© Copyright 2015

Ruixuan Liu

Identification and Inference of Duration Models Driven By Stochastic Processes

Ruixuan Liu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Yanqin Fan, Chair

Jihyung Lee

Jon A. Wellner

Program Authorized to Offer Degree:

Economics

University of Washington

Abstract

Identification and Inference Of Duration Models Driven By Stochastic Processes

by Ruixuan Liu

Chair of Supervisory Committee:

Professor Yanqin Fan
Department of Economics

I study a new class of duration models driven by stochastic processes. In contrast with the standard hazard-based models, the duration outcome or survival time is defined to be the the first time a Lévy subordinator—a stochastic process with stationary, independent and non-negative increments—crosses a random threshold. Such a model is of substantial interest because not only is it related to optimal stopping time models where agents optimally time their discrete actions, but it also applies to the scenario in which the termination of duration is caused by some gradual and irreversible accumulation of damage.

Chapter 1 presents the univariate version of our model. Surprisingly, when the multiplicative effect from covariates is parameterized as a linear index, our model reduces to a single-index Cox proportional model, where the unknown link function captures the characteristic of the latent process. The large sample property of a sieve maximum partial likelihood estimator of the finite dimensional parameter in this single-index Cox proportional hazard model is studied with right censored data. It is shown that the proposed estimator is asymptotically normal with root-n convergence rate and it achieves the semiparametric efficiency bound.

Chapter 2 proposes a new bivariate competing risks model that specifies each individual duration as the first passage time of a marginal Lévy subordinator, and impose dependence between two durations via the dependent bivariate Lévy subordinators. With semiparametric restrictions on the structural multiplicative effects from observable covariates, we identify these effects, the baseline hazard function, and the characteristics of latent Lévy

subordinators separately in a competing risks setting where one does not observe both durations sequentially. For the index model, I present two root-n consistent and asymptotically normal estimators for the finite dimensional parameters in structural multiplicative effects based on certain average derivatives estimation.

Chapter 3 demonstrates our machinery on a real dataset about the joint retirement decisions of married couples, where the existing hazard-based or process-based duration models do not apply because a significant portion of couples choose to retire simultaneously.

TABLE OF CONTENTS

List of Figures **iii**

List of Tables **iv**

Chapter 1: A Single-Index Cox Model Driven By Lévy Subordinators **1**

1.1 Introduction 1

1.2 Literature Review 4

1.3 A Single-Index Cox Model Driven By Lévy Subordinators 6

1.3.1 A Unified Perspective 6

1.3.2 A New Duration Model 10

1.4 Sieve Partial Likelihood Estimation 13

1.5 Theoretical Results 14

1.5.1 Semiparametric Information Bound 16

1.5.2 Large Sample Properties 19

1.5.3 Bootstrap Consistency 22

1.6 Appendix A 25

Chapter 2: A Competing Risks Model with Time-Varying Heterogeneity and Simultaneous Stopping **41**

2.1 Introduction 41

2.2 Related Works 44

2.3 An Extended Marshall-Olkin (EMO) Model 46

2.3.1 Construction of EMO 46

2.3.2 The Original Marshall-Olkin Model and Existing Generalizations . . . 50

2.3.3 Dependence Properties 51

2.4 Semiparametric Identification and Estimation 57
2.4.1 Empirical Content of EMO under Competing Risks 57
2.4.2 A Conditional Moment Based Estimator 60
2.4.3 A Local Partial Likelihood Based Estimator 63
2.5 Appendix B. 67
2.6 Appendix C. 80
2.7 Appendix D. 89
2.7.1 Identifiability 89
2.7.2 Observationally Equivalent Components 96
2.8 Appendix E. A Review of Lévy Processes 98

Chapter 3: An Empirical Application & Future Work **104**

3.1 Joint Retirement of Couples 104
3.2 Conclusion 106

Bibliography **109**

LIST OF FIGURES

1.1 FPT of Gamma Processes 38
1.2 FPT of Stable Processes 39
1.3 FPT of Compound Poisson Processes 40

LIST OF TABLES

2.1 Comparisons of Three Models 103

3.1 Estimates of Joint Retirement Model 108

ACKNOWLEDGMENTS

I am deeply indebted to my mentor and supervisor Prof. Yanqin Fan for her continuing guidance and support, especially during the transition period in my 4th year from Vanderbilt to University of Washington. I also appreciate insightful comments from Jaap Abbring, Irene Botosaru, Yu-Chin Chen, Emmanuel Guerre, Jihyung Lee, Qi Li, Eric Renault, Thomas Richardson, Aman Ullah, and Jon Wellner.

DEDICATION

To My Family

Chapter 1

A SINGLE-INDEX COX MODEL DRIVEN BY LÉVY SUBORDINATORS

1.1 Introduction

Since the seminal work of Cox (1972), the Cox proportional hazard model (PH) has unarguably been the focal point in the duration or survival analysis. PH is semiparametric in nature where the hazard density function of the duration outcome or survival time T with covariate X is specified in a multiplicative form:

$$\begin{aligned} \lambda_T(t|x) &\equiv \lim_{\delta \to 0} \frac{1}{\delta} P\{t \leq T < t+\delta | T \geq t, X = x\} \\ &= \lambda(t) \exp\left(x'\beta\right), \end{aligned} \tag{1.1.1}$$

depending on a nonparametric baseline hazard density function $\lambda(t)$ and a parametric index $x'\beta$ via the exponential link function. Partial likelihood method (Cox, 1975) is used to estimate the regression coefficient β without estimating the baseline hazard function first, given right censored survival data. With the estimator of β in hand, the estimation of the baseline cumulative hazard function $\Lambda(t) = \int_0^t \lambda(s)\, ds$ readily follows from Breslow's proposal (Breslow, 1972). This estimation procedure delivers root-n consistent and asymptotic normal estimators for both the finite dimensional regression parameter and cumulative baseline hazard function, see Andersen and Gill (1982). More importantly, the estimators are semiparametric efficient in the sense of Begun, Hall, Huang and Wellner (1983). Despite its theoretical elegance, Cox PH comes with several notable restrictions. First of all, the covariates' effect is linear in the conditional log-hazard function. Second, standard Cox regression does not incorporate any unobservable heterogeneity term. Last but not least, PH may not be applicable to structural durations triggered by stochastic processes, which nat-

urally arises from optimal stopping time problems in economics, such as in firms' entry/exit choice (Dixit, 1989), worker's unemployment decision (Mortensen and Pissarides, 1994), and real option type investment (McDonold and Siegels, 1986). Indeed, developments have been made in the literature extending the scope of Cox PH in all three directions. Maintaining the proportional hazard structure, the functional form $\exp\left(x^{'}\beta\right)$ can be relaxed to a completely nonparametric version as in Tibshirani and Hastie (1987), Fan, Gijbels and King (1997), or semiparametric variants in Dabrowska (1997), Huang (1999), and Fan, Lin, and Zhou (2006). Introducing a multiplicative heterogeneity term in the conditional hazard density function, one arrives at Lancaster's (1979) mixed proportional hazard model (MPH). Finally, the recent surge of process-based duration models of Abbring (2012) or Botosaru (2013) also borrow several insights from Cox's original construction, while being able to capture certain time-varying heterogeneity.

In this chapter, I propose a new univariate duration model that inherits prominent features from all three branches. Here the structural duration is defined as the first time a Lévy subordinator—a stochastic process with stationary, independent and non-negative increments—crosses an exponential threshold, and the effect from observable covariate is acting multiplicatively on the latent process. Such a model is of substantial interest because not only is it related to optimal stopping time models where agents optimally time their discrete actions, but it also applies to the scenario in which the termination of duration is caused by some gradual and irreversible accumulation of damage. In fact, I provide a unified framework of duration modelling, where the duration outcome is specified as the first passage time of a latent stochasitic process crossing over a random threshold, including both traditional hazard-based models and recent process-based models. Thereafter, our model could be seen as a natural variant of the mixed proportional hazard model from a process point of view. The model is constructed in a highly structural way based on the threshold-crossing behavior of the latent process; nevertheless, its reduced form is very tractable. When the multplicative effect is parameterized as a linear index, our model exhibits a single-index structure via the conditional hazard function:

$$\lambda_T\left(t|x\right) = \lambda\left(t\right)\exp\left[\psi\left(x^{'}\beta\right)\right] \tag{1.1.2}$$

where the unknown smooth function $\psi(\cdot)$ is the so-called link function. This single-index model incorporate nonlinear effect while alleviating the curse of dimension when multivariate covariates are considered. The restrictive parametric link function is avoided, while greater estimation precision for finite dimensional parameter β is still possible, in contrast to fully nonparametric estimation. The unknown link function is not just motivated from technical relaxation, but it is related to the characteristics of the latent stochastic process.

Even though the single-index structure is one of the most popular semiparametric modelling choice for conditional mean (Ichimura, 1993) or conditional quantile function (Chen, 2007), and its extension to the Cox model has already been envisioned earlier (Ichimura, 1993, Chen, 2007, and Ding and Nan, 2011), rigorous large sample theory for (1.1.2) has not been presented in a completely satisfactory fashion when right censored data is available. Chen, Li, and Wang (1999) are among the first to consider the model in (1.1.2) based on sliced inverse regression technique. Wang (2004) suggests to iterate certain kernel based estimator while assuming some preliminary root-n consistent estimator for β is available beforehand. Huang and Liu (2006) have proposed maximizing sieve partial likelihood with ψ being replaced by a polynomial spline function, but the number of knots is held to be fixed. Consequently, they treat the link function exactly to be some spline function under consideration and derive parametric type asymptotics. In this paper, I fill this theoretical gap and show Huang and Liu's (2006) proposal also works nicely when the number of knots is growing with sample size and those spline functions are merely viewed as approximations of the true link function. In particular, the sieve partial likelihood estimator of β is proved to be asymptotically normal with root-n convergence rate and it achieves the semiparametric efficiency bound. Furthermore, I present the efficient score function and information bound in explicit forms, and prove the bootstrap consistency with general exchangable bootstrap weights.

We are now in a position to describe the agenda for this chapter. Section 1.2 contains a brief survey of the related literature. Section 1.3 provides a unified framework of the duration modelling, bridging the conceptual gap between the hazard-based models and the process-based models. Thereafter, I present the single-index Cox model driven by Levy subordinators and its identifiability. Section 1.4 describes the estimation procedure

in Huang and Liu (2006), while Section 1.5 containing main theoretical results. Several technical lemmas are collected in Appendix in Section 1.6. For a vector v, we let $v^{'}$ be the transpose of it and $v^{\otimes 2} = vv^{'}$ in this chapter.

1.2 Literature Review

In this section, we review the related literature extending the scope of Cox model in three categories.

Even though the baseline function is left unspecified in Cox model, the effect of covariates on the (log-)hazard function is restricted to be linear. Obviously, any misspecification of the effect of covariates would lead to inconsistent estimation and misleading inference, which motivates consideration of flexible nonparametric or semiparametric methods. For a nonparametric smooth risk function $\phi(x)$ measuring the covariates' effect, Tibshirani and Hastie (1987) initiate the local partial likelihood for estimating the nonparametric Cox model. Subsequently, Fan, Gijbels and King (1997), Chen and Zhou (2007) have developed large sample theory for estimating the derivative and the difference of the unknown risk function. For various semiparametric siblings, we refer the readers to the partial linear model in Dabrowska (1997), partial linear and additive model in Huang (1999), functional ANOVA model in Huang, Kooperberg, Stone, and Truong (2000), and a varying-coefficient model in Fan, Lin, and Zhou (2006).

The characterization of individual heterogeneity is a central issue in microeconometrics. Ignoring the unobservable heterogeneity would inevitably bias the estimation and interpretation of the effect from explanatory variables. Lancaster's (1979) mixed proportional hazard (MPH) model has become a standard toolkit allowing researchers to incorporate both observed and unobserved variables into the conditional hazard function of the duration outcome. Relaxing the parametric assumption in Lancaster (1979), Elbers and Ridder (1982), Heckman and Singh (1984) have demonstrated the possibility of identifying MPH with a single spell of duration in a completely nonparametric way, further broadening its scope. Assuming the heterogeneity follows the gamma distribution as in Lancaster (1979), Murphy (1995) proposes a semiparametric MLE and proves its root-n consistency, asymp-

totic normality, and semiparametric efficiency for both the proportional effect parameter and the cumulative baseline hazard function. Without parameterizing the heterogeneity term's distribution function, the semiparametric efficiency bound of MPH for the finite dimensional parameters in multiplicative effect and baseline hazard function could be singular as first realized by Hahn (1994), which excludes the possibility of obtaining root-n consistent estimators. Ridder and Woutersen (2003) present sufficient conditions which avoid this singularity. For this regular semiparametric model, Bearse, Canals-Cerda, and Rilstone (2007), Hausman and Woutersen (2014) have proposed different root-n consistent estimators for the finite dimensional parameters.

Although the hazard based duration models play dominant roles in empirical applications, economic theory in general does not lead to the above specification, which complicates the structural interpretation of various reduced-form estimates (van den Berg, 2001). A typical duration of interest to economists very often arises from solving an optimal stopping time problem, where optimizing economic agents make decisions about the time at which to switch from one state to another, such as in firms' entry/exit choice (Dixit, 1989), worker's unemployment decision (Mortensen and Pissarides, 1994), and real option type investment (McDonold and Siegels, 1986). In all those aforementioned cases, the structural duration is related to some threshold crossing behavior of a latent stochastic process (Abbring, 2012). Indeed what is usually disregarded in the traditional approach to duration or survival analysis is that the particular duration is the end point of some process (Aalen, Borgan, and Gjessing, 2007; Lee and Whitmore, 2006). Parallel to the traditional hazard-based modeling strategy, many process-based models have been developed by making use of a specific parametric sub-class of Lévy processes. For example, the Brownian motion appears in Lancaster's (1972) strike model and Whitmore's (1979) job tenure model, while the gamma process has been utilized in Singpurwalla (1995), Lawless and Crowder (2004) to model the degradation process. Abbring (2012) has pioneered the study of identifiability of process-based duration models, without making any parametric assumptions. Abbring (2012) specifies the duration formally as first passage time of a random threshold by the sample path of a latent stochastic process in his MHT. Abbring (2012) has demonstrated that not only MHT incorporates the time-varying heterogeneity represented by the latent

process, but also it arises more naturally from the optimal stopping time problem where the solution could be described by this threshold-crossing behavior. When the process is chosen to be the spectral negative Lévy process, the identifiability of all structural components could be shown, adapted from the strategies analyzing the identifiability of MPH. So far the identification of MHT is restricted to the univariate duration model, and the suggested estimation is fully parametric. Botosaru (2013) also proposes another univariate duration model making use of the Lévy subordinator, which could be either viewed as a hazard-based model with certain random hazard density or a process-based model with the latent process being a double stochastic integral of the Lévy subordinator. Identification in Botosaru's (2013) random hazard density model relates to solving a nonlinear integral equation. For the identified case, Botosaru (2013) presents a sieve type estimator and proves its consistency.

1.3 A Single-Index Cox Model Driven By Lévy Subordinators

We provide a unified perspective on the structural duration model described by the threshold-crossing behavior of some latent process, including MPH in Lancaster (1979), MHT in Abbring (2012), and a random hazard density model in Botosaru (2013). The construction of our model in the univariate case is also presented, which could be seen as a natural variant of MPH from a process point of view.

1.3.1 A Unified Perspective

Now we start with a generic duration model where the duration is related to the threshold crossing behavior of a latent stochastic process. Formally, the structural duration is defined as the first passage time of a latent process $\Lambda^*(t)$ crossing a random threshold e, with structural proportional effect $\phi(X)$ acting on the process:

$$T \equiv \inf\left\{t : \phi(X)\,\Lambda^*(t) \geq e\right\}. \tag{1.3.1}$$

In all models discussed in the paper, X, $\Lambda^*(t)$ and e are assumed to be mutually independent. The rule specified in (1.3.1) resembles the classical binary choice model closely in

a dynamic setup.[1] But intead of making a discrete choice based on static comparisons, it stems from the optimal stopping time model where agent optimally chooses when to switch to another state at time Y^* with full information, as in de Paula and Honore (2010). Suppose the economic agent is solving a utility maximization problem where the current state generates a payoff to the agent with e and switching to another state gives $\phi(X)\Lambda^*(t)$.[2] Therefore the structural duration in (1.3.1) is the optimal solution of

$$\max_T \left\{ \int_0^T e \exp(-\rho s)\, ds + \int_T^\infty \phi(X)\Lambda^*(t) \exp(-\rho s)\, ds \right\}$$

with the discount rate equal to ρ.

Even though the Cox model or MPH is constructed from a regression modeling on the hazard rate, it is interesting to ask whether it could be put into the same framework where the duration is specified as the first passage time of a latent process. Indeed this is possible. This alternative point of view further reveals the restrictions on the latent process and threshold imposed by MPH and it provides the link of our specification to MPH.

Example 1.3.1 *(Cox Model) It turns out that the duration in Cox model could be equivalently defined as in (1.3.1)*

$$T \equiv \inf \left\{ t : \exp\left(X'\beta\right) \Lambda(t) \geq e \right\} \tag{1.3.2}$$

by setting $\Lambda^*(t)$ equal to the cumulative baseline hazard function $\Lambda(t)$. Furthermore, the multiplicative effect is parameterized as $\exp\left(X'\beta\right)$, and the threshold follows unit exponential distribution, i.e., $e \sim Exp(1)$. The assertion is easy to see as

$$\begin{aligned} \Pr\{T > t | X = x\} &= \Pr\left\{ e > \exp\left(x'\beta\right) \Lambda(t) \,|X = x \right\} \\ &= \exp\left(-\Lambda(t) e^{x'\beta}\right). \end{aligned}$$

This alternative view is favored by Singpurwalla (2006) who has advocated this interpretation of the cumulative function $\Lambda(t)$ is more sensible than merely treating it as the primitive of hazard density.

---

[1] In discrete-time setting, Heckman and Navarro (2007) have constructed a general mixture duration model based on a latent process crossing thresholds.

[2] Because only the comparison plays a role from the agent's perspective, it is not restrictive to let one side to become time invariant.

Example 1.3.2 *(MPH) The duration in a completely nonparametric MPH could be also defined as in (1.3.1) by*

$$Y^* \equiv \inf\left\{t : \phi(X)\,\zeta\Lambda(t) \geq e\right\}, \tag{1.3.3}$$

where the latent process is equal to $\Lambda^*(t) = \zeta\Lambda(t)$ *with the cumulative baseline hazard function* $\Lambda(t)$*. Again, the threshold follows unit exponential distribution, i.e.* $e \sim Exp(1)$*.*

In those two classical hazard-based models, the driving stochastic process is either deterministic (Cox Model) or its randomness is pinned down at time zero (the frailty term ζ is static in MPH). In MPH given the realization of ζ, we will just have a a deterministic and nondecreasing trend $\Lambda(t)$ approaching the threshold from below. The individual heterogeneity does not evolve over time along the entire spell of duration, hence it is certainly desirable to construct a model with time varying heterogeneity. Apparently, one has to make some assumptions on the underlying class for $\Lambda^*(t)$, in order to maintain a tractable structure. In the sequel, we shall restrict our attention to general Lévy processes.

Lévy processes constitute a very rich and attractive class of stochastic processes, including the commonly encountered Brownian motion, gamma process and stable process as special cases. They have attracted considerable attention due to the flexibility for a wide variety of modeling issues in finance, insurance and engineering. In the study of durations, parallel to the traditional hazard based modeling, many process based models have been developed by making use of a specific parametric sub-class of Lévy processes. For example, Brownian motion appears in Lancaster's (1972) strike model and Whitmore's (1979) job tenure model, while gamma process has been utilized in Singpurwalla (1995), Lawless and Crowder (2004) to model the degradation process. Abbring (2012) has pioneered the study of identifiability of process based duration models, without making any parametric assumptions. Before reviewing his contribution, it is worthwhile to introduce the formal definition of a $d-$dimensional Lévy process and two important subclasses. To avoid digressions, we refer the readers to Sato (1999) for more detailed discussions on Lévy process theory.

Definition 1.3.3 *A* $d-$*dimensional Lévy process* $\mathbf{L}(t)$ *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* $\mathbf{L}(r)$*.*

Definition 1.3.4 *A $d-$dimensional Lévy subordinator $\mathbf{L}(t)$ is a Lévy process with almost surely nondecreasing sample path, i.e., for $t \geq s$ one has $\mathbf{L}(t) \geq \mathbf{L}(s)$.*

Definition 1.3.5 *A $d-$dimensional spectral negative Lévy process $\mathbf{L}(t)$ is a Lévy process with no positive jumps.*

The most remarkable property of a Lévy process is that even it's purely defined in terms of descriptive features about the sample path, it admits a very concrete analytic characterization via the well-known Lévy-Khintchine representation, see Sato f(1999). Specifically, its Laplace transform could be expressed as

$$E\left\{\exp\left[-zL(t)\right]\right\} = \exp\left[-t\Phi(z)\right],$$

where $\Phi(\cdot)$ is the so-called Lévy-Laplace exponent function, which is nonparametric and *time-invariant*. This elegant analytic characterization plays a crucial role in identification problems of those duration models driven by Lévy processes, including Abbring (2012) and Botosaru (2013).

Example 1.3.6 *Motivated from the optimal stopping time problems, Abbring (2012) starts with a model where the structural duration is formally defined to be*

$$T \equiv \inf\left\{t : \phi(X)\,\tilde{L}(t) \geq e\right\}, \tag{1.3.4}$$

where e is an arbitrary random threshold and $\tilde{L}(t)$ is a spectral negative Lévy process. Assuming independence of three components $\left(e, \tilde{L}, X\right)$, the empirical content of MHT is revealed through the conditional Laplace transform $\mathcal{L}_{Y^*}(\cdot|X)$ of the structural duration Y^*:

$$\mathcal{L}_T(s|X) \equiv E\left\{\exp(-sT)\,|X\right\} = \mathcal{L}_e\left(\tilde{\Lambda}(s)\,\phi(X)\right), \tag{1.3.5}$$

*where $\mathcal{L}_e$ is the Laplace transform of e and $\tilde{\Lambda}(s)$ is the largest root satisfying $\Phi\left(\tilde{\Lambda}(s)\right) = s$. Despite that there are quite different functions (i.e. $\mathcal{L}_T(s|X)$ and $\tilde{\Lambda}(s)$) appearing in (1.3.5), the mathematical structure is almost identical to (**??**) in MPH. Assuming regular variations of certain functions, Abbring (2012) has shown one could identify the triple $\left(\mathcal{L}_e, \tilde{\Lambda}, \phi\right)$ from $\mathcal{L}_T(s|X)$ nonparametrically.*

Example 1.3.7 *Given our discussion on linking the hazard-based models to their threshold-crossing behavior over the unit exponential threshold, it is clear now that the duratoin in Botosaru's (2013) random hazard rate model could be expressed as*

$$T \equiv \inf \left\{ t : \phi \left(X \right) \Lambda^{*} \left(t \right) \geq e \right\}$$

the latent crossing process is a double stochastic integral of the Lévy subordinator, as $\Lambda^{*} \left(t \right) = \int_0^t \int_0^s f \left(u \right) dL \left(u \right) ds$ *with a transformation function* $f \left(u \right)$*. When the process* $L \left(u \right)$ *is taken to be a Lévy subordinator, the conditional survival function could be written as*

$$S_T \left(t|x \right) = \exp \left[- \int_0^t \Phi \left(\phi \left(x \right) f \left(u \right) \left(t - u \right) \right) du \right],$$

where $\Phi \left(\cdot \right)$ *is the Lévy-Laplace exponent function determined by* $L \left(u \right)$*. The nonparametric identification of* $\left(\Phi, f, \phi \right)$ *is related to solving a nonlinear Volterra integral equation of the first kind with unknown kernel, see Botosaru (2013).*

1.3.2 A New Duration Model

Inspired by Abbring (2012) and Botosaru (2013), I construct a new duration model under the framework of (1.3.1), replacing the product of the cumulative baseline hazard function and the frailty term $\zeta \Lambda \left(t \right)$ in MPH all together with a latent stochastic process $L \left(\Lambda \left(t \right) \right)$. Here $L \left(\cdot \right)$ is a Lévy subordinator and $\Lambda \left(t \right)$ is a deterministically increasing time-change function. This time-change function could be seen as a suitable transformation from certain abstract time scale that the process evolves to the rate of economic transactions (Kyprianou, 2006), which turns out to be the cumulative baseline hazard function later on. Hence the duration in our model becomes

$$T \equiv \inf \left\{ t : \phi \left(X \right) L \left(\Lambda \left(t \right) \right) \geq e \right\}, \tag{1.3.6}$$

where the random threshold is still assumed to be unit exponentially distributed. Notice the latent Lévy subordinator $L \left(\cdot \right)$ has replaced the static frailty term ζ and its sample path property inherits the key feature of a monotonic driving trend in MPH. Another close connection to MPH is that the distribution of frailty term is often taken to be infinitely

divisible, see Hougaard (1986a). Correspondingly in our present model, it's well known the Lévy subordinator has non-negative infinitely divisible distribution (Sato, 1999).

The conditional survival function of Y^* in (1.3.6) could be derived in a straightforward way by the Lévy-Khintchine representation:

$$S_T(t|x) = \exp\left[-\Lambda(t)\,\Phi(\phi(x))\right]. \tag{1.3.7}$$

Now we present some examples with the latent Lévy subordinator restricted to be chosen from specific parametric sub-classes.

Example 1.3.8 *When the Lévy subordinator $L(\cdot)$ is gamma process, our model could be seen as a natural variant of MPH with gamma frailty term as originally in Lancaster (1979). One nice property of the gamma frailty model is that the conditional distribution of frailty term ζ, given survival until any time (i.e. conditional on $Y^* > t$), is also gamma with the original shape parameter, see van den Berg (2001). Abbring and van den Berg (2007) provide another justification, showing that the distribution of frailty term among survivors would always converge to a gamma distribution upon suitable normalization when $t \to \infty$. In Figure 1.1, we plot a typical realization of the first passage time or structural duration with gamma process in our model (1.1.2). Its conditional survival function is*

$$S_T(t|x) = \exp\left[-\Lambda(t)\right]\left(1 + \frac{\phi(x)}{\upsilon}\right)^{\rho},$$

where (υ, ρ) collect the scale and shape parameter in the gamma distribution.

Insert Figure 1.1 Here

Example 1.3.9 *When the Lévy subordinator $L(\cdot)$ is stable process, our model could be seen as a natural variant of MPH with stable frailty term as originally in Hougaard (1986a). One advantage of the stable frailty model under MPH is that one still gets a proportional hazard model when integrating out ζ (see Section 5.2 in van den Berg, 2001), which is not the case for the gamma frailty model. In Figure 1.2, we plot a typical realization of the first passage time or structural duration with stable process in our model (1.1.2). Also its conditional survival function is*

$$S_T(t|x) = \exp\left[-\Lambda(t)\,\upsilon\phi(x)^{\rho}\right],$$

where (υ, ρ) *collect the stable index and power parameter in the stable distribution. Compared with gamma process, the sample path of stable process is more irregular and exhibits larger jumping magnitude.*

Insert Figure 1.2 Here

Example 1.3.10 *In Figure 1.3, we plot a typical realization of the first passage time or structural duration with compound Poisson process in our model (1.1.2). It is well known that compound Poisson process has piece-wise constant sample path, and it is the only process whose Lévy measure is finite. In the absence of covariate* X*, our model boils down to the random shock model studied by Esary, Marshall and Proschan (1973). Under this circumstance, the marginal survival function of* T *could be determined explicitly by model primitives without referring to the Lévy-Khintchine representation. Here the Lévy subordinator is* $L(t) = \sum_{i=1}^{N(t)} \xi_i$*, where the arrival of shocks is governed by a (homogeneous) Poisson process* $N(t)$ *with hazard rate* λ *and individual i.i.d.shocks* $\{\xi_i\}$ *are assumed to non-negative. Thus, the marginal survival function of* Y^* *is equal to*

$$\Pr\{T > t\} = \sum_{k=0}^{\infty} \frac{(\lambda t)^k \exp(-\lambda t)}{k!} \overline{P}_k,$$

where $\overline{P}_k$ *is the probability of surviving after* k *shocks:*

$$\overline{P}_k = \Pr\{\xi_1 + ... + \xi_k \leq e\}.$$

Insert Figure 1.3 Here

Notice when the multiplicative effect $\phi(x)$ is parameterized as in the Cox regression $\exp\left(x'\beta\right)$, we actually arrive at a single-index proportional hazard model with unknown link function $\psi(\cdot) = \log \circ \Phi \circ \exp(\cdot)$ in (1.1.2). The identification of all model primitives in (1.3.7) could be achieved without the identification-at-limit strategy, in contrast with Elbers and Ridder (1982), Heckman and Singer (1984), or Abbring (2012). It is well-known the finite dimensional parameter β is only identified upto scale in the single-index model (Ichimura, 1993), so we adopt the following convention by fixing the first element to be 1. Thus, we partition the covariate vector $X = \left(X_1, X_2'\right)'$ with a univariate component X_1,

and stack β with one 1 as $\widetilde{\beta} = \left(1, \beta^{'}\right)^{'}$. Moreover, we denote $X_0 = X_1 + X_2^{'}\beta_o$ with the true parameter β_o. The regularity conditions below required for identification are fairly weak.

Assumption (I1). The support of X is a convex set with at least one interior point.

Assumption (I2). The support of X_0 contains a nonempty open interval.

Assumption (I3). $E\left[\Phi\left(\phi\left(X\right)\right)\right] = 0$.

Proposition 1.3.11 *Let $\phi\left(x\right) = x^{'}\widetilde{\beta}$ in (1.3.7) and assume Assumptions (I1)-(I3), then the triple (β, Φ, Λ) is identifiable.*

1.4 Sieve Partial Likelihood Estimation

In practice, survival data is often right censored due to termination of the study or early withdrawal from the study, so we observe the random sample consisting of $i.i.d$ data of $Z_i = (V_i, \Delta_i, X_i)$ where $V_i = \min\left(T_i, C_i\right)$ and $\Delta_i = I\left[T_i \leq C_i\right]$. The logarithm of partial likelihood function neglecting terms independent of our parameters is

$$l_n\left(\beta, \psi\right) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i\left\{\psi\left(X_{i1} + X_{i2}^{'}\beta\right) - \log\left(\sum_{k:V_k \geq V_i}\exp\left[\psi\left(X_{k1} + X_{k2}^{'}\beta\right)\right]\right)\right\}$$

since the smooth link function ψ is unknown, it is natural to replace it by an approximating B-spline function $\psi_n\left(\cdot\right)$ (Schumaker, 1981). Let $T_{K_n} = \{t_1, ..., t_{K_n}\}$ be a set of partition points of $[a, b]$ with $K_n = O\left(n^{\upsilon}\right)$ and $\max_j\left|t_j - t_{j-1}\right| = O\left(n^{-\upsilon}\right)$ for some constant $\upsilon \in (0, 1/2)$. Let $\mathcal{S}_n\left(T_{K_n}, K_n, p\right)$ be the space of polynomial splines of order $p \geq 1$, then there exists a set of B-spline basis functions $\{B_j, 1 \leq j \leq q_n\}$ with $q_n = K_n + p$ s.t. for any $s \in \mathcal{S}_n\left(T_{K_n}, K_n, p\right)$ we can write $s\left(\cdot\right) = \sum_{j=1}^{q_n}\gamma_j B_j\left(\cdot\right)$.

Now with $\psi_n\left(\cdot\right) = \sum_{j=1}^{q_n}\gamma_j B_j\left(\cdot\right)$ substituting ψ, we are maximizing the following criterion function in terms of parameters (β, γ):

$$l_n\left(\beta, \gamma\right) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i\left\{\sum_{j=1}^{q_n}\gamma_j B_j\left(X_{i1} + X_{i2}^{'}\beta\right) - \log\left(\sum_{k:V_k \geq V_i}\exp\left[\sum_{j=1}^{q_n}\gamma_j B_j\left(X_{k1} + X_{k2}^{'}\beta\right)\right]\right)\right\}.$$

A Newton-Raphson algorithm or any gradient-based search algorithm can be applied to solve for the score equations to get estimators $\left(\widehat{\beta}, \widehat{\gamma}\right)$. Given the identification restriction

of ψ, we will center the estimator as follows. Let

$$\widehat{\psi}_n^*(\cdot) = \sum_{j=1}^{q_n} \widehat{\gamma}_j B_j(\cdot) \text{ and } \overline{\psi}_n^*(\cdot) = \frac{\sum_{i=1}^n \Delta_i \widehat{\psi}_n^*(\cdot)}{\sum_{i=1}^n \Delta_i}.$$

The resulting estimator of ψ is a centered version and defined to be

$$\widehat{\psi}_n(\cdot) \equiv \widehat{\psi}_n^*(\cdot) - \overline{\psi}_n^*(\cdot),$$

so it satisfies $\sum_{i=1}^n \Delta_i \widehat{\psi}_n \left(X_{i1} + X_{i2}' \widehat{\beta} \right) = 0$, see a similar procedure in Huang (1999). We refer readers to Huang and Liu (2006) for detailed description of the computatoin and simulation results.

1.5 Theoretical Results

A close examination of the exisiting literature on semiparametric estimation theory reveals the novelty and technicality associated with (1.1.2). In contrast with Huang (1999), we are dealing with the bundled parameter $(\beta, \eta(\cdot, \beta))$ where $\eta(x, \beta) = \psi\left(x' \widetilde{\beta}\right)$ in the present scenario. This terminology is used by Huang and Wellner (1996) referring to statistical models where the parameter of interest and nuisance parameter are bundled together. General theory for handling semiparametric estimation with bundled parameter has been developed by Ding and Nan (2011), Nan and Wellner (2013). In our estimation, the sample criterion function is chosen to be the logrithm of partial likelihood function, which is not an average of *i.i.d.* terms as in Ding and Nan (2011), nor weakly dependent as in Chen (2007). Compared with Nan and Wellner (2013), we have to approximate the addtional unknown function $\psi(\cdot)$ by spline sieves. All that said, our large sampel theory for (1.1.2) is indeed a successful marriage of the aforementioned work.

Now we introduce some necessary notations. Let P_n be the empirical measure of $Z_i \equiv (V_i, \Delta_i, X_i)$ and let P be the associated probability measure. Let $P_{\Delta n}$ be the subprobability empirical measure of O_i when $\Delta_i = 1$, with its population version $P_{\Delta n}$. The linear functional notation is convenient, as $P_{\Delta n} f = \int \Delta f dP_n = n^{-1} \sum_{i=1}^n \Delta_i f(Z_i)$. Recall that we let the true parameters be denoted as $(\beta_o, \psi_o, \lambda_o)$, and $X = \left(X_1, X_2'\right)'$ and $X_0 = X_1 + X_2' \beta_o$. Furthermore, stack β with one 1 as $\widetilde{\beta} = \left(1, \beta'\right)'$. Some standard counting process notations in Andersen and Gill (1982) would also facilitate the presentation, so

$N_i(t) = I[V_i \leq t, \Delta_i = 1]$, $Y_i(t) = I[V_i > t]$ and

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp\left[\psi_o\left(X_{i1} + X_{i2}'\beta_o\right)\right] \lambda_o(u)\, du, \tag{1.5.1}$$

are the corresponding counting process, at-risk process and associated martingale respectively. The following important observation by Sasieni (1992a) would appear repeatedly in the remaining section

$$E[K(t)\, g(Z)] = E[g(Z)\,|V = t, \Delta = 1] \tag{1.5.2}$$

where

$$K(t) = \frac{Y(t) \exp\left[\psi_o\left(X_{i1} + X_{i2}'\beta_o\right)\right]}{E\left\{Y(t) \exp\left[\psi_o\left(X_{i1} + X_{i2}'\beta_o\right)\right]\right\}}. \tag{1.5.3}$$

The regularity assumptions are listed below.

Assumption (A1). The finite dimensional parameter space $\mathcal{B}$ is a compact subset of $\mathcal{R}^d$ and the true parameter β_o is an interior point of $\mathcal{B}$. Also, β_o has Euclidean norm equal to 1 and its first coordinate is positive.

Assumption (A2). The covariate X has bounded support and for any $\beta \neq \beta_o$ we have $\Pr\left\{X_2'\beta \neq X_2'\beta\right\} > 0$. $EX_2^{\otimes 2}$is a strictly positive definite matrix.

Assumption (A3). The truncation time $\tau < \infty$ satisfies $\Lambda_o(\tau) = \int_0^\tau \lambda_o(t)\, dt < \infty$, moreover $\Pr(\Delta = 1|X) > 0$ and $\Pr(T > \tau|X) > 0$ almost surely with respect to the probability measure of X.

Assumption (A4). Let $0 < c_1 < c_2 < \infty$ be two constants. The joint sub-density $f\left(t, x'\widetilde{\beta}_o, \Delta = 1\right)$ satisfies $c_1 \leq f\left(t, x'\widetilde{\beta}_o, \Delta = 1\right) < c_2$ for all $\left(t, x'\widetilde{\beta}_o\right) \in [0, \tau] \times [a, b]$.

Assumption (A5). For two arbitrary univariate functions f_1, f_2, we have

$$f_1(V) + f_2(X_0) = 0, a.s.\text{-}P_\Delta \text{ if and only if } f_1(V) = f_2(X_0) = 0, a.s.\text{-}P_\Delta$$

Assumption (A6). Let Ψ^p denote the collection of bounded functions ψ on $[a, b]$ with bounded derivatives $\psi^{(j)}$, $j = 1, ..., k$ and the k-th derivative $\psi^{(k)}$ satisfies the following Lipschitz condition:

$$\left|\psi^{(k)}(s) - \psi^{(k)}(t)\right| \leq L\,|s - t|^\alpha \text{ for } \forall s, t \in [a, b]$$

where k is a positive integer and $\alpha \in (0, 1]$ s.t. $p = k + \alpha \geq 3$ and $L < \infty$. The true unknown link function $\psi_o(\cdot) \in \Psi^p$.

1.5.1 Semiparametric Information Bound

In this section, we calculate the semiparametric information bound for the estimation of β. We refer readers to Bickel, Klaassen, Ritov, and Wellner (1993) for an authoritative and book-length treatment of the semiparametric information bound for parameters in infinite dimensional models. In the standard single-index model on the conditional moment, the calculation of the efficient score function is done by projecting onto a single tangent space (Ichimura, 1993; Newey and Stoker, 1993). In contrast, we have two nuisance nonparametric components in the model. Thereafter, we have to consider the projection onto a sum-space of two non-orthogonal tangent spaces. The iterated projection presented below could be viewed as a variant of the classical result of Frisch and Waugh (1933) to obtain the least-squares estimates for a sub-vector, which states the estimated regression parameter of interest is algebraically equal to run the least-squares regression of one set of residuals against the other. We let the true parameters be denoted as $(\beta_o, \psi_o, \lambda_o)$, and $X = \left(X_1, X_2'\right)'$ and $X_0 = X_1 + X_2'\beta_o$. Furthermore, stack β with one 1 as $\widetilde{\beta} = \left(1, \beta'\right)'$.

First, note that the log-likelihood function for a sample of size one is

$$\Delta\psi\left(X'\widetilde{\beta}\right) + \Delta\lambda\left(V\right) - \exp\left[\psi\left(X'\widetilde{\beta}\right)\right]\Lambda\left(V\right),$$

dropping the terms which do not involve parameters of interest. Consider a parametric smooth submodel $\{\lambda_\chi(\cdot) : \chi \in \mathcal{R}\}$ and $\{\psi_\gamma(\cdot) : \gamma \in \mathcal{R}\}$ that runs through the true model, i.e., $\lambda_0(\cdot) = \lambda_o(\cdot)$ and $\psi_0(\cdot) = \psi_o(\cdot)$. Moreover, we let

$$a\left(\cdot\right) = \frac{\partial}{\partial\chi}\log\lambda_\chi\left(\cdot\right) \text{ and } h\left(\cdot\right) = \frac{\partial}{\partial\gamma}\log\psi_\gamma\left(\cdot\right)$$

represent possbile directions that can approach the true model. Thereafter following Sasieni (1992ab), the score vector for regression parameter β_o is

$$\begin{aligned}\dot{l}_\beta\left(Z\right) &= \Delta\psi\left(X'\widetilde{\beta}\right)X_2 - \exp\left[\psi\left(X'\widetilde{\beta}\right)\right]\Lambda\left(V\right)\dot{\psi}\left(X'\widetilde{\beta}\right)X_2 \\ &= \int\dot{\psi}\left(X'\widetilde{\beta}\right)X_2 dM\left(t\right),\end{aligned}$$

where $M(t)$ is the natural martingale in (1.5.1). Likewise, we have the following two score operators for the nonparametric components:

$$\dot{l}_\psi h\left(Z\right) = \int h\left(X'\widetilde{\beta}\right)dM\left(t\right),$$

$$\dot{l}_{\lambda} a(Z) = \int a(t)\, dM(t).$$

where $a(\cdot) = \frac{\partial}{\partial \chi} \log \lambda_{\chi}(\cdot)$ and $h(\cdot) = \frac{\partial}{\partial \gamma} \log \psi_{\gamma}(\cdot)$ are possible directions approaching $\lambda(\cdot)$ and $\psi(\cdot)$ respectively from some index sets. The natural Hilbert spaces where where those functions $a(\cdot)$ and $h(\cdot)$ sit in are

$$L_{2,V} = \left\{ a : E\left[\Delta a^2(V)\right] < \infty \right\},$$

$$L^0_{2,X_0} = \left\{ h : E\left[\Delta h(X_0)\right] = 0, E\left[\Delta h^2(X_0)\right] < \infty \right\}.$$

Hence the two tangent sets are

$$A_{\lambda} = \left\{ \dot{l}_{\lambda} a : a \in L_{2,V} \right\},$$
$$H_{\psi} = \left\{ \dot{l}_{\psi} h : h \in L^0_{2,X_0} \right\}.$$

In this circumstance, there are two nonparametric components whose tangent sets are not orthogonal. We shall rely on the techniques in Sasieni (1992b) to find the efficient score function for the regression parameter β_o.

Theorem 1.5.1 *The efficient score for estimating β in the single-index Cox model is*

$$\dot{l}^*_{\beta}(Z) = \int \left[\dot{\psi}(X_0) X_2 - a^*(t) - h^*(X_0) \right] dM(t)$$

where $a^*(\cdot)$ and $h^*(\cdot)$ are the unique functions minimizing

$$E\Delta \left\| \dot{\psi}(X_0) X_2 - a(V) - h(X_0) \right\|^2.$$

Here they take the following forms with

$$a^*(t) = E\left[\dot{\psi}(X_0) X_2 - h^*(X_0) \,|V = t, \Delta = 1 \right]$$

and $h^*(\cdot)$ satisfies

$$E\left\{ \dot{\psi}(X_0) X_2 - h^*(X_0) - E\left[\dot{\psi}(X_0) X_2 - h^*(X_0) \,|V = t, \Delta = 1 \right] |X_0 = x_0, \Delta = 1 \right\}$$
$$= 0, \quad a.s.\ w.r.t\ P_{X_0}.$$

Moreover, the semiparametric information bound for estimating β_o is

$$I^*(\beta_o) = E\left[\dot{l}^*_{\beta}(Z)^{\otimes 2} \right].$$

Proof. The efficient score function is of the orthogonal projection of $\dot{l}_\beta$ to the closure of the nuisance sum-space $\overline{A_\lambda + H_\psi}$, hence we need to find the least favorable direction (a^*, h^*) such that $\dot{l}_\beta - \dot{l}_\lambda a^* - \dot{l}_\psi h^*$ is orthogonal to the sumspace $A_\lambda + H_\psi$. Assumptions (A3)-(A5) together with Proposition 1 in Sasieni (1992b) guarantee this projection is well-defined. Furthermore, Proposition 3 in Sasieni (1992b) points out an explicit way to calculate this orthogonal projection as

$$\Pi\left[\dot{l}_\beta|\left(A_\lambda + H_\psi\right)^\perp\right] = \Pi\left[\Pi\left[\dot{l}_\beta|\left(A_\lambda\right)^\perp\right]|\Pi\left[H_\psi|\left(A_\lambda\right)^\perp\right]\right].$$

Hence, we shall first eliminate the hazard function by projecting the scores l_β and l_ψ onto the tangent space for the hazard then and projecting the residual of the projection of l_β onto the sumspace generated by the residual score of the projection of l_ψ. By Theorem 1 of Sasieni (1992b), the two residual scores are

$$K_\beta = \int D_X\left(X_0, t\right) dM\left(t\right),$$

and

$$Kh = \int Dh\left(X_0, t\right) dM\left(t\right),$$

where

$$D_X\left(x_0, t\right) = \dot{\psi}\left(x_0\right) x_2 - E\left[\dot{\psi}\left(x_0\right) x_2 | V = t, \Delta = 1\right],$$

and

$$Dh\left(x_0, t\right) = h\left(x_0\right) - E\left[h\left(x_0\right)|V = t, \Delta = 1\right]$$

Thus, the least favorable direction h^* minimizing $E\left\|K_\beta - Kh\right\|^2$ satisfies

$$E\left[\left(K_\beta - Kh^*\right) Kh\right] = 0 \tag{1.5.4}$$

for all $h \in L_{2,W}$. By Lemma 1 in Sasieni (1992b), (1.5.4) is equivalent as

$$D^*\left(D_X - Dh^*\right) = 0 \text{ a.s.}$$

where D^* is the adjoint of operator D. Now equations (ii) and (iii) from Lemma 3 in Sasieni (1992b) give us

$$D^* D_X = E\left\{\dot{\psi}\left(X_0\right) X_2 - E\left[\dot{\psi}\left(X_0\right) X_2 | V = t, \Delta = 1\right] | x_0 = x_0, \Delta = 1\right\}$$

$$D^* Dh^* = E\left\{h\left(X_0\right) - E\left[h\left(X_0\right)|V = t, \Delta = 1\right]|X_0 = x_0, \Delta = 1\right\}$$

Whence it is straightforward to see the efficient direction $h^*(\cdot)$ satisfies the following condition:

$$E\left\{\dot{\psi}(X_0)X_2 - h^*(X_0) - E\left[\dot{\psi}(X_0)X_2 - h^*(X_0)|V=t,\Delta=1\right]|X_0=x_0,\Delta=1\right\}$$
$$=0, \quad \text{a.s. } P_{X_0}.$$

Moreover, the least favorable direction of the corresponding hazard function is

$$a^*(t) = E\left[\dot{\psi}(X_0)X_2 - h^*(X_0)|V=t,\Delta=1\right].$$

■

1.5.2 Large Sample Properties

The following notations are standard in deriving large sample properties in Cox regression, see Andersen and Gill (1982), Huang (1999):

$$S_{0n}(t,\theta) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\exp\left[\psi\left(X_i'\widetilde{\beta}\right)\right],$$

$$S_0(t,\theta) = E\left[Y(t)\exp\left[\psi\left(X'\widetilde{\beta}\right)\right]\right],$$

$$S_{1n}(t,\theta)[h] = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\exp\left[\psi\left(X_i'\widetilde{\beta}\right)\right]h(X_i),$$

$$S_1(t,\theta)[h] = E\left[Y(t)\exp\left[\psi\left(X'\widetilde{\beta}\right)\right]h(X)\right],$$

wieth a real-valued function $h(\cdot)$ depending on x. Furthermore, for $u=(t,x,\delta)$ define

$$s_n(u,\theta)[h] = h(x) - \frac{S_{1n}(t,\theta)[h]}{S_{0n}(t,\theta)[h]}$$

$$s(u,\theta)[h] = h(x) - \frac{S_1(t,\theta)[h]}{S_0(t,\theta)[h]}$$

The sample criterion function is

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} I[V_i\le\tau]\Delta_i\left\{\psi\left(X_i'\beta\right) - \log S_{0n}(V_i,\theta)\right\}.$$

In the sequel, we shall omit writing out $I[V_i\le\tau]$ without confusion. with its population version as

$$M(\theta) = P_\Delta\left\{\psi\left(X_i'\widetilde{\beta}\right) - \log S_0(V_i,\theta)\right\}.$$

Notice $\dot{\zeta}(x,\beta) = \frac{\partial}{\partial\beta}\psi\left(x'\widetilde{\beta}\right) = \dot{\psi}\left(x'\widetilde{\beta}\right)x_2$.

The first result concerns the consistency and rates of convergence of parameter $\theta = (\beta, \zeta(\cdot,\beta))$ in terms of the following metric

$$d(\theta_1,\theta_2) = |\beta_1 - \beta_2| + \|\zeta_1(\cdot,\beta_1) - \zeta_2(\cdot,\beta_2)\|$$

where

$$\|\zeta_1(\cdot,\beta_1) - \zeta_2(\cdot,\beta_2)\|^2 = \int \left[\psi_1\left(x_1'\widetilde{\beta}\right) - \psi_2\left(x_2'\widetilde{\beta}\right)\right]^2 dF_X(x).$$

Let $\Psi_n^p = \mathcal{S}_n(T_{K_n}, K_n, p)$, and denote

$$\mathcal{H}_n^p = \left\{\zeta(\cdot,\beta) : \zeta(x,\beta) = \psi\left(x'\widetilde{\beta}\right), \psi \in \Psi_n^p, x \in \mathcal{X}, \beta \in \mathcal{B}\right\}$$

and $\Theta_n^p = \mathcal{B} \times \mathcal{H}_n^p$. The proof of Theorem 1.5.2 is relegated to the Appendix.

Theorem 1.5.2 *Let $K_n = O(n^\nu)$ where ν satisfies the restriction $\frac{1}{4p-3} < \nu < \frac{1}{2p}$. Then given Assumptions (A1)-(A6), we have (i) $d\left(\widehat{\theta}_n, \theta_o\right) \to_p 0$ and its rate of convergence is*

$$d\left(\widehat{\theta}_n, \theta_o\right) = O_p\left(n^{-\min(p\nu,(1-\nu)/2)}\right).$$

Notice for $\nu = \frac{1}{2p+1}$ we get the optimal convergence rate as $O_p\left(n^{-\frac{p}{2p+1}}\right)$.

Theorem 1.5.3 *Suppose Assumptions (A1)-(A6) hold and the information matrix $I^*(\beta_o)$ is positive semi-definite, then we have:*

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_o\right) = I^*(\beta_o)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{l}_\beta^*(Z_i) + o_p(1)$$
$$\Longrightarrow N\left(0, I^*(\beta_o)^{-1}\right).$$

The proof of Theorem 1.5.3 is adapted from Theorem 2.1 in Ding and Nan (2011). For earlier results without bundled parameter issue, we refer the readers to Huang (1996, 1999) for a similar exposition.

Proof. We will prove the following claims separately in Appendix. First of all, two estimating equations hold:

$$P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)\left[\dot{\zeta}_o(\cdot,\beta_o)\right]\right\} = o_p\left(n^{-1/2}\right), \tag{1.5.5}$$
$$P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[h^*]\right\} = o_p\left(n^{-1/2}\right),$$

Therefore

$$P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[\widetilde{m}_o]\right\} = o_p\left(n^{-1/2}\right), \tag{1.5.6}$$

for the least favorable directoin $\widetilde{m}_o = \dot{\zeta}_o(\cdot,\beta_o) - h^*$. Morevoer, the stochastic equicontinuity holds for those two processes

$$\begin{aligned}
&P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)\left[\dot{\zeta}_o(\cdot,\beta_o)\right] - s_n(,\theta_o)\left[\dot{\zeta}_o(\cdot,\beta_o)\right]\right\} \\
&- P_{\Delta}\left\{s\left(,\widehat{\theta}_n\right)\left[\dot{\zeta}_o(\cdot,\beta_o)\right] - s(,\theta_o)\left[\dot{\zeta}_o(\cdot,\beta_o)\right]\right\} = o_p\left(n^{-1/2}\right),
\end{aligned}$$

$$\begin{aligned}
&P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[h^*] - s_n(,\theta_o)[h^*]\right\} \\
&- P_{\Delta}\left\{s\left(,\widehat{\theta}_n\right)[h^*] - s(,\theta_o)[h^*]\right\} = o_p\left(n^{-1/2}\right),
\end{aligned}$$

leading to

$$\begin{aligned}
&P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[\widetilde{m}_o] - s_n(,\theta_o)[\widetilde{m}_o]\right\} \\
&- P_{\Delta}\left\{s\left(,\widehat{\theta}_n\right)[\widetilde{m}_o] - s(,\theta_o)[\widetilde{m}_o]\right\} = o_p\left(n^{-1/2}\right).
\end{aligned}$$

Given (1.5.6), one arrives at

$$-P_{\Delta}\left\{s\left(,\widehat{\theta}_n\right)[\widetilde{m}_o] - s(,\theta_o)[\widetilde{m}_o]\right\} = P_{\Delta n}\left\{s_n(,\theta_o)[\widetilde{m}_o]\right\} + o_p\left(n^{-1/2}\right).$$

Now we take the following expansion as in Lemma 5.4 of Huang (1999):

$$\begin{aligned}
&P_{\Delta}\left\{s\left(,\widehat{\theta}_n\right)[\widetilde{m}_o] - s(,\theta_o)[\widetilde{m}_o]\right\} \\
&= -P_{\Delta}\left\{s(,\theta_o)[\widetilde{m}_o]\, s(,\theta_o)[\widetilde{m}_o]\right\}\left(\widehat{\beta}_n - \beta_o\right) \\
&- P_{\Delta}\left\{s(,\theta_o)[\widetilde{m}_o]\, s(,\theta_o)\left[\widehat{\psi}_n - \psi_o\right]\right\} \\
&= -P_{\Delta}\left\{s(,\theta_o)[\widetilde{m}_o]\right\}^{\otimes 2}\left(\widehat{\beta}_n - \beta_o\right) + o_p\left(n^{-1/2}\right)
\end{aligned}$$

The final equality above follows from the fact that h^* is the efficient direction. Thus combining all the results above, we arrive at

$$\sqrt{n}P_{\Delta}\left\{s(,\theta_o)[\widetilde{m}_o]\right\}^{\otimes 2}\left(\widehat{\beta}_n - \beta_o\right) = \sqrt{n}P_{\Delta n}\left\{s_n(,\theta_o)[\widetilde{m}_o]\right\} + o_p(1),$$

so we get

$$\begin{aligned}
&\sqrt{n}P_{\Delta n}\left\{s_n\left(,\theta_o\right)\left[\widetilde{m}_o\right]\right\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \left[\psi\left(X_i'\beta\right)X_i - h^*\left(X_i'\beta\right) - \frac{S_{1n}\left(t,\theta_o\right)\left[\widetilde{m}_o\right]}{S_{0n}\left(t,\theta_o\right)}\right] dM_i\left(t\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \left[\psi\left(X_i'\beta\right)X_i - h^*\left(X_i'\beta\right) - \frac{S_1\left(t,\theta_o\right)\left[\widetilde{m}_o\right]}{S_0\left(t,\theta_o\right)}\right] dM_i\left(t\right) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \left[\frac{S_1\left(t,\theta_o\right)\left[\dot{\zeta}_o\left(\cdot,\beta_o\right) - h^*\right]}{S_0\left(t,\theta_o\right)} - \frac{S_{1n}\left(t,\theta_o\right)\left[\widetilde{m}_o\right]}{S_{0n}\left(t,\theta_o\right)}\right] dM_i\left(t\right)
\end{aligned}$$

the second term is negligible by Lenglart's inequality as in Anderson and Gill (1982):

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left[\frac{S_1\left(t,\theta_o\right)\left[\widetilde{m}_o\right]}{S_0\left(t,\theta_o\right)} - \frac{S_{1n}\left(t,\theta_o\right)\left[\widetilde{m}_o\right]}{S_{0n}\left(t,\theta_o\right)}\right]^2 \\
&\times Y_i\left(t\right)\exp\left[\zeta_o\left(X_i,\beta_o\right)\right] d\Lambda_o\left(t\right) = o_p\left(1\right).
\end{aligned}$$

Hence

$$\begin{aligned}
&\sqrt{n}P_{\Delta n}\left\{s_n\left(,\theta_o\right)\left[\widetilde{m}_o\right]\right\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \left[\psi\left(X_i'\beta\right)X_i - h^*\left(X_i'\beta\right) - \frac{S_1\left(t,\theta_o\right)\left[\widetilde{m}_o\right]}{S_0\left(t,\theta_o\right)}\right] dM_i\left(t\right) + o_p\left(1\right)
\end{aligned}$$

also notice that

$$\frac{S_1\left(t,\theta_o\right)\left[\dot{\zeta}_o\left(\cdot,\beta_o\right) - h^*\right]}{S_0\left(t,\theta_o\right)} = a^*\left(t\right)$$

In the end, the desired conclusion follows:

$$\sqrt{n}P_{\Delta n}\left\{s_n\left(,\theta_o\right)\left[\widetilde{m}_o\right]\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_\beta^*\left(V_i, X_i, \Delta_i\right) + o_p\left(1\right).$$

■

1.5.3 Bootstrap Consistency

Just as other semiparametric models, the construction of a valid confidence set of our finite dimensional parameter requires plugging in additional nonparametric estimates, which would deteriorate the finite sample performance to some extent. Cheng and Huang (2010)

prove the bootstrap consistency for a generic semiparametric M-estimator with exchangable weights. They have treated the standard Cox regression as one demonstrating example, but made one simplifying assumption that $s(,\theta)$ directly appears in the estimating equation instead of $s_n(,\theta)$, pertaining to their *i.i.d.* assumption on individual criterion function. In this section, we propose a novel bootstrap procedure consistent with the partial likelihood structure inspired by Nan and Wellner (2013). We establish the bootstrap consistency with general exchangable bootstrap weights $(W_{ni})_{i=1}^n$. The practical usefulness of more general weighting scheme other than Efron's multinomial weights is particularly important, as the former procedure often gives too many dies applied to censored data. Define the weighted empirical process as $P_n^* f = \frac{1}{n}\sum_{i=1}^n W_{ni} f(X_i)$ .We assume the following requirements on the bootstrap weight.

Assumption (W1). The vector $W_n = (W_{n1}, ..., W_{nn})^\top$ is exchangable for all n.

Assumption (W2). $W_{ni} \geq 0$ for all n, i and $\sum_{i=1}^n W_{ni} = n$ for all n.

Assumption (W3). For some positive constant $C < \infty$, $\limsup_{n\to\infty} \|W_{n1}\|_{2,1} \leq C$, where $\|W_{n1}\|_{2,1} = \int_0^\infty \sqrt{\Pr(W_{n1} \geq u)}du$.

Assumption (W4). $\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{t\geq\lambda} t^2 \Pr(W_{n1} > t) = 0$.

Assumption (W5). $(1/n)\sum_{i=1}^n (W_{ni} - 1)^2 \to_p c^2$ for some positive constant c.

The following quantities serve as the natural bootstrap analog of S_{0n}, S_{1n}, and s_n

$$S_{0n}^*(t,\theta) = \frac{1}{n}\sum_{i=1}^n W_{ni} Y_i(t) \exp\left[\psi\left(X_i'\widetilde{\beta}\right)\right],$$

$$S_{1n}^*(t,\theta)[h] = \frac{1}{n}\sum_{i=1}^n W_{ni} Y_i(t) \exp\left[\psi\left(X_i'\widetilde{\beta}\right)\right] h(X_i),$$

$$s_n^*(t,\theta)[h] = h(x) - \frac{S_{1n}^*(t,\theta)[h]}{S_{0n}^*(t,\theta)[h]}.$$

So the bootstrapped estimates $\left(\widehat{\beta}_n^*, \widehat{\gamma}_n^*\right)$ are defined by maximizing

$$l_n^*(\beta,\gamma) = \frac{1}{n}\sum_{i=1}^n W_{ni}\Delta_i \left\{\sum_{j=1}^{q_n}\gamma_j B_j\left(X_{i1} + X_{i2}'\beta\right) - \log\left(\sum_{k:V_k\geq V_i} W_{ni}\exp\left[\sum_{j=1}^{q_n}\gamma_j B_j\left(X_{k1} + X_{k2}'\beta\right)\right]\right)\right\}.$$

Before stating the theoretical results for the bootstrapped quantities, we should be explicit about the underlying probability space and source of randomness. The following set of

definitions and notations are adapted from Cheng and Huang (2010). We have the product probability space

$$(\mathcal{Z}^{\infty} \times \mathcal{W}, \mathcal{A}^{\infty} \times \otimes, P_{ZW})$$

for the joint randomness from observed data and bootstrap weights. Furthermore, the bootstrap weights are independent from the sample observations, i.e., $P_{ZW} = P_Z \times P_W$.

Definition 1.5.4 *For a real-valued random variable Δ_n, we define (i) $\Delta_n = o_{P_W}(1)$, in P_Z-probability if for any $\varepsilon, \delta > 0$*

$$P_Z\left\{P_{W|Z}\left(|\Delta_n| > \varepsilon\right) > \delta\right\} \to 0.$$

(ii) We define $\Delta_n = O_{P_W}(1)$, in P_Z-probability if for any $\delta > 0$ there exists a M s.t.

$$P_Z\left\{P_{W|Z}\left(|\Delta_n| > M\right) > \delta\right\} \to 0.$$

Theorem 1.5.5 *For the bootstrap version $\widehat{\beta}_n^*$ with general exchange weights $\mathcal{W}_n = (W_{n1}, ..., W_{nn})$, we have that*

$$\left\|\widehat{\beta}_n^* - \widehat{\beta}_n\right\| = O_{P_W}\left(n^{-1/2}\right)$$

in P_Z-probability. Furthermore,

$$\sup_{x \in \mathcal{R}^d}\left|P_{\mathcal{W}_n|\mathcal{Z}_n}\left\{\sqrt{n}\left(\widehat{\beta}_n^* - \widehat{\beta}_n\right) \leq x\right\} - P\left\{N\left(0, I^*\left(\beta_o\right)^{-1}\right) \leq x\right\}\right| \to 0,$$

condtional on observations $\mathcal{Z}_n$ almost surely.

A direct consequence of the last theorem is the validity of the bootstrapped confidence interval, i.e.,

$$\lim_{n\to\infty} P_{ZW}\left(\beta_o \in BC\left(\alpha\right)\right) = 1 - \alpha.$$

where the bootstrapped confidence set is defined to be

$$BC\left(\alpha\right) = \left[\widehat{\beta}_n - \frac{c_n^*\left(1-\alpha/2\right)}{\sqrt{n}}, \widehat{\beta}_n + \frac{c_n^*\left(\alpha/2\right)}{\sqrt{n}}\right],$$

with the $\alpha-$th quantile of bootstrap distributribution $c_n^*\left(\alpha\right) = \inf\left\{\epsilon : P_{\mathcal{W}_n|\mathcal{Z}_n}\left(\sqrt{n}\left(\widehat{\beta}_n^* - \widehat{\beta}_n\right) \leq \epsilon\right) \geq \alpha\right\}$, which could be easily simulated.

1.6 Appendix A

We will first prove the identification results. Some lemmas which are instrumental to the main results and then prove the claimed consistency and rates of convergence afterwards. We let c denote a generic finite positive constant, whose value may change from line to line. Moreover, $A \lesssim B$ means there exists a positive constant C which does not depend on n s.t. $A \leq CB$.

The following lemma is a modification of Theorem 1 in Lin and Kulasekera (2007) with the additional univariate covariate X_1.Recall we have partitioned X as $X = \left(X_1, X_2'\right)'$.

Lemma 1.6.1 *The support $\mathcal{X}$ of $m(\cdot)$ is a bounded convex set with at least one interior point, and $m(\cdot)$ is a nonconstant continuous function on $\mathcal{X}$. If*

$$m(x) = \psi\left(x'\widetilde{\beta}\right) = \varphi\left(x'\widetilde{\alpha}\right),$$

for all $x \in \mathcal{X}$ for some continuous functions ψ and φ, then $\beta = \alpha$ and $\psi \equiv \varphi$ on $\left\{x_1 + x_2'\beta | x \in \mathcal{X}\right\}$.

Proof. It sufficies to show that $\beta = \alpha$. Recall that $\widetilde{\beta} = \left(1, \beta'\right)'$, $\widetilde{\alpha} = \left(1, \alpha'\right)'$ and the following two composition functions:

$$\widetilde{\psi}(\cdot) = \psi\left(\left\|\widetilde{\beta}\right\| \times \cdot\right), \text{ and } \widetilde{\varphi}(\cdot) = \varphi\left(\left\|\widetilde{\beta}\right\| \times \cdot\right),$$

where $\|\cdot\|$ is the standard Euclidean distance. Thus one gets

$$\widetilde{\psi}\left(x'\overline{\beta}\right) = \widetilde{\varphi}\left(x'\overline{\alpha}\right), \text{ for all } x \in \mathcal{X},$$

with $\overline{\beta} = \widetilde{\beta}/\left\|\widetilde{\beta}\right\|$ and $\overline{\alpha} = \widetilde{\alpha}/\|\widetilde{\alpha}\|$. Now we are ready to apply the arguments in Lin and Kulasekera (2007), as both $\overline{\beta}$ and $\overline{\alpha}$ have norm equal to 1 and their first coordinates are positive.

Suppose $\beta \neq \alpha$, so $\overline{\beta} \neq \overline{\alpha}$. By the stated assumptions, there exists a sphere $B(\widetilde{x}, r) \subset \mathcal{X}$ for some $\widetilde{x}$ s.t. $m(\cdot)$ is nonconstant on it. For all $t \in (-r, r)$ we have $\widetilde{x} + t\overline{\beta} \in \mathcal{X}$ as $\overline{\beta}^{\top}\overline{\beta} = 1$. So

$$\begin{aligned} \widetilde{\psi}\left(\widetilde{x}'\overline{\beta} + t\right) &= \widetilde{\psi}\left(\overline{\beta}'\left(\widetilde{x} + t\overline{\beta}\right)\right) = \widetilde{\varphi}\left(\overline{\alpha}'\left(\widetilde{x} + t\overline{\beta}\right)\right), \\ \widetilde{\varphi}\left(\widetilde{x}'\overline{\alpha} + t\right) &= \widetilde{\varphi}\left(\overline{\alpha}'\left(\widetilde{x} + t\overline{\alpha}\right)\right) = \widetilde{\psi}\left(\overline{\beta}'\left(\widetilde{x} + t\overline{\alpha}\right)\right). \end{aligned} \tag{A.1}$$

Because the first non-zero components of $\overline{\beta}$ and $\overline{\alpha}$ are positive, we get $\overline{\beta} \neq -\overline{\alpha}$, hence $\left|\overline{\beta}'\overline{\alpha}\right| < 1$. Then we could iterate based on (A.1):

$$\begin{aligned}\widetilde{\psi}\left(z_0^{\top}\overline{\beta}+t\right) &= \widetilde{\varphi}\left(\overline{\alpha}'\left(\widetilde{x}+t\overline{\beta}\right)\right) = \widetilde{\psi}\left(\widetilde{x}'\overline{\beta}+t\left(\overline{\beta}'\overline{\alpha}\right)^2\right) \\ &= \cdots = \widetilde{\psi}\left(\widetilde{x}'\overline{\beta}+t\left(\widetilde{\beta}^{\top}\overline{\alpha}\right)^n\right) = \cdots = \widetilde{\psi}\left(\widetilde{x}'\overline{\beta}\right),\end{aligned}$$

where we have used the continuity of $\widetilde{\psi}$. This indicates $m(\cdot)$ is a constant function on $B(\widetilde{x}, r)$, leading to a contridication. Thereafter we get $\overline{\beta} = \overline{\alpha}$, $\left\|\widetilde{\beta}\right\| = \|\widetilde{\alpha}\|$, and finally $\beta = \alpha$ as desired. ■

Proof. (of Identification) The moment restriction in (I3) allows us to identify $\Lambda(t)$ first. For the single index structure, we shall apply the previous lemma to $m(x) = \psi\left(x'\widetilde{\beta}\right)$, where $\psi(\cdot) = \log \circ \Phi \circ \exp(\cdot)$ with the Lèvy-Laplace exponent function Φ. Φ belongs to the class of Bernstein function whose derivative is completely monotone, thus $\psi(\cdot)$ is real analytic. Given Assumption (I1), we get the identification of β as $\psi(\cdot)$ is indeed a nonconstant continuous function. Finally, the variation of X_0 along an open interval leads to identification of the analytic function $\psi(\cdot)$ by Proposition 1 in Abbring and van den Berg (2003). ■

Now we present several preparatory lemmas. The first two essentially concern the smoothness of the model, and would be used in determining the rate of convergence.

Lemma 1.6.2 *Under Assumptions (A1)-(A6), for a small $\varepsilon > 0$ we have*

$$\sup_{d(\theta,\theta_o)\leq\varepsilon,\theta\in\Theta_n^p} Var\left[m(\cdot,\theta) - m(\cdot,\theta_o)\right] \leq C\varepsilon^2.$$

Proof. By definition of the individual criterion function, we have

$$\begin{aligned}&E\left[m(\cdot,\theta) - m(\cdot,\theta_o)\right]^2 \\ &= E\left[\psi\left(X_i'\widetilde{\beta}\right) - \psi_o\left(X_i'\widetilde{\beta}_o\right) - \left[\log S_0(V_i,\theta) - \log S_0(V_i,\theta_o)\right]\right]^2 \\ &\lesssim E\left[\psi\left(X_i'\widetilde{\beta}\right) - \psi_o\left(X_i'\widetilde{\beta}_o\right)\right]^2 + E\left[S_0(V_i,\theta) - S_0(V_i,\theta_o)\right]^2 \\ &\lesssim |\beta_o - \beta|^2 + \|\zeta_o(\cdot,\beta_o) - \zeta(\cdot,\beta)\|^2.\end{aligned}$$

■

Lemma 1.6.3 *Under Assumptions (A1)-(A6), wfor a small $\varepsilon > 0$ we have*

$$\inf_{d(\theta,\theta_o)\geq\varepsilon,\theta\in\Theta_n^p} \left[M\left(\theta_o\right) - M\left(\theta\right)\right] \geq C\varepsilon^2.$$

Proof. Here we write $M\left(\theta\right) = M\left(\beta, \zeta\left(\cdot,\beta\right)\right)$ to highlight the nature of the bundled parameter case. For any small $\varepsilon > 0$, and the point θ s.t. $d\left(\theta,\theta_o\right) \geq \varepsilon$, we consider the quadratic expansion for

$$\begin{aligned}
&M\left(\beta_o, \zeta_o\left(\cdot,\beta_o\right)\right) - M\left(\beta, \zeta\left(\cdot,\beta\right)\right) \\
&= \left(\beta_o - \beta\right)' M_{\beta\beta}\left(\beta_o - \beta\right) + 2\left(\beta_o - \beta\right)' M_{\beta\psi}\left[\psi_o - \psi\right] \\
&+ M_{\psi\psi}\left[\psi_o - \psi, \psi_o - \psi\right] + o\left(d^2\left(\theta,\theta_o\right)\right),
\end{aligned}$$

with those second order derivatives calculated following the rules in Ding and Nan (2011). A straightforward computation reveals that

$$\begin{aligned}
M_{\beta\beta} &= P_\Delta\left\{\left[\ddot{\psi}^2\left(X'\widetilde{\beta}\right) - \frac{P\left[Y\left(t\right)e^{\psi\left(X'\widetilde{\beta}\right)}\left(\ddot{\psi}^2\left(X'\widetilde{\beta}\right) + \dot{\psi}^2\left(X'\widetilde{\beta}\right)\right)\right]}{P\left[Y\left(t\right)e^{\psi\left(X'\widetilde{\beta}\right)}\right]}\right] X_2^{\otimes 2}\right\} \\
&+ P_\Delta\left\{\left(\frac{P\left[Y\left(t\right)e^{\psi\left(X'\widetilde{\beta}\right)}\dot{\psi}\left(X'\widetilde{\beta}\right)X\right]}{P\left[Y\left(t\right)e^{\psi\left(X'\widetilde{\beta}\right)}\right]}\right)^{\otimes 2}\right\} \\
&= P_\Delta\left\{E\left[K\left(t\right)\dot{\psi}^2\left(X'\widetilde{\beta}\right)X_2^{\otimes 2}\right] - E\left[K\left(t\right)\left(\dot{\psi}\left(X'\widetilde{\beta}\right)X_2\right)\right]^{\otimes 2}\right\},
\end{aligned}$$

where we have used the definition of $K\left(t\right)$ in (1.5.3) and the fact (1.5.2) which leads to

$$P_\Delta\left\{\left[\ddot{\psi}^2\left(X'\widetilde{\beta}\right) - \frac{P\left[Y\left(t\right)e^{\psi\left(X'\widetilde{\beta}\right)}\ddot{\psi}^2\left(X'\widetilde{\beta}\right)\right]}{P\left[Y\left(t\right)e^{\psi\left(X'\widetilde{\beta}\right)}\right]}\right] X_2^{\otimes 2}\right\} = 0.$$

Similar calculations give us the other two derivatives:

$$\begin{aligned}
M_{\beta\psi}\left[h\right] &= P_\Delta\left\{E\left[K\left(t\right)h\left(X'\widetilde{\beta}\right)\dot{\psi}\left(X'\widetilde{\beta}\right)X_2\right] - E\left[K\left(t\right)h\left(X'\widetilde{\beta}\right)\right]E\left[K\left(t\right)\left(\dot{\psi}\left(X'\widetilde{\beta}\right)X_2\right)\right]\right\}, \\
M_{\psi\psi}\left[h_1,h_2\right] &= P_\Delta\left\{E\left[K\left(t\right)h_1\left(X'\widetilde{\beta}\right)h_2\left(X'\widetilde{\beta}\right)\right] - E\left[K\left(t\right)h_1\left(X'\widetilde{\beta}\right)\right]E\left[K\left(t\right)h_2\left(X'\widetilde{\beta}\right)\right]\right\},
\end{aligned}$$

also see Lemma A.4 in Huang (1999).

After completing squares, one arrives at

$$\begin{aligned} & M\left(\beta_o, \zeta_o\left(\cdot, \beta_o\right)\right) - M\left(\beta, \zeta\left(\cdot, \beta\right)\right) \\ & = P_\Delta \left\{ EK\left(t\right)\left[g^*\left(X\right) - EK\left(t\right) g^*\left(X\right)\right]^2 \right\} \end{aligned}$$

where

$$g^*\left(X\right) = \dot{\psi}\left(X'\widetilde{\beta}\right) X'\left(\beta_o - \beta\right) + \psi_o\left(X'\widetilde{\beta}_o\right) - \psi\left(X'\widetilde{\beta}_o\right).$$

Now by the positive definiteness of $X_2^{\otimes 2}$ and

$$0 < c_1 \leq E\left[K\left(t\right)|X\right] \leq c_2 < \infty,$$

we obtain

$$\inf_{d(\theta,\theta_o)\geq\varepsilon} \left[M\left(\beta_o, \zeta_o\left(\cdot, \beta_o\right)\right) - M\left(\beta, \zeta\left(\cdot, \beta\right)\right)\right] \geq cd^2\left(\theta, \theta_o\right).$$

■

The following lemma characterizes the order of approximation bias in sieve estimation, see Schumaker (1981). The additional term $n^{-(1-\upsilon)/2}$ comes from recentering the estimator, see Lemma A.5 in Huang (1999).

Lemma 1.6.4 *For any $\psi_o \in \Psi^p$, there exists a function $\psi_{on} \in \Psi_n^p$ with $P_{\Delta n}\psi_{on} = 0$ s.t.*

$$\left\|\psi_{on} - \psi_o\right\|_\infty = O\left(n^{-p\upsilon} + n^{-(1-\upsilon)/2}\right). \tag{A.2}$$

The next lemma is Lemma 4.2 in Kong and Nan (2014), and it (via Borel-Cantelli argument) ensures that the frequently appearing denominator term $S_{0n}\left(t, \theta\right)$ is uniformly (for $t \in [0, \tau]$ and $\theta \in \Theta_n^p$) bounded away from zero almost surely.

Lemma 1.6.5 *For the truncation point τ, and any $\eta > 0$, we have*

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n 1\left(V_i \geq \tau\right) \leq \frac{\eta}{2}\right) \leq 2e^{-n\eta^2/2}.$$

Now we introduce some standard terminology in empirical process theory (van der Vaart and Wellner, 1996). $\|\cdot\|_\infty$ is the usual L_∞-norm for a function f with $\|f\|_\infty < \infty$. The bracketing number $N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_\infty\right)$ for subclass $\mathcal{F}$ is defined to be the minimum of m such that

$\exists\, f_1^L, f_1^U, ..., f_m^L, f_m^U$ for $\forall f \in \mathcal{F}$, $f_j^L \leq f \leq f_j^U$ for some j, and $\left\| f_j^U - f_j^L \right\|_\infty \leq \epsilon$.What's more the corresponding bracketing entropy integral is $\mathcal{J}_{[]}(\eta, \mathcal{F}, \|\cdot\|_\infty) = \int_0^\eta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_\infty)} d\epsilon$. The following lemma which is a restatement of Lemma 3.4.2 in van der Vaart and Wellner (1996) based on the supremum norm is useful to bound the normalized empirical process $G_n = \sqrt{n}(P_n - P)$ and $\|\mathcal{G}_n\|_\mathcal{F} = \sup_{f\in\mathcal{F}} |G_n(f)|$. Let $\mathcal{F}$ be a uniformly bounded class of measurable functions such that $Pf^2 \leq \delta^2$ and $\|f\|_\infty \leq M_0$, then

$$E_P^* \|\mathcal{G}_n\|_\mathcal{F} \lesssim \mathcal{J}_{[]}(\delta, \mathcal{F}, \|\cdot\|_\infty) \left[1 + \frac{\mathcal{J}_{[]}(\delta, \mathcal{F}, \|\cdot\|_\infty) M_0}{\delta^2 \sqrt{n}}\right]. \tag{A.3}$$

The next lemma collects various entropy bounds for several functional classes involved in the current paper

Lemma 1.6.6 *The functional classes under consideration are:*

$$\begin{aligned}
\mathcal{F}_1 &= \left\{ f_1 = \psi\left(x'\beta\right) - \log S_0(t,\theta) : \theta \in \Theta_n^p, d(\theta,\theta_o) \leq \delta \right\}, \\
\mathcal{F}_2 &= \left\{ f_2 = \exp\left[\psi\left(x'\beta\right)\right] I[t \leq \tau] : \theta \in \Theta_n^p, d(\theta,\theta_o) \leq \delta \right\}, \\
\mathcal{F}_3 &= \left\{ f_3 = \exp\left[\psi\left(x'\beta\right)\right] h\left(x'\beta\right) I[t \leq \tau] : \theta \in \Theta_n^p, d(\theta,\theta_o) \leq \delta, \|h\|_2 \leq \delta \right\}, \\
\mathcal{F}_4 &= \left\{ f_4 = \exp\left[\psi\left(x'\beta\right)\right] \dot{\psi}\left(x'\beta\right) xI[t \leq \tau] : \theta \in \Theta_n^p, d(\theta,\theta_o) \leq \delta, \left\|\dot{\psi}\right\|_2 \leq \delta \right\}
\end{aligned}$$

then we get

$$N_{[]}(\epsilon, \mathcal{F}_j, \|\cdot\|_\infty) \lesssim (\delta/\epsilon)^{cq_n + d},$$

or in terms of the integral:

$$\mathcal{J}_{[]}(\delta, \mathcal{F}, \|\cdot\|_\infty) \lesssim \sqrt{q_n}\delta + \delta\sqrt{\log(1/\delta)}$$

for the dimension of covariate d and some finite positive constant c which may change with $j = 1, .., 4$.

Proof. Due to similarity, we will only prove the assertion for the class $\mathcal{F}_4$ which is the most involved one. The construction here is analogous to Ding and Nan (2011). According to Schumaker (1981), for $\psi \in \Psi_n^p$ with $\|\psi - \psi_o\|_\infty \leq \delta$ and $\dot{\psi} \in \Psi_n^{p-1}$ with $\left\|\dot{\psi} - \dot{\psi}_o\right\|_\infty \leq \delta$ we get pairs of bracket $\left[\psi_i^L, \psi_i^U\right]$ where $i = 1, ..., (\eta/\epsilon)^{cq_n}$ and $\left[\dot{\psi}_j^L, \dot{\psi}_j^U\right]$ where $j = 1, ..., (\delta/\epsilon)^{cq_n}$. Since $\mathcal{B}$ is a compact subset of $\mathcal{R}^d$, it can be covered by $C\epsilon^{-d}$ balls

with radius ϵ. Given the boundedness of the support $\mathcal{X}$, for any $\beta \in \mathcal{B}$, there exists a β_l satisfing $x^{'}\beta \in \left[x^{'}\beta_l - c_0\epsilon, x^{'}\beta_l + c_0\epsilon\right]$, for another constant c_0. Apparently the number of l needed is of order $(\delta/\epsilon)^d$.

Now let c_1^i, c_2^i be two constants that $\psi_i^L\left(x^{'}\beta_l + c_1^i\epsilon\right)$ and $\psi_i^U\left(x^{'}\beta_l + c_2^i\epsilon\right)$ are minimum and maximum values of ψ_i^L and ψ_i^U respectively in the interval $\left[x^{'}\beta_l - c_0\epsilon, x^{'}\beta_l + c_0\epsilon\right]$. Similarly we define c_3^j, c_4^j which make $\dot{\psi}_j^L\left(x^{'}\beta_l + c_3^j\epsilon\right)$ and $\dot{\psi}_j^U\left(x^{'}\beta_l + c_4^j\epsilon\right)$ to attain the corresponding minimum and maximum values in the same interval. Consider the set of brackets as:

$$\left\{\left[f_{ijl}^L, f_{ijl}^U\right], i = 1, ..., (\delta/\epsilon)^{cq_n}; j = 1, ..., (\delta/\epsilon)^{cq_n}, l = 1, ..., d\,(\delta/\epsilon)^d\right\}$$

where

$$f_{ijl}^L = \exp\left[\psi_i^L\left(x^{'}\beta_s + c_1^i\epsilon\right)\right]\dot{\psi}_j^L\left(x^{'}\beta_l + c_3^j\epsilon\right)xI\left[t \le \tau\right],$$
$$f_{ijl}^U = \exp\left[\psi_i^U\left(x^{'}\beta_s + c_2^i\epsilon\right)\right]\dot{\psi}_j^U\left(x^{'}\beta_l + c_4^j\epsilon\right)xI\left[t \le \tau\right].$$

By construction, the lower and upper bound functions are ordered and serve as a valid bracketing pair for the function $f \in \mathcal{F}_4$, and the overall number needed is of order $(\delta/\epsilon)^{cq_n+d}$ as claimed. What's more

$$\left\|f_{ijl}^U - f_{ijl}^L\right\|_\infty \le C\left(\left\|\psi_i^U - \psi_i^L\right\|_\infty + \left\|\dot{\psi}_j^U - \dot{\psi}_j^L\right\|_\infty + \|\beta_l - \beta_s\|\right) \lesssim \epsilon,$$

utilizing the boundedness of $\mathcal{X}$. ■

Lemma 1.6.7 *Given the entropy bounds in Lemma 1.6.6, we get*

(i) for $\Theta_\delta = \{\theta : \|\psi - \psi_n\| \le \delta, \|\psi\|_\infty \le c\}$

$$\sup_{t,\theta\in\Theta_\delta}\left|\frac{S_{0n}(t,\theta)}{S_{0n}(t,\theta_{on})} - \frac{S_0(t,\theta)}{S_0(t,\theta_{on})}\right| = n^{-1/2}O_p\left(\sqrt{q_n}\delta + \delta\sqrt{\log(1/\delta)}\right), \tag{A.4}$$

(ii) for any g *such that* $\|g\|_2 = o\left(q_n^{-1/2}\right)$

$$\sup_t\left|\frac{S_{1n}(t,\theta_{on})[g]}{S_{0n}(t,\theta_{on})} - \frac{S_1(t,\theta_{on})[g]}{S_0(t,\theta_{on})}\right| = o_p\left(n^{-1/2}\right), \tag{A.5}$$

(iii) for any $\zeta(\cdot,\beta) \in \mathcal{H}_n^p$ *and* $g = \dot{\zeta}(\cdot,\beta)$ *or* $g = h^*$.

$$\sup_t\left|\frac{S_{1n}\left(t,\widehat{\theta}_n\right)[g]}{S_{0n}\left(t,\widehat{\theta}_n\right)} - \frac{S_{1n}(t,\theta_o)[g]}{S_{0n}(t,\theta_o)} - \frac{S_1\left(t,\widehat{\theta}_n\right)[g]}{S_0\left(t,\widehat{\theta}_n\right)} + \frac{S_1(t,\theta_o)[g]}{S_0(t,\theta_o)}\right| = o_p\left(n^{-1/2}\right) \tag{A.6}$$

Proof. Since

$$\frac{S_{0n}(t,\theta)}{S_{0n}(t,\theta_{on})}-\frac{S_0(t,\theta)}{S_0(t,\theta_{on})}=\frac{S_{0n}(t,\theta_{on})S_0(t,\theta)-S_{0n}(t,\theta)S_0(t,\theta_{on})}{S_{0n}(t,\theta_{on})S_0(t,\theta_{on})},$$

and the denominator is bounded away from zero with probability 1, it is sufficient to consider the numerator on the r.h.s. of the equality above. This numerator could be expressed as

$$\begin{aligned}&S_0(t,\theta_{on})\left[S_{0n}(t,\theta)-S_{0n}(t,\theta_{on})-S_0(t,\theta)+S_0(t,\theta_{on})\right]\\&-\left[S_{0n}(t,\theta_{on})-S_0(t,\theta_{on})\right]\left[S_0(t,\theta)-S_0(t,\theta_{on})\right].\end{aligned}$$

It is clear that

$$\begin{aligned}&S_{0n}(t,\theta)-S_{0n}(t,\theta_{on})-S_0(t,\theta)+S_0(t,\theta_{on})\\&=(P_n-P)\left\{Y(t)\left[\exp\left(\psi\left(x'\beta\right)\right)-\exp\psi_{on}\left(x'\beta\right)\right]\right\}\\&=n^{-1/2}O_p\left(\sqrt{q_n}\delta+\delta\sqrt{\log(1/\delta)}\right).\end{aligned}$$

Moreover, $S_{0n}(t,\theta_{on})-S_0(t,\theta_{on})=n^{-1/2}\sqrt{q_n}$ and it holds

$$\begin{aligned}\left|S_0(t,\theta)-S_0(t,\theta_{on})\right|&\le EY(t)\left|\exp\left(\psi\left(x'\beta\right)\right)-\exp\psi_{on}\left(x'\beta\right)\right|\\&\le C\left\|\psi-\psi_n\right\|.\end{aligned}$$

Therefore, we get

$$\left|S_{0n}(t,\theta_{on})S_0(t,\theta)-S_{0n}(t,\theta)S_0(t,\theta_{on})\right|=n^{-1/2}O_p\left(\sqrt{q_n}\delta+\delta\sqrt{\log(1/\delta)}\right).$$

For the second conclusion,

$$\begin{aligned}&\frac{S_{1n}(t,\theta_{on})[g]}{S_{0n}(t,\theta_{on})}-\frac{S_1(t,\theta_{on})[g]}{S_0(t,\theta_{on})}\\&=\frac{S_0(t,\theta_{on})\left\{S_{1n}(t,\theta_{on})[g]-S_1(t,\theta_{on})[g]\right\}-S_1(t,\theta_{on})[g]\left\{S_{0n}(t,\theta_{on})[g]-S_0(t,\theta_{on})[g]\right\}}{S_{0n}(t,\theta_{on})S_0(t,\theta_{on})}.\end{aligned}$$

Again, the conclusion follows from the fact that those two terms in the curly brackets are of $o_p\left(n^{-1/2}\right)$ and the denominator is bounded away from zero with probability 1.

After carrying out the straightforward algebra as in Lemma A.7 of Huang (1999), the third conclusion follows from Donsker properties $S_{1n}(t,\theta)-S_1(t,\theta)$ and $S_{0n}(t,\theta)-S_0(t,\theta)$.

■

Proof. (of Consistency) We shall prove the consistency result that $d(\theta, \theta_o) \to_p 0$, by verifying those conditions leading to Theorem 3.1 in Chen (2007). Condition 3.2 and 3.4 regarding the approximation and compactness properties are automatically satisfied by the B-spline sieves we use.

Condition 3.1 (ii) is ensured by our Lemma 1.6.3 while letting $g(\varepsilon) = \varepsilon^2$ and $\delta(k) = C$ in Chen (2007)

$$\inf_{d(\theta,\theta_o)\geq\varepsilon, \theta\in\Theta_n^p} [M(\theta_o) - M(\theta)] \geq C\varepsilon^2.$$

Condition 3.3 regarding the upper semicontinuity property is satisfied because for small enough $\varepsilon > 0$ and any θ s.t. $d(\theta, \theta_o) \leq \varepsilon$ we have

$$|M(\theta_o) - M(\theta)| \leq C\varepsilon^2,$$

just by the upper bound of $E[K(t)|X]$ in our proof of Lemma 1.6.3.

Finally, because the functional class $\mathcal{F}_1$ has a finite bracketing number for any δ as shown in Lemma 1.6.6, it satisfies the uniform law of large number as in Theorem 2.4.1 in van der Vaart and Wellner (1996). Similarly, we have the uniform convergence of $|S_{0n}(t,\theta) - S_0(t,\theta)|$ to zero when θ varies over Θ_n^p. Thus, Condition 3.5 in Chen (2007) is checked by the uniform convergence of $M_n(\theta)$ to $M(\theta)$. ■

Proof. (of Convergence Rate) Given the consistency result, we shall apply Theorem 3.2.5 in van der Vaart and Wellner (1996) to deduce the rate of convergence. We shall prove that

$$\begin{aligned} &E \sup_{d(\theta,\theta_n)\leq\delta} |M_n(\theta) - M_n(\theta_n) - (M(\theta) - M(\theta_n))| \\ &= n^{-1/2}\delta\left(q_n^{1/2} + \log^{1/2}(1/\delta)\right), \end{aligned}$$

and $\inf_{d(\theta,\theta_n)\geq\varepsilon} [M(\theta) - M(\theta_n)] \geq cd^2(\theta, \theta_o)$. So the convergence rate r_n satisfies

$$r_n^2\left(r_n^{-1}q_n^{1/2} + r_n^{-1}\log^{1/2}(r_n)\right) = O\left(n^{1/2}\right),$$

thus $r_n = q_n^{-1}n^{1/2} = n^{(1-\upsilon)/2}$.

$$
\begin{aligned}
&M_n(\theta) - M_n(\theta_n) - (M(\theta) - M(\theta_n)) \\
&= (P_{\Delta n} - P_\Delta)[m(\cdot,\theta) - m(\cdot,\theta_n)] - P_{\Delta n}\left[\log\frac{S_{0n}(\cdot,\theta)}{S_{0n}(\cdot,\theta_n)} - \log\frac{S_0(\cdot,\theta)}{S_0(\cdot,\theta_n)}\right] \\
&\equiv L_{1n}(\theta) + L_{2n}(\theta).
\end{aligned}
$$

For the first term $L_{1n}(\theta)$, we get

$$
E\sup_{d(\theta,\theta_n)\leq\delta}|L_{1n}(\theta)| = Cn^{-1/2}q_n^{1/2}\delta.
$$

While for the second term, we have

$$
\begin{aligned}
\sup_{d(\theta,\theta_n)\leq\delta}|L_{2n}(\theta)| &\leq 2\sup_{d(\theta,\theta_n)\leq\delta,t\in[0,\tau]}\left|\log\frac{S_{0n}(\cdot,\theta)}{S_{0n}(\cdot,\theta_n)} - \log\frac{S_0(\cdot,\theta)}{S_0(\cdot,\theta_n)}\right| \\
&\leq C\sup_{d(\theta,\theta_n)\leq\delta,t\in[0,\tau]}\left|\log\frac{S_{0n}(\cdot,\theta)}{S_{0n}(\cdot,\theta_n)} - \log\frac{S_0(\cdot,\theta)}{S_0(\cdot,\theta_n)}\right| \\
&\leq Cn^{-1/2}\delta\left(q_n^{1/2} + \log^{1/2}(1/\delta)\right).
\end{aligned}
$$

Moreover for the approximation error, it follows upon (A.2)

$$
\begin{aligned}
&|M(\theta_o) - M(\theta)| \\
&\leq c\|\psi_{on} - \psi_o\|_2 = O\left(n^{-p\upsilon}\right),
\end{aligned}
$$

where $\zeta_{on}(x,\beta_o) = \psi_{on}\left(x'\beta_o\right)$. In sum, the rate of convergence is $O_p\left(\max\left\{n^{-(1-\nu)/2}, n^{-p\nu}, n^{-p\upsilon}\right\}\right) = O_p\left(n^{-\min(p\nu,(1-\nu)/2)}\right)$ as claimed. ■

Lemma 1.6.8 *We have the following two estimating equations hold approximately:*

$$
P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[g]\right\} = o_p\left(n^{-1/2}\right),
$$

with $g = \dot{\zeta}_o(\cdot,\beta_o)$ *or* $g = h^*$.

Proof. We only prove the claim for $\dot{\zeta}_o(\cdot,\beta_o)$, as the second one follows from Lemma 5.2 of Huang (1999) upon notational change. By the first order condition for the regression parameter, we already get

$$
P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)\left[\widehat{\dot{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)\right]\right\} = o_p\left(n^{-1/2}\right).
$$

Hence it remains to bound the term

$$\begin{aligned} & P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]\right\} \\ & \equiv I_{1n}+I_{2n}+I_{3n}, \end{aligned}$$

where

$$I_{1n}=\left(P_{\Delta n}-P_{\Delta}\right)\left\{\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)-\frac{S_1\left(t,\widehat{\theta}_n\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]}{S_0\left(t,\widehat{\theta}_n\right)}\right\},$$

$$I_{2n}=P_{\Delta n}\left\{\frac{S_1\left(t,\widehat{\theta}_n\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]}{S_0\left(t,\widehat{\theta}_n\right)}-\frac{S_{1n}\left(t,\widehat{\theta}_n\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]}{S_{0n}\left(t,\widehat{\theta}_n\right)}\right\},$$

$$I_{3n}=P_{\Delta}\left\{\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)-\frac{S_1\left(t,\widehat{\theta}_n\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]}{S_0\left(t,\widehat{\theta}_n\right)}\right\}.$$

Referring to the convergence rate results, we have $\left\|\widehat{\zeta}_n\left(\cdot,\widehat{\beta}_n\right)-\zeta_o\left(\cdot,\beta_o\right)\right\|_2=O_p\left(n^{-\min(p\nu,(1-\nu)/2)}\right)$, thus by Lemma 7 in Stone (1986) we get

$$\begin{aligned} \left\|\widehat{\zeta}_n\left(\cdot,\widehat{\beta}_n\right)-\zeta_o\left(\cdot,\beta_o\right)\right\|_\infty &= q_n^{1/2}O_p\left(n^{-\min(p\nu,(1-\nu)/2)}\right) \\ &= O_p\left(n^{-\min\left(\left(p-\frac{1}{2}\right)\nu,(1-2\nu)/2\right)}\right). \end{aligned}$$

Now resorting to Cor. 6.21 in Schumaker (1981) we have

$$\left\|\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right\|_\infty=O_p\left(n^{-\min\left(\left(p-\frac{3}{2}\right)\nu,(1-4\nu)/2\right)}\right). \tag{A.7}$$

Thus set $\delta_n=n^{-\min\left(\left(p-\frac{3}{2}\right)\nu,(1-4\nu)/2\right)}$ in the maximal inequality in Lemma 3.4.2 of van der Vaart and Wellner (1996), one arrives at

$$I_{1n}=O_p\left(n^{-1/2}q_n^{1/2}\delta_n\right)=n^{-1/2}O_p\left(n^{-\min((p-2)\nu,(1-5\nu)/2)}\right)=o_p\left(n^{-1/2}\right).$$

While $I_{2n}=o_p\left(n^{-1/2}\right)$ is confirmed by (A.5) as $\left\|\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right\|_\infty=o_p\left(q_n^{-1/2}\right)$ given our assumptions on p and ν.

When it comes to the third term, recall that for any function g

$$P_{\Delta}\left\{g-\frac{S_1\left(t,\theta_o\right)[g]}{S_0\left(t,\theta_o\right)}\right\}=E\left\{\Delta g-E\left[\Delta g|V=t,\Delta=1\right]\right\}=0,$$

thus we get the desired conclusion

$$
\begin{aligned}
& I_{3n} \\
& = P_{\Delta}\left\{ \frac{S_1\left(t,\theta_o\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]}{S_0\left(t,\theta_o\right)} - \frac{S_1\left(t,\widehat{\theta}_n\right)\left[\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right]}{S_0\left(t,\widehat{\theta}_n\right)} \right\} \\
& \leq c\left\|\dot{\widehat{\zeta}}_n\left(\cdot,\widehat{\beta}_n\right)-\dot{\zeta}_o\left(\cdot,\beta_o\right)\right\|_{\infty}\left\|\widehat{\zeta}_n\left(\cdot,\widehat{\beta}_n\right)-\zeta_o\left(\cdot,\beta_o\right)\right\|_2 \\
& = O_p\left(n^{-\min\left(\left(p-\frac{3}{2}\right)\nu,(1-4\nu)/2\right)}\right)\times O_p\left(n^{-\min(p\nu,(1-\nu)/2)}\right)=o_p\left(n^{-1/2}\right).
\end{aligned}
$$

■

Lemma 1.6.9 *We have the stochastic equicontinuity condition hold for the process:*

$$
P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[g]-s_n\left(,\theta_o\right)[g]\right\}-P_{\Delta}\left\{s\left(,\widehat{\theta}_n\right)[g]-s\left(,\theta_o\right)[g]\right\}=o_p\left(n^{-1/2}\right),
$$

with $g=\dot{\zeta}_o\left(\cdot,\beta_o\right)$ *or* $g=h^*$.

Proof. Following Huang (1999), wedecompose the term on LHS into two parts:

$$
J_{1n}=\left(P_{\Delta n}-P_{\Delta}\right)\left\{s\left(,\widehat{\theta}_n\right)[g]-s\left(,\theta_o\right)[g]\right\},
$$

and

$$
J_{2n}=P_{\Delta n}\left\{s_n\left(,\widehat{\theta}_n\right)[g]-s_n\left(,\theta_o\right)[g]\right\}-P_{\Delta n}\left\{s\left(,\widehat{\theta}_n\right)[g]-s\left(,\theta_o\right)[g]\right\}.
$$

$J_{2n}=o_p\left(n^{-1/2}\right)$ is proved by (A.6) in our Lemma 1.6.6.

The conclusion of $J_{1n}=o_p\left(n^{-1/2}\right)$ would follow from the $P-$Donsker property of $s\left(,\theta\right)[g]$ indexed by θ, hence the implied stochastic equicontinuity result. This assertion follows from the $P-$Donsker properties of $\mathcal{F}_2$ and $\mathcal{F}_3$ (or $\mathcal{F}_4$) and Donsker permanence property of the convex hull when taking expectations over the class $\mathcal{F}_2$ and $\mathcal{F}_3$ (or $\mathcal{F}_4$), appearing in the ratio term of $s\left(,\theta\right)[g]$, see Theorems 2.10.2 and 2.10.3 in van der Vaart and Wellner (1996). ■

There are two key ingradients in our proof for the convergence rate and asymptotic normality. They are $P-$Donsker property and the maximal inequality (A.3), once the complexity of certain functional class is pinned down. When we move to the bootstrap

world, the first property is guaranteed by the multiplier inequality in Lemma 3.6.7 of Van der Vaart and Wellner (1996) with general exchangeable weights. As for the second part, lemma 1 in Cheng and Huang (2010) would be important to characterize the convergence rate for bootstrapped processes. Let

$$\mathcal{F}_\delta = \{f(\cdot,\theta) - f(\cdot,\theta_o) : d(\theta,\theta_o) \leq \delta\},$$

Let $F_n(Z)$ be the envelope function of this class $\mathcal{F}_{\delta_n}$. If we have

$$\lim_{\lambda\to\infty} \lim \sup_{n\to\infty} \sup_{t\geq\lambda} P_Z(F_n(Z) > t) = 0, \tag{A.8}$$

for each sequence $\delta_n \to 0$. Then we have

$$\sup_{d(\theta,\theta_o)\leq\delta} |G_n^*(f(\cdot,\theta) - f(\cdot,\theta_o))| = O_{P_{ZW}}(\varphi_n^*(\delta)),$$

where

$$\varphi_n^*(\delta) \geq \mathcal{J}_{[]}(\delta,\mathcal{F}_\delta,\|\cdot\|_\infty)\left[1 + \frac{\mathcal{J}_{[]}(\delta,\mathcal{F}_\delta,\|\cdot\|_\infty)}{\delta^2\sqrt{n}}\right].$$

One sufficient condition to (A.8) is ensured by the uniform boundedness of the underlying class. Given our Assumptions (A1)-(A6), we do have the uniform boundedness for all those functional classes in Lemma 1.6.6.

Proof. (of Bootstrap Consistency) Repeating the argument in our proof of Theorem 1.5.2, one could get $d\left(\widehat{\theta}_n^*,\theta_o\right) = o_{P_{ZW}}(1)$. Then given Theorem 3 in Cheng and Huang, we have

$$d\left(\widehat{\theta}_n^*,\theta_o\right) = O_{P_W}\left(n^{-\min(p\nu,(1-\nu)/2)}\right),$$

in P_Z-probability. Following the proof of our Theorem 1.5.3, one arrives at

$$\sqrt{n}P_\Delta\{s(,\theta_o)[\widetilde{m}_o]\}^{\otimes 2}\left(\widehat{\beta}_n^* - \beta_o\right) = \sqrt{n}P_{\Delta n}^*\{s_n^*(,\theta_o)[\widetilde{m}_o]\} + o_{P_W}(1),$$

in P_Z-probability. Notice that

$$P_{\Delta n}^*\{s_n^*(,\theta_o)[\widetilde{m}_o]\} = \frac{1}{n}\sum_{i=1}^n W_{ni}\Delta_i\left\{\widetilde{m}_o(X_i) - \frac{S_{1n}^*(V_i,\theta_o)[\widetilde{m}_o]}{S_{0n}^*(V_i,\theta_o)}\right\}, \tag{A.9}$$

where

$$S_{1n}^*(t,\theta)[h] = \frac{1}{n}\sum_{i=1}^n W_{ni}Y_i(t)\exp\left[\psi\left(X_i'\widetilde{\beta}\right)\right]h(X_i),$$

and

$$S_{0n}^{*}\left(t,\theta\right)h=\frac{1}{n}\sum_{i=1}^{n}W_{ni}Y_{i}\left(t\right)\exp\left[\psi\left(X_{i}^{\prime}\widetilde{\beta}\right)\right].$$

Because $\widetilde{m}_o$ and θ_o is fixed at the true value in (A.9), the convergence rate of $S_{1n}^{*}\left(t,\theta_o\right)\left[\widetilde{m}_o\right]$ and $S_{0n}^{*}\left(t,\theta_o\right)$ to their corresponding limit $S_1\left(t,\theta_o\right)\left[\widetilde{m}_o\right]$ and $S_0\left(t,\theta_o\right)$ would be with root-n for $t\in[0,\tau]$. Hence we could proceed as in the Proposition 3 of Nan and Wellner (2013) by taking their $\Omega_i\left(t\right)=W_i\left(t\right)=W_{ni}$ for all t and i here. Analogous to the last set of equations on Page 1175 in Nan and Wellner (2013), we have

$$\begin{aligned}&\sqrt{n}P_{\Delta}\left\{s\left(,\theta_o\right)\left[\widetilde{m}_o\right]\right\}^{\otimes2}\left(\widehat{\beta}_{n}^{*}-\beta_o\right)\\&=\sqrt{n}P_{n}^{*}\left[\Delta\left(\widetilde{m}_o\left(X\right)-\frac{S_1\left(V,\theta_o\right)\left[\widetilde{m}_o\right]}{S_0\left(V,\theta_o\right)}\right)-\int s\left(t,\theta_o\right)\left[\widetilde{m}_o\right]I\left(Y\geq t\right)e^{\psi\left(X_{i}^{\prime}\widetilde{\beta}_o\right)}d\Lambda_o\left(t\right)\right]+o_{P_W}\left(1\right)\end{aligned}$$

in P_Z-probability. The part associated with the operation P in Nan and Wellner (2013) cancels out in our case, because of the independence of weights with the observation and the fact that $\frac{1}{n}\sum_{i=1}^{n}W_{ni}=1$. Finally, we obtain

$$\begin{aligned}&\sqrt{n}P_{\Delta}\left\{s\left(,\theta_o\right)\left[\widetilde{m}_o\right]\right\}^{\otimes2}\left(\widehat{\beta}_{n}^{*}-\widehat{\beta}_{n}\right)\\&=\sqrt{n}\left(P_{n}^{*}-P_n\right)\left[\Delta\left(\widetilde{m}_o\left(X\right)-\frac{S_1\left(V,\theta_o\right)\left[\widetilde{m}_o\right]}{S_0\left(V,\theta_o\right)}\right)-\int s\left(t,\theta_o\right)\left[\widetilde{m}_o\right]I\left(Y\geq t\right)e^{\psi\left(X_{i}^{\prime}\widetilde{\beta}_o\right)}d\Lambda_o\left(t\right)\right]+o_{P_W}\left(1\right)\end{aligned}$$

in P_Z-probability. The final display of the bootstrapped version converges weakly to the same limit as its asymptotic counterpart conditionally for almost surely observations by Theorem 3.16.3 in van der Vaart and Wellner (1996). ■

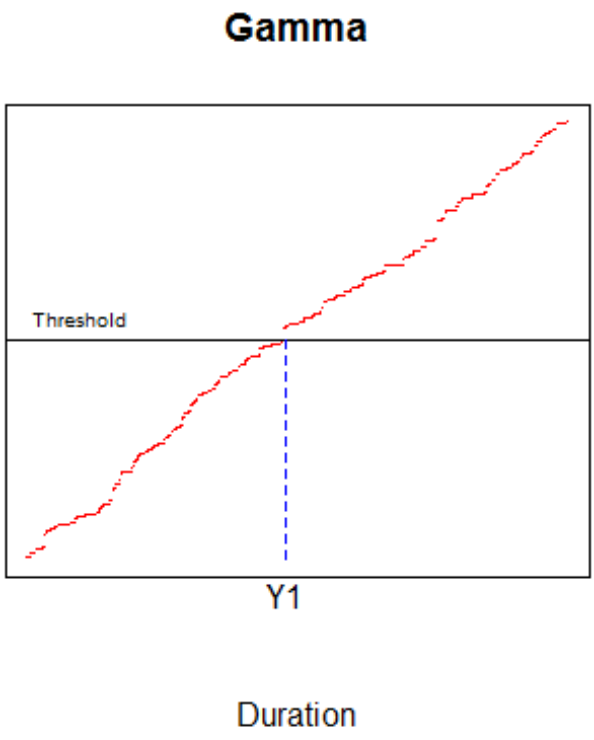

Figure 1.1: FPT of Gamma Processes

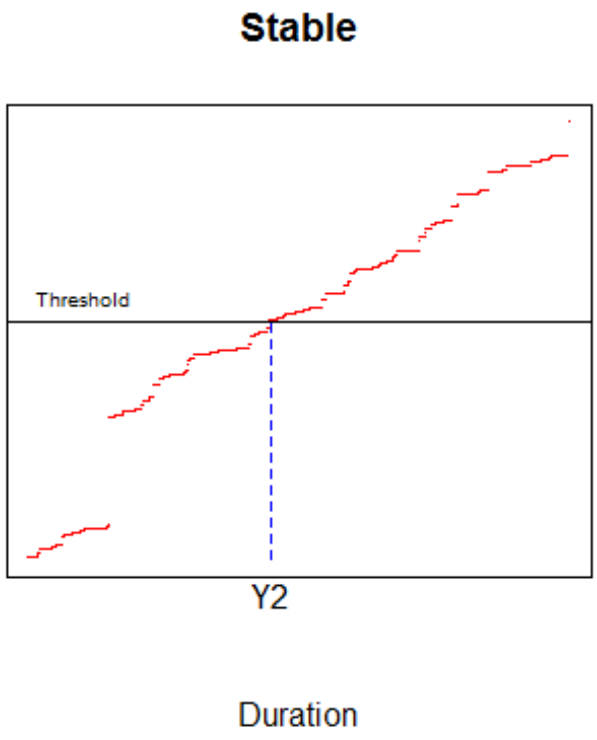

Figure 1.2: FPT of Stable Processes

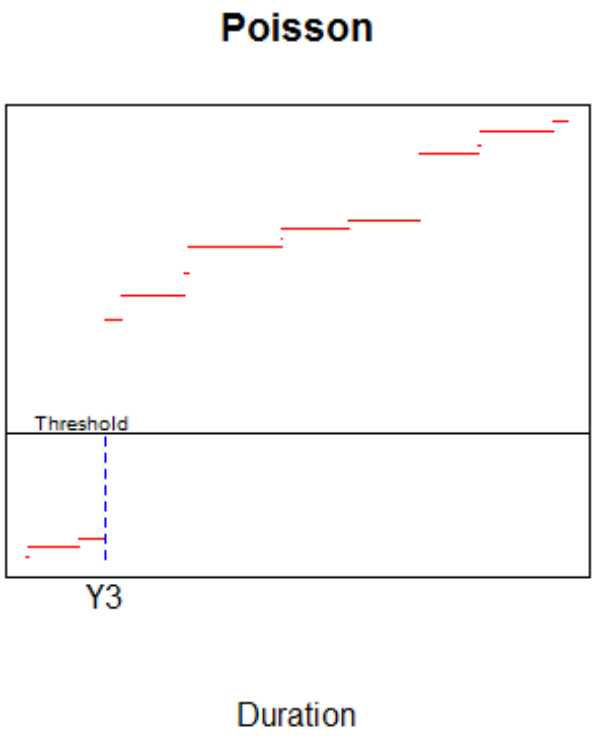

Figure 1.3: FPT of Compound Poisson Processes

Chapter 2

A COMPETING RISKS MODEL WITH TIME-VARYING HETEROGENEITY AND SIMULTANEOUS STOPPING

2.1 Introduction

In typical economic applications, a spell of duration could potentially end for various reasons, whereas only the realized duration is recorded in the dataset. Mathematically, this fits into the competing risks model framework, where two or more types of durations may occur on a subject, but only the one occurring first is observed together with its occurrence time, and other durations are censored. For two latent durations or failure times (Y_1^*, Y_2^*) of interest, economists or econometricians only observe the minimum $V = \min\{Y_1^*, Y_2^*\}$ and the cause of failure as $\arg\min_{\iota\in\{1,2\}}(Y_\iota^*)$. Applications of the competing risks model in economics[1] include Flinn and Heckman's (1982), who investigated the duration of unemployment where an individual can exit unemployment either by finding a job or by leaving the labor market; Katz and Meyer's (1990), who studied the probability of leaving unemployment through recalls and new jobs; and Deng, Quigley, and Van Order's (2000), who investigated mortgage termination by prepayment or default.

This chapter proposes and studies a new bivariate competing risks model that specifies each individual duration as the first time a stochastic process crosses a random threshold, and imposes dependence between two durations via the unobservable dependent bivariate processes. Here each marginal stochastic process is assumed to be a Lévy subordinator, i.e., a continuous-time process with stationary, independent, and non-negative increments, and two random thresholds are assumed to follow independent unit exponential distributions. Moreover, we appeal to the so-called Lévy copula (Kallsen and Tankov, 2006) to generate

[1] Upon a change of sign, the competing risks model nests the classical Roy model as well, where workers self-select the sector that gives them the highest expected earnings. Hence instead of dealing with the identified minimum, $\max\{Y_1^*, Y_2^*\}$ and $\arg\max\{Y_1^*, Y_2^*\}$ are observable, see Heckman and Honore (1990).

dependence between two latent Lévy subordinators without restricting their marginal specifications. The choice of latent processes, thresholds, and dependence structure in our model could be seen as a natural variant of the competing risks version of MPH from a process point of view. MPH could also be put into the same framework, with its duration defined by the first passage time of a deterministically increasing function (multiplied by a static heterogeneity term) crossing a unit exponential threshold, and in its competing risks version the dependence between two durations is caused by unobservable static heterogeneity terms. Another close connection to MPH is that the distribution of non-negative heterogeneity term is often assumed to be infinitely divisible, such as gamma distribution in Lancaster (1979) and stable distribution in Hougaard (1986a). Correspondingly in the present model, it is well known that any Lévy subordinator has non-negative infinitely divisible distribution (Sato, 1999). But in contrast with the traditional hazard-based models including MPH, our model embeds time-varying heterogeneity represented by the latent processes, and it allows simultaneous termination of both durations with positive probability. Surprisingly, the conditional distribution function of two durations in our model is of the same form as that in Marshall and Olkin (1967). Compared with the original and existing generalizations of Marshall-Olkin type model (Klein, Keiding, and Kamby, 1989; An, Christensen, and Gupta, 2004), the durations in our model are defined in a highly structural way and the concurrent termination of both durations is not necessarily driven by some independent and common shock. Thus, I will name the present model the Extended Marshall-Olkin (EMO) model.

EMO exhibits tractable mathematical structures and interesting probablistic properties, and has empirical content and various testable implications under a competing risks scheme where one does not observe both durations sequentially. The conditional joint survival function of structural durations admits a neat characterization, where structural components enter into the model via their natural forms, instead of various inverses (Laplace transform, or inverse of Lévy-Laplace exponent function) as in Abbring (2012). Several popular association measures between the bivariate durations, including Pearson's correlation coeffcient, Kendall's tau and Spearman's rho, are all expressed in closed forms, revealing how the dependence between two latent processes translates into the dependence between their first passage times, or the structural durations. For various semiparametric restric-

tions on the structural multiplicative effects from observable covariates, we could identify these effects, the baseline hazard function, and the characteristics of latent Lévy subordinators separately in a competing risks setting. When the structural multiplicative effects are of linear index form, the strength in identification directly leads to straigthforward semiparametric estimation methods based on average derivatives (Stoker, 1986). We establish root-n consistency and asymptotic normality of two estimators for the finite dimensional parameters in structural multiplicative effects. In comparison, although the competing risks version of MPH is known to be nonparametrically identifiable (Heckman and Honore, 1989; Abbring and van den Berg, 2003), only partial answers have been given regarding non-/semiparametric estimation, see Fermanian (2004) and Chen (2010). When it comes to the existing process-based models, the literature is restricted to the univariate duration models and their identificatiability, notably Abbring's MHT (2012). A numerical procedure inverting the Laplace transformation has been suggested to estimate a fully parametric MHT in Abbring and Salimans (2012).

It is worthwhile highlighting the novelty in our identification and inference strategy as existing methods do not apply here. In our model, the conditional survival copula function of two durations changes with covariates' value, unlike the models in Heckman and Honore (1989), Abbring and van den Berg (2003), and Lee and Lewbel (2013). Furthermore, this conditional survival copula is not Archimedean and does not even admit a density function, which goes beyond the framework in Fan and Liu (2013). But thanks to the Marshall-Olkin structure, we can first identify certain reduced-form regressions which are determined jointly by the structural multiplicative effects from observable covariates and the characteristics of latent Lévy subordinators. The introduction of those reduced-form regressions is not merely for mathmatical convenience, but also facilitates the examination of how the deep structural parameters of theoretical models are related to reduced-form parameters. Thereafter we proceed to identify all model primitives by imposing restrictions on the functional form of structural multiplicative effects with some innocuous normalizations. Along this journey, we explore the analytical properties of the Lévy-Laplace exponent function and the corresponding Lévy copula function. For the linear index model, we provide two root-n consistent and asymptotic normal estimators for the finite dimensional parameters in struc-

tural multiplicative effects derived from conditional moment restriction (Arnold, 1968) and local partial likelihood maximization (Fan, Gijbels, and King, 1997), respectively. Without point identification, I also construct a confidence set for the structural multiplicative effects by inverting a kernel-based test, allowing a wide range of underlying Lévy structures.

2.2 Related Works

Without observing both durations sequentially, the competing risks scheme induces delicate nonidentifiability issues once dependence between two durations is allowed, as established by Tsiatis (1975). Fortunately, the point identification is assured for the competing risks version of MPH by Heckman and Honore (1989), Abbring and van den Berg (2003) where the dependence is generated via the bivariate heterogeneity terms. Unlike their univariate counterpart, only partial answers have been given regarding non-/semiparametric estimation of MPH under competing risks. Fermanian (2004) has proved several consistency results for different model primitives based on kernel estimators, but no asymptotic distribution has been developed. Moreover, it is not clear what the optimal rates of convergence are there. With a *known* survival copula function between the heterogeneity terms, Chen (2010) has studied a semiparametric MLE without restricting the baseline hazard functions and obtained analogous results as in Murphy (1995).

Identification analysis of competing risks models has a long history dating back to Cox (1959). Tsiatis (1975) uses an explicit construction to demonstrate the non-identifiability of the marginal distribution function of Y once the independent censoring mechanism is dispensed with. Crowder (1991) further amplifies this identifiability crisis by proving that even if the two marginal distribution functions of (Y, C) are given, their joint distribution still remains unidentified generally. Peterson (1976) obtains the worst-case bounds for both the marginal distribution function of Y and the joint distribution function of Y and C. Bedford and Meijison (1997) characterize the situation when simultaneous failure is present.

In response to the identifiability crisis, three general approaches have been taken in the literature to achieve point identification of a competing risks model. First, covariate information and specific model structures imposed on the marginal distributions may re-

store point identification, see e.g., the proportional hazards and accelerated failure time models in Heckman and Honore (1989), Abbring and van den Berg (2003), Lee (2006), and Lee and Lewbel (2012); Second, in the spirit of Peterson (1976), worst-case bounds have been adapted to accelerated failure time model with discrete regressors in Honore and Lleras-Muney (2006), and censored quantile regression in Khan and Tamer (2009). Finally, assuming a *known* copula for the individual risks, Zheng and Klein (1995) first extend point identification results for independent risks to dependent risks and propose a copula-graphic estimator of the marginal survival function. When the copula function is Archimedean and *known*, Rivest and Wells (2001) first derive an explicit expression for the copula-graphic estimator of the survival function proposed in Zheng and Klein (1995). In the context of a censored quantile regression where the independent variable is subject to certain dependent censoring, Fan and Liu (2013) consider partial identification where the copula function is allowed to vary within a prespecified class. They demonstrate the possibility of getting more informative insight for the conditional quantile function over the approach taking worst-case bounds.

Our work is also complementary to the recent study on empirical game models involving duration outcomes. For strategic agents in those game models, the optimal durations could also be described via some threshold-crossing behaviors. Honore and de Paula (2010) have studied the identification of a simultaneous equation model where bivariate durations are determined by the optimal choice of two strategic agents. Their system resembles a classical simultaneous equation model in which endogenous variables interact with observable and unobservable exogenous components to characterize an economic environment. Despite the presence of multiple equilibria, all model primitives are shown to be identifiable by observing bivariate durations (not with competing risks) in Honore and de Paula (2010). In extension to allow simultaneous stopping phenomenon, Honore and de Paula (2014) consider the Nash bargaining solution in a cooperative game for joint retirement decisions for married couples. Honore and de Paula (2014) appeal to the indirect inference for structural parameters in their model. However, in their latter work no formal identification results have been proved and certain parameter values are fixed in the estimation.

2.3 An Extended Marshall-Olkin (EMO) Model

In this section, we will present the formal model setup of EMO and derive the conditional joint surivial function of the bivariate structural durations (Y_1^*, Y_2^*), highlighting its connection with the multivariate MPH. We further distinguish our modeling strategy with the original Marshall-Olkin (1967) model and existing generalizations. Pertaining to the specific conditional joint survival function, we discuss various probabilistic properties of EMO, especially its dependence structure which has important implications for its empirical content under competing risks in the next section.

2.3.1 Construction of EMO

We propose a bivariate model where each individual duration is defined as the first passage time of a latent Lévy subordinator crossing unit exponential threshold. And the dependence is generated via a bivariate dependent Lévy subordinator whose two margins are denoted as $L_1(t)$ and $L_2(t)$. Formally those two durations are defined by:

$$Y_\iota^* \equiv \inf\left\{t_\iota : \phi_\iota(X) L_\iota\left[\Lambda(t_\iota)\right] \geq e_\iota\right\}, \tag{2.3.1}$$

where $e_\iota \sim Exp(1)$ for $\iota = 1, 2$. Notice that we assume the time change function $\Lambda(\cdot)$ is common for both durations.

As already discussed in Section 2, not only is the class of Lévy subordinators flexible enough to incorporate a variety of interesting parametric prcesses (see Appendix E), but it also admits a sharp characterization via the well known Lévy-Khintchine representation. For each marginal Lévy subordinator $L_\iota(t)$, its Lévy-Laplace exponent function $\Phi_\iota(\cdot)$ is defined through the relation:

$$E\left\{\exp\left[-zL_\iota(t)\right]\right\} = \exp\left[-t\Phi_\iota(z)\right].$$

In fact $\Phi_\iota(\cdot)$ could be expressed as the following integral with respect to the jump intensity measure $\Pi_\iota(\cdot)$:

$$\Phi_\iota(z) = \int_{\mathcal{R}_+} \left[1 - e^{-zy}\right] \Pi_\iota(dy). \tag{2.3.2}$$

Here the characterizing jump measure $\Pi_\iota(\cdot)$ is known to be a Lévy measure (see the definition in Appendix E) and we denote its corresponding Lévy density function as $\pi_\iota(\cdot)$, if $\Pi_\iota(\cdot)$ is absolutely continuous with respect to the Lebesgue measure.

A key concept capturing the dependence structure of the Lévy processes is the so-called Lévy copula, introduced in Cont and Tankov (2003), Kallsen and Tankov (2006). It is worth emphasizing that the Lévy copula is not a standard copula coupling two marginal distribution functions, rather it is connected with the Lévy measure which measures the jump intensity. Nevertheless, it does share certain similarities with the standard distributional or survival copula functoin, including the Sklar theorem and its role in the joint Lévy-Laplace exponent function $\Psi_{12}(z_1, z_2)$:

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

Here we have

$$\Phi_{12}(u, v) = \iint_{\mathcal{R}_+^2} \left[1 - e^{-y_1 u - y_2 v}\right] d\mathcal{C}_L\left(\overline{\Pi}_1(y_1), \overline{\Pi}_2(y_2)\right), \tag{2.3.3}$$

where $\mathcal{C}_L(\cdot,\cdot)$ is the Lévy copula coupling the two marginal Lévy tail integrals $\overline{\Pi}_\iota(y) = \int_y^\infty \pi_\iota(y)\, d\mu(y)$. We refer the readers to Appendix E for a probablistic interpretation of the tail integral. Theorem 4.1 in Cont and Tankov (2003) ensures that (2.3.3) could actually be marginalized into (2.3.2).

The next theorem contains the conditional joint survival function of (Y_1^*, Y_2^*). Notice that our model exhibits a singularity, meaning there is a positive probability mass assigned on the diagonal $x = y$, whose two-dimensional Lebesgue measure is equal to zero.

Theorem 2.3.1 *Assume that the bivariate durations are defined by (2.3.1) with mutually independent covariates X, Lévy subordinators $(L_1(t), L_2(t))$, and unit exponential thresholds (e_1, e_2). If two thresholds are independent as well, then we have the conditional joint survival function as*

$$\begin{aligned} & \Pr\left\{Y_1^* > t_1, Y_2^* > t_2 | X = x\right\} \\ = \; & \exp\left[-\Psi_1(x)\Lambda(t_1) - \Psi_2(x)\Lambda(t_2) - \Psi_3(x)\Lambda\left[\max(t_1, t_2)\right]\right], \end{aligned} \tag{2.3.4}$$

where

$$\begin{aligned}\Psi_1(x) &= \Phi_{12}(\phi_1(x),\phi_2(x)) - \Phi_2(\phi_2(x)), \\ \Psi_2(x) &= \Phi_{12}(\phi_1(x),\phi_2(x)) - \Phi_1(\phi_1(x)), \\ \Psi_3(x) &= \Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x)) - \Phi_{12}(\phi_1(x),\phi_2(x)),\end{aligned} \tag{2.3.5}$$

moreover all $\Psi_j(x) \geq 0$ *for* $j = 1, 2, 3$.

Proof. Since the processes have independent increments, we consider two cases separately in our proof. When $t_1 \leq t_2$, we get

$$\begin{aligned}&\Pr\{Y_1^* > t_1, Y_2^* > t_2 | X = x\} \\ =\ &\Pr\{e_1 > \phi_1(x) L_1[\Lambda(t_1)], e_2 > \phi_2(x) L_2[\Lambda(t_2)]\} \\ =\ &E\{\exp(-\phi_1(x) L_1[\Lambda(t_1)] - \phi_2(x) L_2[\Lambda(t_2)])\} \\ =\ &E\{\exp[-(\phi_1(x) L_1[\Lambda(t_1)] + \phi_2(x) L_2[\Lambda(t_1)]) - \phi_2(x)(L_2[\Lambda(t_2)] - L_2[\Lambda(t_1)])]\} \\ =\ &\exp[-(\Lambda(t_2) - \Lambda(t_1))\Phi_2(\phi_2(x)) - \Lambda(t_1)\Phi_{12}(\phi_1(x),\phi_2(x))].\end{aligned}$$

The first equality comes from the definition, and the second one utilizes the distributional assumption on the random thresholds. The third equality is prepared for using the independence between increments of Lévy processes, while the final one follows the Lévy-Khintchine representation.

Analogously for the other case $t_1 \geq t_2$, we have:

$$\begin{aligned}&\Pr\{Y_1^* > t_1, Y_2^* > t_2 | X = x\} \\ =\ &\exp[-(\Lambda(t_1) - \Lambda(t_2))\Phi_1(\phi_1(x)) - \Lambda(t_2)\Phi_{12}(\phi_1(x),\phi_2(x))].\end{aligned}$$

Now the unified expression in (2.3.4) just summarizes those two cases and those functions $\Psi_j(x)$ arise from a simple reparameterization. It is not immediately clear that all $\Psi_j(x)$ are non-negative, we present the proof of the non-negativeness of $\Psi_j(x)$ in Appendix A, by making use of the Frechet-Hoeffding type bound for Lévy copula. ■

Recall that in our univariate model, $L(\Lambda(t))$ plays a similar role as $\zeta\Lambda(t)$ in MPH, while allowing time-varying heterogeneity. The multivariate MPH model in Heckman and

Honore (1989), Abbring and van der Berg (2003) starts with a specification on the marginal conditional hazard density function in terms of the proportional effect from covariate, the frailty term, and the baseline hazard function:

$$\begin{aligned}\theta_{Y_1^*}(t|X,\zeta) &= \phi_1(X)\zeta_1\lambda_1(t),\\ \theta_{Y_2^*}(t|X,\zeta) &= \phi_2(X)\zeta_2\lambda_2(t).\end{aligned}$$

And the underlying dependence between (Y_1^*, Y_2^*) is produced via the dependent frailty pair (ζ_1, ζ_2). Analogously our model specifies the structural duration individually as the first passage time via (2.3.1), and the dependence is generated by the bivariate Lévy subordinators $(L_1(t), L_2(t))$, essentially through the Lévy copula. Referring to Example 2.4 and Example 2.5, if one chooses the marginal processes to be gamma or stable, our EMO could be seen as the process-based versions of the correlated gamma frailty model in Parner (1998), or the stable frailty model in Hougaard (1986b).

Example 2.3.2 *In the absence of covariates X and further assuming $(L_1(t), L_2(t))$ is a compound Poisson process with bivariate non-negative shocks, our model boils down to the bivariate random shock model studied by Marshall and Shaked (1979). Under this circumstance, the joint survival function of (Y_1^*, Y_2^*) could be determined explicitly by model primitives without referring to the Lévy-Khintchine representation or Lévy copula. Here each Lévy subordinator is $L_\iota(t) = \sum_{i=1}^{N(t)} W_{i,\iota}$ for $\iota = 1, 2$, and the arrival of shocks is governed by a common Poisson process $N(t)$ with rate λ.*[2] *The joint survival function of (Y_1^*, Y_2^*) is*

$$\begin{aligned}&\Pr\{Y_1^* > t_1, Y_2^* > t_2\}\\ =\ &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\exp(-\lambda t_1)(\lambda t_1)^k}{k!}\frac{\exp(-\lambda(t_2-t_1))(\lambda(t_2-t_1))^k}{l!}E\left[F^{[k,k+l]}(e_1,e_2)\right],\end{aligned}$$

for $t_1 \le t_2$, or

$$\begin{aligned}&\Pr\{Y_1^* > t_1, Y_2^* > t_2\}\\ =\ &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\exp(-\lambda t_2)(\lambda t_2)^k}{k!}\frac{\exp(-\lambda(t_1-t_2))(\lambda(t_1-t_2))^k}{l!}E\left[F^{[k+l,k]}(e_1,e_2)\right],\end{aligned}$$

[2] The single Poisson process is not restrictive because several Poisson processes can always be superimposed to yield a new Poisson process and the several sequences of damage distributions can be replaced by a sequence of appropriate mixture, see Marshall and Shaked (1979).

for $t_1 \geq t_2$. And in the above expression,

$$F^{[k_1,k_2]}(w_1, w_2) \equiv \Pr\left\{\sum_{j=1}^{k_1} W_{j,1} \leq w_1, \sum_{j=1}^{k_2} W_{j,2} \leq w_2\right\}.$$

Remark 2.3.3 *The threshold-crossing behaviors of structural durations also arise from empirical game models. Honore and de Paula (2010) have studied the identification of a simultaneous equation model where bivariate durations are determined by the optimal choice of two strategic agents. Covariates and unobservable heterogeneity terms are employed similar as in the generalized accelerated failure time model:*

$$\begin{aligned} Y_1^* &\equiv \inf\left\{t_1 : \phi_1(X)\Lambda(t_1)\exp\left(\delta 1\left\{Y_2^* \leq t_1\right\}\right) \geq e_1\right\}, \\ Y_2^* &\equiv \inf\left\{t_2 : \phi_2(X)\Lambda(t_2)\exp\left(\delta 1\left\{Y_1^* \leq t_2\right\}\right) \geq e_2\right\}. \end{aligned}$$

In this model, thresholds (e_1, e_2) are allowed to be arbitrary, but the leader-follower effect parameter δ and deterministic trend function $\Lambda(\cdot)$ are assumed to be common for both players. Their system resembles a classical simultaneous equation model in which endogenous variables interact with observable and unobservable exogenous components to characterize an economic environment. Despite the presence of multiple equilibria, all model primitives are shown to be identifiable by observing bivariate durations (not with competing risks) in Honore and de Paula (2010). In extension to incorporate simultaneous stopping phenomenon, Honore and de Paula (2014) consider the Nash bargaining solution in a cooperative game for joint retirement decisions for married couples. In comparison with their simultaneous equition model, our model could be viewed as a seemingly unrelated regression as the dependence is produced via the unobservable dependent processes.

2.3.2 The Original Marshall-Olkin Model and Existing Generalizations

The prominent Marshall-Olkin (1967) model is widely adopted in biostatistics, engineering, and life testing where simultaneous failures of human organs, components, or engines are present. The original model does not involve any Lévy subordinator and is constructed with three types of independent shocks: one leading to joint spell termination and two leading to individual spell termination. The shocks to the two durations are assumed to form

three independent Poisson processes with (positive) parameters λ_1, λ_2, and λ_3; depending on whether the shock kills only component 1, only component 2, or both components simultaneously. The times Z_1, Z_2, and Z_3 of occurrence of these three shocks are independent exponential random variables with parameters λ_1, λ_2, and λ_3, respectively. So $Y_1^* = \min(Z_1, Z_3)$ and $Y_2^* = \min(Z_2, Z_3)$ follow the joint distribution given below:

$$\Pr\{Y_1^* > t_1, Y_2^* > t_2\} = \exp\left[-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max(t_1, t_2)\right].$$

In comparison, our construction specifies $Y_\iota^* = \inf\{t : L_\iota(t) \geq e_\iota\}$ for $\iota = 1, 2$, and those parameters could also be expressed as

$$\begin{aligned}
\lambda_1 &= \Phi_{12}(1,1) - \Phi_2(1), \\
\lambda_2 &= \Phi_{12}(1,1) - \Phi_1(1), \\
\lambda_3 &= \Phi_1(1) + \Phi_2(1) - \Phi_{12}(1,1).
\end{aligned}$$

Maintaining the presence of a common shock, Klein, Keiding and Kamby (1989), An, Christensen and Gupta (2004) have generalized the Marshall-Olkin model with covariates and frailty terms. The observed bivariate durations are still assumed to be $Y_1^* = \min(Z_1, Z_3)$ and $Y_2^* = \min(Z_2, Z_3)$, then a conditional marginal hazard density model like MPH is specified for each Z_j for $j = 1, 2, 3$. Although estimation procedures have been suggested in both papers, neither of them presents any large sample theory for the proposed estimators. Honore and de Paula (2010, 2014) have criticized this type of generalization of Marshall-Olkin model due to a lack of structural interpretation. And they further point out the existence of a common shock to both durations seems too restrictive for economic applications. It is self-evident that our alternative construction has circumvanted those shortcomings.

2.3.3 Dependence Properties

In EMO, dendence between the two duration variables is generated by dependence between the unobserved characteristics of latent stochastic processes. Referring to a bivariate Brownian motion, it is obvious that covariance structure between those two marginal Brownian

motions captures all the dependence. The answer for pure jumping processes is not immediately clear, as the dependence comes from the relationship of their jump intensities. Such a dependence characterization relies on the notion of Lévy copula. Moreover we are actually interested how the dependence between latent processes transfers into the dependence between structural durations. In EMO, both the Pearson's correlation coefficient and two popular rank correlation coefficients, Kendall's tau and Spearman's rho for the bivariate durations are given in closed forms. In this context, the advantage of EMO is borne out, compared with the calculation and bounds analysis for bivariate MPH in van den Berg (1997).

Lévy Copula and Association Measures

Cont and Tankov (2003), Kallsen and Tankov (2006) introduce the concept of Lévy copula and advocate its use in dependence modeling over the standard distributional copula for Lévy processes. Except in special cases such as stable processes, the distributional copula is typically time-varying. For given two marginal infinitely divisible distributions, it is unclear which distributional copula generates a joint infinitely divisible distribution. Furthermore, the laws of multivariate Lévy process are conveniently specified by their Lévy measures, only in a few cases the distribution or probability density function could be given in closed forms.

For Lévy subordinators, any Lévy copula is between the two extreme cases, the independent Lévy copula $\mathcal{C}_{L,\perp}(u,v)$ and the completely dependent Levy copula $\mathcal{C}_{L,\parallel}(u,v)$ with the following expressions (Cont and Tankov, 2003):

$$\begin{aligned} \mathcal{C}_{L,\perp}(u,v) &= u \cdot I\{v=\infty\} + v \cdot I\{u=\infty\}, \\ \mathcal{C}_{L,\parallel}(u,v) &= \min\{u,v\}. \end{aligned}$$

A convenient parametric form which can represent the dependence level between those two extremes is the Lévy-Clayton copula with a single parameter $\alpha > 0$:

$$\mathcal{C}_{L,\alpha}(u,v) = \left(u^{-\alpha} + v^{-\alpha}\right)^{-1/\alpha}. \tag{2.3.6}$$

Admitting that the Lévy copula is suitable to capture dependence between two Lévy

subordinators, our goal is to understand the dependence structure imposed on the resulting first passage times, or the structural durations. A straightforward algebra shows the conditional survival copula function of two durations is:

$$\mathcal{C}_S\left(u,v|x\right)=uv\min\left(u^{-\omega_1(x)},v^{-\omega_2(x)}\right), \tag{2.3.7}$$

where

$$\omega_1\left(x\right)=\frac{\Psi_3\left(x\right)}{\Psi_1\left(x\right)+\Psi_3\left(x\right)}\text{ and }\omega_2\left(x\right)=\frac{\Psi_3\left(x\right)}{\Psi_2\left(x\right)+\Psi_3\left(x\right)}.$$

To see this, note that $S^{-1}_{Y^*_\iota}\left(u|x\right)=\Lambda^{-1}\left(\frac{-\log u}{\Psi_\iota(x)+\Psi_3(x)}\right)$ for $\iota=1,2$, then by monotonicity of $\Lambda\left(\cdot\right)$ and non-negativeness of those coefficients:

$$\begin{aligned}\mathcal{C}_S\left(u,v|x\right)&=S_{Y^*_1,Y^*_2}\left(S^{-1}_{Y^*_1}\left(u|x\right),S^{-1}_{Y^*_2}\left(v|x\right)|x\right)\\&=\begin{cases}uv^{1-\omega_2(x)}, & \text{if }\frac{\log u}{\Psi_1(x)+\Psi_3(x)}\leq\frac{\log v}{\Psi_2(x)+\Psi_3(x)}\\u^{1-\omega_1(x)}v, & \text{if }\frac{\log u}{\Psi_1(x)+\Psi_3(x)}\geq\frac{\log v}{\Psi_2(x)+\Psi_3(x)}\end{cases},\end{aligned}$$

which equals to (2.3.7) indeed. This is known as the Marshall-Olkin copula or the generalized Cuadras-Auge copula. Although this copula is singular without a copula density function, it is an extreme copula with its Pickands' function (see Section 3.3.4 in Nelsen, 2006):

$$A\left(t|x\right)=1-\min\left(\omega_2\left(x\right)t,\omega_1\left(x\right)\left(1-t\right)\right).$$

Notice that the conditional survival copula function (2.3.7) changes with x as well, in contrast with the commonly considered situations with an invariant copula in Heckman and Honore (1989), Abbring and van der Berg (2003). The conditional survival copula reveals that our subsequent analysis will not fit into the framework in Fan and Liu (2013), because it is not Archimedean, neither it admits a copula density function with respect to the two dimensional Lebesgue measure. Nevertheless, the dependence structure still delivers very informative empirical content even under competing risks setting in the next section.

Next proposition summarizes several dependence measures which are directly linked to this (conditional) generalized Cuadras-Auge copula, see Nelson (2006). This should be contrasted with the bivariate MPH in van den Berg (1997), where only certain bounds for Kendall's tau are available.

Proposition 2.3.4 *Let $\tau_K(x)$ stand for (conditional) Kendall's tau and $\rho_S(x)$ for Spearman's rho. Moreover, let $t_U(x)$, $t_L(x)$ denote the (conditional) upper and lower tail dependence parameters. Then they are of closed forms in EMO:*

$$\begin{aligned}
\tau_K(x) &= \frac{\omega_1(x)\,\omega_2(x)}{\omega_1(x)+\omega_2(x)-\omega_1(x)\,\omega_2(x)},\\
\rho_S(x) &= \frac{3\omega_1(x)\,\omega_2(x)}{2\left[\omega_1(x)+\omega_2(x)\right]-\omega_1(x)\,\omega_2(x)},\\
t_U(x) &= \min\left(\omega_2(x),\omega_1(x)\right),\ t_L(x)=0.
\end{aligned}$$

One nice property related to the Lévy-Clayton copula is that the joint Lévy exponent function (2.3.3) can be futher simplified as:

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

where

$$\kappa(u,v;\alpha)=uv\iint_{\mathcal{R}_+^2}e^{-y_1u-y_2v}\mathcal{C}_{L,\alpha}\left(\overline{\Pi}_1(y_1),\overline{\Pi}_2(y_2)\right)dy_1dy_2, \tag{2.3.8}$$

see Proposition 1 in Epifani and Lijoi (2010). Note that the function $\kappa(u,v;\alpha)$ summarizes all the dependence structure when the Lévy-Clayton copula is adopted. Given the representation (2.3.8), we have the following expressions for the coefficients in the extended Marshall-Olkin model:

$$\begin{aligned}
\Psi_1(x) &= \Phi_1(\phi_1(x))-\kappa(\phi_1(x),\phi_2(x);\alpha), \qquad (2.3.9)\\
\Psi_2(x) &= \Phi_2(\phi_2(x))-\kappa(\phi_1(x),\phi_2(x);\alpha),\\
\Psi_3(x) &= \kappa(\phi_1(x),\phi_2(x);\alpha).
\end{aligned}$$

Consequently, the expressions for those exponents in (2.3.7) simplify to

$$\omega_1(x)=\frac{\kappa(\phi_1(x),\phi_2(x);\alpha)}{\Phi_1(\phi_1(x))}\text{ and }\omega_2(x)=\frac{\kappa(\phi_1(x),\phi_2(x);\alpha)}{\Phi_2(\phi_2(x))}.$$

Moreover, the dependence of two Lévy processes could be translated into the (conditional) Pearson's correlation coefficient between those two passage times Y_1^* and Y_2^*. For bivariate MPH with Weibull baseline hazard functions, van den Berg (1997) obtains a closed form expression for Pearson's correlation coefficient, see equation (6) in van den Berg. In comparison, our general expression (2.3.10) in the next proposition does not restrict the form of baseline hazard function.

Proposition 2.3.5 *(Epifani and Lijoi, 2010) Let* $\rho_\alpha\left(Y_1^*, Y_2^*|x\right)$ *stand for the conditional correlation coefficient when the Lévy copula is (2.3.6), then*

$$\rho_\alpha\left(Y_1^*, Y_2^*|x\right) = \frac{\iint_{\mathcal{R}_+^2} e^{-\Lambda(t)\Phi_1(\phi_1(x))-\Lambda(t)\Phi_2(\phi_2(x))}\left[e^{\Lambda(s\wedge t)\kappa(\phi_1(x),\phi_2(x);\alpha)}-1\right]dsdt}{\prod_{j=1,2}\sqrt{2\int_0^\infty te^{-\Lambda(t)\Phi_j\left(\phi_j(x)\right)}dt-\left(\int_0^\infty e^{-\Lambda(t)\Phi_j\left(\phi_j(x)\right)}dt\right)^2}}. \quad (2.3.10)$$

In the special case when $\Lambda(t) = t$*, we have the classical result as in Marshall and Olkin (1967):*

$$\rho_\alpha\left(Y_1^*, Y_2^*|x\right) = \frac{\kappa\left(\phi_1(x), \phi_2(x); \alpha\right)}{\Phi_1\left(\phi_1(x)\right) + \Phi_2\left(\phi_2(x)\right) - \kappa\left(\phi_1(x), \phi_2(x); \alpha\right)}.$$

More Testable Implications and Applications

Regarding the Marshall-Olkin structure, we discuss the bivariate memory-less property and conditional indpendence proporty for duration outcomes in EMO. For the independent case, we point out its potential applicability in the second-price auction model (or the $k-$out of$-p$ system).

Remark 2.3.6 *According to the characterization in Muliere and Scarsini (1987), (2.3.4) is the only bivariate function (fixing* x*) which satisfies the following weak bivariate memory-less property:*

$$\Pr\left\{Y_1 > t_1 * t, Y_2 > t_2 * t | Y_1 > t, Y_2 > t, X = x\right\} = \Pr\left\{Y_1 > t_1, Y_2 > t_2 | X = x\right\},$$

given the marginals satisfy the memory-less property for the same operation $*$*. The notation* $*$ *is the biniary operation assumed to be associative and reducible, for example* $*$ *could be addition* $+$*, see Muliere and Scarsini (1987). The interpretation would be that given both components have survived to time* t*, then the conditional probability of both components surviving an additional time* (t_1, t_2) *is the same as the unconditional probability of surviving to time* (t_1, t_2) *starting at the origin.*

Remark 2.3.7 *In EMO we actually have conditional independence between* Y *and* D*. Kochar and Proschan (1991) obtained a complete characterization of this independence between* Y *and* D*, which asserts that the conditional cause specific hazard density functions of different failure causes are proportional to each other. Indeed, in our model (2.3.4), the sub-hazard density functions are all proportional to the derivative of the shape function* $\Lambda(t)$*.*

Remark 2.3.8 *When two latent Lévy subordinators are independent, there is no simultaneous failure. Because here the Lévy copula is the independent one* $\mathcal{C}_{L,\perp}(u,v)$*. Hence we have* $\Phi_{12}(u,v) = \Phi_1(u) + \Phi_2(v)$ *and those reduced-form regressions are* $\Psi_1(x) = \Phi_1(\phi_1(x))$, $\Psi_2(x) = \Phi_2(\phi_2(x))$, *and* $\Psi_3(x) \equiv 0$*. Therefore the conditional joint survival function is simply the product of two marginals:*

$$\begin{aligned} & \Pr\{Y_1^* > t_1, Y_2^* > t_2|x\} = \exp\left[-\Psi_1(x)\Lambda(t_1) - \Psi_2(x)\Lambda(t_2)\right] \\ = \;& \Pr\{Y_1^* > t_1|x\}\Pr\{Y_2^* > t_2|x\}. \end{aligned}$$

The independent bivariate exponential distribution (when $\Lambda(t) = t$*) is the only one satisfying the above strong bivariate memory-less property as noted by Marshall and Olkin (1967):*

$$\Pr\{Y_1^* > s_1 + t_1, Y_2^* > s_2 + t_2|x\} = \Pr\{Y_1^* > s_1, Y_2^* > t_2|x\} \times \Pr\{Y_1^* > t_1, Y_2^* > t_2|x\},$$

for all $s_1, s_2, t_1, t_2 > 0$*.*

Remark 2.3.9 *In the independence case, our model is also useful in application of the* $k-$*out of*$-p$ *system or its special case, the second price auction where the second highest bid and the identity of bidders being observed. Now we have* p *durations defined by*

$$Y_j^* \equiv \inf\left\{t_j : \phi_j(X) L_j(t_j) \geq e_j\right\},\ j = 1, .., p, \tag{2.3.11}$$

where all the components are independent and the Lévy subordinators $L_j(t)$ *are assumed to be mutually independent. Instead of observing all the durations, one could observe* $Y_{(r)}^*$*, the* $r-$*th order statistic of* $\left\{Y_j^*\right\}_{j=1}^p$ *with* $r = p-k+1$ *and the identity of this failed component. In reliability studies, this is the so-called* $k-$*out of*$-p$ *system where the system fails if and only if the first* k *components fail. Now by the definition in (2.3.11) we have conditional independent exponential distributions:*

$$\Pr\left\{Y_j^* > t|x\right\} = \exp\left[-\Psi_j(x)\, t\right],$$

with $\Psi_j(x) = \Phi_j\left(\phi_j(x)\right)$*. According to Basu and Ghosh (1983), all* $\Psi_j(x)$ *could be identified from the* $k-$*out of*$-p$ *system. So our estimation and inference procedure presented below may apply in this context as well.*

2.4 Semiparametric Identification and Estimation

In this section we consider the empirical content of EMO under competing risks. Without observing both durations sequentially, we only get $V = \min(Y_1^*, Y_2^*)$ and the cause of failure D defined slightly different from the one in a standard competing risks model without simultaneous failures. Here $D = 1$ if $Y_1^* < Y_2^*$, $D = 2$ if $Y_2^* < Y_2^*$, and $D = 3$ if $Y_1^* = Y_2^*$. From the discussion in Section 3, (Y_1^*, Y_2^*) are still conditionally dependent. In this chapter, we denote the transpose of a vector v by v^T.

2.4.1 Empirical Content of EMO under Competing Risks

Despite the difficulty associated with the general competing risks problem, we get a very concrete characterization of the empirical content of EMO. Thanks to the specific structure of Marshall-Olkin model, we could first identify certain reduced-form regressions which are determined jointly by the structural multiplicative effects from observable covariates and the characteristics of latent Lévy subordinators. The introduction of those reduced form regressions facilitates the examination of how the deep structural parameters of some important theoretical models can be related to reduced-form parameters. Thereafter we proceed to identify all model primitives by imposing restrictions on the functional form of structural regressions with some innocuous normalizations. The following table summaries the comparsion made for three models. Clearly, our EMO embeds time-varying heterogeneity and inherits nice properties from MPH in terms of identifiability and flexiable semiparametric estimation under competing risks.

Insert Table 2.1 here

The following main theorem states the identifiability of all model primitives given additional restrictions on the functional form of $(\phi_1(x), \phi_2(x))$. We first list all the regularity assumptions needed for identification.

Assumption (D0) (i) The time-change function $\Lambda : \mathcal{R}_+ \to \mathcal{R}_+$, is nondecreasing and absolutely continuous with its density denoted by $\lambda(\cdot)$. Moreover, there exist a point x_0 s.t. $\Phi_{12}(\phi_1(x_0), \phi_2(x_0)) = 1$.

(ii) The two marginal Lévy measures satisfy $\int_{\mathcal{R}_+} y_\iota \Pi_\iota (dy_\iota) < \infty$ for $\iota = 1, 2$ and the joiny Lévy measure satisfies $\int_{\mathcal{R}_+^2} y_1 y_2 \Pi_{12} (dy_1, dy_2) < \infty$. Moreover, two marginal tail integrals $\overline{\Pi}_\iota$ are continuous for $\iota = 1, 2$.

(iii) The support of $(\phi_1 (X), \phi_2 (X))$ contains an open rectangle in $\mathcal{R}_+^2$.

Assumption (D1) (Index Model) For $\iota = 1, 2$ assume $\phi_\iota (x) = \exp\left(x^T \beta_\iota\right)$, and the support of X is not contained in any linear subspace of $\mathcal{R}^d$, moreover $\beta_\iota \in \mathcal{S}_+^{d-1}$ defined as

$$\mathcal{S}_+^{d-1} \equiv \{\beta : \|\beta\|_2 = 1, \text{ and } \beta\text{'s first non-zero coordinate is positive}\}.$$

Assumption (D2) (Additive Model) For $\iota = 1, 2$ assume $\phi_\iota (x) = \phi_{\iota,1} (x_1) + \cdots + \phi_{\iota,d} (x_d)$, and $\phi_{\iota,l} (x_l)$ are nonconstant for at least two values of l and the following normalization restrictions hold

$$\int \phi_{\iota,l} (x_l) \, dx_l = 1, \text{ and } \sum_{l=1}^{d} \int \left[\phi_{\iota,l} (x_l)\right]^2 dx_l = d.$$

Assumption (D3) (Multiplicative Model) Assume the covariate X consists of $\left(X_o, \widetilde{X}\right)$ with a univariate X_o and $\phi_\iota (x) = x_o \times \widetilde{\phi}_\iota (\widetilde{x})$ for $\iota = 1, 2$. Moreover 0 is a limit point of the support of X_o, and there exists a normalizing point $\widetilde{x}_*$ s.t. $\widetilde{\phi}_\iota (\widetilde{x}_*) = 1$, for $\iota = 1, 2$.

Theorem 2.4.1 *(i) Let Assumption (D0) holds, then given a unique pair of* $(\phi_1 (x), \phi_2 (x))$, $(\pi_\iota, \mathcal{C}_L)$ *with* $\iota = 1, 2$ *are also unique in the competing risks setting. (ii) In addition if either one of Assumption (D1), (D2) or (D3) holds, then all the structural components are point identified.*

The following proposition collects the expressions for the sub-density functions and already appears in Marshall and Olkin (1967). This convenient form is also utilized by Basu and Ghosh (1978) to show the identifiability of standard MO distribution under competing risks. A short proof is given in the Appendix A for completeness.

Proposition 2.4.2 *(Basu and Ghosh, 1978) Assume* $\Lambda (\cdot)$ *is absolutely continuous with its*

density function $\lambda(\cdot)$*, then conditional sub-density functions take the following forms:*

$$\begin{aligned} f_{Y,D=1}(t|x) &= \lambda(t)\,\Psi_1(x)\exp\left[-\Lambda(t)\,\Phi_{12}(\phi_1(x),\phi_2(x))\right], \\ f_{Y,D=2}(t|x) &= \lambda(t)\,\Psi_2(x)\exp\left[-\Lambda(t)\,\Phi_{12}(\phi_1(x),\phi_2(x))\right], \\ f_{Y,D=3}(t|x) &= \lambda(t)\,\Psi_3(x)\exp\left[-\Lambda(t)\,\Phi_{12}(\phi_1(x),\phi_2(x))\right], \end{aligned} \tag{2.4.1}$$

where $D=1$ *if* $Y_1^* < Y_2^*$*,* $D=2$ *if* $Y_2^* < Y_2^*$*, and* $D=3$ *if* $Y_1^* = Y_2^*$*.*

The normalization assumption in (D0)(i) is innocuous since we do not observe the magnitude of the latent process anyway. The variation restriction in (D0)(iii) on structural regressions is rather mild in view of Abbring and van den Berg (2003). Assumption (D0)(ii) requires the marginal Lévy measure has a finite first moment, which turns out to be equivalent as the corresponding moment restriction on $L_1(1)$ or $L_2(1)$, see Wolfe (1971). Correspondingly, assuming the finiteness of the first moment of a frailty term is widely adopted in the identification results of MPH, see Elbers and Ridder (1982), Heckman and Honore (1989), and Abbring and van den Berg (2003). In fact, such an assumption is almost necessary to identify MPH, see Ridder (1990). Again this confirms that a Lévy subordinator is a natural generalization of frailty term from a process point of view. Here we also need the assumption $E[L_1(1)\,L_2(1)] < \infty$, restricting the joint Lévy measure. Assumptions (D1)-(D3) are standard normalization conditions pertaining to those model specifications.

Given the identifiability of the reduced-form regressions under competing risks, the identification of all model primitives boils down to deal with various composite functions involving structural regression functions and Lévy-Laplace exponent function. Some mild restrictions on the functional forms of the structural proportional effects in all three model specifications would lead to point identification. Taking the index model in Assumption (D1) for example, the reduced-form regressions belong to the multiple indices models, whose identification typically rely on exclusion restrictions as in Ichimura and Lee (1991). But in our context, such an assumption could be circumvented since we could first identify certain single-index model given the relations in (2.3.5). In those compositions, the unknown link function (marginal Lévy-Laplace exponent function) could be identified without extra regularity assumptions, as those Lévy-Laplace exponent functions are automatically very

smooth, see Appendix E. In the end, after all the marginal components are identified, the Lévy copula is also uniquely determined given the Sklar's theorem in Kallsen and Tankov (2006). The same strategy applies to the additive model in Horowitz and Mammen (2007) and the multiplicative model inspired by Abbring and van den Berg (2003).

Referring to semiparametric estimation, we will carry out the detailed analysis for the index model in Assumption (D1) and leave other semiparametric specifications for future research. In the next two subsections, I offer two semiparametric estimators for the structural multiplicative effect parameters given *i.i.d.* obervations $\{O_i\}_{i=1}^n$ with $O_i = (V_i, D_i, X_i)$. The first one explores the conditional moment restriction with an known time-change function $\Lambda(t)$, while the second one comes from local partial likelihood maximization without this restriction.

2.4.2 A Conditional Moment Based Estimator

We shall set $\Lambda(t) = t$ in this subsection. The later restriction on the deterministic shape function is not crucial, but it does allow for a simple estimation procedure through the conditional moment functions of observables (V, D), inspired by Arnold (1968). For our index model, upon simple algebra one gets:

$$\Phi_\iota\left(\exp\left(x'\beta_\iota\right)\right) = \Psi_\iota(x) + \Psi_3(x), \text{ for } \iota = 1, 2.$$

Now the identification of β_ι for $\iota = 1, 2$ follows from those standard techniques in single index model upon normalization as in Ichimura (1993). We define two auxiliary random variables which would be more convenient for the presentation in the sequel:

$$D_{i,\iota} = I\{D_i = \iota \text{ or } 3\} \text{ for } \iota = 1, 2.$$

Our key observation is essetianlly in line with Arnold (1968) who has considered the minimum-variance unbiased estimator for the parameters in the original Marshall-Olkin distribution:

$$\begin{aligned} E[V|X = x] &= \frac{1}{\Psi_1(x) + \Psi_2(x) + \Psi_3(x)}, \\ \Pr[D_\iota = 1|X = x] &= \frac{\Psi_\iota(x) + \Psi_3(x)}{\Psi_1(x) + \Psi_2(x) + \Psi_3(x)} \text{ for } \iota = 1, 2. \end{aligned} \tag{2.4.2}$$

For the index model, they could be expressed in the following equivalent forms:

$$\begin{aligned} E\left[V|X=x\right] &= \frac{1}{\Phi_{12}\left(\exp\left(x'\beta_1\right),\exp\left(x'\beta_2\right)\right)}, \\ \Pr\left[D_\iota=1|X=x\right] &= \frac{\Phi_\iota\left(\exp\left(x'\beta_\iota\right)\right)}{\Phi_{12}\left(\exp\left(x'\beta_1\right),\exp\left(x'\beta_2\right)\right)} \text{ for } \iota=1,2. \end{aligned} \tag{2.4.3}$$

Differentiating the results in Arnold (1968) gives us:

$$\left[\frac{\partial}{\partial x}\Phi_\iota\left(\exp\left(x'\beta_\iota\right)\right)\right]\times E\left[V|X=x\right]=\frac{\partial}{\partial x}\Pr\left[D_\iota=1|x\right]-\Phi_\iota\left(\exp\left(x'\beta_\iota\right)\right)\times\frac{\partial}{\partial x}E\left[V|X=x\right].$$

Hence we could identify parameters $\beta_\iota^{\#}$ which are scale-equivalent as the original structural parameters β_ι according to Stoker (1986), Powell, Stock, and Stoker (1989):

$$\beta_\iota^{\#} = \int \frac{\partial}{\partial x}\Phi_\iota\left(\exp\left(x'\beta_\iota\right)\right) f_X\left(x\right)dx \tag{2.4.4}$$

$$= E\left[\frac{b_{D_\iota}\left(X_i\right)}{a_V\left(X_i\right)}-\frac{a_{D_\iota}\left(X_i\right)b_V\left(X_i\right)}{\left(a_V\left(X_i\right)\right)^2}\right], \tag{2.4.5}$$

where $a_Z\left(x\right)=E\left[Z|X=x\right]$, and $b_Z\left(x\right)=\frac{\partial}{\partial x}E\left[Z|X=x\right]$ for $Z=V$ or $Z=D_\iota$. All those involved here are conditional moment functions or their first-order partial derivatives.

Starting from Fan and Gijbels (1996), the local polynomial estimator has become a default choice if one plans to estimate either the function value or its various partial derivatives. Li, Lu, and Ullah (2003) have studied the average derivative estimators from a general multivariate local polynomial fitting, and demonstrated its better finite sample performance, compared with the local constant based approaches in Hardle and Stoker (1989), Powell, Stock, and Stoker (1989). Here we follow this route taken by Li, Lu, and Ullah (2003), since our "average derivative" involves both the function values and its first order derivatives all together. We will consider a local polynomial fitting with order p for the following conditional mean specification:

$$Z_i = m_Z\left(X_i\right)+\varepsilon_{i,Z},$$

with $E\left[\varepsilon_{i,Z}|X_i\right]=0$ holds. The generic variable would be chosen as $Z_i=V_i$ or $Z_i=D_{i,\iota}$ below. Let us pin down some notations for the local polynomial estimator with order p (adapted from Tsybakov, 2008), with d dimensional covariates. Let $\mathbf{v}=(v_1,\cdots,v_d)$ with

$|\mathbf{v}| = v_1 + \cdots + v_d$, and let P be the total number of $\mathbf{v}$ with $|\mathbf{v}| \leq p$. The following vector is in $\mathcal{R}^P$:

$$U(z)^T = \left(\frac{z^{\mathbf{v}}}{\mathbf{v}!}, |\mathbf{v}| \leq p \right), \tag{2.4.6}$$

where $\mathbf{v}! = \prod_{i=1}^{d} v_i!$, the vectors of $\mathbf{v} \in \mathcal{N}^d$ being ordered lexicographically, and $z^{\mathbf{v}} = \prod_{i=1}^{d} z_i^{v_i}$.

The local polynomial estimation is carried out by minimizing the following weighted least squares problem:

$$\widehat{\theta}_Z(x) = \arg\min_{\theta} \frac{1}{nh_n^d} \sum_{i=1}^{n} \left[Z_i - \theta^T U\left(\frac{X_i - x}{h_n} \right) \right]^2 K\left(\frac{X_i - x}{h_n} \right). \tag{2.4.7}$$

Further we define two types of selection vectors $e \in \mathcal{R}^P$ to pick up the estimators for the function value and its first order derivatives respectively:

$$\begin{aligned} e_0^T &= (1, 0, 0, ..., 0), \\ e_t^T &= \left(0, \underbrace{0, ..., 1...., 0}_{d}, 0..., 0 \right), \end{aligned}$$

where $1 \leq t \leq d$. Let the estimators for the intercept and slope coefficients be denoted as:

$$\begin{aligned} \widehat{a}_Z(x) &= e_{\mathbf{0}}^T \widehat{\theta}_Z(x), \\ \widehat{b}_{Z,t}(x) &= h_n^{-1} e_t^T \widehat{\theta}_Z(x), \end{aligned}$$

and $\widehat{b}_Z(x) = \left(\widehat{b}_{Z,1}(x), ..., \widehat{b}_{Z,d}(x) \right)^T$. Whenever we evaluate the estimate at a sample point X_i, we shall use its leave-one-out version:

$$\widehat{\theta}(X_i) = \arg\min_{\theta} \frac{1}{nh_n^d} \sum_{j \neq i} \left[Z_j - \theta^T U\left(\frac{X_j - X_i}{h_n} \right) \right]^2 K\left(\frac{X_j - X_i}{h_n} \right),$$

and the corresponding versions of $\widehat{a}_Z(X_i)$ and $\widehat{b}_Z(X_i)$ are defined in the same way. Finally we define our estimators for $\beta_\iota^{\#}$ as:

$$\widehat{\beta}_\iota^{\#} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\widehat{a}_V(X_i)} \left[\widehat{b}_{D_\iota}(X_i) - \frac{\widehat{a}_{D_\iota}(X_i)}{\widehat{a}_V(X_i)} \widehat{b}_V(X_i) \right], \tag{2.4.8}$$

for $\iota = 1, 2$. The following theorem shows the desired asymptotic normality of our estimator, and we spell out the form of its asymptotic covariance in Appendix B.

Theorem 2.4.3 *Under Assumptions (B1)-(B5) in Appendix B, we have the following asymptotic normality holds for* $\iota = 1, 2$*:*

$$\sqrt{n}\left(\widehat{\beta}_{\iota}^{\#} - \beta_{\iota}^{\#}\right) \Longrightarrow N\left(0, \Omega_{\iota}^{\#}\right).$$

Utilizing the bivariate indices structure as those reduced-form regressions depend on x only through $\left(x^T\beta_1, x^T\beta_2\right)$, further iteration could be carried out as in Hristache, Juditsky, Polzehl, and Spokoiny (2001). Given the preliminary estimator $\widehat{\beta}_{\iota}^{(0)} \equiv \widehat{\beta}_{\iota}^{\#}$, we could define the following local linear estimators from $l-$th iteration:

$$\begin{aligned} & \left(\widehat{a}_{Z,l}(x), \widehat{b}_{Z,l}(x)\right) \\ = & \arg\min_{(a,b)} \sum_{i=1}^{n} \left[Z_i - a - b^T (X_i - x)\right]^2 \mathbf{K}_{h_n}\left((X_i - x)^T \widehat{\beta}_1^{(l-1)}, (X_i - x)^T \widehat{\beta}_2^{(l-1)}\right), \end{aligned}$$

with a bivariate kernel function $\mathbf{K}_{h_n}(\cdot, \cdot)$. Then $\left(\widehat{\beta}_1^{(l)}, \widehat{\beta}_2^{(l)}\right)$ are formed as in (2.4.8) but with those $\left(\widehat{a}_{Z,l}(\cdot), \widehat{b}_{Z,l}(\cdot)\right)$ for $Z = V$ or $Z = D_{\iota}$. Hristache, Juditsky, Polzehl, and Spokoiny (2001) have demonstrated stable finite sample performances for their iterated estimator, but unfortunately no asymptotic distribution has been established there.

2.4.3 A Local Partial Likelihood Based Estimator

Despite its simplicity, the previous estimator relies on two rather unrealistic assumptions, namely a known baseline hazard $\Lambda(t)$ and absence of additional censoring due to early termination of the study or failure of follow-up. The estimator developed in this sub-section fills the gap even with the aforementioned obstacles. A notable practical advantage of the current procedure is demonstrated when there is some additional censoring which is typically the case. Although independent censoring can be handled by traditional hazard-based models in duration or survival analysis, its incorporation into the recent structural duration model calls for extra attention. For example, even though Abbring's (2012) MHT is naturally expressed in terms of the conditional Laplace transform, one has to invert the Laplace transform back to the probability density function as in Abbring and Salimans (2012) due to the presence of additional independent censoring. In contrast, such a complication is simply avoided in our EMO.

Regardless of the unknown function $\Lambda(t)$, those conditional probabilities $\Pr[D_\iota = 1|X = x]$ still take the following forms

$$\Pr[D_\iota = 1|X = x] = \frac{\Phi_\iota\left(\exp\left(x'\beta_\iota\right)\right)}{\Phi_{12}(\phi_1(x), \phi_2(x))} \text{ for } \iota = 1, 2, \tag{2.4.9}$$

which are readily to be estimated from the sample. It is interesting that in EMO, the random variable $V = \min(Y_1^*, Y_2^*)$ satisfies

$$\Pr[V > t|X = x] = \exp[-\Lambda(t)\Phi_{12}(\phi_1(x), \phi_2(x))],^3 \tag{2.4.10}$$

with its conditional hazard function is of proportional structure,

$$\lambda_V(t|x) = \lambda(t)\Psi(x) = \lambda(t)\exp\psi(x).$$

where $\exp\psi(\cdot) = \Psi(\cdot) = \Phi_{12}(\phi_1(\cdot), \phi_2(\cdot))$. Hence we could adapt the methods estimating the nonparametric proportional hazard model in Fan, Gijbels, and King (1997), Chen and Zhou (2007). Combining (2.4.9) and (2.4.10), our ultimate goal is to extract relevant intercept and slope estimates from the local partial likelihood maximization and form proper average derivatives, as in Section 4.2. Due to the reparameterization $\Psi(x) = \exp\psi(x)$, our structural parameter are defined to be

$$\begin{aligned} \beta_1^* &= E[a_{D_1}(X_i) b_V(X_i) + a_V(X_i) b_{D_1}(X_i)], \\ \beta_2^* &= E[a_{D_2}(X_i) b_V(X_i) + a_V(X_i) b_{D_3}(X_i)], \end{aligned} \tag{2.4.11}$$

where $a_V(X_i) = \exp(\alpha(X_i))$ and $b_V(X_i) = \exp(\alpha(X_i))\beta_1(X_i)$, with $\alpha(\cdot) = \psi(\cdot)$ and $\beta_1(\cdot) = \frac{\partial}{\partial x}\psi(\cdot)$.

Recall that $V = \min(Y_1^*, Y_2^*)$ and the cause of failure D distinguishes three settings, i.e., $D = 1$ if $Y_1^* < Y_2^*$, $D = 2$ if $Y_2^* < Y_1^*$, and $D = 3$ if $Y_1^* = Y_2^*$. In this sub-section, we allow that the failure time V is subject to additional independent censoring from random variable C, so we let $\widetilde{V} = \min(V, C)$ and $\Delta = I\left[\widetilde{V} < C\right]$. First of all, the baseline hazard function could be estimated by the Breslow type estimator placing mass on the points $\left(\widetilde{V}_i\right)_{i=1}^n$ via

$$\widehat{\lambda}\left(\widetilde{V}_i\right) = \frac{1}{\sum_{l \in R\left(\widetilde{V}_i\right)} \Psi(X_l)},$$

[3] In view of Ai and Chen (2003), it is expected that one could replace the unknown $\Lambda(t)$ with a sieve type estimator in our setting. We will not explore such a possibility as one could circumvant the estimation of this time-change function (or profile it out) as we do here

where $R\left(\widetilde{V}_i\right) \equiv \left\{l : \widetilde{V}_l \geq \widetilde{V}_i\right\}$ collects the observations at risk. After profiling out the baseline hazard function, the (log-)partial likelihood is

$$L_{n,P} = \sum_{i=1}^{n} \left(\psi\left(X_{(i)}\right) - \log \left[\sum_{l \in R(\widetilde{V}_i)} \exp \psi\left(X_l\right) \right] \right). \tag{2.4.12}$$

Now it is clear that one could replace $\psi(\cdot)$ with its local polynomial version in Fan, Gijbels, and King (1997). One more complication is that only derivative terms of $\psi(\cdot)$ appear in the local partial likelihood in Fan, Gijbels, and King (1997) while the intercept term is cancelled out in (2.4.12), because $\psi(\cdot)$ is only identified upto a scale normalization. Our procedure would be first estimating the first order derivatives of $\psi(\cdot)$, and then the difference of $\psi(x) - \psi(x_0)$ by Chen and Zhou's (2007) method assuming x_0 is a normalization point such that $\psi(x_0) \equiv 0$.

Some standard notations of counting processes would facilitate our presentation. The natural counting process is

$$N_i^{(n)}(t) = I\left[\widetilde{V}_i \leq t, \Delta_i = 1\right]$$

with the at-risk process $Y_i^{(n)}(t) = I\left[\widetilde{V}_i > t\right]$. The local martingale attached to the counting process is

$$M_i^{(n)}(t) = N_i^{(n)}(t) - \int Y_i^{(n)}(t)\, \Psi\left(X_i\right) d\Lambda(t).$$

We define $\beta(x)$ to be the vector stacking all the partial derivatives of $\psi(x)$ upto order p, and the first order derivatives are denoted by $\beta_1(x)$. Moreover, we set $U^*(u)$ be the vector of $U(u)$ without the first element 1, and $H^* \equiv Diag\left(h_n^{|\mathbf{v}|}, 1 \leq |\mathbf{v}| \leq p\right)$. In practice we have to stay away from the support of $\widetilde{V}$ and all the estimation is carried out upto a truncation point τ such that $\Lambda(\tau) < \infty$ (Andersen and Gill, 1982). So the local partial likelihood with kernel weights takes the form of

$$\begin{aligned}
&\sum_{i=1}^{n} \int_0^{\tau} K_{h_n}\left(X_i - x\right) U^*\left(\frac{X_i - x}{h_n}\right)^T \beta(x)\, dN_i^{(n)}(t) \\
&- \sum_{i=1}^{n} \int_0^{\tau} \log \sum_{i=1}^{n} Y_i^{(n)}(t) \exp\left(U^*\left(\frac{X_i - x}{h_n}\right)^T \beta(x)\right) K_{h_n}\left(X_i - x\right) dN_i^{(n)}(t).
\end{aligned}$$

By another slight abuse of notation, we denote $X_i^*(x) = U^*(X_i - x)$ and $U_i^*(x) = U^*\left(\frac{X_i - x}{h_n}\right)$. Furthermore let

$$S_{n,k}(t, \zeta(x), x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(X_i - x) Y_i^{(n)}(t) \exp\left(X_i^*(x)^T \beta_0(x) + U_i^*(x)^T \zeta(x)\right) U_i^*(x)^{\otimes k}.$$

Whence the maximum partial likelihood estimator is a solution to $\dot{l}_n\left(H^*\left[\widehat{\beta}(x) - \beta(x)\right]\right) = 0$ which could be computed by the Newton-Raphson algorithm for any x:

$$\begin{aligned} & \dot{l}_n(\zeta(x)) \\ = \; & \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_{h_n}(X_i - x)\left[U_i^*(x) - \frac{S_{n,1}(t, \zeta(x), x)}{S_{n,0}(t, \zeta(x), x)}\right] dN_i^{(n)}(t). \end{aligned}$$

In order to estimate the difference of the intercept term, we resort to Chen and Zhou (2007) replacing $K_{h_n}(X_i - x)$ with $K_{h_n}(X_i - x) + K_{h_n}(X_i - x_0)$. The sample criterion function is $\widetilde{L}_n(\alpha(x), \beta(x), \beta(x_0)) = \widetilde{L}_{n1}(\alpha(x), \beta(x), \beta_0(x_0)) + \widetilde{L}_{n2}(\alpha(x), \beta(x), \beta(x_0))$ where

$$\begin{aligned} & \widetilde{L}_{n1}(\alpha(x), \beta(x), \beta(x_0)) \\ = \; & \sum_{j=1}^{n} K_{h_{0n}}\left(X_{(j)} - x_0\right) U_j^*(x_0)^T \beta(x_0) \\ & - \sum_{j=1}^{n} K_{h_{0n}}\left(X_{(j)} - x_0\right) \log\left\{\sum_{i \in R_j} K_{h_{0n}}(X_i - x_0) e^{U_i^*(x_0)^T \beta(x_0)} + K_{h_{0n}}(X_i - x) e^{\left[\alpha(x) + U_i^*(x)^T \beta(x)\right]}\right\}, \end{aligned}$$

and

$$\begin{aligned} & \widetilde{L}_{n2}(\alpha(x), \beta(x), \beta(x_0)) \\ = \; & \sum_{j=1}^{n} K_{h_{0n}}\left(X_{(j)} - x\right)\left[\alpha(x) + U_j^*(x_0)^T \beta(x)\right] \\ & - \sum_{j=1}^{n} K_{h_{0n}}\left(X_{(j)} - x_0\right) \log\left\{\sum_{i \in R_j} K_{h_{0n}}(X_i - x_0) e^{U_i^*(x_0)^T \beta(x_0)} + K_{h_{0n}}(X_i - x) e^{\left[\alpha(x) + U_i^*(x)^T \beta(x)\right]}\right\}. \end{aligned}$$

We proceed with two step estimation by plugging $\widehat{\beta}(\cdot)$ in $\widetilde{L}_n(\alpha(x), \beta(x), \beta(x_0))$ from previous derivative estimation, so $\widehat{\alpha}(x) = \arg\max_{\alpha(x)} \widetilde{L}_n\left(\alpha(x), \widehat{\beta}(x), \widehat{\beta}(x_0)\right)$.

In the end, our final estimators for the structural parameters are

$$\widehat{\beta}_\iota^* \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{a}_{D_\iota}(X_i) \exp\left(\widehat{\alpha}(X_i)\right) \widehat{\beta}_1(X_i) + \exp\left(\widehat{\alpha}(X_i)\right) \widehat{b}_{D_\iota}(X_i)\right\}, \tag{2.4.13}$$

for $\iota = 1, 2$. The following theorem shows the desired asymptotic normality of our estimator, and we spell out the form of its asymptotic covariance in Appendix C.

Theorem 2.4.4 *Under Assumptions (C1)-(C7) in Appendix C, we have the following asymptotic normality holds for* $\iota = 1, 2$*:*

$$\sqrt{n}\left(\widehat{\beta}_{\iota}^{*} - \beta_{\iota}^{*}\right) \Longrightarrow N\left(0, \Omega_{\iota}^{*}\right).$$

The intercept estimation could be collected to estimate the function itself, which is used in the following Breslow estimator for the time-change function:

$$\widehat{\Lambda}\left(t\right) = \int_{0}^{t} \frac{\sum_{i=1}^{n} dN_{i}^{(n)}\left(s\right)}{\sum_{i=1}^{n} \exp\left(\widehat{\alpha}\left(X_{i}\right)\right) Y_{i}^{(n)}\left(s\right)}.$$

The baseline hazard function is typically not of primary interest, so we only give its convergence rate in the next theorem. The asymptotic normality is expected to follow Theorem 4 in Chen, Guo, Sun, and Wang (2010).

Theorem 2.4.5 *Suppose the Assumptions (C1)-(C7) in Appendix C hold, then for any bounded measurable function* $B\left(t\right)$*,*

$$\sqrt{n}\int_{0}^{t} B\left(t\right) d\left[\widehat{\Lambda}\left(t\right) - \Lambda\left(t\right)\right] = O_{p}\left(1\right).$$

2.5 Appendix B.

In the following three appendices we let C, M be a generic finite positive constant whose value might change from line to line, and occasionally we use subscript when more than one constant appear in the display.

It's clear from Section 2.4 that our estimators for the structural regression parameter would be based on various conditional moments and corresponding first order derivatives. We write a generic nonparametric regression function as

$$Z_{i} = m_{Z}\left(X_{i}\right) + \varepsilon_{i,Z},$$

where $E\left[\varepsilon_{i,Z} | X_{i}\right] = 0$ holds almost surely. The LHS variable would be substituted with $Z_{i} = V_{i}$ or $Z_{i} = D_{i,\iota}$ later on. The local polynomial estimator takes a closed form as:

$$\widehat{\theta}_Z(x) = B_{nx}^{-1}\left[\frac{1}{nh_n^d}\sum_{i=1}^{n} Z_i K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right)\right], \tag{B.1}$$

where

$$B_{nx} = \frac{1}{nh_n^d}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right) U^T\left(\frac{X_i - x}{h_n}\right).$$

Now we list the regularity assumptions needed. We denote the boundary of the covariate's support as $\partial\mathcal{X}$, and define $\mathcal{X}_\delta = \{x \in \mathcal{X} : \min_{y\in\partial\mathcal{X}} d(x,y) \leq \delta\}$.

Assumption (B1). The conditional mean function is $p+1$ order differentiable with bounded $(p+1)$-th order derivative uniformly in $x \in \mathcal{X}$, thus

$$|m^{(\mathbf{v})}(x) - m^{(\mathbf{v})}\left(x'\right)| \leq L||x - x'||,$$

for all $\mathbf{v} \in \mathcal{N}^d$ such that $|\mathbf{v}| = p$, one has $m^{(\mathbf{v})}(x) = \frac{\partial^v}{\partial x_1^{v_1}\cdots\partial x_d^{v_d}} m(x)$, stands for partial derivative and $||\cdot||$ for the Euclidean norm. Furthermore, for every $h > 0$ and every point $x \in \mathcal{X}_{\sqrt{h}}$, the set $B_{\sqrt{h}}(x) \cap \mathcal{X}$ contains a nonempty interior.

Assumption (B2). The marginal density function $f_X(x)$ exist and is second order continuously differentiable. Moreover, it is bounded up and below uniformly, i.e. $\forall x \in \mathcal{X}$, $f_X(x) \geq M_0 > 0$ and $f_X(x) \leq M_1 < \infty$ hold.

Assumption (B3). The conditional mean function $a_V(x) = E[V|X = x]$ is bounded away from zero uniformly in x, $a_V(x) \geq M_0 > 0$. Also we impose the following higher moment bound

$$\sup_{x\in\mathcal{X}} E[V^\gamma|X = x] \leq M_1 < \infty,$$

for some $\gamma > 2$.

Assumption (B4). The kernel density function K is a compactly supported (w.l.o.g let it be $[-1,1]^d$), bounded kernel (up by $K_{\max}$). Furthermore, $\mathcal{K} = \left\{K\left(\frac{x-\cdot}{h_n}\right) : h_n > 0, x \in \mathcal{R}^d\right\}$ is pointwisely measurable and of V C type. Also for $\mu_p(K) = \int K(u) U(u) U^T(u)\, du$, $\vartheta_p(K) = \int K^2(u) U(u) U^T(u)\, du$, with $\mu_p(K)$,$\vartheta_p(K)$ positive definite $P \times P$ matricies.

Assumption (B5). The bandwidth sequence satisfies: $h \to 0$, $\frac{nh^{d+2}}{\log n} \to \infty$ and $nh^{2p+2} \to 0$.

The proof of main theorem is patterned after Hardle and Stoker (1989), Li, Lu, and Ullah (2003). Namely, we decompose the statistic into three parts: the first is the centered $i.i.d$ average of the infeasible estimator while the second is a U-statistic. The third remainder term is shown to be negligible by applying simple supremum bounds on various local polynomial estimators. The second U-statistic would further be linearized via Hoeffding decomposition with linear terms dominating the second-order canonical U-statistics. Thereafter the desired asymptotic normality follows from the behavior of those centered linear terms. Because we are dealing with various intercept and slope estimates all together, the detail is slightly more involved.

Proof. (of Theorem 2.4.3) We will only show the asymptotic normality for one chosen coordinate value, say $t = 1$, as the multivariate case could be checked by the standard Cramer-Wold device as in Lemma 4 in Li, Lu, and Ullah (2003). By a slight abuse of notation, we drop the subscript of t as well. Recall our average derivative type estimators for $\beta_\iota^\#$ are

$$\widehat{\beta}_\iota^\# = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\widehat{a}_V(X_i)}\left[\widehat{b}_{D_\iota}(X_i) - \frac{\widehat{a}_{D_\iota}(X_i)}{\widehat{a}_V(X_i)}\widehat{b}_V(X_i)\right],$$

with $\iota = 1, 2$. We further decompose it into two parts $\beta_\iota^\# = \beta_{\iota,1}^\# + \beta_{\iota,2}^\#$ where $\beta_{\iota,1}^\# = E\left[\frac{b_{D_\iota}(X_i)}{a_V(X_i)}\right]$ and $\beta_{\iota,2}^\# = E\left[\frac{a_{D_\iota}(X_i)b_V(X_i)}{(a_V(X_i))^2}\right]$. Our analysis could be carried out separately for those two terms naturally:[4]

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{b}_{D_\iota}(X_i)}{\widehat{a}_V(X_i)} - \beta_{\iota,1}^\# \\ = \; & \frac{1}{n}\sum_{i=1}^{n} \frac{b_{D_\iota}(X_i)}{a_V(X_i)} - \beta_{\iota,1}^\# + t_{n1} + r_{n1}, \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{a}_{D_\iota}(X_i)\,\widehat{b}_V(X_i)}{(\widehat{a}_V(X_i))^2} - \beta_{\iota,2}^\# \\ = \; & \frac{1}{n}\sum_{i=1}^{n} \frac{a_{D_\iota}(X_i)\,b_V(X_i)}{(a_V(X_i))^2} - \beta_{\iota,2}^\# + t_{n2} + r_{n2}, \end{aligned}$$

[4]To ease the notational burden, we omit the subcript ι on those statistics t_{n1} or r_{n1}, although they do depend on whether $\iota = 1$ or 2.

The two statistics are defined by:

$$t_{n1} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{b}_{D_\iota}(X_i)}{a_V(X_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{a}_V(X_i)\, b_{D_\iota}(X_i)}{a_V^2(X_i)},$$

and

$$t_{n2} = \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{a}_{D_\iota}(X_i)\, b_V(X_i)}{(a_V(X_i))^2} + \frac{1}{n}\sum_{i=1}^{n}\frac{a_{D_\iota}(X_i)\,\widehat{b}_V(X_i)}{(a_V(X_i))^2} - \frac{2}{n}\sum_{i=1}^{n}\frac{a_{D_\iota}(X_i)\, b_V(X_i)\,\widehat{a}_V(X_i)}{(a_V(X_i))^3}.$$

Meanwhile, the two remainder terms are:

$$\begin{aligned} r_{n1} \equiv\ & \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\widehat{a}_V(X_i)} - \frac{1}{a_V(X_i)}\right]\left[\widehat{b}_{D_\iota}(X_i) - b_{D_\iota}(X_i)\right] \\ & + \frac{1}{n}\sum_{i=1}^{n}\frac{b_{D_\iota}(X_i)}{a_V(X_i)}\left[\frac{1}{\widehat{a}_V(X_i)} - \frac{1}{a_V(X_i)}\right]\left[a_V(X_i) - \widehat{a}_V(X_i)\right], \end{aligned}$$

and

$$\begin{aligned} r_{n2} =\ & \frac{1}{n}\sum_{i=1}^{n}\frac{\left[\widehat{a}_{D_\iota}(X_i) - a_{D_\iota}(X_i)\right]\left[\widehat{b}_V(X_i) - b_V(X_i)\right]}{(a_V(X_i))^2} - \frac{1}{n}\sum_{i=1}^{n}\frac{a_{D_\iota}(X_i)\left[a_V(X_i) - \widehat{a}_V(X_i)\right]^2}{(a_V(X_i))^4} \\ & + \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{a}_{D_\iota}(X_i)\,\widehat{b}_V(X_i) - a_{D_\iota}(X_i)\, b_V(X_i)\right]\left[\frac{1}{(\widehat{a}_V(X_i))^2} - \frac{1}{(a_V(X_i))^2}\right] \\ & + \frac{1}{n}\sum_{i=1}^{n}\frac{a_{D_\iota}(X_i)\left[(a_V(X_i))^2 - (\widehat{a}_V(X_i))^2\right]}{(a_V(X_i))^2}\left[\frac{1}{(\widehat{a}_V(X_i))^2} - \frac{1}{(a_V(X_i))^2}\right]. \end{aligned}$$

Hence in the remaining part, we show the negligibility of the remainder terms, i.e., $r_{n1} = o_p\left(n^{-1/2}\right)$ and $r_{n2} = o_p\left(n^{-1/2}\right)$. And the following linearizations are proved

$$t_{nj} = \frac{1}{n}\sum_{i=1}^{n}\varphi_j(O_i) + o_p\left(n^{-1/2}\right), \text{ for } = 1, 2.$$

The exact expressions for the influence functions φ_1 and φ_2 could be found in Lemma 2.5.5 below.

In the end we pin down the influence function

$$\begin{aligned} & \widehat{\beta}_\iota^{\#} - \beta_\iota^{\#} \\ =\ & \frac{1}{n}\sum_{i=1}^{n}\frac{b_{D_\iota}(X_i)}{a_V(X_i)} - \beta_{\iota,1}^{\#} + \frac{1}{n}\sum_{i=1}^{n}\frac{a_{D_\iota}(X_i)\, b_V(X_i)}{(a_V(X_i))^2} - \beta_{\iota,2}^{\#} \\ & + \frac{1}{n}\sum_{i=1}^{n}\varphi_1(O_i) + \frac{1}{n}\sum_{i=1}^{n}\varphi_2(O_i) + o_p\left(n^{-1/2}\right). \end{aligned}$$

Hence the desired asymptotic normality follows from standard Lindeberg CLT *i.i.d.* averages given the moment restrictions in Assumption (B3). The asymptotic variance is the one of the influence function above. ■

It calls for a uniform asymptotic analysis of the derivative estimators from local polynomial fitting. We drop the subscript Z in $m_Z(\cdot)$ without confusion in the remaining. An estimator for $m^{(\mathbf{s})}(x)$ with $|\mathbf{s}| = s$, and $0 \leq s \leq p$ is defined as $\widehat{m}^{(\mathbf{s})}(x) = h_n^{-s} e_{\mathbf{s}}^T \hat{\theta}(x)$, where $e_{\mathbf{s}}^T$ is the particular selection vector picking up $\mathbf{s}$-th partial derivatives. Write it as a linear smoother

$$\widehat{m}^{(\mathbf{s})}(x) = \sum Z_i W_{ni}^{\mathbf{s}}(x), \tag{B.2}$$

with

$$W_{ni}^{\mathbf{s}}(x) = \frac{1}{nh_n^{d+s}} e_{\mathbf{s}}^T B_{nx}^{-1} K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right). \tag{B.3}$$

Consider the p-th order Taylor expansion for a generic smooth function $m(\cdot)$ locally around x, where the window size is controlled by h_n.

$$\begin{aligned} m(X_i) &= \sum_{0 \leq |\mathbf{s}| \leq p} \frac{m^{(\mathbf{s})}(x)}{\mathbf{s}!} (X_i - x)^{\mathbf{s}} + R(X_i, x) \\ &= U^T\left(\frac{X_i - x}{h_n}\right) q(x) + R(X_i, x), \end{aligned}$$

where $R(X_i, x) = \sum_{|s|=p} |r_{\mathbf{s}}(X_i, x)|$, the total number of the sum equals to $N_p = \binom{p+d-1}{d-1}$, similarly one could define $N_s, 0 \leq s \leq p$. So the local polynomial estimation could reproduce the p-th order polynomial functions in the sense:

$$\begin{aligned} & \widehat{\theta}(x) \\ = \; & q(x) + B_{nx}^{-1}\left[\frac{1}{nh_n^d} \sum K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right) R(X_i, x)\right] \\ & + B_{nx}^{-1}\left[\frac{1}{nh_n^d} \sum \varepsilon_i K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right)\right]. \end{aligned}$$

Thus we could decompose it into two parts:

$$\begin{aligned}
& \widehat{m}^{(\mathbf{s})}(x) - m^{(\mathbf{s})}(x) \\
= \; & e_{\mathbf{s}}^T B_{nx}^{-1} \left[\frac{1}{n h_n^{d+s}} \sum K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right) R(X_i, x) \right] \\
& + h_n^{-s} e_{\mathbf{s}}^T B_{nx}^{-1} \left[\frac{1}{n h_n^{d}} \sum \varepsilon_i K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right) \right] \\
= \; & \sum R(X_i, x) W_{ni}^{\mathbf{s}}(x) + \sum \varepsilon_i W_{ni}^{\mathbf{s}}(x).
\end{aligned}$$

We first bound the bias term $B^{\mathbf{s}}(x) \equiv \sum R(X_i, x) W_{ni}^{\mathbf{s}}(x)$ uniformly in covariates' value.

Lemma 2.5.1 *Under Assumptions (B1)-(B5), we get* $\sup_x |B^{\mathbf{s}}(x)| = O_p\left(h_n^{p-s+1}\right)$.

Proof. Recall $R(X_i, x) = \sum_{|s|=p} |r_{\mathbf{s}}(X_i, x)|$, the total number of the sum equals to $N_p = \binom{p+d_x-1}{d_x-1}$, similarly one could define $N_s, 0 \le s \le p$.

$$\begin{aligned}
|r_{\mathbf{s}}(X_i, x)| &= \left| \frac{m^{(\mathbf{s})}(\Delta_x)}{p!} (X_i - x)^p - \frac{m^{(\mathbf{s})}(x)}{p!} (X_i - x)^p \right| \\
&\le \frac{L h_n}{p!} h_n^{\mathbf{s}} = \frac{L}{p!} h_n^{p+1},
\end{aligned}$$

and $q^T(x) = \left(m^{(\mathbf{s})}(x) h_n^{\mathbf{s}}, |\mathbf{s}| \le p\right)$, Δ_x is some intermediate value in neighborhood around x within a distance of order h_n.Therefore, the reminder term is controlled (uniformly in x) of the following order:

$$|R(X_i, x)| \le \frac{L N_p}{p!} h_n^{p+1}, \tag{B.4}$$

hence the bias term is hence Bias is bounded up by

$$\begin{aligned}
|B^{\mathbf{s}}(x)| &\le \sup |R(X_i, x)| \sum |W_{ni}^{\mathbf{s}}(x)| \\
&\le \frac{L N_p}{p!} h_n^{p+1-s} ||e_{\mathbf{s}}|| \times ||B_{nx}^{-1} \frac{1}{h_n^d} K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right)||,
\end{aligned}$$

B_{nx} is asymptotically positive semidefinite and symmetric, and let $\lambda_{\min}(B_{nx})$ be its smallest eigenvalue. Then for any vector v, we get $||B_{nx}^{-1} v|| \le \frac{||v||}{\lambda_{\min}(B_{nx})} \le C||v||$.

Given x we could bound the individual weight by

$$\begin{aligned}
|W_{ni}^{s}(x)| &\le h_n^{-s} ||B_{nx}^{-1} \frac{1}{n h_n^d} K\left(\frac{X_i - x}{h_n}\right) U\left(\frac{X_i - x}{h_n}\right)|| \\
&\le \frac{C K_{\max}}{n h_n^{s+d}} ||I\left(\left|\frac{X_i - x}{h_n}\right| \le 1\right) U\left(\frac{X_i - x}{h_n}\right)|| \le \frac{C K_{\max}}{n h_n^{s+d}}.
\end{aligned}$$

Finally the bias term could be bounded up uniformly by

$$|B^{\mathbf{s}}(x)| \leq \frac{LN_p}{p!} h_n^{p+1} \sum |W_{ni}^{\mathbf{s}}(x)| \leq \frac{2LCN_p K_{\max}}{p!} h_n^{p-s+1}.$$

■

For the stochastic term, we rely on Doney, Einmahl, and Mason (2006) to handle local polynomial estimators with uniform in bandwidth asymptotics. First recall some standard definitions in Einmahl and Mason (2005). Let $\mathcal{F}$ collects some functions $f : \mathcal{R}^d \to \mathcal{R}$. We say the class $\mathcal{F}$ is of VC type with respect to an envelope F if the covering number $N(\mathcal{F}, L_2(Q), \varepsilon)$, the smallest number of $L_2(Q)$ open balls of radius ε required to cover $\mathcal{F}$, statisfies

$$N(\mathcal{F}, L_2(Q), \varepsilon) \leq \left(\frac{M \|F\|_{L_2(Q)}}{\varepsilon} \right)^v, \text{ for } 0 < \varepsilon \leq 2 \|F\|_{L_2(Q)},$$

for some universal positive constant M, v for every probability measure Q on the underlying space. Also by pointwise measurable class $\mathcal{F}$, we mean there exists a countable subclass $\mathcal{F}_0$ of $\mathcal{F}$ s.t. we can find for any $f \in \mathcal{F}$ a sequence of functions $\{f_m\}$ in $\mathcal{F}_0$ for which $f_m(x) \to f(x)$ for $x \in \mathcal{R}^d$. We shall apply the following lemma when f is the identity map and the generalized kernel function $\mathbf{K}\left(\frac{x-\cdot}{h_n}\right)$ is $K\left(\frac{x-\cdot}{h_n}\right) U\left(\frac{x-\cdot}{h_n}\right)$. The independent variable is chosen bo be $Z_i = D_{i,\iota}$ for (B.5), or $Z_i = V_i$ for (B.6).

Lemma 2.5.2 *For i.i.d. observations, we consider the following generalized kernel type estimator*

$$g_{nh_n}(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} c_f(x) \mathbf{K}\left(\frac{x - X_i}{h_n}\right) f(Z_i),$$

where the classes $\mathcal{F} = \{f\}$ *and* $\mathcal{K} = \left\{\mathbf{K}\left(\frac{x-\cdot}{h_n}\right) : h_n > 0, x \in \mathcal{R}^{d_x}\right\}$ *are both pointwisely measurable and of VC type. Moreover,* $\mathcal{C} \equiv \{c_f : f \in \mathcal{F}\}$ *is relatively compact w.r.t the supnorm topology and either*

$$\exists M > 0 : F(Z) 1\{x \in \mathcal{J}\} \leq M < \infty, \tag{B.5}$$

or

$$\sup_{x \in \mathcal{J}} E[F^\gamma(Z) | X = x] = \overline{M} < \infty, \tag{B.6}$$

with $\gamma > 2$ and the envelope function F for class $\mathcal{F}$. we have

$$\lim \sup_{n\to\infty} \sup_{h\in\mathcal{H}_n} \sup_{f\in\mathcal{F}} \frac{\sqrt{nh_n^d}\left\|g_{nh_n}(x) - Eg_{nh_n}(x)\right\|}{\sqrt{d\log h_n \vee \log\log n}} < \infty,$$

where

$$\mathcal{H}_n = \left\{h_n : C\left(\frac{\log n}{n}\right)^{\eta} \leq h_n \leq h_0\right\},$$

with $\eta = 1$ in the bounded case (B.5), or $\eta = 1 - 2/\gamma$ under the moment restriction in (B.6).

In the current context, we need to expand the denominator B_{nx} into two terms similarly as Lemma A.1 in Li, Lu, and Ullah (2003).

Lemma 2.5.3 *Under Assumptions (B1)-(B5), we have*

$$\begin{aligned} &\sup_x \left|B_{nx}^{-1} - f_X^{-1}(x)\,\mu_p^{-1}(K) - h_n G(x)\right| \\ = \;& O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right), \end{aligned}$$

where

$$G(x) \equiv f_X^{-2}(x)\,\mu_p^{-1}(K)\left[\int u^T f_X^{(1)}(x)\,K(u)\,U(u)\,U^T(u)\,du\right]\mu_p^{-1}(K).$$

Proof. We shall handle B_{nx} first and then its inverse. Note that

$$\begin{aligned} EB_{nx} &= \frac{1}{h_n^d}\int K\left(\frac{X_i - x}{h_n}\right)U\left(\frac{X_i - x}{h_n}\right)U^T\left(\frac{X_i - x}{h_n}\right)dF_X(X_i) \\ &= \int f_X(x + uh_n)\,K(u)\,U(u)\,U^T(u)\,du \\ &= f_X(x)\int K(u)\,U(u)\,U^T(u)\,du + h_n\int u^T f_X^{(1)}(x)\,K(u)\,U(u)\,U^T(u)\,du + O\left(h_n^2\right). \end{aligned}$$

And by the results in Doney, Einmahl, and Mason (2006), we get

$$\sup_x |B_{nx} - EB_{nx}| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$$

Hence we use von Neumann series expansion for the inverse matrix:

$$\begin{aligned} B_{nx}^{-1} &= \frac{1}{f_X(x)}\left[\int K(u)\,U(u)\,U^T(u)\,du\right]^{-1} \\ &- \frac{h_n}{f_X^2(x)}\left[\int K(u)\,U(u)\,U^T(u)\,du\right]^{-1}\left[\int u^T f_X^{(1)}(x)\,K(u)\,U(u)\,U^T(u)\,du\right]\times \\ &\left[\int K(u)\,U(u)\,U^T(u)\,du\right]^{-1}, \end{aligned}$$

upto the order of $O_p\left(h_n^2+\sqrt{\frac{\log n}{nh_n^d}}\right)$. Now the conclusion follows from the definition of $\mu_p(K)$ and $G(x)$. ■

Thus the linear representation we are seeking is

$$\begin{aligned} & \widehat{m}^{(\mathbf{s})}(x)-m^{(\mathbf{s})}(x) \\ = & \left[\frac{1}{nh_n^{d+s}f_X(x)}\sum e_{\mathbf{s}}^T\mu_p^{-1}(K)\,\varepsilon_i K\left(\frac{X_i-x}{h_n}\right)U\left(\frac{X_i-x}{h_n}\right)\right] \\ & -\left[\frac{1}{nh_n^{d+s-1}}\sum e_{\mathbf{s}}^T G(x)\,\varepsilon_i K\left(\frac{X_i-x}{h_n}\right)U\left(\frac{X_i-x}{h_n}\right)\right] \\ & +O_p\left(h_n^2+\sqrt{\frac{\log n}{nh_n^d}}\right)O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right). \end{aligned} \tag{B.7}$$

From the previous two lemmas we immediately obtain

$$\sup_{x\in\mathcal{X}}\left|\widehat{m}^{(\mathbf{s})}(x)-m^{(\mathbf{s})}(x)\right|=O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\frac{1}{h_n^s}\right),\ \text{for } s=0,1. \tag{B.8}$$

Thereafter the two remainder terms are indeed negligible:

$$r_{nj}=O_p\left(\frac{\log n}{nh_n^{d+1}}\right)=o_p\left(n^{-1/2}\right)\ \text{for } j=1,2,$$

under the assumption $\frac{\log n}{\sqrt{n}h_n^{d+1}}\to 0$.

To handle the U-processes which are related to t_{nj}, we need to introduce addtional terminologies and notations, see de la Pena and Gine (1999). For a kernel function f of k variables, we denote

$$U_n^{(k)}(f)=\frac{(n-k)!}{n!}\sum_{i\in I_n^k}f(X_{i_1},\cdots,X_{i_k}),$$

where $I_n^m=\{(i_1,\cdots,i_m):1\le i_j\le n, i_j\ne i_k \text{ if } j\ne k\}$. Now suppose f is symmetric in its entries, we have the well-known Hoeffding decomposition:

$$U_n^{(m)}(f)-Ef=\sum_{k=1}^m U_n^{(k)}(\pi_k f),$$

where

$$\pi_k f=(\delta_{x_1}-P)\times\cdots\times(\delta_{x_k}-P)\times P^{m-k}f.$$

Moreover let σ^2 (which we call maximal variance) be any number satisfying

$$\left\|P^m f^2\right\|_{\mathcal{F}}\le\sigma^2\le M^2.$$

Lemma 2.5.4 *(Gine and Mason, 2007, Theorem 8) Let $\mathcal{F}$ be a collection of measurable symmetric functions $f : S^m \to \mathcal{R}$, bounded up by M in absolute values, and let P be any probability measure on $(S, \mathcal{S})$. Assume $\mathcal{F}$ is of VC type with envelope function $F \equiv M$ and with characteristics A and v. Then for every $m \in \mathcal{N}$, and $A \geq e^m, v \geq 1$, there exist constants C_1, C_2, s.t. for any $k = 1, ..., m$,*

$$n^k E \left\| U_n^{(k)} (\pi_k f) \right\|_{\mathcal{F}}^2 \leq C_1 2^k \sigma^2 \left(\log \left(\frac{A}{\sigma} \right) \right)^k, \tag{B.9}$$

assuming $n\sigma^2 \geq C_2 \log\left(\frac{A}{\sigma}\right)$.

The following lemma gives the expression for the influence function in those terms $t_{n,1}$ and $t_{n,2}$.

Lemma 2.5.5 *Under Assumptions (B1)-(B5), we have for $j = 1, 2$:*

$$t_{nj} = \frac{1}{n} \sum_{i=1}^{n} \varphi_j (O) + o_p \left(n^{-1/2} \right),$$

where the influence functions are defined by

$$\begin{aligned} \varphi_1 (O_i) &= e_t^T \mu_p^{-1} (K) \frac{\varepsilon_{i,D}}{\left[a_Y' (X_i) \right]^2} \int K (u) U (u) u du \\ &\quad + \frac{\varepsilon_{i,V} b_{D_\iota} (X_i)}{a_V^2 (X_i)} + \frac{\varepsilon_{i,D_\iota} f_X (X_i)}{a_V (X_i)} e_t^T G (X_i) \int K (u) U (u) du, \end{aligned}$$

and

$$\begin{aligned} \varphi_2 (O_i) &= \frac{b_V (X_i) \varepsilon_{i,D_\iota}}{a_V^2 (X_i)} + 2 \frac{a_{D_\iota} (X_i) b_V (X_i) \varepsilon_{i,V}}{a_V^3 (X_i)} \\ &\quad + \frac{a_{D_\iota} (X_i)}{a_V^2 (X_i)} e_t^T G (X_i) \varepsilon_{i,V} \int K (u) U (u) du \qquad + \left[\frac{a_{D_\iota} (X_i)}{a_V^2 (X_i)} \right]' e_t^T \mu_p^{-1} (K) \varepsilon_{i,V} \int K (u) U (u) u du. \end{aligned}$$

Proof. The goal is to collect the dominating linear terms from various U-statistics. We only carry out a detailed demonstration with $\widehat{\beta}_{\iota,1}^{\#}$, and state the results for the part related to $\widehat{\beta}_{\iota,2}^{\#}$. By the linear representation in (B.7), it follows

$$\begin{aligned} t_{n1} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{b}_{D_\iota} (X_i) - b_{D_\iota} (X_i)}{a_V (X_i)} - \frac{1}{n} \sum_{i=1}^{n} \frac{\left[\widehat{a}_V (X_i) - a_V (X_i) \right] b_{D_\iota} (X_i)}{a_V^2 (X_i)} \\ &= \frac{1}{h_n^{d+1}} U_n^{(2)} (f_1) - \frac{1}{h_n^d} U_n^{(2)} (f_2) - \frac{1}{h_n^d} U_n^{(2)} (f_3) + o_p \left(n^{-1/2} \right). \end{aligned}$$

where

$$
\begin{aligned}
U_n^{(2)}(f_1) &= \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{a_V(X_i)\,f_X(X_i)}\left(\sum_{j\neq i}e_t^T\mu_p^{-1}(K)\,\varepsilon_{j,D_\iota}K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right)\right),\\
U_n^{(2)}(f_2) &= \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{a_V(X_i)}\left(\sum_{j\neq i}e_t^T G(X_i)\,\varepsilon_{j,D_\iota}K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right)\right),\\
U_n^{(2)}(f_3) &= \frac{1}{n^2}\sum_{i=1}^{n}\frac{b_{D_\iota}(X_i)}{a_V^2(X_i)\,f_X(X_i)}\left(\sum_{j\neq i}e_{\mathbf{0}}^T\mu_p^{-1}(K)\,\varepsilon_{j,V}K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right)\right).
\end{aligned}
$$

These three terms could be analyzed via U-statistics by a simple symmetrization, say for the first one

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{nh_n^{d+1}}\sum_{j\neq i}e_t^T\mu_p^{-1}(K)\,\frac{\varepsilon_{j,D_\iota}}{a_V(X_i)\,f_X(X_i)}K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right)\right)\\
&= \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}e_t^T\mu_p^{-1}(K)\,\frac{1}{h_n^{d+1}}K\left(\frac{X_j-X_i}{h_n}\right)\left[\frac{\varepsilon_{j,D_\iota}}{a_V(X_i)\,f_X(X_i)}U\left(\frac{X_j-X_i}{h_n}\right)+\frac{\varepsilon_{i,D_\iota}}{a_V(X_j)\,f_X(X_j)}U
\end{aligned}
$$

The classes of functions that come into play are

$$
\begin{aligned}
\mathcal{F}_1 &= \left\{f_1 : f_1=\frac{\varepsilon_{j,D_\iota}}{a_V(X_i)\,f_X(X_i)}e_t^T\mu_p^{-1}(K)\,K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right) : h_n\in\mathcal{H}\right\},\\
\mathcal{F}_2 &= \left\{f_2 : f_2=\frac{\varepsilon_{j,D_\iota}}{a_V(X_i)}e_t^T G(X_i)\,K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right) : h_n\in\mathcal{H}\right\},\\
\mathcal{F}_3 &= \left\{f_3 : f_3=\frac{\varepsilon_{j,V}b_{D_\iota}(X_i)}{a_V^2(X_i)\,f_X(X_i)}e_{\mathbf{0}}^T\mu_p^{-1}(K)\,K\left(\frac{X_j-X_i}{h_n}\right)U\left(\frac{X_j-X_i}{h_n}\right) : h_n\in\mathcal{H}\right\}.
\end{aligned}
$$

The first order term in the Hoeffding decomposition for f_1 is

$$\begin{aligned}
&\frac{1}{h_n^{d+1}} U_n^{(1)}(\pi_1 f_1) \\
= &\frac{1}{n}\sum_{i=1}^{n} E_j\left[e_t^T \mu_p^{-1}(K)\frac{1}{h_n^{d+1}} K\left(\frac{X_i - X_j}{h_n}\right)\frac{\varepsilon_{i,D}}{a_V(X_j) f_X(X_j)} U\left(\frac{X_i - X_j}{h_n}\right)\right] \\
= &\frac{1}{n}\sum_{i=1}^{n} \varepsilon_{i,D} e_t^T \mu_p^{-1}(K)\int \frac{1}{a_V(x_j) h_n^{d+1}} K\left(\frac{x_j - X_i}{h_n}\right) U\left(\frac{X_i - x_j}{h_n}\right) dF_X(x_j) \\
= &\frac{1}{n}\sum_{i=1}^{n} \varepsilon_{i,D} e_t^T \mu_p^{-1}(K)\int \frac{1}{a_V(X_i - u h_n) h_n} K(u) U(u)\, du \\
= &\frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_{i,D}}{a_Y(X_i) h_n} e_t^T \mu_p^{-1}(K)\int K(u) U(u)\, du \\
&+\frac{1}{n}\sum_{i=1}^{n} \varepsilon_{i,D} e_t^T \mu_p^{-1}(K)\frac{1}{\left[a_V'(X_i)\right]^2}\int K(u) U(u)\, u du + s.o. \\
= &\frac{1}{n}\sum_{i=1}^{n} e_t^T \mu_p^{-1}(K)\frac{\varepsilon_{i,D}}{\left[a_V'(X_i)\right]^2}\int K(u) U(u)\, u du + s.o.
\end{aligned}$$

since $e_t^T \mu_p^{-1}(K)\int K(u) U(u)\, du = \left(\mu_p^{-1}(K)\mu_p(K)\right)_{t+1,1} = I_{t+1,1} = 0$ for $1 \le t \le d$ as in Lemma A.3 of Li, Lu, and Ullah (2003). For f_2, f_3, we get

$$\begin{aligned}
&\frac{1}{h_n^d} U_n^{(1)}(\pi_1 f_2) \\
= &\frac{1}{n}\sum_{i=1}^{n} E_j\left[\frac{\varepsilon_{i,D_\iota}}{a_V(X_j)} e_{\mathbf{1}}^T G(X_j)\frac{1}{h_n^d} K\left(\frac{X_i - X_j}{h_n}\right) U\left(\frac{X_i - X_j}{h_n}\right)\right] \\
= &\frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_{i,D_\iota} f_X(X_i)}{a_V(X_i)} e_{\mathbf{1}}^T G(X_i)\int K(u) U(u)\, du + s.o.,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{h_n^d} U_n^{(1)}(\pi_1 f_3) \\
= &\frac{1}{n}\sum_{i=1}^{n} E_j\left[\frac{\varepsilon_{i,V} b_{D_\iota}(X_j)}{a_V^2(X_j) f_X(X_j)} e_{\mathbf{0}}^T \mu_p^{-1}(K)\frac{1}{h_n^d} K\left(\frac{X_i - X_j}{h_n}\right) U\left(\frac{X_i - X_j}{h_n}\right)\right] \\
= &\frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_{i,V} b_{D_\iota}(X_i)}{a_V^2(X_i)}\left[e_0^T \mu_p^{-1}(K)\int K(u) U(u)\, du\right] + s.o. \\
= &\frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_{i,V} b_{D_\iota}(X_i)}{a_V^2(X_i)} + s.o.,
\end{aligned}$$

where the final equality follows $e_0^T \mu_p^{-1}(K)\int K(u) U(u)\, du = \left(\mu_p^{-1}(K)\mu_p(K)\right)_{1,1} = 1$.

Meanwhile, the second order tem of the Hoeffding decomposition is of smaller order

$$h_n^{-d-1}U_n^{(2)}(\pi_2 f_1) = O_p\left(\frac{\log n}{nh_n}\right),$$

by the maximal moment bound in Lemma 2.5.4, once we bound the maximal variance as

$$E\left(\left[\frac{\varepsilon_{j,D_\iota}}{a_V(X_i)}e_t^T G(X_i) K\left(\frac{X_j - X_i}{h_n}\right) U\left(\frac{X_j - X_i}{h_n}\right)\right]^2\right) = O(h_n).$$

Similarly, for the other second order terms, we have $h_n^{-d}U_n^{(2)}(\pi_2 f_2) = O_p\left(\frac{\log n}{n}\right)$, and $h_n^{-d}U_n^{(2)}(\pi_2 f_3) = O_p\left(\frac{\log n}{n}\right)$.

For t_{n2}, the decomposition is carried out as

$$\begin{aligned} t_{n2} &= \frac{1}{n}\sum_{i=1}^n \frac{\widehat{a}_{D_\iota}(X_i) b_V(X_i)}{(a_V(X_i))^2} + \frac{1}{n}\sum_{i=1}^n \frac{a_{D_\iota}(X_i)\widehat{b}_V(X_i)}{(a_V(X_i))^2} - \frac{2}{n}\sum_{i=1}^n \frac{a_{D_\iota}(X_i) b_V(X_i)\widehat{a}_V(X_i)}{(a_V(X_i))^3} \\ &= \frac{1}{h_n^d}U_n^{(2)}(g_1) + \frac{1}{h_n^{d+1}}U_n^{(2)}(g_2) + \frac{1}{h_n^d}U_n^{(2)}(g_3) - \frac{1}{h_n^d}U_n^{(2)}(g_4) + o_p\left(n^{-1/2}\right). \end{aligned}$$

where

$$\begin{aligned} U_n^{(2)}(g_1) &= \frac{1}{n^2}\sum_{i=1}^n \frac{b_V(X_i)}{a_V^2(X_i) f_X(X_i)}\left(\sum_{j\neq i} e_{\mathbf{0}}^T \mu_p^{-1}(K)\varepsilon_{j,D_\iota} K\left(\frac{X_j - X_i}{h_n}\right) U\left(\frac{X_j - X_i}{h_n}\right)\right), \\ U_n^{(2)}(g_2) &= \frac{1}{n^2}\sum_{i=1}^n \frac{a_{D_\iota}(X_i)}{a_V^2(X_i) f_X(X_i)}\left(\sum_{j\neq i} e_t^T \mu_p^{-1}(K)\varepsilon_{j,V} K\left(\frac{X_j - X_i}{h_n}\right) U\left(\frac{X_j - X_i}{h_n}\right)\right), \\ U_n^{(2)}(g_3) &= \frac{1}{n^2}\sum_{i=1}^n \frac{a_{D_\iota}(X_i)}{a_V^2(X_i)}\left(\sum_{j\neq i} e_t^T G(X_i)\varepsilon_{j,V} K\left(\frac{X_j - X_i}{h_n}\right) U\left(\frac{X_j - X_i}{h_n}\right)\right), \\ U_n^{(2)}(g_4) &= \frac{2}{n^2}\sum_{i=1}^n \frac{a_{D_\iota}(X_i) b_V(X_i)}{a_V^3(X_i) f_X(X_i)}\left(\sum_{j\neq i} e_{\mathbf{0}}^T \mu_p^{-1}(K)\varepsilon_{j,V} K\left(\frac{X_j - X_i}{h_n}\right) U\left(\frac{X_j - X_i}{h_n}\right)\right). \end{aligned}$$

A completely analogous derivation of the Hoeffding decomposition would spell out its influence function as

$$\begin{aligned} \varphi_2(O_i) &= \frac{b_V(X_i)\varepsilon_{i,D_\iota}}{a_V^2(X_i)} + 2\frac{a_{D_\iota}(X_i) b_V(X_i)\varepsilon_{i,V}}{a_V^3(X_i)} \\ &\quad + \frac{a_{D_\iota}(X_i)}{a_V^2(X_i)} e_t^T G(X_i)\varepsilon_{i,V}\int K(u) U(u)\, du + \left[\frac{a_{D_\iota}(X_i)}{a_V^2(X_i)}\right]' e_t^T \mu_p^{-1}(K)\varepsilon_{i,V}\int K(u) U(u)\, u du. \end{aligned}$$

■

2.6 Appendix C.

In this appendix, we abbreviate Chen, Guo, Sun, and Wang (2010) as CGSW. For any vector $z \in \mathcal{R}^p$, we denote $z^{\otimes 0} = 1$, $z^{\otimes 1} = z$ and $z^{\otimes 2} = zz^T$ as in Anderson and Gill (1982). It will prove to be convenient to introduce additional notations:

$$\begin{aligned} P(t,x) &= E[Y(t)|X=x], \\ \rho(t,x) &= E[Y(t)|X=x]\exp[\psi(x)], \\ \varsigma(t,x) &= \frac{\partial}{\partial x}[f_X(x)\rho(t,x)] \times \rho^{-1}(t,x). \end{aligned}$$

Moreover we definte

$$S_{n,k}(t,\zeta(x),x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x) Y_i(u) \exp\left(X_i^*(x)^T \beta_0(x) + U_i^*(x)^T \zeta(x)\right) U_i^*(x)^{\otimes k},$$

$$S_{n,k}^*(t,x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x) Y_i(u) \Psi(X_i) U_i^*(x)^{\otimes k},$$

with their corresponding limits as

$$\begin{aligned} s_k(t,\zeta(x),x) &= f_X(x)\rho(t,x)\int \exp\left(U^*(u)^T \zeta(x)\right) K(u) U^{*\otimes k}(u)\, du, \\ s_k^*(t,x) &= f_X(x)\rho(t,x)\int K(u) U^{*\otimes k}(u)\, du, \end{aligned}$$

for $k = 0, 1, 2$. Notice that $s_k(t,0,x) = s_k^*(t,x)$. Furthermore, let

$$J_n(t,\zeta(x),x) = \frac{S_{n,2}(t,\zeta(x),x)}{S_{n,0}(t,\zeta(x),x)} - \left[\frac{S_{n,1}(t,\zeta(x),x)}{S_{n,0}(t,\zeta(x),x)}\right]^{\otimes 2}.$$

Now we list the regularity conditions adapted from Fan, King, and Gijbels (1997), and CGSW (2010).

Assumption (C1). The function $\psi(x)$ is $(p+1)-$th order continuously differentiable with bounded $(p+1)$-th order derivative uniformly in $x \in \mathcal{X}$ for all $x \in \mathcal{X}$. Furthermore, for every $h > 0$ and every point $x \in \mathcal{X}_{\sqrt{h}}$, the set $B_{\sqrt{h}}(x) \cap \mathcal{X}$ contains a nonempty interior.

Assumption (C2). X is bounded and has compact support on $\mathcal{X} \subset \mathcal{R}^d$ with its density function f_X is bounded away from zero and infinity, moreover it is 2nd order differentiable for all $x \in \mathcal{X}$.

Assumption (C3). $P(t,x)$ is uniformly continuous and $\varsigma(t,x)$ is uniformly bounded for $(t,x) \in [0,\tau] \times \mathcal{X}$. Moreover, $s_k(t,\zeta(x),x)$ are uniformly continuous in (t,x) for $k = 0,1,2$.

Assumption (C4). τ is finite such that $\Pr(V > \tau|X) > 0$ almost surely.

Assumption (C5). The following matrix is positive definite for all $x \in \mathcal{X}$:

$$\int_0^\tau \left[\frac{s_2(t,\zeta(x),x)}{s_0(t,\zeta(x),x)} - \left(\frac{s_1(t,\zeta(x),x)}{s_0(t,\zeta(x),x)} \right)^{\otimes 2} \right] s_0^*(t,x)\lambda(t)\,dt.$$

Assumption (C6). Those two bandwidth sequences satisfy: $h \to 0$, $\frac{nh^{d+2}}{\log n} \to \infty$ and $nh^{2p+2} \to 0$ with $h = h_n$ and $h = h_{0n}$. Moreover, $h_{0n}/h_n \to 0$.

Assumption (C7). The kernel density function K is a compactly supported (w.l.o.g let it be $[-1,1]^{d_x}$), bounded kernel (up by $K_{\max}$) and $\int K(u)\,udu = 0$. Furthermore, $\mathcal{K} = \left\{ K\left(\frac{x-\cdot}{h_n}\right) : h_n > 0, x \in \mathcal{R}^d \right\}$ is pointwisely measurable and of VC type. Also for $\mu_p^*(K) = \int [K(u) - v] U^{*\otimes 2}(u)\,du$, $\vartheta_p^*(K) = \int [K(u) - v]^{\otimes 2} U^{*\otimes 2}(u)\,du$, with $\mu_p^*(K)$,$\vartheta_p^*(K)$ positive definite $P \times P$ matricies.

The idea is essentially the same as the conditional moment based estimator since the structural regression parameters are estimated by certain average derivatives. The extra complication stems from the non-*i.i.d.* structure when dealing with the partial likelihood structure. Fortunately, the proof could be streamlined by using the counting process and martingale theory in Anderson and Gill (1982). Recall an n-component multivariate counting process $N^{(n)} = \left(N_1^{(n)}, ..., N_n^{(n)}\right)$ where $N_i^{(n)}$ counts the observed events in the life of ith individual. The sample path of $N_1^{(n)}, ..., N_n^{(n)}$ are step functions, zero at time 0, with jumps of size 1 only, no two component processes jumping at the same time. For each n, $N^{(n)}$ has random intensity process $\lambda^{(n)} = \left(\lambda_1^{(n)}, ..., \lambda_n^{(n)}\right)$ specified as

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\lambda(t)\Psi(X_i).$$

Thus

$$M_i^{(n)}(t) = N_i^{(n)}(t) - \int_0^t \lambda_i^{(n)}(u)\,du,$$

are local martingales satisfying

$$\left\langle M_i^{(n)}(t), M_j^{(n)}(t) \right\rangle = \begin{cases} \int_0^t \lambda_i^{(n)}(u)\,du, & \text{if } i = j \\ 0. & \text{otherwise} \end{cases}.$$

The (maximal) tail inequalities presented in the next two lemmas allow us to obtain sharp control of various stochastic integrals w.r.t. the local martingale (uniformly). In the sequel we shall suppress the superscript (n) for notational brevity.

Lemma 2.6.1 *(Lenglart's Inequalities) (a) Let N be a univariate counting process with intensity process λ, then $\forall \delta, \eta > 0$*

$$\Pr\left[N\left(1\right) > \eta\right] \leq \frac{\delta}{\eta} + \Pr\left[\int_0^1 \lambda\left(t\right) dt > \delta\right].$$

(b) Let M be a local square integrable marginale, then $\forall \delta, \eta > 0$

$$\Pr\left[\sup_{t\in[0,1]} \left|M\left(t\right)\right| > \eta\right] \leq \frac{\delta}{\eta^2} + \Pr\left[\left\langle M\left(1\right), M\left(1\right)\right\rangle > \delta\right].$$

Lemma 2.6.2 *(van de Geer, 1995) Consider a general counting process $\{N_t\}_{t\geq 0}$ with continuous compensator $\{\Lambda_t\}_{t\geq 0}$. Let $\mathcal{G}$ be a class of predictable functions with $g_t \geq -L$ for all $t \geq 0$ and $g \in \mathcal{G}$ and let $H\left(\delta, b, B\right)$ be the entropy of $\mathcal{G}$, where $B \subset \left\{A_T \leq \sigma_T^2\right\}$. There exists constants C_1, C_2, C_3 depending on L s.t. for ε*[5]

$$\frac{\varepsilon b^2}{C_1} \geq \int_0^b \sqrt{H\left(x, b, B\right)} dx \vee b, \tag{C.1}$$

then for a fixed time T, we have

$$\begin{aligned} &\Pr\left(\sup_{g\in\mathcal{G}} \left|\int_0^T g d\left(N - \Lambda\right)\right| \geq \varepsilon b^2, \text{ and } \sup_{g\in\mathcal{G}} d_T^2\left(g, 0\right) \leq b^2 \cap B\right) \qquad \text{(C.2)} \\ \leq\ & C_3 \exp\left(-\frac{\varepsilon^2 b^2}{C_2}\right), \end{aligned}$$

where

$$d_T\left(g, \widetilde{g}\right) = \left(\frac{1}{2}\int_0^T \left[\exp\left(g\right) - \exp\left(\widetilde{g}\right)\right]^2 d\Lambda\right)^{1/2}.$$

A necessary task working with the partial likelihood structure is to handle the stochastic terms like $S_{n,k}\left(t, \zeta\left(x\right), x\right)$ and $S_{n,k}^*\left(t, x\right)$. The following lemma allows us to readily verify

$$\begin{aligned} \sup_{t\in[0,\tau]} \sup_{x\in\mathcal{X}} \left\|S_{n,k}\left(t, \zeta\left(x\right), x\right) - s_k\left(t, \zeta\left(x\right), x\right)\right\| &= o_p\left(1\right), \qquad \text{(C.3)} \\ \sup_{t\in[0,\tau]} \sup_{x\in\mathcal{X}} \left\|S_{n,k}^*\left(t, x\right) - s_k^*\left(t, x\right)\right\| &= o_p\left(1\right). \end{aligned}$$

[5] In van de Geer (1995), the theorem is stated for $\varepsilon \in [0, 1]$ which is actually taken to be $1/2$ in her subsequent applications. However any ε would work as long as (C.1) is satisfied.

Lemma 2.6.3 *(CGSW Lemma 4) Under Assumptions (C1)-(C7), if $g(z,u,x)$ is a continuous function, then*

$$\sup_{t\in[0,\tau],x\in\mathcal{X}}\left|c_n(x,t)-Ec_n(x,t)\right|=O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right),$$

where

$$c_n(x,t)=\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\,g\left(X_i,\frac{X_i-x}{h_n},x\right)K_h(X_i-x),$$

and

$$c(x,t)=f_X(x)\,P(t,x)\int g(x,u,x)\,K(u)\,du.$$

The claimed results in (C.3) follow upon choosing $g(z,u,x)$ to be either

$$g\left(X_i,\frac{X_i-x}{h_n},x\right)=\exp\left(X_i^*(x)^T\beta_0(x)+U_i^*(x)^T\zeta(x)\right)U^*\left(\frac{X_i-x}{h_n}\right)^{\otimes k},$$

or

$$g\left(X_i,\frac{X_i-x}{h_n},x\right)=\exp\left(\psi(X_i)\right)U^*\left(\frac{X_i-x}{h_n}\right)^{\otimes k}.$$

Proof. (of Theorem 2.4.5) We will prove the root-n consistency and asymptotic normality for each coordinate (say for $t=1$) from the average derivatives because the multivariate version follows standard Cramer-Wold device. Moreover the structural parameter is the sum of relevant terms in (2.4.11), it is enough to consider $\iota=1$.

Step 1. We show that $\sup_x\left|\widehat{\zeta}(x)\right|\to 0$ in probability where $\widehat{\zeta}(x)=H^*\left[\widehat{\beta}(x)-\beta(x)\right]$, where $H^*\equiv Diag\left(h_n^{|\mathbf{v}|},1\le|\mathbf{v}|\le p\right)$. This is done by proving the sample criterion function converges uniformly to a function having a well separated point of maximum as detailed in Lemma 2.6.4.

Step 2. Given the uniform consistency, we obtain the linear representation for $\widehat{\zeta}(x)$ uniformly over x

$$H\left[\widehat{\beta}(x)-\beta(x)\right]=-\left[\ddot{l}_n(0,x)\right]^{-1}\dot{l}_n(0,x)+O_p\left(\frac{\log n}{nh_n^d}\right).$$

The stochastic order of the score and Hessian functions need to be pinned down to ensure the remainder term is negligible for the root-n consistency result.

Step 3. Similar as the linearization in Appendix B, one obtains

$$\begin{aligned}\widehat{\beta}_1^* - \beta_1^* &= \frac{1}{n}\sum_{i=1}^{n} e^{\alpha(X_i)}\left[a_{D_1}(X_i)\,\beta_1(X_i) + b_{D_1}(X_i)\right] - Ee^{\alpha(X_i)}\left[a_{D_1}(X_i)\,\beta_1(X_i) + b_{D_1}(X_i)\right] \\ &\quad + \sum_{j=1}^{4} t_{nj} + o_p\left(n^{-1/2}\right),\end{aligned}$$

where

$$\begin{aligned}t_{n1} &\equiv \frac{1}{n}\sum_{i=1}^{n}\left\{e^{\alpha(X_i)} a_{D_1}(X_i)\left[\widehat{\beta}_1(X_i) - \beta_1(X_i)\right]\right\}, \\ t_{n2} &\equiv \frac{1}{n}\sum_{i=1}^{n}\left\{e^{\alpha(X_i)}\left[a_{D_1}(X_i)\,\beta_1(X_i) + b_{D_1}(X_i)\right]\left[\widehat{\alpha}(X_i) - \alpha(X_i)\right]\right\}, \\ t_{n3} &= \frac{1}{n}\sum_{i=1}^{n}\left\{e^{\alpha(X_i)}\left[\widehat{b}_{D_1}(X_i) - b_{D_1}(X_i)\right]\right\} \\ t_{n4} &= \frac{1}{n}\sum_{i=1}^{n}\left\{e^{\alpha(X_i)}\beta_1(X_i)\left[\widehat{a}_{D_1}(X_i) - a_{D_1}(X_i)\right]\right\},\end{aligned}$$

Because the regressor is time-invariant, the local martingale has zero mean conditional on all $\{X_i\}_{i=1}^n$. Hence the asymptotic normality of t_{n1}, t_{n2} could be checked apart from the average of *i.i.d.* terms on RHS of the decomposition. Handling t_{n3} and t_{n4} is identical to the argument in Section 2.4.2, and will be omitted. Since t_{n2} stands for the average of intercept terms in local polynomial fitting, its asymptotic normality follows from Theorem 3 in CGSW (2010) by taking their $\phi(\cdot)$ function to be $\phi(x) = e^{\alpha(x)}\left[a_{D_1}(x)\,\beta_1(x) + b_{D_1}(x)\right]$. Below we will show

$$t_{n1} = \int V_1(x)\, f_X(x)\, dx + o_p\left(n^{-1/2}\right),$$

where

$$V_1(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\int e_1^T \ddot{l}_n^{-1}(0,x)\left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right] dM_i(t). \tag{C.4}$$

An stochastic expansion of the function $V_1(x)$ is carried out to get its influence function. Specifically it holds that

$$\int V_1(x)\, f_X(x)\, dx = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau g_i^{(n)}(t)\, dM_i(t) + o_p\left(n^{-1/2}\right), \tag{C.5}$$

with an appropriate choice of $g_i^{(n)}(t)$ function. Then the asymptotic normality follows by verifying the conditions in Rebollebo's CLT (Andersen and Gill, 1982) for

$$W^{(n)}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau g_i^{(n)}(t)\, dM_i(t).$$

In the end we show

$$\left\langle W^{(n)}(\tau), W^{(n)}(\tau)\right\rangle \to_p A(\tau),$$

and the Lindeberg condition is satisfied for $\forall \varepsilon > 0$:

$$\frac{1}{n}\int_0^\tau \sum_{i=1}^{n}\left[g_i^{(n)}(t)\right]^2 \lambda(t)\,\Psi(X_i)\, I\left\{\left|g_i^{(n)}(t)\right| > \sqrt{n}\varepsilon\right\} dt \to_p 0.$$

The asymptotic variance is the one of the influence function containing both i.i.d. averages and stochastic integrals with respect to the local martingales. ■

Lemma 2.6.4 *Under Assumptions (C1)-(C7),* $\sup_x \left|\widehat{\zeta}(x)\right| \to 0$ *in probability where* $\widehat{\zeta}(x) = H^*\left[\widehat{\beta}(x) - \beta(x)\right]$.

Proof. To prove the uniform consistency, we show the sample criterion function (namely, the recentered partial log-likelihood function) converges uniformly to a function having a well separated point of maximum. In the end, we apply Lemma A.1 in Sun, Gilbert, and McKeague (2009) to conclude.

Note that $\widehat{\zeta}(x)$ maximizes the difference of criterion function:

$$l_n(\zeta(x), x) - l_n(0, x) = Q_n(\zeta(x), x) + X_n(\zeta(x), x),$$

with

$$\begin{aligned} Q_n(\zeta(x), x) &= \int_0^\tau S_{n,1}^*(t, x)^T \zeta(x)\lambda(t)\, dt \\ &\quad - \int_0^\tau \log\left[\frac{S_{n,0}(t, \zeta(x), x)}{S_{n,0}(t, 0, x)}\right] S_{n,0}^*(t, x)\lambda(t)\, dt, \end{aligned}$$

and

$$X_n(\zeta(x), x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(X_i - x)\left\{U_i^*(x)^T\zeta(x) - \log\left[\frac{S_{n,0}(t, \zeta(x), x)}{S_{n,0}(t, 0, x)}\right]\right\} dM_i(t).$$

By Lemma 2.6.3, we get $Q_n(\zeta(x), x) \to_p Q(\zeta(x), x)$ uniformly with

$$\begin{aligned} Q(\zeta(x), x) &= \int_0^\tau s_1^*(t,x)^T [\beta(x) - \beta^*(x)] \lambda(t)\, dt \\ &\quad - \int_0^\tau \log\left[\frac{s_0(t, \zeta(x), x)}{s_0(t,0,x)}\right] s_0^*(t,x)\lambda(t)\, dt. \end{aligned}$$

The function $Q(\zeta(x), x)$ possesses a unique well separated maximum point at 0 as

$$\begin{aligned} \frac{\partial}{\partial \zeta(x)} Q(\zeta(x), x) &= \int_0^\tau \left[\frac{s_1^*(t,x)}{s_0^*(t,x)} - \frac{s_1(t,\zeta(x),x)}{s_0(t,\zeta(x),x)}\right] s_0^*(t,x)\lambda(t)\, dt, \\ \frac{\partial}{\partial \zeta(x)\partial \zeta(x)^T} Q(\zeta(x), x) &= \int_0^\tau \left[\frac{s_2(t,\zeta(x),x)}{s_0(t,\zeta(x),x)} - \left(\frac{s_1(t,\zeta(x),x)}{s_0(t,\zeta(x),x)}\right)^{\otimes 2}\right] s_0^*(t,x)\lambda(t)\, dt. \end{aligned}$$

Hence $\frac{\partial}{\partial \zeta(x)} Q(\zeta(x), x)|_{\zeta(x)=0} = 0$ and $\frac{\partial}{\partial \zeta(x)\partial \zeta(x)^T} Q(\zeta(x), x) > 0$ uniformly for $x \in \mathcal{X}$.

For the other stochastic term, we shall apply van de Geer's maximal inequality in Lemma 2.6.2 to show

$$\sup_x |X_n(\zeta(x), x)| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$$

Notice $X_n(\zeta(x), x)$ is a locally square integrable martingale with predictable variation as

$$\begin{aligned} &\langle X_n(\zeta(x), x), X_n(\zeta(x), x)\rangle \\ &= \frac{1}{n^2}\sum_{i=1}^n \int_0^\tau K_h^2(X_i - x)\left\{U_i^*(x)^T \zeta(x) - \log\left[\frac{S_{n,0}(t,\zeta(x),x)}{S_{n,0}(t,0,x)}\right]\right\}^{\otimes 2} Y_i(t)\exp(\psi(X_i))\, d\Lambda(t) \\ &= O_p\left(\frac{1}{nh_n^d}\right). \end{aligned}$$

Hence $d_T(X_n(\zeta(x), x), 0) = O\left(\frac{1}{nh_n^d}\right)$. If we choose $b = \frac{M}{nh_n^d}$ and $\varepsilon = \sqrt{\log n \times nh_n^d}$ in (C.2), then

$$\Pr\left\{\sup_x |X_n(\zeta(x), x)| \geq M\sqrt{\frac{\log n}{nh_n^d}}\right\} \leq \exp(-M \log n),$$

for large enough n. ■

Recall the score function admits the following decomposition $\dot{l}_n(\zeta(x), x) = V_n(\zeta(x), x) + B_n(\zeta(x), x)$ and

$$\begin{aligned} V_n(\zeta(x), x) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(X_i - x)\left[U_i^*(x) - \frac{S_{n,1}(t,\zeta(x),x)}{S_{n,0}(t,\zeta(x),x)}\right] dM_i(t), \\ B_n(\zeta(x), x) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(X_i - x)\left[U_i^*(x) - \frac{S_{n,1}(t,\zeta(x),x)}{S_{n,0}(t,\zeta(x),x)}\right] Y_i(t)\exp(\psi(X_i))\, d\Lambda(t). \end{aligned}$$

We will handle the bias term first at $\zeta(x) = 0$.

Lemma 2.6.5 *Suppose Assumptions (C1)-(C7) hold, we get the uniform order of the bias term as* $\sup_x |B_n(0,x)| = O_p\left(h^{p+1}\right)$.

Proof. Recall that

$$B_n(0,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(X_i - x)\left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right] Y_i(t)\exp(\psi(X_i))\, d\Lambda(t).$$

Thus

$$\begin{aligned} B_n(0,x) = {} & \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(X_i - x) Y_i(u)\left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right] \\ & \times\left[\exp(\psi(X_i)) - \exp\left(X_i^*(x)^T\beta\right)\right] d\Lambda(u), \end{aligned}$$

since the following relationship holds

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(X_i - x) Y_i(u)\left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right]\exp\left(X_i^*(x)^T\beta\right) d\Lambda(u) \\ = {} & \int_0^\tau \left[S_{n,1}(t,0,x) - S_{n,1}(t,0,x)\right] d\Lambda(u) = 0. \end{aligned}$$

A standard Taylor expansion leads to

$$\begin{aligned} & \exp(\psi(X_i)) - \exp\left(X_i^*(x)^T\beta\right) \\ = {} & \exp(\psi(x))\frac{\psi^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1} + o_p\left(h^{p+1}\right). \end{aligned}$$

for $|X_i - x| \le h_n$ given the compact support of the kernel function. Therefore we get

$$\begin{aligned} B_n(0,x) = {} & h_n^{p+1} f_X(x)\exp(\psi(x))\frac{\psi^{(p+1)}(x)}{(p+1)!}\Lambda(\tau,x) \\ & \times\int K(u)(u^* - v)u^{p+1}du + o_p\left(h_n^{p+1}\right). \end{aligned}$$

■

Lemma 2.6.6 *Suppose Assumptions (C1)-(C7) hold, the stochastic term is bounded uniformly as*

$$\sup_x |V_n(0,x)| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$$

Proof. A straightforward computation of the quadratic variation of $V_n(0,x)$ shows

$$\begin{aligned} & \langle V_n(0,x), V_n(0,x)\rangle \\ = & \frac{1}{n^2 h_n^{2d}} \sum_{i=1}^{n} \int_0^\tau K_h^2(X_i - x) \left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right]^{\otimes 2} Y_i(t) \exp(\psi(X_i))\, d\Lambda(t) \\ = & \frac{1}{nh_n^d} f_X(x)\, \Psi(x)\, \Lambda(\tau, x) \int K^2(u)\, (u^* - v)^{\otimes 2}\, du. \end{aligned}$$

Whence we could apply (C.2) as handling the term $X_n(\zeta(x), x)$ to conclude, once the additional $\sqrt{\log n}$ factor is incorporated. ■

We have already argued via applying Lemma 5 in CGSW (2010) to get

$$\begin{aligned} & \sqrt{n} t_{n1} \\ = & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau \int e_1^T \ddot{l}_n^{-1}(0,x) K_h(X_i - x) \left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right] f_X(x)\, dx dM_i(t) + o_p(1), \end{aligned}$$

In order to apply Rebollebo's CLT, we need to pin down the influence function.

Lemma 2.6.7 *Suppose Assumptions (C1)-(C7) hold, with $V_1(x)$ defined in (C.4) we get*

$$\int V_1(x) f_X(x)\, dx = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau g_i^{(n)}(t)\, dM_i(t) + o_p\left(n^{-1/2}\right).$$

with $g_i^{(n)}(t)$ given in (C.6) below. Moreover the Lindeberg condition in Rebollebo's CLT is satisfied.

Proof. As in the proof of Lemma 2.5.5, we have the following expansion:

$$\begin{aligned} & \int e_1^T \ddot{l}_n^{-1}(0,x) K_h(X_i - x) \left[U_i^*(x) - \frac{S_{n,1}(t,0,x)}{S_{n,0}(t,0,x)}\right] f_X(x)\, dx \\ = & e_1^T \mu_p^{*-1}(K) \int K(u) [U^*(u) - v]\, du + h_n g_i^{(n)}(t) + o(h_n) \\ = & h_n g_i^{(n)}(t) + o(h_n), \end{aligned}$$

where the first term is identically equal to zero because $e_1^T \mu_p^{*-1}(K) \int K(u)[U^*(u) - v]\, du = \left(\mu_p^{*-1}(K)\, \mu_p^*(K)\right)_{2,1} = I_{2,1} = 0$. The expression for $g_i^{(n)}(t)$ comes from expanding both the numerator and denominator upto the first order

$$g_i^{(n)}(t) = e_1^T \mu_p^{*-1}(K) \int u^T \varsigma(t,x) K(u) U^*(u)\, du dx \quad + \tag{C.6}$$

$$e_1^T \mu_p^{*-1}(K) \left[\int u^T \varsigma(t,x) K(u) U^{*\otimes 2}(u) - (K(u) U^*(u) u)^{\otimes 2}\, du dx\right] \quad \mu_p^{*-1}(K) \int K(u)[U^*(u) - v]\, du.$$

Hence under Assumptions (C3) and (C7), $g_i^{(n)}(t)$ satisfies the Lindeberg condition in Rebollebo's CLT due to its uniform boundeness. ■

Proof. (of Theorem 2.4.6) The following string of equality is straightforward:

$$\begin{aligned}
& \int_0^\tau B(t)\, d\left[\widehat{\Lambda}(t) - \Lambda(t)\right] \\
= & \int_0^\tau \frac{B(t)}{\sum_j \exp\left(\widehat{\alpha}(X_j)\right) Y_j(t)} \sum_i d\left[N_i(t) - \exp\left(\widehat{\alpha}(X_i)\right) Y_i(t)\, d\Lambda(t)\right] \\
= & \int_0^\tau \frac{B(t)}{\sum_j \exp\left(\widehat{\alpha}(X_j)\right) Y_j(t)} \sum_i dM_i(t) \\
& - \int_0^\tau \frac{B(t)}{\sum_j \exp\left(\widehat{\alpha}(X_j)\right) Y_j(t)} \sum_i \left[\exp\left(\widehat{\alpha}(X_i)\right) - \exp\left(\psi(X_i)\right)\right] Y_i(t)\, d\Lambda(t) \\
= & \int_0^\tau \frac{B(t)}{\sum_j \exp\left(\widehat{\alpha}(X_j)\right) Y_j(t)} \sum_i dM_i(t) \\
& - \sum_i \left[\widehat{\alpha}(X_i) - \psi(X_i)\right] \int_0^\tau \frac{B(t)\, Y_i(t)\, d\Lambda(t)}{\sum_j \exp\left(\widehat{\alpha}(X_j)\right) Y_j(t)} + o_p\left(n^{-1/2}\right) \\
= & O_p\left(n^{-1/2}\right).
\end{aligned}$$

where the first term is $O_p\left(n^{-1/2}\right)$ following the same line as Lemma 2.6.7. The change is rather notation-wise from the slope term to the intercept term in a local polynomial fitting. The second term is also $O_p\left(n^{-1/2}\right)$ as in Theorem 3 in CGSW (2010). ■

2.7 Appendix D.

We shall first present proofs of the point identification results in Section 3 and discuss related extensions of the Marshall-Olkin structure using alternative constructions. Then we proceed to present some observational equivalent pairs of model primitives if the latent processes and structural multiplicative effects are unrestricted.

2.7.1 Identifiability

Proof. (of Proposition 4.2) Recall that conditional joint survival function is

$$\begin{aligned}
& S(t_1, t_2|x) \\
= & \exp\left[-\Psi_1(x)\Lambda(t_1) - \Psi_2(x)\Lambda(t_2) - \Psi_3(x)\Lambda\left[\max(t_1, t_2)\right]\right],
\end{aligned}$$

whence for $t_1 < t_2$, it simplifies to $\exp\left[-\Psi_1(x)\Lambda(t_1) - [\Psi_2(x) + \Psi_3(x)]\Lambda(t_2)\right]$. Upon differentiation, it has a density of the following form:

$$\Psi_1(x)[\Psi_2(x) + \Psi_3(x)]\lambda(t_1)\lambda(t_2)\exp\left[-\Psi_1(x)\Lambda(t_1) - [\Psi_2(x) + \Psi_3(x)]\Lambda(t_2)\right].$$

Integrating the above density over the range $[t, +\infty] \times (t_1, +\infty]$ delivers the conditional sub-survival function for $D = 1$:

$$\begin{aligned} & S_{V,D=1}(t|x) \\ = & \int_t \int_{t_1} \Psi_1(x)[\Psi_2(x) + \Psi_3(x)]\lambda(t_1)\lambda(t_2)\exp\left[-\Psi_1(x)\Lambda(t_1) - [\Psi_2(x) + \Psi_3(x)]\Lambda(t_2)\right] dt_2 dt_1 \\ = & \Psi_1(x)\int_t \exp\left[-[\Psi_1(x) + \Psi_2(x) + \Psi_3(x)]\Lambda(t_1)\right] d\Lambda(t_1) \\ = & \frac{\Psi_1(x)}{\Psi_1(x) + \Psi_2(x) + \Psi_3(x)}\exp\left[-[\Psi_1(x) + \Psi_2(x) + \Psi_3(x)]\Lambda(t)\right]. \end{aligned}$$

A similar derivation gives one the conditional sub-survival function for $D = 2$:

$$\begin{aligned} & S_{V,D=2}(t|x) \\ = & \frac{\Psi_2(x)}{\Psi_1(x) + \Psi_2(x) + \Psi_3(x)}\exp\left[-[\Psi_1(x) + \Psi_2(x) + \Psi_3(x)]\Lambda(t)\right]. \end{aligned}$$

Hence subtracting the sum of those two sub-survival functions from the joint survival function evaluated at $t_1 = t_2 = t$, we get

$$\begin{aligned} & S_{V,D=3}(t|x) \\ = & \frac{\Psi_3(x)}{\Psi_1(x) + \Psi_2(x) + \Psi_3(x)}\exp\left[-[\Psi_1(x) + \Psi_2(x) + \Psi_3(x)]\Lambda(t)\right]. \end{aligned}$$

Notice that the following equation holds

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

given the reparameterization in (2.3.5). Finally the claimed expressions of those conditional sub-density functions follow by differentiating the corresponding sub-survival function. ■

Proof. (of Theorem 2.4.1) Invoking the normalization assumption on $\Phi_{12}(\phi_1(x_0), \phi_2(x_0))$, we could identify $\Phi_{12}(\phi_1(x), \phi_2(x))$ and the baseline hazard function $\Lambda(t)$ separately from the proportional structure of the random variable V. The those reduced-form functions

$\Psi_j(x)$ could be identified from the conditional probabilities as $\Pr\{D=j|x\}$ for $j=1,2,3$ given that we have already identified $\Phi_{12}(\phi_1(x),\phi_2(\cdot)) = \Psi_1(x)+\Psi_2(x)+\Psi_3(x)$.

For the first part of the theorem, we assume that $(\phi_1(x),\phi_2(x))$ are known and set $(u,v) \equiv (\phi_1(x),\phi_2(x))$. The goal is to identify the characteristics related to the latent processes, given we could identify those reduced-form regressions under competing risks. When it comes to the two marginal Lévy-Laplace exponent functions, we have

$$\begin{aligned}\Phi_1(u) &= \int_{\mathcal{R}_+}\left[1-e^{-yu}\right]\Pi_1(dy),\\ \Phi_2(v) &= \int_{\mathcal{R}_+}\left[1-e^{-yv}\right]\Pi_2(dy).\end{aligned}$$

In fact, $\Phi_1(\cdot),\Phi_2(\cdot)$ belong to the class of so-called Bernstein functions, meaning their derivatives are completely monotone as in Theorem 3.2 in Schilling, Song, and Vondracek (2012). The operator $\mathcal{L}$ denotes the Laplace transform of a positive measure and we define two positive measures $\widetilde{\Pi}_\iota(dy) = y\Pi_\iota(dy)$ on the non-negative real line for $\iota=1,2$. Thus, we get

$$\begin{aligned}\frac{\partial}{\partial u}\Phi_1(u) &= \int_{\mathcal{R}_+} e^{-yu}y\Pi_1(dy) = \mathcal{L}\circ\widetilde{\Pi}_1,\\ \frac{\partial}{\partial v}\Phi_2(v) &= \int_{\mathcal{R}_+} e^{-yv}y\Pi_2(dy) = \mathcal{L}\circ\widetilde{\Pi}_2,\end{aligned}$$

by differentiating w.r.t u and v respectively. Assumption (D0)(ii) guarantees that $\widetilde{\Pi}_\iota(\cdot)$ are finite measures for $\iota=1,2$. Hence, by Theorem 12b in Widder (1947) and Prop.1 in Abbring and van den Berg (2003) we could identify $\left(\widetilde{\Pi}_1,\widetilde{\Pi}_2\right)$ from $\left(\mathcal{L}\circ\widetilde{\Pi}_1,\mathcal{L}\circ\widetilde{\Pi}_2\right)$, as long as we get enough variation on some non-empty open sets as in Assumption (A0)(iii). Thereafter, Π_ι are also identified for $\iota=1,2$.

It suffcies to identify the Lévy copula function to fully pin down the characteristics of the Lévy subordinators. Notice that

$$-\frac{\partial^2}{\partial u\partial v}\Phi_{12}(u,v) = \int_{\mathcal{R}_+^2} e^{-y_1u-y_2v}y_1y_2\Pi_{12}(dy_1\times dy_2) = \mathcal{L}\circ\widetilde{\Pi}_{12}.$$

Whence, we could identify $\widetilde{\Pi}_{12}(dy) = y_1y_2\Pi_{12}(dy_1\times dy_2)$ by the bivariate Laplace inversion theorem. Given the joint Lévy measure, its copula function $\mathcal{C}_L$ is unique by the Sklar theorem in Kallsen and Tankov (2006).

Now we proceed to the idenfication of structural multiplicative effects $(\phi_1(x), \phi_2(x))$. The mathematical structure boils down to deal with simple composition of functions by noting that

$$\Phi_\iota(\phi_\iota(x)) = \Psi_\iota(x) + \Psi_3(x),$$

for $\iota = 1, 2$.

(Index Model): The identifiability follows from Theorem 1 in Lin and Kulasekera (2007) given the normalization restriction in D.1 and the smoothness of Lévy-Laplace exponent functions.

(Additive Model): Notice $\Phi_\iota(u)$ is monotone and continuously differentiable when the first moment exists. Also its first order derivative is bounded away from zero. Hence by proper normalization restriction, one could identify all the structural components due to Prop. 3.1 in Howowitz and Mammen (2007).

(Multiplicative Model): The mathematical structure here is essentially the same as in MPH. Note that we could identify

$$\Phi_\iota(\phi_\iota(x)) = \Phi_\iota\left(x_o\widetilde{\phi}_\iota(\widetilde{x})\right) = \int_{\mathcal{R}_+}\left[1 - e^{-x_o\widetilde{\phi}_\iota(\widetilde{x})y}\right]\Pi_\iota(dy),$$

from the competing risks setting. Thereafter if we take the differentiation w.r.t x_o,

$$\frac{\partial}{\partial x_o}\Phi_\iota(\phi_\iota(x)) = \int_{\mathcal{R}_+}\left[e^{-x_o\widetilde{\phi}_\iota(\widetilde{x})y}\right]\widetilde{\phi}_\iota(\widetilde{x})\, y\Pi_\iota(dy),$$

Thus as in Abbring and van den Berg (2003), we arrives at

$$\widetilde{\phi}_\iota(\widetilde{x}) = \lim_{x_o\to 0}\frac{\frac{\partial}{\partial x_o}\Phi_\iota(\phi_\iota(x_o, \widetilde{x}))}{\frac{\partial}{\partial x_o}\Phi_\iota(\phi_\iota(x_o, \widetilde{x}_*))},$$

by the normalizing restriction and existence of first moment. ■

Lemma 2.7.1 *For non-negative (u, v), we have the following inequalities*

$$\max\{\Phi_1(u), \Phi_2(v)\} \leq \Phi_{12}(u, v) \leq \Phi_1(u) + \Phi_2(v),$$

for the marginal and joint Lévy-Laplace exponent functions of Levy subordinators $(L_1(t), L_2(t))$.

Proof. For an arbitrary Lévy copula $\mathcal{C}_L$ defined on $\mathcal{R}_+^2$, we have the Frechet-Hoeffding type inequality (Cont and Tankov, 2004) hold $\mathcal{C}_{L,\perp} \leq \mathcal{C}_L \leq \mathcal{C}_{L,\parallel}$ where

$$\begin{aligned} \mathcal{C}_{L,\perp}(u,v) &= u \cdot I\{v = +\infty\} + v \cdot I\{u = +\infty\}, \\ \mathcal{C}_{L,\parallel}(u,v) &= \min\{u,v\}. \end{aligned}$$

$\mathcal{C}_{L,\perp}(u,v)$ is the independence copula and $\mathcal{C}_{L,\parallel}(u,v) = \min\{u,v\}$ is the complete dependence copula function, respectively. Notice that $1 - e^{-y_1 u - y_2 v}$ is the distribution function of a bivariate exponential distribution function with parameters (u,v), which we denote by $F_{u,v}(\cdot,\cdot)$. For the second inequality in Lemma 2.7.1,

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

where integration-by-parts is used in the second equality is due to the Frechet-Hoeffding type inequality for Lévy copula. Similarly, we have

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

where the final inequality follows from the fact that

$$\begin{aligned} & 1 - e^{-y_1 u - y_2 v} \\ \geq & \max\left\{1 - e^{-y_1 u}, 1 - e^{-y_2 v}\right\}, \end{aligned}$$

and the integration is carried out on the line $\left\{(y_1, y_2) : \overline{\Pi}_1(y_1) = \overline{\Pi}_2(y_2)\right\}$. ■

One might wonder whether our new construction using Lévy subordinators would lead to a multivariate Marshall-Olkin model for any finite dimensional duration variables. I provide a positive answer for any finite dimensional vector Y_j^*, based on a single Lévy subordinator $L(t)$. This restriction should not be too surprising, the original MO model for general multivariate case involves overall $2^N - 1$ mutually independent latent exponential variables, and each observable failure time is the minimum of 2^{N-1} of those exponential variables, see Section 4 in Marshall and Olkin (1967). To the best of my knowledge, this construction has never been stated explicitly in the literature before, although the idea is from Theorem 3.3 in Mai and Scherer (2009). The restriction to one common latent process intead of a general multivariate process stems from a technical requirement on certain sequence to be completely monotone. First some notations and terminology would be introduced to facilitates the presentation. Let ∇ stand for the difference operator defined on sequence $(a_k)_{k\in N}$ by $\nabla a_k = a_k - a_{k+1}$.

Definition 2.7.2 *The sequence $(a_k)_{k\in\mathcal{N}}$ is said to be completely monotone if $\nabla^j a_k \geq 0$ for all k, $j \in \mathcal{N}$, where the $j-th$ order of difference operator is defined formally by*

$$\nabla^j a_k = \sum_{i=0}^{j} (-1)^i \binom{j}{i} a_{k+i}.$$

Proposition 2.7.3 *For any finite positive integer $N \geq 2$, the following random variables follow the Marshall-Olkin distribution:*

$$Y_j^* = \inf\left\{t : L(t) \geq \varepsilon_j\right\}, \text{ for } j = 1, ..., N$$

where the Levy subordinator $L(t)$ and unit exponential thresholds ε_j are all independent. The joint survival function takes the following form:

$$\begin{aligned} & \Pr\left\{Y_1^* > y_1, ..., Y_N^* > y_N\right\} \qquad (A.1) \\ = \; & \exp\left[-\sum_{j=1}^{N} \lambda_j y_j - \sum_{i<j} \lambda_{ij} \max(y_i, y_j) - ... - \lambda_{12...N} \max(y_1, ..., y_N)\right], \end{aligned}$$

with coefficients λ. all being non-negative.

Proof. Let $y_{(1)} \leq \cdots \leq y_{(N)}$ be the ordered values, then

$$\begin{aligned}
& \Pr\left\{Y_1^* > y_1, ..., Y_N^* > y_N\right\} \\
= \; & E\left[\exp\left(-\sum_{j=1}^{N} L\left(y_j\right)\right)\right] \\
= \; & E\left[\exp\left\{-\left(L\left(y_{(N)}\right) - L\left(y_{(N-1)}\right)\right) - 2\left(L\left(y_{(N-1)}\right) - L\left(y_{(N-2)}\right)\right) - \cdots - NL\left(y_{(1)}\right)\right\}\right] \\
= \; & \exp\left\{-\left[y_{(N)} - y_{(N-1)}\right]\Phi\left(1\right) - \left[y_{(N-1)} - y_{(N-2)}\right]\Phi\left(2\right) - \cdots - y_{(1)}\Phi\left(N\right)\right\} \\
= \; & \exp\left\{-\sum_{j=1}^{N} a_{N-j} y_{(j)}\right\},
\end{aligned}$$

where in the last equation we set $a_k \equiv \Phi\left(k+1\right) - \Phi\left(k\right)$. By Theorem 2.9 in Mai and Scherer (2009), this sequence $\left(a_k\right)_{k\in\mathcal{N}}$ is completely monotone. Thus, proceed from the last line in Lemma A.2 in Mai and Scherer (2009):

$$\begin{aligned}
& \exp\left\{-\sum_{j=1}^{N} a_{N-j} y_{(j)}\right\} \\
= \; & \exp\left\{-\sum_{j=0}^{N-1}\left(\sum_{1\leq i_0<\cdots<i_j\leq N}\left[\max\left(y_{i_0}, ..., y_{i_j}\right)\right]\right)\nabla^j a_{N-j-1}\right\},
\end{aligned}$$

which is indeed of the form of (A.1). Finally, those coefficients are all non-negative guaranteed by the complete monotonicity of sequence $\left(a_k\right)_{k\in\mathcal{N}}$. ■

Here we provide an alternative construction of EMO resorting to further subordinating two independent processes. It is well-known that subordinating independent processes is a common approach to generate dependent Lévy processes, see Sato (1999). Assume we have three mutually independent Lévy subordinators $L_1\left(t\right)$, $L_2\left(t\right)$, and $\mathcal{T}\left(t\right)$. Two structural durations are now defined by

$$Y_\iota^* \equiv \inf\left\{t_\iota : \phi_\iota\left(X\right) L_\iota\left(\mathcal{T}\left(t_\iota\right)\right) \geq \varepsilon_\iota\right\}, \; \iota = 1 \text{ or } 2. \tag{A.2}$$

Proposition 2.7.4 *Assume the structural durations are defind as in (A.2), where $\mathcal{T}\left(t\right)$ has its Lévy-Laplace exponent function $\Xi\left(\cdot\right)$. The conditional joint distribution of two durations*

is of the Marshall-Olkin forms with the reduced-form regressions:

$$\begin{aligned}
\Psi_1(x) &= \Xi(\Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x))) - \Xi(\Phi_2(\phi_2(x))), \\
\Psi_2(x) &= \Xi(\Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x))) - \Xi(\Phi_1(\phi_1(x))), \\
\Psi_3(x) &= \Xi(\Phi_1(\phi_1(x))) + \Xi(\Phi_2(\phi_2(x))) \\
&\quad -\Xi(\Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x))).
\end{aligned}$$

Proof. The proof is similar as our Theorem 2.3.1 in the main text, with the aid of Lemma 2.15 in Kyprianou (2006) which states $(L_1(\mathcal{T}(t)), L_2(\mathcal{T}(t)))$ is still a bivariate Levy subordinator. Moreover, the corresponding marginal Lévy-Laplace exponent functions are $\Xi \circ \Phi_1$ and $\Xi \circ \Phi_2$ respectively. Hence for (i) $t_1 \leq t_2$ we get:

$$\begin{aligned}
&\Pr\{Y_1^* > t_1, Y_2^* > t_2 | X = x\} \\
= \; & E\{\exp[-(\phi_1(x) L_1[\mathcal{T}(t_1)] + \phi_2(x) L_2[\mathcal{T}(t_1)]) - \phi_2(x)(L_2[\mathcal{T}(t_2)] - L_2[\mathcal{T}(t_1)])]\} \\
= \; & \exp[-(t_2 - t_1)\Xi(\Phi_2(\phi_2(x)))] \times E\{\exp[-\mathcal{T}(t_1)[\Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x))]]\} \\
= \; & \exp[-(t_2 - t_1)\Xi(\Phi_2(\phi_2(x))) - t_1 \Xi(\Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x)))].
\end{aligned}$$

For (ii) $t_1 \geq t_2$, a similar derivation leads to

$$\begin{aligned}
&\Pr\{Y_1^* > t_1, Y_2^* > t_2 | X = x\} \\
= \; & \exp[-(t_1 - t_2)\Xi(\Phi_1(\phi_1(x))) - t_2 \Xi(\Phi_1(\phi_1(x)) + \Phi_2(\phi_2(x)))].
\end{aligned}$$

The claimed expression follows upon straightforward algebra. ■

2.7.2 Observationally Equivalent Components

We present a detailed construction of model primitives which are observationally equivalent in a competing risks setting, if the latent processes and multiplicative effects remain unrestricted. Since the time-change function $\Lambda(\cdot)$ could be separately identified, the confounding issue arises due to the composition of the Lévy-Laplace exponent function and structural multiplicative effect. We will first investigate the problem for the univariate duration model. Let $\Phi_S(u) = \alpha u^\beta$ be the Lévy-Laplace exponent function of a stable process with index α and shape parameter β and $\Phi_G(u) = \rho \log\left(\frac{u+v}{v}\right)$ be the Lévy-Laplace exponent function

of a gamma process with two parameters (ρ, υ). Denote the corresponding multiplicative functions as $\phi_S(x)$ and $\phi_G(x)$. Simple algebra shows that if the following equations hold

$$\begin{aligned} \phi_G(x) &= \upsilon\left\{\exp\left[\frac{\alpha}{\rho}\left(\phi_S(x)\right)^{\beta}\right]-1\right\}, \text{ or} \\ \phi_S(x) &= \left[\frac{\rho}{\alpha}\log\left(\frac{\phi_G(x)}{\upsilon}+1\right)\right]^{1/\beta}, \end{aligned} \tag{A.3}$$

then we have the same compositions of those functions:

$$\Phi_S\left(\phi_S(x)\right) = \Phi_G\left(\phi_G(x)\right).$$

The relationship between $\phi_G(x)$ and $\phi_S(x)$ as in (A.3) can hardly be reconciled as identification up to location/scale/power transformation. Even if we restrict the attention to compound Poisson process, an observationally equivalent pair of structural components could still be constructed if the distribution of an increment W is unrestricted. Recall that a compound Poisson process has its Lévy-Laplace exponent function as

$$\Phi_{CP}\left(\phi(x)\right) = \lambda\left[\mathcal{L}_W\left(\phi(x)\right)-1\right].$$

Varying the distribution or Laplace tranform of the increment variable W leads to a rich family of choices with tractable formulas.

When it comes to the bivariate setting, the same strategy would apply. For $j = 1, 2$, we set $\Phi_{S,j}(u) = \alpha_j u^{\beta_j}$, $\Phi_{G,j}(u) = \rho_j \log\left(\frac{u+\upsilon_j}{\upsilon_j}\right)$, then the following choice

$$\phi_{G,j}(x) = \upsilon_j\left\{\exp\left[\frac{\alpha_j}{\rho_j}\left(\phi_{S,j}(x)\right)^{\beta_j}\right]-1\right\}, \tag{A.4}$$

would generate the same marginal information via

$$\Phi_{S,j}\left(\phi_{S,j}(x)\right) = \Phi_{G,j}\left(\phi_{G,j}(x)\right).$$

Considering the empirical content of the joint Lévy-Laplace exponent function, $\Phi_{S,12}$ is determined for a given Lévy copula $\mathcal{C}_{L.S}$ for the stable process (this copula is known to be homogeneous of order 1, see Kallsen and Tankov, 2006). Denote the inverse transformation in (A.4) $\mathcal{T}$ via $\phi_{S,j} = \mathcal{T} \circ \phi_{G,j}$, then we could find a bivariate function $\Phi_{G,12}$ from the

restriction $\Phi_{G,12}\left(\phi_{G,1}\left(x\right),\phi_{G,2}\left(x\right)\right)=\Phi_{S,12}\left(\phi_{S,1}\left(x\right),\phi_{S,2}\left(x\right)\right)$ by noting that

$$\begin{aligned}&\Phi_{G,12}\left(u,v\right)=\Phi_{S,12}\left(\mathcal{T}\circ\phi_{G,1}\circ\phi_{G,1}^{-1}\left(u\right),\mathcal{T}\circ\phi_{G,2}\circ\phi_{G,2}^{-1}\left(v\right)\right)\\=\;&\Phi_{S,12}\left(\mathcal{T}\left(u\right),\mathcal{T}\left(v\right)\right),\end{aligned}$$

if $\phi_{G,j}\left(x\right)$ are assumed to beinvertible for $j=1,2$. With the constructed bivariate Levy exponent function $\Phi_{G,12}$, a Levy copula function $\mathcal{C}_{G.S}$ exists from Sklar Theorem in Kallsen and Tankov (2006). But $\mathcal{C}_{G.S}$ is not necessarily homogeneous of order 1, unlike $\mathcal{C}_{L.S}$.

In sum, we have demonstrated the possibility of getting observationally equivalent structural components $\left(\phi_{S,j},\pi_{S,j},\mathcal{C}_{L.S}\right)$, $\left(\phi_{G,j},\pi_{G,j},\mathcal{C}_{G.S}\right)$ with distinct latent processes, various shapes of multiplicative effects, and different dependence structure.

2.8 Appendix E. A Review of Lévy Processes

Lévy process, a process with independent and stationary increments, is the natural generalization of random walk in continuous time and it encompasses many well studied ones, such as Brownian motion, (compound) Poisson process, gamma process, inverse Gaussian process. The distribution of Lévy process at a fixed time coincides with the infinitely divisible law, which also arises as the most general limit law in the classical limit theorems for null arrays. Recall that random vector ξ in $\mathcal{R}^d$ is said to be infinitely divisible, iff. $\sum_j \xi_{nj}\to^d \xi$ for some $i.i.d$ array $\left\{\xi_{nj}\right\}$, further iff. $\xi=^d L\left(1\right)$, where $L\left(t\right)$ is a Lévy process in $\mathcal{R}^d$, see Sato (1999). Two sub-families of the general Lévy process are of particular interest. One is the so-called spectral negative Lévy process used in Abbring (2012), which cannot have any positive jump. While the other is the Lévy subordinator, which has non-decreasing sample path. The Gaussian component would necessarily be ruled out in a Lévy subordinator. Below I collect some characterization and terminology about Levy process, and refer the interested readers to Sato (1999) and Cont and Tankov (2003) for further information.

Despite the richness of Lévy class, it admits a concise characterization via the celebrated Lévy-Khintchine representation. A $d-$dimensional Lévy process $\mathbf{L}\left(t\right)$ is completely characterized by the triplet $\left(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Pi}\right)$ through it characteristic function

$$E\left\{\exp\left[\mathbf{i}z\mathbf{L}\left(t\right)\right]\right\}=\exp\left[t\boldsymbol{\Psi}\left(z\right)\right],z\in\mathcal{R}^d,$$

with

$$\mathbf{\Psi}(z) = \iota \langle \boldsymbol{\mu}, z \rangle - \frac{1}{2} z' \mathbf{\Sigma} z + \int_{\mathcal{R}^d} \left[e^{\mathbf{i}\langle y,z \rangle} - 1 - \mathbf{i} \langle y, z \rangle 1_{\{|y| \leq 1\}} \right] \mathbf{\Pi}(dy),$$

where $\mathbf{i} = \sqrt{-1}$, $\mathbf{\Pi}$ is the so-called Lévy measure.

Definition 2.8.1 *A positive Radon measure* $\mathbf{\Pi}$ *on* $\mathcal{R}^d \backslash \{0\}$ *is called Lévy measure if it satisfies the following condition:*

$$\int_{\mathcal{R}^d} \left(1 \wedge |y|^2 \right) \mathbf{\Pi}(dy) < \infty.$$

Recall that a Lévy subordinator $\mathbf{L}(t)$ has almost surely nondecreasing sample path, i.e., for $t \geq s$ one has $\mathbf{L}(t) \geq \mathbf{L}(s)$. The following lemma gives an equivalent characterization of the subordinator in terms of the triplet $(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Pi})$.

Proposition 2.8.2 $\mathbf{L}(t)$ *is a Lévy subordinator iff.* $\boldsymbol{\mu} \geq \mathbf{0}, \mathbf{\Sigma} = \mathbf{0}$,*and* $\mathbf{\Pi}$ *assigns zero measure outside the positive quadrant with* $\int_{\mathcal{R}_+^d} (1 \wedge y) \mathbf{\Pi}(dy) < \infty$.

The restriction $\int_{\mathcal{R}_+^d} (1 \wedge y) \mathbf{\Pi}(dy) < \infty$ arises because the nondecreasing process has bounded variation. We could simplify its Lévy-Laplace exponent function in this case a bit:

$$\mathbf{\Psi}(z) = \mathbf{i} \left\langle \boldsymbol{\mu}', z \right\rangle + \int_{\mathcal{R}^d} \left[e^{\mathbf{i}\langle y,z \rangle} - 1 \right] \mathbf{\Pi}(dy),$$

where

$$\boldsymbol{\mu}' = \boldsymbol{\mu} - \int_{|y| \leq 1} y \mathbf{\Pi}(dy).$$

In this paper we restrict our attention to the case where $\boldsymbol{\mu}' = 0$ without any deterministic linear trend. Otherwise, we need to take second order derivatives to identify $\mathbf{\Pi}$ and $\boldsymbol{\mu}$. By Theorem 4.1 in Cont and Tankov, there is no linear trend for each marginal Lévy process either.

The next two lemmas are about the equivalence between moment conditions for random variables $\mathbf{L}(t)$ and moment conditions for the corresponding Lévy measure. For the absolute moment conditions, see Corollary 25.8 in Sato (1999).

Proposition 2.8.3 *(Theorem 25.17 Sato, 1999)* $\forall$ *non-negative* u*, the exponential moment* $E\{\exp[-u\mathbf{L}(t)]\}$ *is finite for some* t,*iff. it is finite for all* t*, iff.*

$$\int_{|y| \geq 1} e^{-\langle y,u \rangle} \mathbf{\Pi}(dy) < \infty.$$

In this situation, we have

$$E\left\{\exp\left[-u\mathbf{L}\left(t\right)\right]\right\}=\exp\left[t\mathbf{\Psi}\left(\mathbf{i}u\right)\right].$$

Proposition 2.8.4 *(Wolfe, 1971) The existence of moment for marginal distribution* $\mathbf{L}\left(t\right)$ *is equivalent as the existence of moment for the Lévy measure* $\mathbf{\Pi}$*.*

For most Lévy processes, the jump intensity would grow or explode as the jump size converges to zero. The resulting implication is that the *survival function* version of Lévy measure is more tractable than its *distribution function* version. Such *survival function* attached to the Lévy measure is the so-called tail integral which we define next.

Definition 2.8.5 *(Tail Integral) A two-dimensional tail integral is a function* $\overline{\Pi}:\mathcal{R}_{+}^{2}\rightarrow\mathcal{R}_{+}$ *s.t.*

(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.*

The probabilistic interpretation of the tail integral measure $\overline{\Pi}\left(x\right)$ is the following. Let $N(t,x)$ denote the number of jumps of magnitude $\geq x>0$ during the time interval $[0,t]$. Then $N(t,x)$ is a Poisson process with parameter $\overline{\Pi}\left(x\right)t$.

Definition 2.8.6 *(Lévy Copula) A two dimensional Lévy copula for Lévy subordinators is a* $2-$*increasing grounded function* $\mathcal{C}_{L}\left(u,v\right):\left[0,\infty\right]^{2}\rightarrow\left[0,\infty\right]$ *with uniform margins, i.e.*

$$\mathcal{C}_{L}\left(u,\infty\right)=\mathcal{C}_{L}\left(\infty,u\right)=u.$$

Theorem 2.8.7 *(Sklar Theorem) Let* $\overline{\Pi}$ *be a two-dimensional tail integral with margins* $\overline{\Pi}_{1},\overline{\Pi}_{2}$*, then there exists a positive Levy copula* $\mathcal{C}_{L}$ *s.t.*

$$\overline{\Pi}\left(u,v\right)=\mathcal{C}_{L}\left(\overline{\Pi}_{1}\left(u\right),\overline{\Pi}_{2}\left(v\right)\right), \tag{E.1}$$

if $\overline{\Pi}_{1},\overline{\Pi}_{2}$ *are continuous, the copula function* $\mathcal{C}_{L}$ *is unique. Conversely for a given Lévy copula* $\mathcal{C}_{L}$ *and two marginal tail integrals* $\overline{\Pi}_{1},\overline{\Pi}_{2}$*, (E.1) defines a two-dimensional tail integral.*

Recall the Clayton-Lévy copula takes the form $\mathcal{C}_{L,\alpha}(u,v) = (u^{-\alpha}+v^{-\alpha})^{-1/\alpha}$, with a single parameter $\alpha \in (0,\infty)$. A straightforward yet slightly tedious computation leads to the following expressions of its partial derivative w.r.t the arguments u or v:

$$\begin{aligned}\frac{\partial}{\partial u}\mathcal{C}_{L,\alpha}(u,v) &= u^{-\alpha-1}\left(u^{-\alpha}+v^{-\alpha}\right)^{-\frac{1}{\alpha}-1},\\ \frac{\partial}{\partial v}\mathcal{C}_{L,\alpha}(u,v) &= v^{-\alpha-1}\left(u^{-\alpha}+v^{-\alpha}\right)^{-\frac{1}{\alpha}-1},\end{aligned}$$

and its derivative w.r.t he parameter α:

$$\begin{aligned}&\frac{\partial}{\partial \alpha}\mathcal{C}_{L,\alpha}(u,v)\\ =\ &\left(u^{-\alpha}+v^{-\alpha}\right)^{-\frac{1}{\alpha}-1}\left[\frac{\log\left(u^{-\alpha}+v^{-\alpha}\right)}{\alpha^2}-\frac{u^{-\alpha}\log\frac{1}{u}+v^{-\alpha}\log\frac{1}{v}}{\alpha\left(u^{-\alpha}+v^{-\alpha}\right)}\right].\end{aligned}$$

When the marginal Lévy measure has a density function $\pi(\cdot)$ with respect to the Lebesgue measure and the Lévy copula function is differentiable, the joint Lévy-Laplace exponent function admits the following alternative expression:

$$\Phi_{12}(z_1,z_2) = \iint_{\mathcal{R}_+^2}\left[1-e^{-z_1y_1-z_2y_2}\right]\frac{\partial^2}{\partial u\partial v}\mathcal{C}_L(\cdot,\cdot)\,|_{\substack{u=\overline{\Pi}_1(y_1)\\ v=\overline{\Pi}_2(y_2)}}\pi_1(y_1)\,\pi_2(y_2)\,dy_1dy_2,$$

Next we list some commonly used univariate Lévy densities, from Cont and Tankov (2003), Gjessing, Aalen and Hjort (2003).

Example 2.8.8 *The stable process has a power parameter $\rho \in (0,1)$ and stable index $\upsilon \in (0,2]$, the marginal Lévy density is*

$$\pi_S(y) = \frac{\upsilon\rho}{\Gamma(1-\rho)\,y^{1+\rho}},$$

and the Lévy exponent function

$$\Phi_S(u) = \upsilon u^{\rho}.$$

Example 2.8.9 *The gamma process has a shape parameter ρ and scale parameter υ, the marginal Lévy density is*

$$\pi_G(y) = \frac{\rho e^{-\upsilon y}}{y},$$

and the Lévy exponent function

$$\Phi_G(u) = \rho\log\left(1+\frac{u}{\upsilon}\right).$$

Example 2.8.10 *The power variance function process is a general process nesting the above gamma process as a special case, the marginal Lévy density is*

$$\pi_S(y) = \frac{n\rho\upsilon^n}{\Gamma(n+1)} y^{n-1} e^{-\upsilon y},$$

with $\upsilon > 0$, $n > -1$ *and* $n\rho > 0$ *and the Lévy exponent function is*

$$\Phi_{PVF}(u) = \rho\left\{1 - \left(1 + \frac{\upsilon}{u}\right)^n\right\}.$$

The gamma process arises as a borderline case when $n = 0$.

Finally we would like to mention the compound Poissson process $\mathbf{L}(t) = \sum_{i \le N(t)} W_i$, because it is the only one having piece-wise constant sample path (almost surely) and its correponding Lévy measure is finite, i.e. $\int_{\mathcal{R}^d} \mathbf{\Pi}(dy) < \infty$. In addition, when W_i has a density function $\mathbf{f}(w)$ and the arrival rate of $N(t)$ is λ, the Lévy measure has a density function $\boldsymbol{\pi}(w) = \lambda \mathbf{f}(w)$ and the Lévy exponent function simplifies to

$$\mathbf{\Psi}_{CP}(z) = \lambda \int_{\mathcal{R}^d} \left[e^{\mathbf{i}\langle w,z\rangle} - 1\right] \mathbf{f}(w)\, dw.$$

Table 2.1: Comparisons of Three Models

	MHT	MPH	EMO
Structural Component			
Threshold	Unrestricted	Unit Exponential	Unit Exponential
Latent Process	S.N. Lévy	Cum. Baseline Hazard	Lévy Subordinator
Sample Path	No Positive Jump	Increasing	Increasing
Heterogeneity	Time-varying	Static	Time-varying
Univariate Case			
Identification	Yes	Yes	Yes
Semiparametric Est.	N.A.	Yes	Yes
Competing Risks			
Identification	N.A.	Yes	Yes
Semiparametric Est.	N.A.	N.A.	Yes

Chapter 3

AN EMPIRICAL APPLICATION & FUTURE WORK

3.1 Joint Retirement of Couples

Understanding the retirement mechanism helps to guide optimal design of employer-provided and government benefit programs (Blau, 1998; Blau and Gilleskie, 2006; Gustman and Steinmeier, 2004). Our focus on the married couple stems from the fact that most people approaching retirement age are married and a significant portion of them choose to retire at the same time. Empirical studies documenting the joint retirement of couples abound based on different datasets (Honore and de Paula, 2014). Standard hazard-based bivariate models (or the existing process-based univariate modeling strategy imposing independence between two durations) do not apply in this scenario, because the simultaneous failure is ruled out from the very beginning in those models. The retirement of couples is no doubt a joint decision process, but the individual's response to the spousal retirement could be ambiguous, theoretically speaking. Indeed competing hypotheses exist about whether the person would retire earlier if the spouse does. On one hand, the person could postpone retirement to counter-act the decrease of total family income due to the spousal withdrawal from the labor market. Furthermore, the increased time at home of the retired spouse may lower the opportunity cost of work for the other spouse if there is substitution in home production (Coile, 2004). On the other hand, the simultaneous retirement may result in complementarities in leisure time, or if the retired spouse requires assistance with acitivities of daily living, see An, Christensen, and Gupta (2004). In this application, the Lévy subordinator with its increasing sample path is more suitable to represent the latent aging process or the degradation of health condition of the elderly, as opposed to the spectrally negative Lévy process in Abbring (2012) which fluctuates up and down. There is no positive jump in the spectrally negative Lévy process by definition, but unexpected health shocks,

such as heart attacks or new cancer diagnoses, are quite common for people near retirement (Coile, 2004). The machinery developed in this paper provides a nice complementary tool compared with existing methods that have been used. The literature on the structural estimation of a joint retirement decision model is entirely parametric and is silent about the identifiability. While in those reduced form estimation, it is a common practice to treat the spouse's retirement date as a time-varying regressor.

Similar to Honore and de Paula (2014), we estimated the model using eight waves of the Health and Retirement Study (every two years from 1992 to 2006) and kept households where at least one individual was 60 years old or more.[1] To avoid redundancy, we refer the readers to Honore and de Paula (2014) for a detailed discussion on the descriptive statistics. The retirement duration was observed at a monthly frequency. There were 1,284 couples in total. Even though we did observe bivariate durations for some couples, the partial likelihood in Section 2.4.3 is very convenient in the presence of independent censoring. In this sample, over 40% of the observations were subject to additional censoring, meaning that when the survey was conducted, either one of the family members had not retired yet. This explains why we carried out the estimation as if we were given with competing risks data. Two continuous covariates are the total health expenditure per individual[2] X_1 (inflation adjusted using the CPI to Jan/2000 dollars) and the financial wealth[3] X_2 (inflation adjusted using the CPI to Jan/2000 dollars). Both covariates are standardized into the unit interval based on the empirical probablity integral transformation. We compare estimates among different groups of people with their own and spouse's self-reported health condition

[1] The selection critiria is the same as in Honore and de Paula (2014). We classify as retired a respondent who is not working and not looking for work and one for whom there is any mention of retirement through the employment status or the questions that ask the respondent whether he or she considers him- or herself to be retired. To avoid left-censoring, selected households also had both partners working at the initial period. Right-censoring occurs when someone dies or has his or her last interview before the end of the survey. We excluded individuals who were part of the military. Finally, we exclude households with multiple spouses and/or couples throughout the period of analysis, couples with conflicting information over marital status or other joint variables, and couples of the same gender.

[2] This is the total health expenditure per individual in the previous 12 months for the first two waves and the previous 2 years for the subsequent years

[3] This measure includes the value of checking and savings accounts, stocks, mutual funds, investment trusts, CDs, government bonds, Treasury bills and all other savings minus the value of debts such as credit card balances, life insurance policy loans or loans from relatives. It does not include housing wealth or private pension holdings.

(very good, good, or poor), with or without health insurance, with or without pension. Here we adopted the methodology in Li and Racine (2007) which involves smoothing over category data. It would induce some bias but comes with a much smaller variance especially when different categories are not well balanced.

Insert Table 3.1 Here

Table 3.1[4] reports estimates of the index model using partial likelihood estimation. We normalize the direction by fixing the first component to be 1 (the effect of health spending). Consistent with earlier findings, greater wealth induces people to retire earlier across all the considerations. It is particularly interesting to note that when the husband is not in good health, the wife is more willing to sacrifice and retire earlier, but not the inverse.

3.2 Conclusion

This dissertation contains a systematic study of duration models characterized by threshold-crossing behavior of latent Lévy subordinators. With its distributional property and non-decreasing sample path, the Lévy subordinator could be seen as the natural variant of the static heterogeneity terms in MPH from a process point of view, and it resembles the notion of usage and wear-out effect (Singpurwalla, 1995) closely. Hence, besides those optimal stopping time problems in economics, our model also applies to the scenario concerning the life/failure time in biology, medical science, and engineering, where the traditional hazard-based models dominate. Furthermore, it is transparent how the deep structural parameters of theoretical models can be related to reduced-form parameters, and all model primitives could be point identified for various semiparametric specifications. In spite of the sophisticated construction, our models exhibit very tractable mathematical structures, inducing flexible semiparametric estimation procedures with rigorously established large sample properties.

In the univariate case, our modelling strategy leads to a novel presentation of the single-index Cox proportional hazard model. It is proved that the maximum sieve partial likelihood

[4]The standard errors are computed by Bayesian bootstrap with 100 replications as in Cheng and Huang (2010). It's widely known the standard Efron's bootstrap tend to be unstable when there is censoring.

estimator for the finite dimensional parameter is asymptotically normal with root-n convergence rate and it achieves the semiparametric efficiency bound. When it comes to the competing risks model, we have utilized the Levy copula to generate dependence between two durations without making any parametric assumptions. With semiparametric restrictions on the structural multiplicative effects from observable covariates, we identify these effects, the baseline hazard function, and the characteristics of latent Lévy subordinators separately in a competing risks setting where one does not observe both durations sequentially. For the index model, I present two root-n consistent and asymptotically normal estimators for the finite dimensional parameters in structural multiplicative effects based on certain average derivatives estimation. Interestingly, the model generates equal durations with positive probability, motivated from an empirical application on the joint retirement decisions of married couples. The latter feature is ruled out from the very beginning in the traditional duration or survival analysis.

Numerious extensions are possible. First, one could incorporate endogenous and time-varying covariates by modeling additional stochastic processes explicitly. Renault, van den Heijden, and Werker (2014) present a structural model for transaction time (duration) and asset prices (associated marks) driven by multivariate Brownian motion. In their model, successive passage times of one latent Gaussian component relative to random boundaries define durations and the other correlated components generate the marks. It is both interesting and challenging to search for other processes which lead to tractable structure. Second, it is desirable to relax the parametric assumption on the threshold variable and to consider a more general Lévy process, i.e., the spectral positive Lévy process (Kluppelberg, Kyprianou, and Maller, 2004; Boyarchenko and Levendorskii, 2012). Last but not least, it is worthwhile to study the empirical content of stochastic game models driven by a Lévy process; see Chapter 11 in Kyprianou (2006).

Table 3.1: Estimates of Joint Retirement Model

Est.(S.E.)	Health Condition			Health Insurance		Pension (DB)	
Own	V Good	Good	Poor	Yes	No	Yes	No
Husband	0.31 (0.070)	0.55 (0.086)	1.14 (0.113)	0.48 (0.036)	0.53 (0.072)	0.95 (0.076)	0.78 (0.035)
Wife	0.67 (0.058)	0.82 (0.079)	1.71 (0.137)	0.56 (0.039)	0.61 (0.085)	0.73 (0.064)	0.59 (0.041)
Spouse	V Good	Good	Poor	Yes	No	Yes	No
Husband	0.44 (0.084)	0.32 (0.095)	0.10 (0.108)	0.51 (0.040)	0.65 (0.081)	0.83 (0.072)	0.82 (0.035)
Wife	0.83 (0.064)	0.89 (0.083)	1.91 (0.121)	0.58 (0.046)	0.79 (0.094)	0.70 (0.066)	0.63 (0.042)

BIBLIOGRAPHY

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

[2] Abbring, J.H. (2012), Mixed hitting-time models, *Econometrica,* 80, 783–819.

[3] 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 B,* 65 701-710.

[4] Abbring, J. H. and G.J. van den Berg (2007), The unobserved heterogeneity distribution in duration analysis, *Biometrika,* 94, 87-99.

[5] Abbring, J.H. and T. Salimans (2012), The likelihood of mixed hitting times, working paper.

[6] Ai, C. and X. Chen (2003), Efficient estimation of models with conditional moment restrictions containing unknown functions, *Econometrica,* 71, 1795-1843.

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

[8] Andersen, P.K. and R.D. Gill (1982), Cox's regression model for counting processes: a large sample study, *The Annals of Statistics,* 10, 1100-1120.

[9] Arnold, B.C. (1968), Parameter estimation for a multivariate exponential distribution, *Journal of the American Statistical Association,* 63, 848-852.

[10] Basu, A.P. and J.K. Ghosh (1978), Identifiability of the Multinomial and other distributions under competing risks model, *Journal of Multivariate Analysis,* 8, 413-429.

[11] Basu, A.P. and J.K. Ghosh (1983), Identifiability results for a k-out-of-p system, *Communications in Statistics-Theory and Methods,* 12, 199-205.

[12] Bearse, P., Canals-Cerda, J., and P. Rilstone (2007), Efficient semiparametric estimation of duration models with unobserved heterogeneity, *Econometric Theory,* 23, 281-308.

[13] 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.

[14] Begun, J.M., Hall, W.J., Huang, W.M., and Wellner J.A. (1983), Information and asymptotic efficiency in parametric-nonparametric models, *The Annals of Statistics* 11, 432-452.

[15] Bickel, P.J., Ritov, Y., Klaassen, J., and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins University Press.

[16] Blau, D. (1998), Labor force dynamics of older married couples, *Journal of Labor Economics,* 16, 595-629.

[17] Blau, D. and D. Gilleskie (2004), Health insurance and retirement of married couples, *Journal of Applied Econometrics,* 21, 935–953.

[18] Botosaru, I. (2013), A duration model with dynamic unobserved heterogeneity, working paper.

[19] Boyarchenko, S. and S. Levendorskii (2012), Preemption games under Lévy uncertainty, working paper.

[20] Botosaru, I. (2013), A duration model with dynamic unobserved heterogeneity, working paper.

[21] Breslow, N.E. (1972). Contribution to the discussion of "Regression models and life-tables" by D. R. Cox. *Journal of the Royal Statistical Society, Series B* 34, 216-217.

[22] Chen, C.-H., Li, K.-C. and J.-L. Wang (1999), Dimension reduction for censored regression data, *The Annals of Statistics,* 27, 1-23.

[23] Chen, K., Guo, S., Sun L., and J.-L. Wang (2010), Global partial likelihood for nonparametric proportional hazards models, *Journal of the American Statistical Association,* 105, 750-760.

[24] Chen, S. and Zhou, L. (2007). Local partial likelihood estimation in proportional hazards regression. *The Annals of Statistics,* 888-916.

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

[26] Chen, Y.-H. (2010), Semiparametric marginal regression analysis for dependent competing risks under an assumed copula, *Journal of the Royal Statistical Society B,* 72 235-251.

[27] Cheng, G. and J.Z. Huang (2010), Bootstrap consistency for general semiparametric M-estimation, *The Annals of Statistics,* 38, 2884-2915.

[28] Coile, C. (2004), Health shocks and couples' labor supply decisions, NBER working paper 10810.

[29] Cox, D.R. (1959), The analysis of exponentially distributed life-times with two types of failure, *Journal of the Royal Statistical Society B,* 21, 411–421.

[30] Cox, D.R. (1972). Regression models and life-table (with discussion). *Journal of the Royal Statistical Society, Series B* 34, 187-220.

[31] Cox, D.R. (1975). Partial likelihood. *Biometrika* 62, 269-276.

[32] Cont, R. and P. Tankov (2004), *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, Boca Raton, FL.

[33] Crowder, M. (1991), On the identifiability crisis in competing risk analysis, *Scandinavian Journal of Statistics* 18, 223-233.

[34] Dabrowska, D.M. (1997). Smoothed Cox regression. *The Annals of Statistics* 25, 1510-1540.

[35] de la Pena, V.H. and E. Gine (1999), *Decoupling: From Dependence to Independence.* Springer.

[36] Deng, Y., Quigley, J.M., and R. Van Order (2000), Mortgage terminations, heterogeneity and the exercise of mortgage options, *Econometrica,* 68, 275-307.

[37] Ding, Y. and Nan, B. (2011). A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *The Annals of Statistics* 39, 3032–3061.

[38] Dixit, A.K. (1989), Entry and exit decisions under uncertainty, *Journal of Political Economy,* 97, 620–638.

[39] Dony, J., Einmahl, U., and D.M. Mason (2006), Uniform in bandwidth consistency of local polynomial regression function estimators, *Austrian Journal of Statistics,* 35, 105-120.

[40] Einmahl, U. and D.M. Mason (2005), Uniform in bandwidth consistency of kernel-type function estimators, *The Annals of Statistics,* 33, 1380-1403.

[41] Elbers, C. and G. Ridder (1982), True and spurious duration dependence: The identifiability of the proportional hazard model, *Review of Economic Studies,* 49, 403-409.

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

[43] Esary, J.D., Marshall, A.W., and F. Proschan (1973), Shock models and wear processes, *The Annals of Probability,* 1, 627-649.

[44] Fan, J. and I. Gijbels, (1996), *Local Polynomial Modeling and Its Applications,* Chapman & Hall, London.

[45] Fan, J., Gijbels, I., and M. King (1997), Local likelihood and local partial likelihood in hazard regression, *The Annals of Statistics,* 25, 1661-1690.

[46] Fan, J., Lin, H., and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *The Annals of Statistics,* 34, 290-325.

[47] Fan, Y. and R. Liu (2013), Partial identification and inference in censored quantile regression: a sensitivity analysis, working paper.

[48] Frisch, R. and Waugh, F.V. (1933). Partial time regressions as compared with individual trends. *Econometrica,* 387-401.

[49] Fermanian, J.-D. (2003), Nonparametric estimation of competing risks models with covariates, *Journal of Multivariate Analysis,* 85, 156–191.

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

[51] Gine, E. and D.M. Mason (2007), On local U-statistic processes and the estimation of densities of functions of several sample variables, *The Annals of Statistics,* 35, 1105-1145.

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

[53] Gustman, A. and T. Steinmeier (2004), Social security, pensions and retirement behaviour with the family, *Journal of Applied Econometrics,* 19, 723-737.

[54] Hahn, J. (1994), The efficiency bound of the mixed proportional hazard model, *Review of Economic Studies,* 61, 607-629.

[55] Hardle, W. and T.M. Stoker (1989), Investigating smooth multiple regression by the method of average derivatives, *Journal of The American Statistical Association,* 84, 986–995.

[56] Hausman, J.A. and T. Woutersen (2014), Estimating a semi-parametric duration model without specifying heterogeneity, *Journal of Econometrics,* 178, 114–131.

[57] Heckman, J.J. and B. Honore (1989), The identifiability of competing risks model, *Biometrika,* 76, 325-330.

[58] Heckman, J.J. and B. Honore (1990), The empirical content of the Roy model, *Econometrica,* 58, 1121-1149.

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

[60] Heckman, J.J. and B. Singer (1984), The identifiability of the proportional hazard model, *Review of Economic Studies,* 51, 231-241.

[61] Honore, B. and A. de Paula (2010), Interdependent durations, *Review of Economic Studies,* 77, 1138-1163.

[62] Honore, B. and A. de Paula (2014), Interdependent durations in joint retirement, working paper.

[63] Honore, B., and A. Lleras-Muney (2006), Bounds in competing risks models and the war on cancer, *Econometrica,* 74, 1675-1698.

[64] Hougaard, P. (1986a), Survival models for heterogeneous populations derived from stable distributions, *Biometrika,* 73, 387-396.

[65] Hougaard, P. (1986b), A class of multivanate failure time distributions, *Biometrika,* 73, 671-678.

[66] Horowitz, J.L. and E. Mammen (2007), Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions, *The Annals of Statistics,* 35, 2589-2619.

[67] Hristache, M., Juditsky, A., Polzehl, J., and V. Spokoiny (2001), Structure adaptive approach for dimension reduction, *The Annals of Statistics,* 29, 1537-1566.

[68] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics* 24 540–568.

[69] Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics* 27 1536–1563.

[70] Huang, J. and Wellner, J.A. (1997). Interval censored survival data: A review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis Lecture Notes in Statistics* 123 123–169. Springer: New York.

[71] Huang, J. Z., Kooperberg, C., Stone, C.J., and Truong, Y.K. (2000). Functional ANOVA modeling for proportional hazards regression. *The Annals of Statistics,*, 961-999.

[72] Huang, J. Z. and Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics,* 62, 793-802.

[73] Ichimura, H. (1993), Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics,* 58, 71-120.

[74] Ichimura, H. and L.F. Lee (1991), Semiparametric least squares estimation of multiple index models: single equation estimation, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, edited by William A. Barnett, James Powell, George E. Tauchen.

[75] Kallsen, J. and P. Tankov, (2006), Characterization of dependence of multidimensional Levy processes using Levy copulas, *Journal of Multivariate Analysis,* 97, 1551–1572.

[76] Katz, L.E. and B.D. Meyer (1990), Unemployment insurance, recall expectations, and unemployment outcomes, *Quarterly Journal of Economics,* 105, 973-1002.

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

[78] Klein, J.P., Keiding, N., and C. Kamby (1989), Semiparametric Marshall-Olkin models applied to the occurrence of metastases at multiple sites after breast cancer, *Biometrics*, 45, 1073-1086.

[79] Kluppelberg, C., Kyprianou, A.E., and R.A. Maller (2004), Ruin probabilities and overshoots for general Lévy insurance risk processes, *Annals of Applied Probability,* 14, 1766-1801.

[80] Kochar, S.C. and F. Proschan (1991), Independence of time and cause of failure in the multiple dependent competing risks model, *Statistica Sinica,* 1, 295-299.

[81] Kong, S. and Nan, B. (2014). Non-asymptotic oracle inequalities for the high-dimensional Cox regression via Lasso. *Statistica Sinica,* 24, 25-42.

[82] Kyprianou A.E. (2006), *Introductory Lectures on Fluctuations of Levy Processes with Applications,* Springer.

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

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

[85] Lancaster, T. (1990), *Econometric Analysis of Transition Data,* Cambridge University Press.

[86] Lawless, J. and M. Crowder (2004), Covariates and random effects in a gamma process model with application to degradation and failure, *Lifetime Data Analysis*, 10, 213–227.

[87] Lee, M.-L.T. and G.A. Whitmore (2006), Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary, *Statistical Science*, 21, 501–513.

[88] Lee, S. and A. Lewbel (2013), Nonparametric identification of accelerated failure time competing risks model, *Econometric Theory,* 29, 905-919.

[89] Li, Q., Lu, X. and A. Ullah (2003), Multivariate local polynomial regression for estimating average derivatives, *Journal of Nonparametric Statistics,* 15, 607-624.

[90] Li, Q. and J.S. Racine (2007), *Nonparametric Econometrics: Theory and Practice,* Princeton Press.

[91] Lin, W. and K.B. Kulasekera (2007), Identifiability of single-index models and additive-index models, *Biometrika,* 94, 496-501.

[92] Mai, J.-F. and M. Scherer (2009), Lévy-frailty Copulas, *Journal Multivariate Analysis,* 100, 1567-1585.

[93] Major, P. (2006), An estimate on the supremum of a nice class of stochastic integrals and U-statistics, *Probability Theory and Related Fields,* 134, 489-537.

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

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

[96] McDonald, R. And D. Siegel (1986), The value of waiting to invest, *Quarterly Journal of Economics,* 101, 707–728.

[97] Mortensen, D.T. and C.A. Pissarides (1994), Job creation and job destruction in the theory of unemployment, *Review of Economic Studies*, 61, 397–415.

[98] Muliere, P. and M. Scarsini (1987), Characterization of a Marshall-Olkin type class of distributions, *Annals of the Institute of Statistical Mathematics,* 39, 429-441.

[99] Murphy, S.A. (1995), Asymptotic theory for the frailty model, *The Annals of Statistics,* 23, 182–198.

[100] Nan, B. and Wellner, J.A. (2013), A general semiparametric Z-estimation approach for case-cohort studies. *Statistica Sinica* 23, 1155-1180.

[101] Nelson, R.B. (2006), *An Introduction to Copulas*, 2nd Edition, Springer.

[102] Newey, W.K. and Stoker, T.S. (1993), Efficiency of weighted average derivative estimators and index models. *Econometrica* 61, 1199-1223.

[103] Parner, E. (1998), Asymptotic theory for the correlated gamma-frailty model, *The Annals of Statistics,* 26, 183-214.

[104] Peterson, A.V. (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.

[105] Powell, J., Stock, J., and T.M. Stoker (1989), Semiparametric estimation of index coefficients, *Econometrica,* 57, 1403-1430.

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

[107] Ridder, G. (1990), The non-parametric identification of generalized accelerated failure-time models, *Review of Economic Studies,* 57, 167-181.

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

[109] Rivest, L.P. and M.T. Wells (2001), A martingale approach to the copula-graphic estimator for the survival function under dependent censoring, *Journal of Multivariate Analysis,* 79, 138-155.

[110] Sasieni, P. (1992a). Information bounds for the conditional hazard ratio in a nested family of regression models. *Journal of the Royal Statistical Society, Series B* 54, 617-635.

[111] Sasieni, P. (1992b). Non-orthogonal projections and their application to calculating the information in a partly linear Cox model. *Scandinavian Journal of Statistics* 19, 215-233.

[112] Sato, K. (1999), *Levy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

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

[114] Schumaker, L.L. (1981). *Spline Functions: Basic Theory.* New York: Wiley.

[115] Singpurwalla, N.D. (1995), Survival in dynamic environments, *Statistical Science*, 10, 86–103.

[116] Singpurwalla, N.D. (2006), The hazard potential: introduction and overview, *Journal of the American Statistical Association,* 101, 1705-1717.

[117] Stoker, T.M. (1986), Consistent estimation of scaled coefficients, *Econometrica,* 54, 1461-1481.

[118] Sun, Y., Gilbert, P.B., and I.W. McKeague (2009), Proportional hazards models with continuous marks, *The Annals of Statistics,* 37, 394-426.

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

[120] Tsybakov, A. (2008), *Introduction to Nonparametric Estimation,* Springer, New York.

[121] van den Berg, G.J. (1997), Association measures for durations in bivariate hazard rate models, *Journal of Econometrics,* 79, 221-245.

[122] van den Berg, G.J. (2001), Duration models: specification, identification, and multiple durations, in *Handbook of Econometrics, Vol. 5*, ed. by J.J. Heckman and E. Leamer. Amsterdam: Elsevier Science, Chapter 55, 3381–3460.

[123] van de Geer, S. (1990), Estimating a regression function, *The Annals of Statistics,* 18, 907-924.

[124] van de Geer, S. (1995), Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes, *The Annals of Statistics,* 23, 1779-1801.

[125] Van der Vaart, A.W. and Wellner J.A. (1996), *Weak Convergence and Empirical Processes: With Application to Statistics.* Springer-Verlag: New York.

[126] Wang, W. (2004). Proportional hazards regression models with unknown link function and time-dependent covariates. *Statistica Sinica* 14, 885–905.

[127] Whitmore, G.A. (1979), An inverse Gaussian model for labour turnover, *Journal of the Royal Statistical Society A,* 142, 468–478.

[128] Widder, D.V. (1946), *The Laplace Transform,* Princeton University Press.

[129] Wolfe, S.J. (1971), On moments of infinitely divisible distribution functions, *The Annals of Mathematical Statistics,* 42, 2036-2043.

[130] Zheng, M. and J.P. Klein (1995), Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika,* 82, 127-138.