]

Honoré, B. E. and Á. de Paula (2010), "Interdependent durations." *Review of Economic Studies*, 77, 1138–1163. [535, 536, 540, 547, 548, 557]

Honoré, B. E. and Á. de Paula (2018), "A new model for interdependent durations with an application to joint retirement." *Quantitative Economics*, 9, 1299–1333. [535, 537, 542, 543, 547, 548, 556, 557, 559, 561, 571, 572]

Honoré, B. E. and A. Lleras-Muney (2006), "Bounds in competing risks models and the war on cancer." *Econometrica*, 74, 1675–1698. [536, 551]

Horowitz, J. L. (1996), "Semiparametric estimation of a regression model with an unknown transformation of the dependent variable." *Econometrica*, 64, 103–137. [552, 553]

Horowitz, J. L. (1999), "Semiparametric estimation of a proportional hazard model with unobserved heterogeneity." *Econometrica*, 67 (5), 1001–1028. [542]

Horowitz, J. L. and W. Hardle (1996), "Direct semiparametric estimation of single-index models with discrete covariates." *Journal of the American Statistical Association*, 91, 1632–1640. [557]

Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models." *Journal of Econometrics*, 58. [551]

Jacho-Chavez, D., A. Lewbel, and O. Linton (2010), "Identification and nonparametric estimation of a transformed additively separable model." *Journal of Econometrics*, 156 (2), 392–407. [550, 551, 552, 557, 558, 563, 568]

Jensen, R. (1982), "Adoption and diffusion of an innovation of uncertain profitability." *Journal of Economic Theory*, 27. [537, 543]

Joe, H. (2014), *Dependence Modeling With Copulas*. CRC Press. [545]

Kallsen, J. and P. Tankov (2006), "Characterization of dependence of multidimensional Lévy processes using Lévy copulas." *Journal of Multivariate Analysis*, 97 (7), 1551–1572. [536, 538, 540, 551, 564, 565]

Khan, S. and E. Tamer (2009), "Inference on endogenously censored regression models using conditional moment inequalities." *Journal of Econometrics*, 152, 104–119. [536, 551]

Lancaster, T. (1972), "A stochastic model for the duration of a strike." *Journal of the Royal Statistical Society: Series A*, 135, 257–271. [535, 538]

Lancaster, T. (1979), "Econometric methods for the duration of unemployment." *Econometrica*, 47, 939–956. [535, 567]

Lee, S. and A. Lewbel (2013), "Identification of a competing risks model with unknown transformations of latent failure times." *Econometric Theory*, 29 (5), 905–919. [536, 548]

Liu, R. (2020), "Supplement to 'A competing risks model with time-varying heterogeneity and simultaneous failure'." *Quantitative Economics Supplemental Material*, 11, https://doi.org/10.3982/QE1159. [557]

Lundberg, S. and R. A. Pollak (1994), "Noncooperative bargaining models of marriage." *The American Economic Review*, 84, 132–137. [557]

Marshall, A. W. and I. Olkin (1967), "A multivariate exponential distribution." *Journal of the American Statistical Association*, 62, 30–44. [536, 546]

Marshall, A. W. and I. Olkin (1988), "Families of multivariate distributions." *Journal of the American Statistical Association*, 83, 834–841. [536, 545]

Marshall, A. W. and M. Shaked (1979), "Multivariate shock models for distributions with increasing hazard rate average." *Annals of Probability*, 7, 343–358. [541, 546]

Martinussen, T. and T. H. Scheike (2002), "A flexible additive multiplicative hazard model." *Biometrika*, 89 (2), 283–298. [555]

McFadden, D. (1974), "Conditional logit analysis of qualitative choice analysis." In *Frontiers in Econometrics*, 105–142, Academic Press. [544]

Nelsen, R. B. (2006), *An Introduction to Copulas*, second edition. Springer. [540, 544, 564]

Oakes, D. (1989), "Bivariate survival models induced by frailties." *Journal of the American Statistical Association*, 84, 487–493. [536, 546, 547]

Peterson, A. (1976), "Bounds for a joint distribution function with fixed sub-distribution functions: Applications to competing risks." *Proceedings of the National Academy of Sciences USA*, 73, 11–13. [551]

Powell, J., J. Stock, and T. Stoker (1989), "Semiparametric estimation of index coefficients." *Econometrica*, 57, 1403–1430. [551, 552, 553]

Reinganum, J. (1981), "Dynamic games of innovation." *Journal of Economic Theory*, 25, 21–41. [537, 538, 542, 543, 549]

Reinganum, J. (1982), "A dynamic game of R and D: Patent protection and competitive behavior." *Econometrica*, 50, 671–688. [537, 542, 543, 549]

Renault, E., T. van den Heijden, and B. Werker (2014), "The dynamic mixed hitting-time model for multiple transaction prices and times." *Journal of Econometrics*, 180, 233–250. [561]

Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models." *Review of Economic Studies*, 57, 167–181. [542, 549, 563, 567, 568]

Ridder, G. and I. Tunali (1999), "Stratified partial likelihood estimation." *Journal of Econometrics*, 92, 193–232. [554]

Ridder, G. and T. M. Woutersen (2003), "The singularity of the information matrix of the mixed proportional hazard model." *Econometrica*, 71 (5), 1579–1589. [542, 550]

Rubin, D. (1981), "Bayesian bootstrap." *The Annals of statistics*, 9, 130–134. [559, 572]

Ruppert, D., S. J. Sheather, and M. P. Wand (1995), "An effective bandwidth selector for local least squares regression." *Journal of the American Statistical Association*, 90, 1257–1270. [558, 559]

Ryu, K. (1993), "Structural duration analysis of management data." *Journal of Econometrics*, 57, 91–115. [538, 542]

Sato, K. (2013), *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press. [536, 539, 564, 565]

Schilling, R., R. Song, and Z. Vondracek (2012), *Bernstein Functions: Theory and Applications*. SIAM. [564]

Sheather, S. J. and M. C. Jones (1991), "A reliable data-based bandwidth selection method for kernel density estimation." *Journal of the Royal Statistical Society: Series B (Methodological)*, 53, 683–690. [558]

Stock, J. (1988), "Estimating continuous-time processes subject to time deformation: An application to postwar US GNP." *Journal of the American Statistical Association*, 83, 77–85. [541]

Stoker, T. (1986), "Consistent estimation of scaled coefficients." *Econometrica*, 54, 1461–1481. [551]

Tsiatis, A. A. (1975), "A nonidentifiability aspect of the problem of competing risks." *Proceedings of the National Academy of Sciences USA*, 72, 20–22. [548, 549]

van den Berg, G. (2001), "Duration models: Speciifcation, identification, and multiple durations." In *Handbook of Econometrics*, Vol. 5, 3381–3460, Elsevier. [536, 538, 545, 546]

Wood, S. (2003), "Thin plate regression splines." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 95–114. [553, 560]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 20 June, 2018; final version accepted 16 September, 2019; available online 25 September, 2019.