Essays in Multivariate Duration Models

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors

der Wirtschaftswissenschaften der Universität Mannheim

submitted by Georgios Effraimidis

May 2012

Dean: Professor Dr. Martin Peitz

Referee: Professor Dr. Gerard J. van den Berg

Co-Referee: Professor Dr. Enno Mammen

Day of the Defense: 19 July 2012

Acknowledgements

I would like to express how deeply indebted I am towards my supervisor, Gerard van den Berg who gave me his support, encouragement, and advice since I started working on my research five years ago. Gerard provided me with the foundation to become a researcher on duration analysis which is my true research passion. Above all, Gerard never forced me to work on a particular research project as he always let me free decide about my research directions. I am also grateful to him as when he moved from Vrije Universiteit in Amsterdam to University of Mannheim, he asked me joining him which I really appreciated. The kindness and guidance of Enno Mammen throughout parts of my research are definitely acknowledged. Talking with Enno Mammen about statistics is a unique experience as the feedback he provides during a meeting is equivalent to reading a chapter of statistics textbook. Moreover, I would like to express my gratitude to Geert Ridder for giving me the opportunity to visit twice the University of Southern California. I really enjoyed my discussions with him which were always challenging as well as stimulating. Geert is among the pioneers on the topic of duration analysis and he clearly changed my way of thinking about general problems that arise in econometrics.

Moreover, I would like to mention a few colleagues from the three places I have visited during my PhD. First, I would like to thank Betty Wu who is currently Assistant Professor at Yonsei University and helped me during the first steps of my PhD program in Amsterdam. Second, I would like to thank Eleonora Granziera, Martin Weidner, Carlos-Hernandez Mireles, Dimitrios Pipinis, and Vanessa Smith from University of Southern California for their hospitality during my two visits in Los Angeles. It was a great experience to meet and spend some time with them. Special thanks to Dimitris Pipinis who has been a great friend and we have been discussing topics as diverse as economics, political economy, and biostatistics. Finally, I would like to thank Petyo Bonev, Ina Drepper, and Pia Pinger from University of Mannheim with whom I had several inspirational discussions about my research.

Finally, and most importantly, I would like to deeply thank my mother and my father who have supported me in all my pursuits, made me what I am, and also my sister.

Abstract

Duration analysis, which is also known as survival analysis, is a core subject of applied statistics and econometrics. Application of duration analysis techniques can be found in actuarial science, demography, economics, finance, marketing, and many other scientific fields. In the univariate case, the tools of duration analysis are used for the study of the distribution of a certain duration variable which is possibly associated with a set of explanatory covariates. This variable measures the time to the occurrence of an event of interest such as transition from unemployment to employment, retirement time, onset of a disease, purchase of a product. The main difference between duration analysis and standard regression analysis is that sometimes the duration variable is right-censored, namely, the only available information we have is that its realization exceeds a certain value.

Multivariate duration analysis is the natural extension of the univariate analysis. In this set up, multiple duration variables, which specify the time to the occurrence of multiple events, are considered and their joint distribution is analyzed for describing the association among them. These variables can be either parallel or sequential. Parallel duration variables refer to cases in which the multiple duration variables are measured by using the same reference point of time. On the other hand, sequential duration variables refer to cases in which the measurement of each duration variable starts after the realization of some other duration variable. Death times of twins or the corresponding time of onset of several diseases are multivariate examples with parallel duration variables. On the other hand, unemployment duration and the subsequent employment duration is an example with sequential durations.

The current PhD dissertation deals with multivariate duration models. In particular, it consists of three independent essays on multivariate duration models. In the next three paragraphs, a synopsis of each essay is given.

The first essay, written jointly with Gerard J. van den Berg, considers bivariate frailty models in which the frailty terms enter multiplicatively on the corresponding hazard rates. The frailty terms capture unobserved or nonmeasurable characteristics that affect the duration outcomes. We assume that the joint distribution of the frailty terms is characterized by gamma marginals. In

particular, the gamma distribution is widely used in empirical analysis for modelling the distribution for the unobserved heterogeneity terms. Both analytical and graphical arguments have been developed in the past which rationalize this specific choice. First, the focus of the paper is on the concepts of negative quadrant dependence and positive quadrant dependence between the duration variables. Second, two measures of association between the duration variables are considered; the Pearson's correlation coefficient and the Kendall's tau. In particular, (sharp) bounds for these measures are derived and the necessary conditions are discussed discussion is provided about the conditions which should be satisfied so that the bounds are approached very well.

The second essay, written jointly with Carlos Hernandez Mireles and Gerard Tellis, is concerned with a new trivariate hazard rate model which can be applied to study the relationship among the timing of the corresponding events. The suggested model allows for three types of dependence among the timing of the underlying events: due to unobserved heterogeneity, lagged dependence, and due to causality. As shown in the paper, this model can be nonparametrically identified and as consequence the three different types of dependence are disentangled. The new model is adopted to study the endogenous relationship between the timing of three important events in the sales and prices of new products. Specifically, we investigate the causal relationship between the sales crash, price crash, and sales recovery. A sales crash is a significant and permanent cut in the sales of a new product. On the other hand, the sales recovery is a sales peak which is realized after the crash. Finally, the price crash is a deep and permanent reduction in the price of a new product.

The last essay deals with competing risks models which are very popular in the scientific field of duration analysis. Such models deal with cases in which we observe only the minimum duration among several multiple durations for each individual unit under study. The goal of this paper is the development of statistical properties of the cumulative incidence function. This function, which is common in the empirical practice, specifies the probability that a particular duration variable will be realized by a certain point of time and before the other duration variables. The proposed estimator is nonparametric, that is, no parametric assumptions are made regarding the data generating process. In addition, the estimator allows for Missing At Random observations. More precisely, for some observations we have information about the value of the minimum duration variable, but not information about which duration variable is the one smallest value of realization.

Association Measures in Bivariate Gamma Frailty Models

1 Introduction

Frailty models are widely applied for the description of the distribution of time that a subject spends in a certain state of interest. The main building block of these models is the hazard rate function that specifies the instantaneous rate at which the subject transits from the state of interest to some other state, given that the transition has not occurred yet. The hazard rate is expressed as a function of *i*) the elapsed duration in the state of interest, *ii*) observed characteristics, and *iii*) a frailty term which captures characteristics that are not measurable or observable due to the nature of the problem. In the bivariate frailty model, two duration variables are considered that may describe two parallel, possibly competing, durations of the same subject (e.g., duration until merger and duration until bankruptcy of a firm) or single durations of two subjects which belong to the same cluster (e.g., death times of twins). The bivariate frailty model has several applications in economics (An et al., 2004; Røed and Westlie, 2007; Andrén, 2005; Hujer et al., 2006; Tatsiramos, 2004; Benítez-Silva and Heiland, 2008), biostatistical studies (Clayton, 1978; Paik et al., 1994; Shih and Louis, 1995a; Vu and Knuiman, 2002; Zhong and Li, 2002; Vu, 2004; Xiang et al., 2007; Rondeau et al., 2007; Gorfine and Hsu, 2011) and many other fields such as actuarial sciences, demography and engineering.

The main feature of the bivariate frailty model is the dependence between the two duration variables which is caused by dependence between the two frailty terms that enter the underlying hazard rates. In this paper, we first study the notions of negative quadrant dependence and positive quadrant dependence for the joint distribution of the duration variables, if the frailty terms act multiplicatively on the hazard rates. Second, we turn our attention to the level of dependence between the duration variables by imposing the extra condition that the frailty terms are gamma distributed. To quantify the degree of dependence between the duration outcomes, we will employ two association measures, specifically the Pearson's correlation coefficient and the Kendall's tau. The former, which measures the strength of the linear relationship between two random variables, is commonly used in empirical analysis for statistical inference. On the other hand, the latter measures the strength of any monotonic relationship between two random variables and consequently it is characterized by the rank-invariant property. For both association measures we provide lower and upper bounds for the range of the corresponding values. We also compare our findings with the results of Van Den Berg (1997) who provides nonparametric bounds for these two measures.

Our choice to model the marginal distribution of the frailty terms as gamma is based on the literature of the conventional univariate frailty model. Abbring and Van Den Berg (2007) show that if the frailty term enters multiplicatively on the hazard rate, the distribution of the frailty term, among the subjects who are still in the state of interest, converges quite rapidly to a gamma distribution. The only weak requirement they impose is that the actual distribution should be regularly varying at zero with nonnegative exponent. Examples of continuous distributions with this property are the beta, exponential, uniform and in general all distributions with density function that has finite strictly positive limit at zero. For the case of exponent equal to zero all distributions, including finitely discrete distributions, with a positive mass at zero satisfy the property of regular variation. In addition, several authors have proposed graphical and numerical procedures to check for the adequacy of the assumption of gamma distribution for the frailty term. For example, Shih and Louis (1995b) apply a graphical method to test the assumption of gamma frailty by calculating the average of the posterior mean of the frailty given the observed data. Cui and Sun (2004) propose a supremum-type test statistic, whose asymptotic critical values are calculated by Monte Carlo simulation, and apply a numerical method as well as a graphical approach to test the validity of the gamma assumption.

As explained above, we focus on negative as well as positive dependence between the duration variables. In biostatistical applications, the duration variables under study are usually positively dependent as the corresponding hazard rates share same unobserved or nonmeasurable characteristics (e.g., environmental, genetic). On the other hand, in social sciences, besides a few cases where the underlying duration variables are positively dependent, there are also numerous examples where the data reveal negative dependence between the duration variables. In labor economics, for instance, consider an individual who is unemployed and faces two competing exits from the unemployment state: employment and dropping out of the labor force. If the unemployed individual is strongly motivated (which is not observed) to get a job, then the duration until employment will be negatively associated with the duration until nonparticipation in the labor force, as high motivation to work mainly results in selection into the state of employment.

The results of this paper are useful for practitioners and researchers who work with the bivariate frailty model in which the frailty terms enter multiplicatively on the hazard rates. First of all, we provide a general result about the dependence structure of the bivariate distribution of the duration variables given the bivariate dependence of the frailty terms. Notice that for the derivation of this result we do not impose any restriction on the shape of the marginal distribution of the frailty terms. Second, by calculating bounds for the Pearson's correlation coefficient we can get insights about the flexibility of the bivariate gamma frailty model with Weibull baseline hazards. Moreover, the results on the bounds for the Kendall's tau are more general concerning the flexibility of the bivariate gamma frailty model as we do not make use of any parametric assumption regarding the effect of time and explanatory variables on the hazard rates. The main criterion for assessing the flexibility in the two above cases is the range of possible values that the two measures of association can attain. Finally, we compare different families of bivariate distributions with gamma marginals and we examine which of them can fit data with negative and/or positive dependence between the duration variables.

The rest of the paper is structured as follows. Section 2 briefly introduces the bivariate frailty model and discusses dependence properties of the joint survival function (equivalently, joint distribution) of the duration variables given the dependence structure of the distribution of the frailty terms. In Section 3, we discuss the properties of different bivariate distributions with gamma marginals which can be used for modeling the bivariate distribution of the two frailty terms. Section 4 focuses on the bounds for the Pearson's correlation coefficient and Section 5 studies the bounds for the Kendall's tau. Section 6 concludes and discusses possible extensions. The mathematical proofs are deferred to Appendix A. In Appendix B, we consider the dependence properties of some popular bivariate copulas. Finally, for notational convenience, we will omit the transpose symbol for vectors throughout the paper. We hope this slight abuse of notation will not create any confusion for the reader.

2 Quadrant dependence in the bivariate frailty model

Let T_1 and T_2 represent the nonnegative stochastic durations of interest and X be a vector of observable characteristics with support $\mathcal{X} \subseteq \mathbf{R}^d$, where d is a finite positive integer number. Denote by $x \in \mathcal{X}$ the realization of X. In addition, introduce two frailty terms $V_1 \in \mathbf{R}_+$ and $V_2 \in \mathbf{R}_+$ that are independent of the vector X and directly affect the realization of T_1 and T_2, respectively. The random variables V_1 and V_2 capture unobserved or nonmeasurable time-invariant characteristics. The corresponding hazard rate of the duration variables $T_1|x, V_1$ and $T_2|x, V_2$ is expressed as follows

$$\begin{aligned} \theta_1(t|x,V_1) &= \lambda_1(t,x)V_1, \\ \theta_2(t|x,V_2) &= \lambda_2(t,x)V_2, \end{aligned} \tag{1}$$

with $\lambda_1 : \mathbf{R}_+ \times \mathcal{X} \to (0,\infty)$ and $\lambda_2 : \mathbf{R}_+ \times \mathcal{X} \to (0,\infty)$. We shall assume that the functions $\lambda_1(.,x)$ and $\lambda_2(.,x)$ are integrable on bounded intervals of the positive real line, that is, the quantities $\Lambda_1(t,x) = \int_0^t \lambda_1(\omega,x)d\omega$ and $\Lambda_2(t,x) = \int_0^t \lambda_2(\omega,x)d\omega$ exist for each $(t,x) \in \mathbf{R}_+ \times \mathcal{X}$.

We denote by G the distribution of the bivariate random vector (V_1, V_2) and by G_1 and G_2 the marginal distribution of V_1 and V_2, respectively. The main assumption that will hold throughout this paper is $T_1 \perp T_2|x,V_1,V_2$. In words, the duration variables are stochastically independent of each other given the observable characteristics and the frailty terms. Let $i = 1,2$, and consider the survival functions $S_i(t|x,V_i) = \mathbf{P}(T_i > t|x,V_i)$ and $S(t_1,t_2|x,V_1,V_2) = \mathbf{P}(T_1 > t_1, T_2 > t_2|x,V_1,V_2)$. The specification (1) implies $S_i(t|x,V_i) = \exp(-\Lambda_i(t,x)V_i)$ and therefore $S(t_1,t_2|x,V_1,V_2) = \exp(-\Lambda_1(t_1,x)V_1 - \Lambda_2(t_2,x)V_2)$ by using the conditional independence property $T_1 \perp T_2|x,V_1,V_2$. Also, introduce the survival functions $S_i(t|x) = \mathbf{P}(T_i > t|x)$ and $S(t_1,t_2|x) = \mathbf{P}(T_1 > t_1, T_2 > t_2|x)$. The survival function of $T_i|x$ can be explicitly calculated by a mixture of exponential distributions in the following way

$$S_i(t|x) = \int_{\mathbf{R}_+} \exp(-\Lambda_i(t,x)v)dG_i(v) = \mathcal{L}_{G_i}\left(\Lambda_i(t,x)\right), \quad i = 1, 2, \tag{2}$$

where the generic symbol $\mathcal{L}$ denotes the Laplace Transform (LT) of the corresponding probability measure. Likewise, the survival function of $(T_1, T_2)|x$ can be represented by a mixture of bivariate exponential distributions as follows

$$\begin{aligned} S(t_1, t_2|x) =& \int_{\mathbf{R}_+^2} \exp(-\Lambda_1(t_1,x)v_1 - \Lambda_2(t_2,x)v_2)dG(v_1, v_2) \\ =& \mathcal{L}_G\left(\Lambda_1(t_1,x), \Lambda_2(t_2,x)\right). \end{aligned} \tag{3}$$

If $V_1 \perp V_2$ we get $\mathcal{L}_G(s_1, s_2) = \mathcal{L}_{G_1}(s_1)\mathcal{L}_{G_2}(s_2)$ for all $(s_1, s_2) \in \mathbf{R}_+^2$ and thus we have, by (2) and (3), $T_1 \perp T_2|x$ for any $x \in \mathcal{X}$. On the other hand, if $T_1 \perp T_2|x$ for some $x \in \mathcal{X}$, then $V_1 \perp V_2$ by noting, in view of (1), that $\ln V_i = -\ln \Lambda_i(T_i, x) + \epsilon_i$ for $i = 1, 2$, where ϵ_1, ϵ_2 are independent random variables with probability density function $f_i(\epsilon) = e^{\epsilon} \exp\left(-e^{\epsilon}\right).$

We first begin with the definitions of negative quadrant dependence and positive quadrant dependence Lehmann (1966).

Definition 1 *An $\mathbf{R}_+^2-$valued bivariate random vector (W_1, W_2) and its distribution function are said to be negative $($positive$)$ quadrant dependent if*

$$\mathbf{P}(W_1 \le w_1, W_2 \le w_2) \le (\ge)\mathbf{P}(W_1 \le w_1)\mathbf{P}(W_2 \le w_2) \textit{ for all } (w_1, w_2) \in \mathbf{R}_+^2.$$

Equivalently, an $\mathbf{R}_+^2-$bivariate random vector (W_1, W_2) and its survival function are said to be negative $($positive$)$ quadrant dependent if

$$\mathbf{P}(W_1 > w_1, W_2 > w_2) \le (\ge)\mathbf{P}(W_1 > w_1)\mathbf{P}(W_2 > w_2) \textit{ for all } (w_1, w_2) \in \mathbf{R}_+^2.$$

In the sequel, we use the acronyms NQD and PQD for the terms negative quadrant dependent and positive quadrant dependent, respectively. These two dependence concepts are the weakest for describing the dependence structure between two random variables. In particular, the density function of a bivariate random vector is reverse rule of order two, the strongest notion of negative dependence, only if the underlying distribution function is NQD. Likewise, the density function of a bivariate random vector is totally positive of order two, the strongest concept of positive dependence, only if the corresponding distribution is PQD. Next, we recall the definition of the concordance ordering $\prec_C$ that can be found in Joe (1997).

Definition 2 *Suppose $\mathcal{P}^a$ and $\mathcal{P}^b$ are bivariate distribution functions on $\mathbf{R}_+^2$ or bivariate survival functions on $\mathbf{R}_+^2$ with specific marginals $\mathcal{P}_1$ and $\mathcal{P}_2$. If $\mathcal{P}^a(w_1, w_2) \leq \mathcal{P}^b(w_1, w_2)$ for all $(w_1, w_2) \in \mathbf{R}_+^2$, then we say that $\mathcal{P}^b$ is more concordant than $\mathcal{P}^a$, written as $\mathcal{P}^a \prec_C \mathcal{P}^b$.*

We first obtain the following result which states that any concordance ordering between two different distributions of (V_1, V_2) will result in the same concordance ordering between the corresponding survival functions of $(T_1, T_2)|x$.

Proposition 1 *Let G^a and G^b represent two different distributions of the random vector (V_1, V_2) with $G^a \prec_C G^b$. Also, denote by S^a and S^b the corresponding mixtures of bivariate exponential distributions as defined in (3). Then, $S^a \prec_C S^b$ for each $x \in \mathcal{X}$.*

An important remark about Proposition 1 is that its result can be extended to any arbitrary bivariate hazard model in which the $S(t_1, t_2|x,v_1, v_2)$ is bounded, continuous and $2-$positive function in (v_1, v_2) for all $(t_1, t_2, x) \in \mathbf{R}_+^2 \times \mathcal{X}$ (see Appendix A). The next corollary directly follows from Proposition 1 by setting $G^a(v_1, v_2) \leq G_1(v_1)G_2(v_2) = G^b(v_1, v_2), (v_1, v_2) \in \mathbf{R}_+^2$, for the NQD result and $G^a(v_1, v_2) = G_1(v_1)G_2(v_2) \leq G^b(v_1, v_2)$, $(v_1, v_2) \in \mathbf{R}_+^2$, for the PQD result.

Corollary 1 *Let T_1 and T_2 be the duration variables that are generated by the bivariate frailty model (1). If (V_1, V_2) is NQD (PQD), then $(T_1, T_2)|x$ is NQD (PQD) for every $x \in \mathcal{X}$.*

In Section 4 and 5 we shall consider bounds for the values of the Pearson's correlation coefficient and Kendall's tau, respectively. The former quantitatively describes the strength of the linear relationship between T_1 and T_2, whereas the latter is a rank correlation coefficient between T_1 and T_2. According to Corollary 1, the type of quadrant dependence of the random vector (V_1, V_2) determines the type of quadrant dependence of the random vector $(T_1, T_2)|x$ for any $x \in \mathcal{X}$ and thereby the sign of these two association measures.

Assuming $\mathbf{E}(T_i|x) < \infty$ and $\mathbf{E}(T_i^2|x) < \infty$ for $x \in \mathcal{X}$ and $i = 1, 2$, the conditional on x Pearson's correlation coefficient between T_1 and T_2 is expressed as

$$\rho(T_1, T_2|x) = \frac{\text{Cov}(T_1, T_2|x)}{[\text{Var}(T_1|x)\text{Var}(T_2|x)]^{\frac{1}{2}}}. \tag{4}$$

By Hoeffding's identity we have

$$\text{Cov}(T_1, T_2|x) = \int_{\mathbf{R}_+^2} [S(t_1, t_2|x) - S_1(t_1|x)S_2(t_2|x)]\, dt_1 dt_2.$$

Therefore, if $(T_1, T_2)|x$ is NQD for all $x \in \mathcal{X}$ it will hold $S(t_1, t_2|x) - S_1(t_1|x)S_2(t_2|x) \leq 0$ for all $(t_1, t_2, x) \in \mathbf{R}_+^2 \times \mathcal{X}$ and therefore $\rho(T_1, T_2|x) \leq 0$ for any $x \in \mathcal{X}$. The previous inequalities will be reversed in case $(T_1, T_2)|x$ is PQD.

The main drawback of the Pearson's correlation coefficient is that it is not rank-invariant, that is, generally $\rho(T_1, T_2|x) \neq \rho(h_1(T_1), h_2(T_2)|x)$ for any nonlinear strictly monotone transformations h_1 and h_2. A measure that satisfies this property is the Kendall's tau which has attracted the interest of researchers who work on duration analysis (Wang et al., 2000; Martin and Betensky, 2005; Beaudoin et al., 2007; Oakes, 2008). To be more precise, consider two independent copies $(T_1^A, T_2^A)|x$ and $(T_1^B, T_2^B)|x$ of the bivariate random vector $(T_1, T_2)|x$. The value of $\tau(T_1, T_2|x)$ for any $x \in \mathcal{X}$ is calculated by the following difference

$$\tau(T_1, T_2|x) = \mathbf{P}\left[(T_1^A - T_1^B)(T_2^A - T_2^B) > 0|x\right] - \mathbf{P}\left[(T_1^A - T_1^B)(T_2^A - T_2^B) < 0|x\right],$$

which gives

$$\tau(T_1, T_2|x) = 2\mathbf{P}\left[(T_1^A - T_1^B)(T_2^A - T_2^B) > 0|x\right] - 1. \tag{5}$$

Clearly, it holds $-1 \leq \tau(T_1, T_2|x) \leq 1$ for all $x \in \mathcal{X}$ and it is also easy to see that the value of $\tau(T_1, T_2|x)$ is equal to -1 $(+1)$ if and only if $T_2 = h(T_1)$, with h to be a strictly decreasing (increasing) transformation. Some further elaboration of (5) gives

$$\tau(T_1, T_2|x) = 4 \int_{\mathbf{R}_+^2} S(t_1, t_2|x) dS(t_1, t_2|x) - 1. \tag{6}$$

Note that we have chosen to express $\tau(T_1, T_2|x)$ as a functional of $S(t_1, t_2|x)$ and not of $F(t_1, t_2|x)$, where $F(t_1, t_2|x) = 1 - S_1(t_1|x) - S_2(t_2|x) + S(t_1, t_2|x)$, as we find it more convenient for the analysis in the sequel. Hence, in case $(T_1, T_2)|x$ is NQD for all $x \in \mathcal{X}$, it will hold $S(t_1, t_2|x) \leq S_1(t_1|x)S_2(t_2|x)$ for all $(t_1, t_2, x) \in \mathbf{R}_+^2 \times \mathcal{X}$ and by using the result of Theorem 2 of Tchen (1980) it can be readily shown that $\tau(T_1, T_2|x) \leq 0$ for any $x \in \mathcal{X}$. On the other hand, if $(T_1, T_2)|x$ is PQD the previous inequalities will go in the opposite direction.

3 Bivariate frailty distribution with gamma marginals

To derive bounds for the values of the two association measures, we assume $V_i \sim Gamma(k_i, \mu_i)$ for $i = 1, 2$, where the parameters k_i and μ_i are defined as shape parameter and scale parameter, respectively, and we assume that they are strictly positive. More precisely, the probability density of V_i is given by

$$g_i(v) = \frac{1}{\mu_i^{k_i}\Gamma(k_i)} v^{k_i-1} \exp(-\frac{v}{\mu_i}),\ \ v_i > 0, k_i > 0, \mu_i > 0,$$

where the Eulerian gamma function Γ is computed by $\Gamma(k) = \int_0^\infty \omega^{k-1} \exp(-\omega) d\omega$ for $k > 0$. In the next two subsections we shall discuss possible parameterizations of G and the dependence structure they induce on $(T_1, T_2)|x$.

3.1 Bivariate gamma distributions

Before proceeding to the description of two bivariate gamma distributions, recall that if T_1 and T_2 are generated by (1), then $S(t_1, t_2|x) = \mathcal{L}_G\left(\Lambda_1(t_1, x), \Lambda_2(t_2, x)\right)$ and $S_i(t|x) = \mathcal{L}_{G_i}\left(\Lambda_i(t, x)\right)$ for $i = 1, 2$. The first distribution that we study as a candidate for the parameterization of G is the double bivariate gamma Kotz et al. (2000), which has the following stochastic representation

$$V_i = \mu_i(V_0 + V_{0i}), \quad i = 1, 2, \tag{7}$$

with $V_0 \sim Gamma(k_0, 1)$ and $V_{0i} \sim Gamma(k_{0i}, 1)$ to be independent gamma variates. The marginal distribution of V_i is gamma distribution with shape parameter $k_0 + k_{0i}$ and scale parameter μ_i.

Cherian (1941) studied the above distribution for $\mu_1 = \mu_2 = 1$ and $k_{01} = k_{02}$. The use of the double bivariate gamma distribution is widespread in applications in the field of biostatistics (Korsgaard and Andersen, 1998; Zhong and Li, 2002; Jonker et al., 2009). By using Bayes' law we can deduce that the vector (V_1, V_2) is PQD and consequently, by Corollary 1, the vector $(T_1, T_2)|x$ is PQD for each $x \in \mathcal{X}$. There is an alternative way to view that $(T_1, T_2)|x$ is PQD for this case. The LT of the double gamma distribution is expressed as $\mathcal{L}_G(s_1, s_2) = \mathcal{L}_{G_0}(\mu_1 s_1 + \mu_2 s_2)\mathcal{L}_{G_{01}}(\mu_1 s_1)\mathcal{L}_{G_{02}}(\mu_2 s_2)$ for any $(s_1, s_2) \in \mathbf{R}_+^2$. Provided that $V_0 \sim Gamma(k_0, 1)$, it holds $\mathcal{L}_{G_0}(\mu_1 s_1 + \mu_2 s_2) \geq \mathcal{L}_{G_0}(\mu_1 s_1)\mathcal{L}_{G_0}(\mu_2 s_2)$ for all $(s_1, s_2) \in \mathbf{R}_+^2$, from which it is straightforward to infer that $(T_1, T_2)|x$ is PQD. Finally, note that for the limiting case $k_{0i} \to 0$ we get $\mathbf{P}(\mu_2 V_1 = \mu_1 V_2) \to 1$.

The second bivariate gamma distribution that we consider for modelling G is mostly known by its LT which is expressed as follows

$$\mathcal{L}_G(s_1, s_2) = (1 + \mu_1 s_1 + \mu_2 s_2 + \mu_{12} s_1 s_2)^{-k}, \quad (s_1, s_2) \in \mathbf{R}_+^2, \tag{8}$$

with $k > 0$, $\mu_1 > 0$, $\mu_2 > 0$ and $\mu_1\mu_2 - \mu_{12} \geq 0$. The above LT corresponds to a bivariate gamma distribution with $V_1 \sim Gamma(k, \mu_1)$ and $V_2 \sim Gamma(k, \mu_2)$. Kotz et al. (2000) call it Kibble and Moran bivariate distribution. This bivariate gamma distribution is used by Henderson and Shimakura (2003) who apply a Poisson-gamma model in longitudinal data to account for individual-random effects and within-individual serial correlation. The case $\mu_{12} = 0$ corresponds to $\mathbf{P}(\mu_2 V_1 = \mu_1 V_2) = 1$ and the case $\mu_1\mu_2 - \mu_{12} = 0$ corresponds to independence between V_1 and V_2. It is easy to verify that $\mathcal{L}_G(s_1, s_2) \geq \mathcal{L}_{G_1}(s_1)\mathcal{L}_{G_2}(s_2)$ for all $(s_1, s_2) \in \mathbf{R}_+^2$ and therefore the random vector $(T_1, T_2)|x$ is PQD for each $x \in \mathcal{X}$.

Parameterization of G by using one of the two above distributions is convenient; although the corresponding densities have quite complicated expressions, the LT for each distribution has closed form expression which in turn gives closed form expression for $S(t_1, t_2|x)$ as well. The main drawback of using one of these two bivariate distributions is that the $(T_1, T_2)|x$ is PQD and consequently, the $\rho(T_1, T_2|x)$ and $\tau(T_1, T_2|x)$ will be nonnegative for all $x \in \mathcal{X}$. To overcome this limitation we consider in the next subsection the notion of copula for parameterizing G and as we shall discuss, negative values of these two association measures can be attained.

3.2 Copula with gamma marginals

The advantage of using the copula approach is that it allows us to separate the bivariate distribution G into the marginals G_1, G_2 and an $\mathbf{R}$-valued pure dependence parameter ψ which captures the level of dependence between V_1 and V_2. Nelsen (2006) provides a detailed exposition of the important concept of copula.

According to the celebrated Sklar's theorem (Sklar, 1959) and given that the distributions G_1, G_2 are continuous functions, there exists a unique copula $C_\psi : [0,1]^2 \to [0,1]$ such that $G(v_1, v_2) = C_\psi(G_1(v_1), G_2(v_2))$ for all $(v_1, v_2) \in \mathbf{R}_+^2$. It is not difficult to see that the C_ψ is the distribution of the random vector $(G_1(V_1), G_2(V_2))$. Conversely, for any given bivariate distribution G we can construct the corresponding copula by considering the quantity $G(G_1^{-1}(v_1), G_2^{-1}(v_2))$, where $G_i^{-1}(v) = \inf\{\omega \in \mathbf{R} : G_i(\omega) \geq v\}$ for $i = 1, 2$. Hence, we have $\psi = k_0$ for the double bivariate gamma and $\psi = \mu_{12}$ for the Kibble and Moran bivariate distribution.

It is well-known that the following Frechet bounds apply

$$\max\{G_1(v_1) + G_2(v_2) - 1, 0\} \leq C_\psi(G_1(v_1), G_2(v_2)) \leq \min\{G_1(v_1), G_2(v_2)\} \tag{9}$$

for every $(v_1, v_2) \in \mathbf{R}_+^2$. When $C_\psi(G_1(v_1), G_2(v_2)) = \max\{G_1(v_1) + G_2(v_2) - 1, 0\}$ for each $(v_1, v_2) \in \mathbf{R}_+^2$, it holds $G_1(V_1) + G_2(V_2) - 1 = 0$ with probability one, and the random variables V_1 and V_2 are called countermonotonic. When $C_\psi(G_1(v_1), G_2(v_2)) = \min\{G_1(v_1), G_2(v_2)\}$ for all $(v_1, v_2) \in \mathbf{R}_+^2$, it holds $G_1(V_1) = G_2(V_2)$ with probability one, and the random variables V_1 and V_2 are called comonotonic. Equivalently, if C_ψ equals the lower (upper) Frechet bound, the random variable V_1 is a strictly decreasing (increasing) function of V_2. Note that when $\mathbf{P}(\mu_2 V_1 = \mu_1 V_2) = 1$, which implies $k_1 = k_2$, the G coincides with the upper Frechet bound and thus both of the two bivariate gamma distributions that were studied in the previous subsection allow, in the limit, this probabilistic behavior.

The family of bivariate copulas that we could use to parameterize C_ψ is, for instance, either the Archimedean family or the Farlie-Gumbel-Morgenstern (FGM) family. In Appendix B, we provide a discussion about the functional form and dependence properties of three Archimedean copulas, Clayton, Frank, Gumbel, and also the FGM copula. Note that the Clayton copula we describe in Appendix B is a simple extension of the copula introduced by Clayton (1978). The three aforementioned Archimedean copulas are quite flexible in terms of positive dependence between V_1 and V_2 (and consequently, between T_1 and T_2) in the sense that they can be, in the limit, equal to the upper Frechet bound (9). Regarding negative dependence, Gumbel copula does not admit a representation such that V_1 and V_2 are negatively dependent. However, Clayton copula and Frank copula permit for negative dependence, with the Frank copula to converge towards the lower Frechet bound (9) for limiting values of the dependence parameter ψ. Note that the Clayton copula equals the lower Frechet bound for some certain value of the parameter ψ; however, if ψ converges towards this particular value the copula does not converge to the lower Frechet bound. On the other hand, the FGM copula does allow for both negative and positive dependence. But, its shortcoming is that it does not allow for strong (either positive or negative) dependence, that is, for any values of the parameter ψ, the Frechet bounds (9) cannot be approached.

4 Pearson's correlation coefficient

In this section we focus our attention on the Pearson's correlation coefficient under the assumption of $\lambda_i(t,x) = \alpha_i t^{\alpha_i - 1}\varphi_i(x)$ for $i = 1, 2$, with $\alpha_i > 0, t \in \mathbf{R}_+$ and $\varphi_i : \mathcal{X} \to (0, \infty)$. Namely, the hazard rates of the bivariate frailty model (1) are expressed as

$$\begin{aligned} \theta_1(t|x,V_1) &= \alpha_1 t^{\alpha_1 - 1}\varphi_1(x)V_1, \\ \theta_2(t|x,V_2) &= \alpha_2 t^{\alpha_2 - 1}\varphi_2(x)V_2. \end{aligned} \tag{10}$$

The specification (10), which is widely known as Weibull bivariate frailty model, is a special case of the bivariate frailty model $\theta_i(t|x,V_i) = \tilde{\lambda}_i(t)\varphi_i(x)V_i$, where the $\tilde{\lambda}_i$ is called baseline hazard and the φ_i is known as regressor function. Identification of this type of frailty models is provided by Elbers and Ridder (1982), Ridder and Woutersen (2003), Abbring and Ridder (2009) for the univariate case and Honoré (1993) for the bivariate case.

Next, we recall that

$$\rho(T_1, T_2|x) = \frac{\mathrm{Cov}(T_1, T_2|x)}{[\mathrm{Var}(T_1|x)\mathrm{Var}(T_2|x)]^{\frac{1}{2}}}, \quad x \in \mathcal{X}.$$

The covariance and the variance formulas are given by

$$\mathrm{Cov}(T_1, T_2|x) = \mathbf{E}\left[\mathbf{E}(T_1T_2|x,V_1, V_2)\right] - \prod_{i=1}^{2}\mathbf{E}\left[\mathbf{E}(T_i|x,V_i)\right] \tag{11}$$

and

$$\mathrm{Var}(T_i|x) = \mathbf{E}\left[\mathrm{Var}(T_i|x,V_i)\right] + \mathrm{Var}\left[\mathbf{E}(T_i|x,V_i)\right] \tag{12}$$

for $i = 1, 2$, where the outer expectations and variance in the right-hand side of the two above equations are taken with respect to the distribution of the frailty terms. The term $\mathbf{E}\left[\mathrm{Var}(T_i|x,V_i)\right]$ captures the autonomous variation, whereas the term $\mathrm{Var}\left[\mathbf{E}(T_i|x,V_i)\right]$ captures the variation due to the presence of the frailty term. Under the specification (10), the variable $T_i|x,V_i$ follows a Weibull distribution with shape parameter α_i and scale parameter $(\varphi_i(x)V_i)^{-\frac{1}{\alpha_i}}$ and thus $\mathbf{E}(T_i|x,V_i)$ and $\mathrm{Var}(T_i|x,V_i)$ are proportional to $V_i^{-\frac{1}{\alpha_i}}$ and $V_i^{-\frac{2}{\alpha_i}}$, respectively. Denote by ρ_{12} the Pearson's correlation coefficient between $V_1^{-\frac{1}{\alpha_1}}$ and $V_2^{-\frac{1}{\alpha_2}}$. Assuming

$$\mathbf{E}\left(V_i^{-\frac{1}{\alpha_i}}\right) < \infty,\ \mathbf{E}\left(V_i^{-\frac{2}{\alpha_i}}\right) < \infty \text{ for each } \alpha_i > 0,$$

we write

$$\rho_{12} = \frac{\mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}}V_2^{-\frac{1}{\alpha_2}}\right) - \mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}}\right)\mathbf{E}\left(V_2^{-\frac{1}{\alpha_2}}\right)}{\prod_{i=1}^{2}\left[\mathbf{E}\left(V_i^{-\frac{2}{\alpha_i}}\right) - \left[\mathbf{E}\left(V_i^{-\frac{1}{\alpha_i}}\right)\right]^2\right]^{\frac{1}{2}}}. \tag{13}$$

After doing some algebra we can rewrite $\rho(T_1, T_2|x)$ as follows

$$\rho(T_1, T_2|x) = \rho_{12}\prod_{i=1}^{2}\left[\delta(\alpha_i) + (\delta(\alpha_i) - 1)\frac{\left[\mathbf{E}\left(V_i^{-\frac{1}{\alpha_i}}\right)\right]^2}{\mathbf{E}\left(V_i^{-\frac{2}{\alpha_i}}\right) - \left[\mathbf{E}\left(V_i^{-\frac{1}{\alpha_i}}\right)\right]^2}\right]^{-\frac{1}{2}}, \tag{14}$$

with

$$\delta(\alpha_i) = \frac{\left[\Gamma(1+\frac{1}{\alpha_i})\right]^2}{\Gamma(1+\frac{2}{\alpha_i})}, \qquad \alpha_i > 0. \tag{15}$$

The function δ is strictly decreasing function in α_i, with $\lim_{\alpha_i \to 0} \delta(\alpha_i) = \infty$ and $\lim_{\alpha_i \to \infty} \delta(\alpha_i) = 1$. One important observation from (14) is that the value of $\rho(T_1, T_2|x)$ for fixed α_1 and α_2 depends on the strength of the linear relationship between the random variables $V_1^{-\frac{1}{\alpha_1}}$ and $V_2^{-\frac{1}{\alpha_2}}$ and not between the random variables V_1 and V_2. The latter is a consequence of the nonlinearity of the model (10).

Recall from Section 2 that if (V_1, V_2) is NQD (PQD) the $(T_1, T_2)|x$ is NQD (PQD) for any $x \in \mathcal{X}$ and therefore the $\rho(T_1, T_2|x)$ is nonpositive (nonnegative). This works in the formula (14) by way of the term ρ_{12}. In particular, if (V_1, V_2) is NQD (PQD) the $(V_1^{-\frac{1}{\alpha_1}}, V_2^{-\frac{1}{\alpha_2}})$ is NQD (PQD) as well due the monotonic relationship between V_i and $V_i^{-\frac{1}{\alpha_i}}$ for each $\alpha_i > 0$, which in turn implies that ρ_{12} is nonpositive (nonnegative).

Define for $(\alpha_1, \alpha_2) \in (0, \infty)^2$

$$b_l(\alpha_1, \alpha_2) = -\frac{1}{[\delta(\alpha_1)\delta(\alpha_2)]^{\frac{1}{2}} + [(\delta(\alpha_1) - 1)(\delta(\alpha_2) - 1)]^{\frac{1}{2}}} \tag{16}$$

$$b_u(\alpha_1, \alpha_2) = \frac{1}{[\delta(\alpha_1)\delta(\alpha_2)]^{\frac{1}{2}}}. \tag{17}$$

As shown by Van Den Berg (1997), for Weibull baseline hazards and any arbitrary joint distribution function of the random vector (V_1, V_2) the bounds for $\rho(T_1, T_2|x)$ are the following

$$b_l(\alpha_1, \alpha_2) < \rho(T_1, T_2|x) < b_u(\alpha_1, \alpha_2) \tag{18}$$

for each pair $(\alpha_1, \alpha_2) \in (0, \infty)^2$. The bounds are tight for certain bivariate distributions of (V_1, V_2) with discrete support, that is, they are approached arbitrarily closely. Given that $\delta(\alpha_i)$ is strictly decreasing in α_i, it is obvious from the above result that the range of possible values of $\rho(T_1, T_2|x)$ is increasing in α_i and thus the extreme values -1 and 1 are possible to be attained for $\alpha_i \to \infty$. This result can be explained as follows. To obtain maximum correlation it is required that first type of variation $\mathbf{E}\left[\mathrm{Var}(T_i|x,V_i)\right]$ be minimal relative to the second type of variation $\mathrm{Var}\left[\mathbf{E}(T_i|x,V_i)\right]$

for each $= 1, 2$, and the correlation between $V_1^{-\frac{1}{\alpha_1}}$ and $V_2^{-\frac{1}{\alpha_2}}$ be maximal. For $\alpha_i \to \infty$ the first type of variation decreases and is dominated by the second type and thus it is possible to attain any value in the interval $(-1, 1)$. Reverse statement will hold for $\alpha_i \to 0$.

Given that $V_i \sim Gamma(k_i, \mu_i)$, it can be easily shown that

$$\mathbf{E}\left(V_i^{-\frac{1}{\alpha_i}}\right) = \frac{\Gamma\left(k_i - \frac{1}{\alpha_i}\right)}{\Gamma(k_i)}\mu_i^{-\frac{1}{\alpha_i}},\ \mathbf{E}\left(V_i^{-\frac{2}{\alpha_i}}\right) = \frac{\Gamma\left(k_i - \frac{2}{\alpha_i}\right)}{\Gamma(k_i)}\mu_i^{-\frac{2}{\alpha_i}} \tag{19}$$

and therefore, the restriction $k_i > \frac{2}{\alpha_i}$ is imposed so that the the first two moments of $V_i^{-\frac{1}{\alpha_i}}$ are defined for $i = 1, 2$. Then we can express $\rho(T_1, T_2|x)$ as follows

$$\rho(T_1, T_2|x) = \rho_{12}\prod_{i=1}^{2}\left[\delta(\alpha_i) + (\delta(\alpha_i) - 1)\frac{\Gamma^2\left(k_i - \frac{1}{\alpha_i}\right)}{\Gamma\left(k_i - \frac{2}{\alpha_i}\right)\Gamma\left(k_i\right) - \Gamma^2\left(k_i - \frac{1}{\alpha_i}\right)}\right]^{-\frac{1}{2}}. \tag{20}$$

In the next two subsections we shall investigate how the assumption of gamma distributed frailties affect the behavior of $\rho(T_1, T_2|x)$. In particular, our interest is in studying whether the lower and upper bound of (18) can be arbitrarily approached in case the distribution of (V_1, V_2) has gamma marginals.

4.1 Lower bound for the Pearson's correlation coefficient

We first fix our attention on the lower bound for the linear correlation coefficient. The next proposition establishes a nonsharp (i.e., not necessarily attained) lower bound for the $\rho(T_1, T_2|x)$.

Proposition 2 *Suppose T_1 and T_2 are the duration variables that are generated by the bivariate frailty model (10), with $(\alpha_1, \alpha_2) \in (0, \infty)^2, V_1 \sim Gamma(k_1, \mu_1)$ and $V_2 \sim Gamma(k_2, \mu_2)$. Then, the following inequality holds*

$$\rho(T_1, T_2|x) \geq b_{gl}(\alpha_1, \alpha_2), \quad x \in \mathcal{X},$$

with

$$b_{gl}(\alpha_1,\alpha_2) = \min_{\substack{k_1>\frac{2}{\alpha_1}\\ k_2>\frac{2}{\alpha_2}}} \left[\prod_{i=1}^{2}\frac{\Gamma(k_i)}{\Gamma\left(k_i+\frac{1}{\alpha_i}\right)} - \prod_{i=1}^{2}\frac{\Gamma\left(k_i-\frac{1}{\alpha_i}\right)}{\Gamma(k_i)}\right]\prod_{i=1}^{2}\left[\frac{\Gamma\left(k_i-\frac{2}{\alpha_i}\right)}{\Gamma(k_i)} - \frac{\Gamma^2\left(k_i-\frac{1}{\alpha_i}\right)}{\Gamma^2(k_i)}\right]^{-\frac{1}{2}}$$

$$\times\prod_{i=1}^{2}\left[\delta(\alpha_i)+(\delta(\alpha_i)-1)\frac{\Gamma^2\left(k_i-\frac{1}{\alpha_i}\right)}{\Gamma\left(k_i-\frac{2}{\alpha_i}\right)\Gamma(k_i)-\Gamma^2\left(k_i-\frac{1}{\alpha_i}\right)}\right]^{-\frac{1}{2}}.$$

The next table lists the bounds $b_l(\alpha_1,\alpha_2)$ and $b_{gl}(\alpha_1,\alpha_2)$ for different values of α_1,α_2. To make the comparison between $b_l(\alpha_1,\alpha_2)$ and $b_{gl}(\alpha_1,\alpha_2)$ more transparent, all numbers have been rounded off to three decimal digits.

(α_1,α_2)	$b_l(\alpha_1,\alpha_2)$	$b_{gl}(\alpha_1,\alpha_2)$
$(0.5,1)$	-0.175	-0.125
$(0.5,2)$	-0.254	-0.233
$(1,1)$	-0.333	-0.220
$(1,2)$	-0.472	-0.366
$(1,3)$	-0.535	-0.520
$(1.5,2)$	-0.582	-0.397
$(2,2)$	-0.647	-0.451
$(2,3)$	-0.719	-0.580
$(4,4)$	-0.860	-0.590
$(5,5)$	-0.860	-0.599

Table 1: $b_l(\alpha_1,\alpha_2)$ and $b_{gl}(\alpha_1,\alpha_2)$ values.

In view of the results of Table 1, we can claim that the bound $b_{gl}(\alpha_1,\alpha_2)$ is generally closer to zero than the bound $b_l(\alpha_1,\alpha_2)$. These results reveal a limitation of the the bivariate Weibull gamma frailty model to fit data with relatively large negative dependence between the duration variables. Note that the bound $b_{gl}(\alpha_1,\alpha_2)$ is not expected to be tight as three successive inequalities were employed to derive it. Namely, there could be values of α_1, α_2 such that $b_l(\alpha_1, \alpha_2) > b_{gl}(\alpha_1, \alpha_2)$; however, this is clearly due to the use of the three inequalities as the $b_l(\alpha_1, \alpha_2)$ covers all the bivariate distributions with support on $\mathbf{R}_+^2$ and trivially all the bivariate distributions with gamma marginals.

To improve the lower bound for the exponential case (i.e., $\alpha_1 = \alpha_2 = 1$) we carry out Monte Carlo simulation. For the exponential model we have $\rho(T_1, T_2|x) = \rho_{12}\left(\sqrt{k_1 k_2}\right)^{-1}$. For given marginals G_1 and G_2, the ρ_{12} will be minimized if and only if the distribution of (V_1^{-1}, V_2^{-1}) is equal to the lower Frechet bound. However, due to the fact that V_i^{-1} is strictly decreasing transformation of V_i, the ρ_{12} will be minimized for fixed G_1 and G_2 if and only if $G(v_1, v_2) = \max\{G_1(v_1) + G_2(v_2) - 1, 0\}$ for each $(v_1, v_2) \in \mathbf{R}_+^2$. In case G is parameterized by the Frank copula, the lower Frechet bound can be approached very well for limiting values of the dependence parameter. To derive an estimation of the minimum value of ρ_{12} for fixed k_1 and k_2, we draw gamma random variables V_1 and V_2 by using the relationship $G_1(V_1) + G_2(V_2) - 1 = 0$.

For the study of the values of ρ_{12} and $\rho(T_1, T_2|x)$ we present two figures. The first figure shows values of ρ_{12} as a function of k_1 and k_2. Note that we have reversed the axes with the values of k_1 and k_2 so that we have a more clear picture.

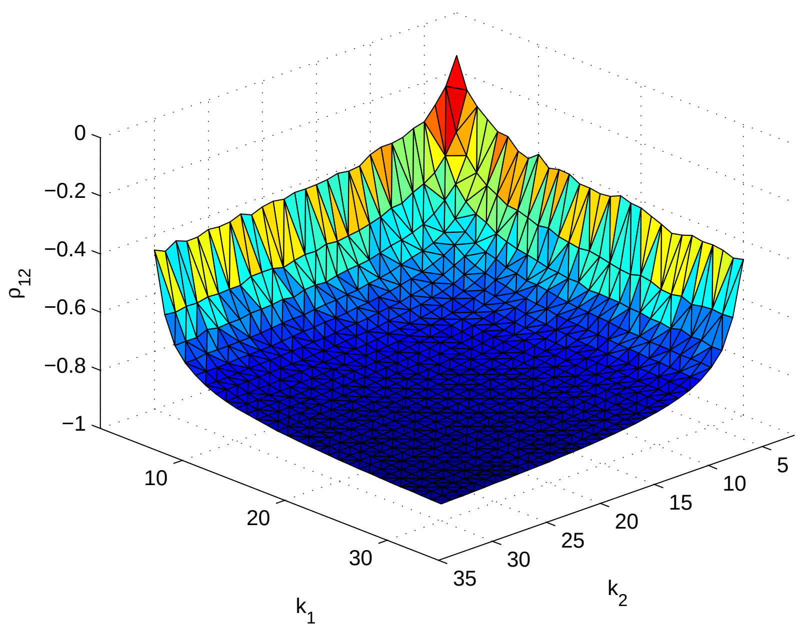


Figure 1: Plot of ρ_{12} as a function of k_1 and k_2, if $\alpha_1 = \alpha_2 = 1$.

The second figure displays the values of $\rho(T_1, T_2|x)$ as a function of k_1 and k_2.

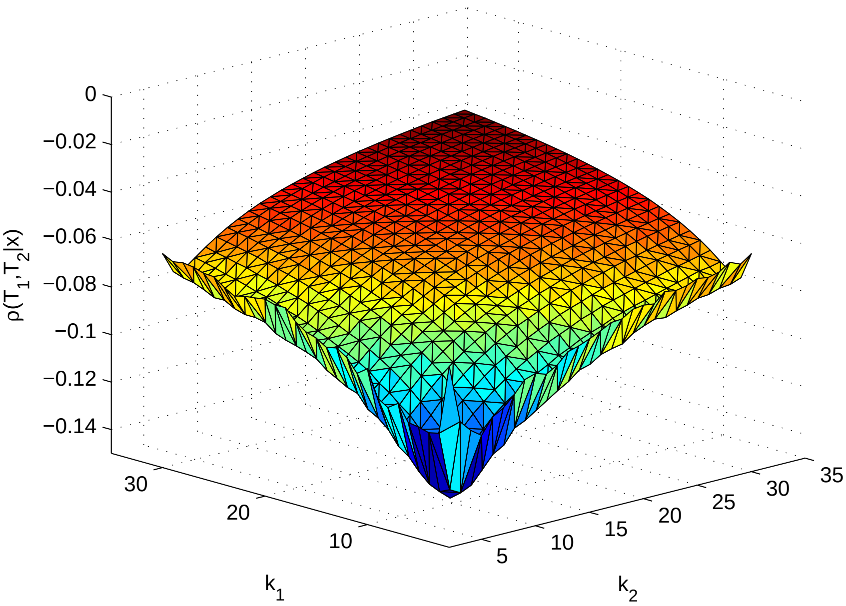


Figure 2: Plot of $\rho(T_1, T_2|x)$ as a function of k_1 and k_2, if $\alpha_1 = \alpha_2 = 1$.

The estimated value of the lower bound is about -0.14, which is clearly much closer to zero than the tight bound $-\frac{1}{3}$. From the two above graphs we can easily notice the two opposite effects of the value of the shape parameters on the values of ρ_{12} and $\rho(T_1, T_2|x)$. More precisely, the ρ_{12} approaches arbitrarily closely the value -1 for large values of k_1 and k_2. However, large values of the shape parameters weaken the linear relationship between the duration variables as the variation of the random variable $T_i|x$ due to the presence of the frailty is negligible with respect to the autonomous variation. To see this, consider for simplicity the case $k_1 = k_2 = k$. Then, we obtain $\mathbf{E}\left[\text{Var}(T_i|x,V_i)\right] = ((k-1)k)^{-1} = O(k^{-2})$ and $\text{Var}\left[\mathbf{E}(T_i|x,V_i)\right] = ((k-1)^2(k-2)) = O(k^{-3}) = o(k^{-2})$ for $k \to \infty$ and $i = 1, 2$.

Next, we consider three other possible families of distributions for G with marginals different from gamma. In particular, Mardia (1970) shows that if the random vector (V_1^{-1}, V_2^{-1}) follows the Filon-Isserk bivariate Beta distribution, the $\rho(T_1, T_2|x)$ can attain any values in the interval $(-\frac{1}{3}, 0]$. Moreover, Van Den Berg (1997) shows that if $(V_i)^{-1} = \sum_{j=1}^{k} U_{ij}^2$ for $i = 1, 2$ and some finite positive integer k, where the vector (U_{1j}, U_{2j}) follows a bivariate normal distribution, the lower bound of $\rho(T_1, T_2|x)$ is about -0.23. Finally, Van Den Berg (1997) shows that if $V_i = \exp(\eta_{i0} + \eta_{i1}\mathcal{N})$, where $\eta_{i0} \in \mathbf{R}$ and $\eta_{i1} \in \mathbf{R}\backslash\{0\}$ for $i = 1, 2$ and $\mathcal{N}$ is a normally distributed random variable, the lower bound of $\rho(T_1, T_2|x)$ is about -0.17. In view of these results and using as criterion the bounds for the Pearson's correlation coefficient, the assumption that the distribution of (V_1, V_2) is characterized by gamma marginals seems quite restrictive for attaining large negative values.

4.2 Upper bound for the Pearson's correlation coefficient

We now concentrate on the bivariate frailty model that has the property $\mathbf{P}(\mu_2 V_1 = \mu_1 V_2) = 1$, which in turn implies $G(v_1, v_2) = \min\{G_1(v_1), G_2(v_2)\}$ for all $(v_1, v_2) \in \mathbf{R}^2_+$ and $k_1 = k_2 = k$. Under the assumption of identical Weibull baseline hazards, that is, $\alpha_1 = \alpha_2 = \alpha$, we have $\rho_{12} \to 1$ for any $k > \frac{2}{\alpha}$. Also, for $k \to \frac{2}{\alpha}$ and given that $\lim_{k \to \frac{2}{\alpha}} \Gamma^2\left(k - \frac{2}{\alpha}\right) \to \infty$, we get by (20)

$$\rho(T_1, T_2|x) \to \frac{\left[\Gamma(1 + \frac{1}{\alpha})\right]^2}{\Gamma(1 + \frac{2}{\alpha})} = b_u(\alpha, \alpha).$$

Therefore, if $\alpha_1 = \alpha_2 = \alpha$ the upper bound of (18) can be arbitrarily approached in case G is equal either to one of the two bivariate gamma distributions of Section 3.1 or to one of the three Archimedean copulas described in detail in Appendix B.

Next, we turn our attention to the case $\alpha_1 \neq \alpha_2$ and $V_i \sim Gamma(k, \mu_i)$ for $i = 1, 2$, that is, $k_1 = k_2 = k$. Although imposing the assumption that both marginals have the same shape parameter may seem restrictive, it is rather general. In particular, it includes as special cases the bivariate frailty model in which $\mathbf{P}(\mu_2 V_1 = \mu_1 V_2) = 1$ that we described above for $\alpha_1 = \alpha_2$ and also the bivariate frailty model for which the (V_1, V_2) is distributed according to the Kibble and Moran bivariate gamma distribution. The next proposition analytically establishes a nonsharp bound for this case that is strictly smaller than the nonparametric upper bound (18).

Proposition 3 *Let T_1 and T_2 be the duration variables that are generated by the bivariate frailty model (10), with $\alpha_1 > \alpha_2 > 0, V_1 \sim Gamma(k, \mu_1)$ and $V_2 \sim Gamma(k, \mu_2)$. Then, the following inequality holds*

$$\rho(T_1, T_2|x) < b_{gu}(\alpha_1, \alpha_2), \quad x \in \mathcal{X},$$

with

$$b_{gu}(\alpha_1,\alpha_2)=\frac{1}{[\delta(\alpha_2)]^{\frac{1}{2}}}\left[\delta(\alpha_1)+(\delta(\alpha_1)-1)\frac{\Gamma^2\left(\frac{2\alpha_1-\alpha_2}{\alpha_1\alpha_2}\right)}{\Gamma\left(\frac{2(\alpha_1-\alpha_2)}{\alpha_1\alpha_2}\right)\Gamma\left(\frac{2}{\alpha_2}\right)-\Gamma^2\left(\frac{2\alpha_1-\alpha_2}{\alpha_1\alpha_2}\right)}\right]^{-\frac{1}{2}}<b_u(\alpha_1,\alpha_2).$$

The next table reports the bounds $b_u(\alpha_1,\alpha_2)$ and $b_{gu}(\alpha_1,\alpha_2)$ for different values of α_1,α_2, with $\alpha_1>\alpha_2$. Like in the case with the lower bound, we have rounded all the numbers off to three decimal points.

(α_1,α_2)	$b_u(\alpha_1,\alpha_2)$	$b_{gu}(\alpha_1,\alpha_2)$
$(0.5,0.25)$	0.049	0.037
$(0.75,0.25)$	0.071	0.041
$(1,0.5)$	0.289	0.204
$(2,0.5)$	0.362	0.194
$(2,1)$	0.627	0.469
$(5,1)$	0.689	0.423
$(5,2)$	0.864	0.704
$(10,2)$	0.880	0.669
$(10,5)$	0.968	0.921
$(20,10)$	0.999	0.976

Table 2: $b_u(\alpha_1,\alpha_2)$ and $b_{gu}(\alpha_1,\alpha_2)$ values.

The reason that $b_{gu}(\alpha_1,\alpha_2)<b_u(\alpha_1,\alpha_2)$ is that the shape parameter k is bounded from below by the maximum between the values of the ratios $\frac{2}{\alpha_1}$ and $\frac{2}{\alpha_2}$ so that the the first two moments of $V_i^{-\frac{1}{\alpha_i}}$ for $i=1,2$ are defined. Moreover, the bound of Proposition 3 is not attained as the gamma distribution is not closed under power transformation. In particular, if $V_1\sim Gamma(k,\mu_1)$ the random variable $V_1^{\frac{\alpha_2}{\alpha_1}}$, for any fixed positive α_1,α_2 with $\alpha_1\neq\alpha_2$, does not follow a gamma distribution and this implies that we cannot have $\rho_{12}=1$ such that V_1 and V_2 are gamma distributed. Hence, even if $\mathbf{P}(\mu_2V_1=\mu_1V_2)=1$ we will always have $\rho_{12}<1$ for any fixed values of α_1,α_2, with $\alpha_1\neq\alpha_2$.

5 Kendall's tau

We proceed now with the derivation of bounds for the range of values of the Kendall's tau as the results of Van Den Berg (1997) do not directly carry over to the bivariate gamma frailty model. As explained in Section 2, for two independent copies $(T_1^A, T_2^A)|x$ and $(T_1^B, T_2^B)|x$ of the bivariate random vector $(T_1, T_2)|x$ we have

$$\tau(T_1, T_2|x) = 2\mathbf{P}\left[(T_1^A - T_1^B)(T_2^A - T_2^B) > 0|x\right] - 1, \quad x \in \mathcal{X}. \tag{21}$$

In contrast to the Pearson's coefficient case, we will not assume anything about the range of values of the shape parameters. Also, we will not impose any condition on the functional form of λ_i except for the limiting result $\lim_{t\to\infty}\int_0^t \lambda_i(\omega, x)d\omega = \infty$ $(i = 1, 2)$. We will make use of the equality

$$\ln V_i = -\ln \Lambda_i(T_i, x) + \epsilon_i, \quad i = 1, 2, \tag{22}$$

with ϵ_1, ϵ_2 to be independent random variables that have probability density function $f_i(\epsilon) = e^{\epsilon}\exp\left(-e^{\epsilon}\right)$. The above equation is an equivalent representation of (1). Also, recall that $S_i(t|x) = \mathcal{L}_{G_i}\left(\Lambda_i(t, x)\right)$ for $(t, x) \in \mathbf{R}_+ \times \mathcal{X}$. Provided that $V_i \sim Gamma(k_i, \mu_i)$, it follows $S_i(t|x) = (1 + \mu_i\Lambda_i(t, x))^{-k_i}$. Therefore, the stochastic duration T_i can be expressed in structural form as follows

$$T_i = \Lambda_i^{-1}\left(\frac{1}{\mu_i}U_i^{-\frac{1}{k_i}} - \frac{1}{\mu_i}, x\right), \qquad U_i \sim Uniform(0, 1), \quad i = 1, 2. \tag{23}$$

We first focus on the lower bound of the values of $\tau(T_1, T_2|x)$. We assume $G(v_1, v_2) = \max\{G_1(v_1) + G_2(v_2) - 1, 0\}$ for each $(v_1, v_2) \in \mathbf{R}_+^2$. This implies $G_1(V_1) + G_2(V_2) - 1 = 0$ with probability one. Hence, V_2 is a strictly decreasing transformation of V_1 and we can write, by (22),

$$\ln \Lambda_2(T_2, x) = \mathcal{H}(T_1, \epsilon_1, \epsilon_2, x), \tag{24}$$

where $\mathcal{H}(., \epsilon_1, \epsilon_2, x)$ is strictly decreasing function, $\lim_{t\to\infty}\mathcal{H}(t, \epsilon_1, \epsilon_2, x) = h(t, x)$ for all $(\epsilon_1, \epsilon_2, x) \in \mathbf{R}^2 \times \mathcal{X}$, and $h(., x)$ is strictly decreasing function. By using the rank-invariant property of Kendall's tau and combining (21) and (24) we have

$$\tau(T_1,T_2|x) = 2\mathbf{P}\left[(T_1^A - T_1^B)(\mathcal{H}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{H}(T_1^B,\epsilon_1^B,\epsilon_2^B,x)) > 0\right] - 1. \tag{25}$$

Clearly, the $\tau(T_1,T_2|x)$ can be also written as follows

$$\begin{aligned}\tau(T_1,T_2|x) = {} & 2\mathbf{P}\left[\left\{(T_1^A - T_1^B) > 0\right\} \cap \left\{\mathcal{H}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{H}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) > 0\right\}\right] + \\ & 2\mathbf{P}\left[\left\{(T_1^A - T_1^B) < 0\right\} \cap \left\{\mathcal{H}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{H}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) < 0\right\}\right] - 1. \end{aligned} \tag{26}$$

For $k_1 \to 0$ and $\mu_1 = O(k_1^{-1})$ we have $T_1^A \to \infty$ and $T_1^B \to \infty$ which yield $\{\mathcal{H}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{H}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) > 0\} \to \{h(T_1^A,x) - h(T_1^B,x) > 0\} = \{T_1^A - T_1^B < 0\}$ and $\{\mathcal{H}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{H}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) < 0\} \to \{h(T_1^A,x) - h(T_1^B,x) < 0\} = \{T_1^A - T_1^B > 0\}$. By making use of these limiting statements, it is obvious, by using (26), that $\tau(T_1,T_2|x) \to -1$.

To derive the conditions needed to be satisfied for the upper bound of the $\tau(T_1,T_2|x)$ values, we require that G be equal to the upper Frechet bound, namely, $G(v_1,v_2) = \min\{G_1(v_1),G_2(v_2)\}$ for each $(v_1,v_2) \in \mathbf{R}_+^2$. Under this scenario, $G_1(V_1) = G_2(V_2)$ with probability one. Thus, V_2 is a strictly increasing transformation of V_1 and therefore we can write, by (22),

$$\ln \Lambda_2(T_2,x) = \mathcal{Y}(T_1,\epsilon_1,\epsilon_2,x), \tag{27}$$

where $\mathcal{Y}(.,\epsilon_1,\epsilon_2,x)$ is a strictly increasing function and $\lim_{t\to\infty}\mathcal{Y}(t,\epsilon_1,\epsilon_2,x) = y(t,x)$ for all $(\epsilon_1,\epsilon_2,x) \in \mathbf{R}^2 \times \mathcal{X}$, and $y(.,x)$ is some strictly increasing function. Performing identical calculations to the ones of the previous paragraph we obtain

$$\begin{aligned}\tau(T_1,T_2|x) = {} & 2\mathbf{P}\left[\left\{(T_1^A - T_1^B) > 0\right\} \cap \left\{\mathcal{Y}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{Y}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) > 0\right\}\right] + \\ & 2\mathbf{P}\left[\left\{(T_1^A - T_1^B) < 0\right\} \cap \left\{\mathcal{Y}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{Y}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) < 0\right\}\right] - 1. \end{aligned} \tag{28}$$

For $k_1 \to 0$ and $\mu_1 = O(k_1^{-1})$ we obtain $T_1^A \to \infty$ and $T_1^B \to \infty$ which in turn gives $\{\mathcal{Y}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{Y}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) > 0\} \to \{y(T_1^A,x) - y(T_1^B,x) > 0\} = \{T_1^A - T_1^B > 0\}$ and $\{\mathcal{Y}(T_1^A,\epsilon_1^A,\epsilon_2^A,x) - \mathcal{Y}(T_1^B,\epsilon_1^B,\epsilon_2^B,x) < 0\} \to \{y(T_1^A,x) - y(T_1^B,x) < 0\} = \{T_1^A - T_1^B < 0\}$. Given the equality $\mathbf{P}\left[(T_1^A - T_1^B) > 0|x\right] = \mathbf{P}\left[(T_1^A - T_1^B) < 0|x\right] = \frac{1}{2}$ for all $x \in \mathcal{X}$ and making use of (28), the limiting result $\tau(T_1, T_2|x) \to 1$ is obtained.

We summarize the above discussion to the next proposition.

Proposition 4 *Suppose T_1 and T_2 are the duration variables that are generated by the bivariate frailty model (1), with $V_1 \sim Gamma(k_1, \mu_1)$ and $V_2 \sim Gamma(k_2, \mu_2)$. Then, the following double inequality holds*

$$-1 < \tau(T_1, T_2|x) < 1, \quad x \in \mathcal{X}.$$

The extreme bounds -1 and 1 are tight in the sense that they can be approached arbitrarily closely. More precisely, if $G(v_1, v_2) = \max\{G_1(v_1) + G_2(v_2) - 1, 0\}$ for each $(v_1, v_2) \in \mathbf{R}_+^2$ and $k_1 \to 0$ with $\mu_1 = O(k_1^{-1})$, or $k_2 \to 0$ with $\mu_2 = O(k_2^{-1})$, we obtain $\tau(T_1, T_2|x) \to -1$. On the other hand, if $G(v_1, v_2) = \min\{G_1(v_1), G_2(v_2)\}$ for each $(v_1, v_2) \in \mathbf{R}_+^2$ and $k_1 \to 0$ with $\mu_1 = O(k_1^{-1})$, or $k_2 \to 0$ with $\mu_2 = O(k_2^{-1})$, then $\tau(T_1, T_2|x) \to 1$.

Therefore, by assuming gamma marginals for the distribution of (V_1, V_2) a necessary condition for approaching the lower bound of $\tau(T_1, T_2|x)$ is the distribution of (V_1, V_2) to be equal to the Frank copula. On the other hand, the upper bound of $\tau(T_1, T_2|x)$ can be approached arbitrarily closely if the bivariate distribution is modelled by the two bivariate gamma distributions of Section 3.1 or one of the three Archimedean copulas presented in Appendix B. Note that if $\mathbf{P}(\mu_2 V_1 = \mu_1 V_2) = 1$, which clearly gives $G(v_1, v_2) = \min\{G_1(v_1), G_2(v_2)\}$ for each $(v_1, v_2) \in \mathbf{R}_+^2$, we will have $k_1 = k_2 \to 0$.

By applying results of Embrechts et al. (2002), we have $\tau(T_1, T_2|x) \to -1$ if and only if $S(t_1, t_2|x) \to \max\{S_1(t_1|x) + S_2(t_2|x) - 1, 0\}$ for all $(t_1, t_2, x) \in \mathbf{R}_+^2 \times \mathcal{X}$, or equivalently, $S_1(T_1|x) + S_2(T_2|x) - 1 = 0$ for all $x \in \mathcal{X}$ with probability approaching one. On the other hand, $\tau(T_1, T_2|x) \to 1$ if and only if $S(t_1, t_2|x) \to \min\{S_1(t_1|x), S_2(t_2|x)\}$ for all $(t_1, t_2, x) \in \mathbf{R}_+^2 \times \mathcal{X}$, or equivalently, $S_1(T_1|x) = S_2(T_2|x)$ for all $x \in \mathcal{X}$ with probability approaching one. Hence, in view of Proposition 1, the condition in Proposition 4 that G is equal to the lower (upper) Frechet bound is indispensable. We should also point out here that $S(t_1, t_2|x)$ can be written in a copula form as function only of $S_1(t_1|x)$ and $S_2(t_2|x)$ and not of x because

$$S(t_1, t_2|x) = \mathcal{L}_G\left(\mathcal{L}_{G_1}^{-1}(S_1(t_1|x)), \mathcal{L}_{G_2}^{-1}(S_2(t_2|x))\right), \quad (t_1, t_2, x) \in \mathbf{R}_+^2 \times \mathcal{X},$$

where $\mathcal{L}^{-1}$ denotes the inverse of the LT of the corresponding probability measure.

6 Conclusions

We examine the dependence structure in bivariate frailty models in which the duration variables are dependent by way of the frailty terms. We first show that if the distribution of the frailty terms is negative (positive) quadrant dependent, then the conditional, on observed characteristics, joint survival function of the duration outcomes is negative (positive) quadrant dependent as well. To quantify the level of dependence between the duration variables, we consider the Pearson's correlation coefficient and the Kendall's tau. We provide bounds for the range of values of these measures under the assumption of gamma distributed frailty terms. To model the dependence structure between the frailty terms, we can use either standard bivariate gamma distributions or copulas with gamma marginals. The former induce only positive dependence between the duration variables, whereas the latter can induce positive and/or negative dependence. Strong negative (positive) dependence between the duration outcomes can be generated by bivariate distributions of the frailty terms which can be, in the limit, equal to the lower (upper) Frechet bound.

We calculate bounds for the values of the Pearson's correlation coefficient, if the baseline hazards have Weibull specification. Regarding the negative values, we analytically provide a nonsharp lower bound. We improve the lower bound for the exponential case by means of Monte Carlo simulation. The resulting lower bound is closer to zero than its nonparametric analogue which is derived by Van Den Berg (1997). For positive values of the Pearson's coefficient we show that the upper bound of Van Den Berg (1997) can be approached arbitrarily closely in case the Weibull specifications are identical. Moreover, we provide an upper bound for different Weibull specifications which is strictly smaller than the nonparametric bound. The resulting bound cannot be attained due to the fact that the gamma distribution is not closed under power transformation.

In contrast to the Pearson's correlation coefficient, Kendall's tau can take any value in the interval $(-1, 1)$ regardless of the functional form specification about the hazard rates. If the bivariate distribution of the frailty terms approaches the lower (upper) Frechet bound and the first moment of the frailty term(s) is finite, then the lower (upper) bound can be approached arbitrarily closely. In particular, we should impose the condition that one of the two shape parameters converges towards zero.

Fruitful topic for future research is the bounds for the two association measures in bivariate duration models where the two duration variables are parallel and the realization of one of these two variables affects the hazard rate of the other. Moreover, promising topic for investigation is the range of values for local measures of dependence such as the cross-ratio function **?**. Finally, it is of practical relevance to consider the concepts of lower tail and upper tail dependence between the duration variables. In particular, if the data display dependence between extreme values of the duration variables, we should know which bivariate distributions for the frailty terms allow such dependence pattern.

Appendix A

This appendix presents the mathematical proofs for the first three propositions in the main text.

Proof of Proposition 1. By definition

$$S^j(t_1,t_2|x)=\int_{\mathbf{R}_+^2} S(t_1,t_2|x,v_1,v_2)dG^j(v_1,v_2), \quad j=a,b, \tag{A-1}$$

where $S(t_1,t_2|x,v_1,v_2) = \exp\left(-\Lambda_1(t_1,x)v_1 - \Lambda_2(t_2,x)v_2\right)$ for $(t_1,t_2,x,v_1,v_2) \in \mathbf{R}_+^2 \times \mathcal{X} \times \mathbf{R}_+^2$. The integrand is a continuous bounded function in (v_1,v_2) for any $(t_1,t_2,x) \in \mathbf{R}_+^2 \times \mathcal{X}$. Moreover, it holds $\frac{\vartheta^2}{\vartheta v_1 \vartheta v_2} S(t_1,t_2|x,v_1,v_2) > 0$ for each $(t_1,t_2,x,v_1,v_2) \in \mathbf{R}_+^2 \times \mathcal{X} \times (0,\infty)^2$ (i.e., the $S(t_1,t_2|x,v_1,v_2)$ is $2-$positive function in v_1,v_2). Given that $G^a \prec_C G^b$, we obtain the inequality $S^a(t_1,t_2|x) \leq S^b(t_1,t_2|x)$ for all $(t_1,t_2,x) \in \mathbf{R}_+^2 \times \mathcal{X}$ by Theorem 2 of Tchen (1980). Recall also that

$$S_i(t|x) = \int_{\mathbf{R}_+} \exp\left(-\Lambda_i(t,x)v\right) dG_i(v), \quad i=1,2. \tag{A-2}$$

Provided that G^a and G^b are characterized by the fixed marginals G_1 and G_2, it follows that the bivariate survival functions S^a and S^b are characterized by the same marginals, S_1 and S_2. This in turn implies that $S^a \prec_C S^b$ for each $x \in \mathcal{X}$. ∎

Define for each $\varepsilon > 0$ the digamma function

$$\psi(\varepsilon) = \frac{\Gamma'(\varepsilon)}{\Gamma(\varepsilon)} \tag{A-3}$$

and the polygamma function

$$\psi^{(n)}(\varepsilon) = \frac{d^n \psi(\varepsilon)}{d\varepsilon^n}, \quad n \in \mathbf{N}, \tag{A-4}$$

with $\psi^{(0)}(.) = \psi(.)$. Moreover, it holds

$$\psi^{(n)}(\varepsilon) = (-1)^{n+1} \int_{\mathbf{R}_+} \frac{t^n}{1-e^{-t}} e^{-\varepsilon t} dt, \quad \varepsilon > 0. \tag{A-5}$$

We state Lemma 1 which is needed for the proof of Proposition 2 and 3. Its simple proof, which makes use of (A-4) and (A-5), is omitted.

Lemma 1 *Let* $(\varepsilon_1, \varepsilon_2, \varepsilon_3) \in (0, \infty)^3$. *Then,*

$$\Gamma(\varepsilon_1)\Gamma(\varepsilon_1+\varepsilon_2+\varepsilon_3) - \Gamma(\varepsilon_1+\varepsilon_2)\Gamma(\varepsilon_1+\varepsilon_3) > 0.$$

Proof of Proposition 2. Recall that

$$\rho(T_1, T_2|x) = \rho_{12} \prod_{i=1}^{2} \left[\delta(\alpha_i) + (\delta(\alpha_i) - 1) \frac{\Gamma^2\left(k_i - \frac{1}{\alpha_i}\right)}{\Gamma\left(k_i - \frac{2}{\alpha_i}\right)\Gamma(k_i) - \Gamma^2\left(k_i - \frac{1}{\alpha_i}\right)} \right]^{-\frac{1}{2}} \tag{A-6}$$

for $x \in \mathcal{X}$, where

$$\rho_{12} = \frac{\mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}} V_2^{-\frac{1}{\alpha_2}}\right) - \mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}}\right)\mathbf{E}\left(V_2^{-\frac{1}{\alpha_2}}\right)}{\prod_{i=1}^{2}\left[\mathbf{E}\left(V_i^{-\frac{2}{\alpha_i}}\right) - \left[\mathbf{E}\left(V_i^{-\frac{1}{\alpha_i}}\right)\right]^2\right]^{\frac{1}{2}}}. \tag{A-7}$$

Note that by Lemma 1 we get $\Gamma\left(k_i - \frac{2}{\alpha_i}\right)\Gamma(k_i) - \Gamma^2\left(k_i - \frac{1}{\alpha_i}\right) > 0$ for $\varepsilon_1 = k_i - \frac{2}{\alpha_i}$ and $\varepsilon_2 = \varepsilon_3 = \frac{1}{\alpha_i}$, with $k_i > \frac{2}{\alpha_i}$. Given also that $\delta(\alpha_i) > 1$ for each $\alpha_i > 0$, our problem reduces to bound from below, for fixed marginals G_1, G_2, the numerator of (A-7).

Denote by $\mathbf{E}_l$ the expectation with respect to the probability measure $\max\{G_1(v_1) + G_2(v_2) -$

$1, 0\}$. By using the formula for the covariance and employing Hoeffding's identity we get

$$\mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}} V_2^{-\frac{1}{\alpha_2}}\right) \geq \mathbf{E}_l\left(V_1^{-\frac{1}{\alpha_1}} V_2^{-\frac{1}{\alpha_2}}\right). \tag{A-8}$$

The mapping $\omega \mapsto (\omega)^{-1}$ is strictly convex and thus Jensen's inequality entails

$$\mathbf{E}_l\left(V_1^{-\frac{1}{\alpha_1}} V_2^{-\frac{1}{\alpha_2}}\right) \geq \left[\mathbf{E}_l\left(V_1^{\frac{1}{\alpha_1}} V_2^{\frac{1}{\alpha_2}}\right)\right]^{-1}, \tag{A-9}$$

which together with (A-8) implies

$$\mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}} V_2^{-\frac{1}{\alpha_2}}\right) \geq \left[\mathbf{E}_l\left(V_1^{\frac{1}{\alpha_1}} V_2^{\frac{1}{\alpha_2}}\right)\right]^{-1} \tag{A-10}$$

For $G = \max\{G_1(v_1) + G_2(v_2) - 1, 0\}$, the random vector (V_1, V_2) is NQD, which in turn gives that the $(V_1^{\frac{1}{\alpha_1}}, V_2^{\frac{1}{\alpha_2}})$ is NQD as well due to the fact that $V_i^{\frac{1}{\alpha_i}}$ is strictly increasing transformation of V_i for $i = 1, 2$. Using again the formula of the covariance and Hoeffding's identity we get

$$\mathbf{E}_l\left(V_1^{\frac{1}{\alpha_1}} V_2^{\frac{1}{\alpha_2}}\right) \leq \mathbf{E}\left(V_1^{\frac{1}{\alpha_1}}\right)\mathbf{E}\left(V_2^{\frac{1}{\alpha_2}}\right). \tag{A-11}$$

Therefore, combining (A-10) and (A-11) we deduce

$$\mathbf{E}\left(V_1^{-\frac{1}{\alpha_1}} V_2^{-\frac{1}{\alpha_2}}\right) \geq \left[\mathbf{E}\left(V_1^{\frac{1}{\alpha_1}}\right)\mathbf{E}\left(V_2^{\frac{1}{\alpha_2}}\right)\right]^{-1}. \tag{A-12}$$

For $m < k_i$, the $m-$th moment of V_i and V_i^{-1} is given by

$$\mathbf{E}\left(V_i^m\right) = \frac{\Gamma\left(k_i + m\right)}{\Gamma\left(k_i\right)}\mu_i^m, \ \mathbf{E}\left(V_i^{-m}\right) = \frac{\Gamma\left(k_i - m\right)}{\Gamma(k_i)}\mu_i^{-m}. \tag{A-13}$$

Hence, use of the formulas (A-6), (A-7), (A-12) and (A-13) for $m = \frac{1}{a_i}$ and $m = \frac{2}{a_i}$ and some algebra yields the thesis of the proposition. ■

Proof of Proposition 3. For $i = 1, 2$ the ratio within the brackets in the formula of $\rho(T_1, T_2|x)$, see (A-6), can be rewritten as $\frac{1}{\mathcal{F}(y_i(k), \alpha_i^{-1}) - 1}$ for $y_i(k) = k - \frac{2}{\alpha_i}$, where $\mathcal{F}(\varepsilon_1, \varepsilon_2) = \frac{\Gamma(\varepsilon_1)\Gamma(\varepsilon_1 + 2\varepsilon_2)}{\Gamma^2(\varepsilon_1 + \varepsilon_2)}$, $\varepsilon_1 > 0, \varepsilon_2 > 0$. We first show that $\mathcal{F}(\varepsilon_1, \varepsilon_2)$ is strictly decreasing in ε_1 for each $\varepsilon_2 \in (0, \infty)$, which in turn will imply that $\mathcal{F}(y_i(k), \alpha_i^{-1})$ is strictly decreasing in k for any positive α_i. Taking the logarithm of $\mathcal{F}(\varepsilon_1, \varepsilon_2)$ and then differentiating with respect to ε_1 we obtain

$$\frac{\vartheta \log \mathcal{F}(\varepsilon_1, \varepsilon_2)}{\vartheta \varepsilon_1} = \psi(\varepsilon_1) + \psi(\varepsilon_1 + 2\varepsilon_2) - 2\psi(\varepsilon_1 + \varepsilon_2), \tag{A-14}$$

Differentiating $\frac{\vartheta \log \mathcal{F}(\varepsilon_1,\varepsilon_2)}{\vartheta \varepsilon_1}$ with respect to ε_2 it follows

$$\frac{\vartheta}{\vartheta \varepsilon_2}\left[\frac{\vartheta \log \mathcal{F}(\varepsilon_1, \varepsilon_2)}{\vartheta \varepsilon_1}\right] = 2\psi^{(1)}(\varepsilon_1 + 2\varepsilon_2) - 2\psi^{(1)}(\varepsilon_1 + \varepsilon_2). \tag{A-15}$$

Clearly, $\psi^{(2)}(\varepsilon) \leq 0$ for $\varepsilon > 0$, which in turn implies $\frac{\vartheta}{\vartheta \varepsilon_2}\left[\frac{\vartheta \log \mathcal{F}(\varepsilon_1,\varepsilon_2)}{\vartheta \varepsilon_1}\right] \leq 0$ for all $(\varepsilon_1, \varepsilon_2) \in (0,\infty)^2$ by using the (A-5). Hence, given that $\frac{\vartheta \log \mathcal{F}(\varepsilon_1,0)}{\vartheta \varepsilon_1} = 0$ it follows $\frac{\vartheta \log \mathcal{F}(\varepsilon_1,\varepsilon_2)}{\vartheta \varepsilon_1} \leq 0$ for all $(\varepsilon_1, \varepsilon_2) \in (0,\infty)^2$. Therefore, given that $\rho(T_1, T_2|x)$ is strictly decreasing in $\mathcal{F}(y_i(k), \alpha_i^{-1})$ for $\rho_{12} > 0$, it follows that it is strictly decreasing in k for every positive α_i, and consequently, for $\rho_{12} = 1$, $k \to \max\{\frac{2}{\alpha_1}, \frac{2}{\alpha_2}\} = \frac{2}{\alpha_2}$, and by continuity of $\Gamma(.)$ the bound is obtained. By Lemma 1, we have $\log \mathcal{F}(\varepsilon_1, \varepsilon_2) > 0$ for all $\varepsilon_1 > 0, \varepsilon_2 > 0$ and thus $\Gamma\left(\frac{2(\alpha_1-\alpha_2)}{\alpha_1\alpha_2}\right)\Gamma\left(\frac{2}{\alpha_2}\right) - \Gamma^2\left(\frac{2\alpha_1-\alpha_2}{\alpha_1\alpha_2}\right) > 0$ for all $\alpha_1, \alpha_2 > 0$,with $\alpha_1 > \alpha_2$. Using also the property $\delta(\alpha_1) > 1$ for each $\alpha_1 > 0$, the inequality $b_{gu}(\alpha_1, \alpha_2) < b_u(\alpha_1, \alpha_2)$ is shown. ■

Appendix B

In this appendix we provide a brief discussion about the Archimedean family, the FGM family of copulas, and their corresponding properties. The Archimedean family is constructed according to $C_\psi(\omega_1, \omega_2) = \xi_\psi^{[-1]}(\xi_\psi(\omega_1) + \xi_\psi(\omega_2))$ with $\xi_\psi : [0,1] \to [0,\infty), \xi_\psi'(\omega) < 0, \xi_\psi''(\omega) > 0$ for each $\omega \in (0,1)$ and $\xi_\psi(1) = 0$. The function $\xi_\psi^{[-1]}(\omega)$ is called pseudo-inverse and is equal to $\xi_\psi^{-1}(\omega)$ if $\omega < \xi_\psi(0)$ and 0 elsewhere. In case $\xi_\psi^{-1}(\omega) = \xi_\psi^{[-1]}(\omega)$ for every $\omega \in [0,\infty)$, both the copula and the respective generator are called strict. The case of $\xi_\psi(\omega) = -\ln\omega$ corresponds to independence between the underlying random variables. Nelsen (2006) describes this important class of copulas. We first describe the three most popular copulas which belong to the Archimedean family.

Clayton Copula: For $\xi_\psi(\omega) = \frac{1}{\psi}(\omega^{-\psi} - 1)$ we obtain the Clayton copula which is given by

$$C_\psi(\omega_1, \omega_2) = \max\left\{\left(\omega_1^{-\psi} + \omega_2^{-\psi} - 1\right), 0\right\}^{-\frac{1}{\psi}}, \ \psi \in [-1, \infty)\backslash 0. \tag{B-1}$$

If $\psi \in [-1, 0)$ the C_ψ is NQD and for every $\psi \in (0, \infty)$ the C_ψ is PQD. Additionally, $C_{-1}(\omega_1, \omega_2)$ $= \max\{\omega_1 + \omega_2 - 1, 0\}, \lim_{\psi\to\infty} C_\psi(\omega_1, \omega_2) = \min\{\omega_1, \omega_2\}$ and $\lim_{\psi\to 0} C_\psi(\omega_1, \omega_2) = \omega_1\omega_2$ for every $(\omega_1, \omega_2) \in [0, 1]^2$. Note that $\lim_{\psi\to -1} C_\psi(\omega_1, \omega_2) \neq \max\{\omega_1 + \omega_2 - 1, 0\}$, which implies that C_ψ is not right-continuous at -1.

Frank Copula: If we apply $\xi_\psi(\omega) = -\ln \frac{e^{-\psi\omega}-1}{e^{-\psi}-1}$ as generator, we get the Frank copula

$$C_\psi(\omega_1, \omega_2) = -\frac{1}{\psi} \ln\left[1 + \frac{(e^{-\psi\omega_1} - 1)(e^{-\psi\omega_2} - 1)}{e^{-\psi} - 1}\right], \ \psi \in (-\infty, \infty)\backslash 0. \tag{B-2}$$

For any $\psi \in (-\infty, 0)$ the C_ψ is NQD and for every $\psi \in (0, \infty)$ the C_ψ is PQD. Additionally, $\lim_{\psi\to-\infty} C_\psi(\omega_1, \omega_2) = \max\{\omega_1+\omega_2-1, 0\}, \lim_{\psi\to\infty} C_\psi(\omega_1, \omega_2) = \min\{\omega_1, \omega_2\}$ and $\lim_{\psi\to 0} C_\psi(\omega_1, \omega_2)$ $= \omega_1\omega_2$ for all $(\omega_1, \omega_2) \in [0, 1]^2$.

Gumbel Copula: For $\xi_\psi(\omega) = (-\ln \omega)^\psi$ we get the Gumbel copula which is expressed as

$$C_\psi(\omega_1, \omega_2) = \exp\left[-\left((-\ln \omega_1)^\psi + (-\ln \omega_2)^\psi\right)^{\frac{1}{\psi}}\right], \ \psi \in [1, \infty). \tag{B-3}$$

The $C_\psi(\omega_1, \omega_2)$ is PQD for any $\psi \in (1, \infty)$. Moreover, $\lim_{\psi\to\infty} C_\psi(\omega_1, \omega_2) = \min\{\omega_1, \omega_2\}$, $C_1(\omega_1, \omega_2) =$ $\omega_1\omega_2$ for any $(\omega_1, \omega_2) \in [0, 1]^2$.

Finally, another copula that we could employ for parameterizing G is the Farlie-Gumbel-Morgenstern (FGM) copula.

Farlie-Gumbel-Morgenstern Copula: This family of distributions is expressed as

$$C_\psi(\omega_1, \omega_2) = \omega_1\omega_2 + \psi\omega_1\omega_2(1 - \omega_1)(1 - \omega_2), \ \psi \in [-1, 1]. \tag{B-4}$$

If $\psi \in [-1, 0)$ the C_ψ is NQD, if $\psi \in (0, 1]$ the C_ψ is PQD, and $C_0(\omega_1, \omega_2) = \omega_1\omega_2$ for any $(\omega_1, \omega_2) \in [0, 1]^2$.

References

Abbring, J. H. and Ridder, G. (2009), "A note on the non-parametric identification of generalized accelerated failure-time models," *Unpublished manuscript, Tilburg University.*

Abbring, J. H. and Van Den Berg, G. J. (2007), "The unobserved heterogeneity distribution in duration analysis," *Biometrika*, 94, 87–99.

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

Andrén, D. (2005), "Never on a Sunday: Economic incentives and short-term sick leave in Sweden," *Applied Economics*, 37, 327–338.

Beaudoin, D., Duchesne, T., and Genest, C. (2007), "Improving the estimation of Kendall's tau when censoring affects only one of the variables," *Computational Statistics & Data Analysis*, 51, 5743–5764.

Benítez-Silva, H. and Heiland, F. (2008), "Early claiming of social security benefits and labour supply behaviour of older Americans," *Applied Economics*, 40, 2969–2985.

Cherian, F. (1941), "A bivariate gamma type distribution function," *Journal of the Indian Mathematical Society*, 5, 133–144.

Clayton, D. G. (1978), "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika*, 65, 141–151.

Cui, S. and Sun, Y. (2004), "Checking for the gamma frailty distribution under the marginal proportional hazards frailty model," *Statistica Sinica*, 14, 249–267.

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

Embrechts, P., McNeil, A., and Straumann, D. (2002), "Correlation and dependence in risk management: properties and pitfalls," *Risk management: value at risk and beyond*, 176–223.

Gorfine, M. and Hsu, L. (2011), "Frailty-based competing risks model for multivariate survival data," *Biometrics*, 67, 415–426.

Henderson, R. and Shimakura, S. (2003), "A serially correlated gamma frailty model for longitudinal count data," *Biometrika*, 90, 355–366.

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

Hujer, R., Thomsen, S. L., and Zeiss, C. (2006), "The effects of short-term training measures on the individual unemployment duration in West Germany," *Center for European Economic Research Discussion Paper*.

Joe, H. (1997), *Multivariate models and dependence concepts*, vol. 73, Chapman & Hall/CRC.

Jonker, M., Bhulai, S., Boomsma, D., Ligthart, R., Posthuma, D., and Van Der Vaart, A. (2009), "Gamma frailty model for linkage analysis with application to interval-censored migraine data," *Biostatistics*, 10, 187–200.

Korsgaard, I. R. and Andersen, A. H. (1998), "The additive genetic gamma frailty model," *Scandinavian Journal of Statistics*, 25, 225–269.

Kotz, S., Johnson, N., and Balakrishnan, N. (2000), *Continuous multivariate distributions: models and applications*, vol. 1, Wiley-Interscience.

Lehmann, E. L. (1966), "Some concepts of dependence," *The Annals of Mathematical Statistics*, 1137–1153.

Mardia, K. (1970), *Families of bivariate distributions*, Griffin London.

Martin, E. and Betensky, R. (2005), "Testing quasi-independence of failure and truncation times via conditional Kendall's tau," *Journal of the American Statistical Association*, 100, 484–492.

Nelsen, R. (2006), *An introduction to copulas*, Springer Verlag.

Oakes, D. (2008), "On consistency of Kendall's tau under censoring," *Biometrika*, 95, 997–1001.

Paik, M. C., Tsai, W. Y., and Ottman, R. (1994), "Multivariate survival analysis using piecewise gamma frailty," *Biometrics*, 975–988.

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

Røed, K. and Westlie, L. (2007), "Unemployment insurance in welfare states: Soft constraints and mild sanctions," *Department of Economics, University of Oslo, Memorandum.*

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., and Soubeyran, P. (2007), "Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events," *Biostatistics*, 8, 708–721.

Shih, J. H. and Louis, T. A. (1995a), "Assessing gamma frailty models for clustered failure time data," *Lifetime Data Analysis*, 1, 205–220.

— (1995b), "Inferences on the association parameter in copula models for bivariate survival data," *Biometrics*, 1384–1399.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges." *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231.

Tatsiramos, K. (2004), "The effect of unemployment insurance on unemployment duration and the subsequent employment stability," *IZA Discussion Paper No. 1163.*

Tchen, A. (1980), "Inequalities for distributions with given marginals," *The Annals of Probability*, 8, 814–827.

Van Den Berg, G. J. (1997), "Association measures for durations in bivariate hazard rate models," *Journal of Econometrics*, 79, 221–245.

Vu, H. T. V. (2004), “Estimation in semiparametric conditional shared frailty models with events before study entry,” *Computational statistics & data analysis*, 45, 621–637.

Vu, H. T. V. and Knuiman, M. W. (2002), “A hybrid ML-EM algorithm for calculation of maximum likelihood estimates in semiparametric shared frailty models,” *Computational statistics & data analysis*, 40, 173–187.

Wang, W., Wells, M. T., et al. (2000), “Estimation of Kendall's tau under censoring,” *Statistica Sinica*, 10, 1199–1216.

Xiang, L., Yau, K., Tse, S., and Lee, A. (2007), “Influence diagnostics for random effect survival models: Application to a recurrent infection study for kidney patients on portable dialysis,” *Computational Statistics & Data Analysis*, 51, 5977–5993.

Zhong, X. and Li, H. (2002), “An additive genetic gamma frailty model for two-locus linkage analysis using sibship age of onset data,” *Statistical Applications in Genetics and Molecular Biology*, 1, 1–24.

A Causal Trivariate Hazard Model: Investigating the Sales-Price Crash of New Products

1 Introduction

The first iPhone was introduced on September 2007 at a price of \$599. However, just 66 days after its launch, Apple decided to drop the iPhone's price by \$200 and set its new price to \$399. It was hypothesized that Apple's price cut was a response to sluggish sales (BusinessWeek Online, 2007; Wired Magazine, 2007). How likely are firms to cut new product prices after experiencing a sales slowdown? Are such dramatic price reductions effective for halting sales slowdowns?

In this paper we propose a novel triple hazard model that is suitable for studying the endogenous relationship between three events, and simultaneously, for identifying the causal effects among them. That is, our model is useful for identifying how the hazard rate of one event is affected causally by the occurrence of preceding events. Current developments of hazard models incorporate association (correlation) among the timing of multiple events. For example, Chintagunta and Haldar (1998) propose a copula approach (Danaher and Smith, 2011) to model the joint density of the purchase timing of two related product categories. More recently, Park and Fader (2004) develop a duration model that allows for observed and unobserved association across the timing of visits to multiple websites. Finally, Schweidel et al. (2008) propose a bivariate duration model that allows correlation between consumers' acquisition and retention times. Investigating dependence across multiple events is academically and managerially important (Chintagunta and Dong, 2006) because very often marketers need to analyze multiple dependent duration variables.

Moreover, marketers have the need to evaluate the effectiveness of their marketing mix and to uncover whether marketing actions causally lead to expected outcomes such as higher revenues, lower consumer churn (Braun and Schweidel, 2011), or lower inter-purchase timing (Chintagunta, 1998). In this respect, econometric models with "treatment effects" (or "causal effects") come in handy because they are useful for identifying causal relationships among variables or events and they have been applied, for example, for evaluating policy interventions like new regulations, job assistance programs, or new medical treatments. Specifically in the context of duration analysis, models with causal effects have been used in health and labor studies. For example, Van Ours (2003) and Van Ours and Williams (2009) test whether the timing of the first use of cannabis is a gateway to the later use of harder drugs or to school drop-out. Abbring et al. (2005) analyze whether the timing of unemployment sanctions and unemployment are causally linked. Heckman and Borjas (1980) test whether current unemployment causes future unemployment. Finally Van den Berg and Drepper (2011) studies the causal interdependence between the survival rate of twins.

We develop a new hazard model that incorporates both causal effects and three dependent duration variables and we apply our method for studying the sales crash, the price crash, and the sales recovery of new products. Two central research questions in our case study are whether price crashes are an effective tool for deferring or halting a sales crash, and whether a sales crash has a clear effect on the timing of price crashes. In addition, we also model the timing of a modest sales recovery that usually follows the price crash and we test whether price is causally linked to the recovery.

Previous marketing literature has addressed the relationship between prices and the timing of turning point in sales of new products. For example, Golder and Tellis (1997) find that a 1% decrease in the price of a new product is associated with a 4.5% increase in the probability of its takeoff while Golder and Tellis (2004) find that a 1% decrease in price is associated with a 5% decrease in the probability of a slowdown in the sales of a new product. The main limitation of previous studies lays in their assumption that price is an exogenous factor affecting the timing of turning point in sales. We address this limitation by allowing both statistical correlation and causality between the timing of important turning points in the sales and prices of new products.

We also contribute to the marketing literature by documenting the heterogeneity of the sales and price crashes. Previous studies have documented extensively the heterogeneity in diffusion patterns across countries and product categories (Talukdar et al., 2001; Tellis et al., 2003; Van Everdingen et al., 2009) and the plausible determinants of inflection points in sales, like the sales takeoff (Chandrasekaran and Tellis, 2008; Tellis et al., 2003; Agarwal and Bayus, 2002; Golder and Tellis, 1997) or the sales saddle (Chandrasekaran and Tellis, 2011). However, crash points have not been documented before and they are a critical moment in the diffusion of new products that needs further exploration. For instance, Peres et al. (2010), Hauser et al. (2006), and Golder and Tellis (1997) call for further research regarding the heterogeneity and determinants of slowdowns in the sales of new products.

Slowdowns, or saddle points, are crucial because revenues may be reduced dramatically at these inflection points and managers need to plan or modify their marketing strategies accordingly (Chandrasekaran and Tellis, 2011; Van Everdingen et al., 2009; Van den Bulte, 2000; Goldenberg et al., 2002; Vakratsas and Kolsarici, 2008). Moreover, sales slowdowns are of utmost importance for deciding upon the pricing policy of fast moving consumer goods and high-tech products, both typically having very short life cycles after launch. That is, our empirical application is of great relevance for marketing managers in the high-tech, fast-fashion, and information good industries.

The plan of the paper is as follows. In Section 2 we describe our data and the phenomena that we study. In Section 3 we present our model, its parametrization, and the intuition behind its identification. In Section 4 we present estimation results. Finally, in Section 5 we summarize our findings and discuss managerial implications and further research topics. The model identification results, its likelihood, and estimation method are introduced in the Appendices A, B, and C.

2 Defining the Big Crash

One of the most popular video-game titles in our dataset is the video-game “Star Wars: Shadow of the Empire”. This title sold 300 thousand units at its market introduction while by its sixth month it was selling less than 50 thousand units. Would the timing of this dramatic sales decrease, its sales crash, have a causal impact on the timing of its price crash? Could its price crash aid in the recovery of its sales?

We make use of a market-level "natural experiment" that is occurring in the market of video-games. In this experiment we observe market outcomes in terms of the timing of the sales crash, the timing of the price crash, and the timing of the sales recovery. We may observe, for example, the realization of a sales crash at time t_s and the subsequent realization of a price crash at a time t_p (with $t_p > t_s$). A second plausible market outcome consists of observing the realization of a price crash followed by the realization of a sales crash, hence observing $t_p < t_s$. The realization of a sales recovery may be any time after observing a sales crash at t_s and it may occur before or after a price crash. In summary, every time a new video game is introduced to the market, we may observe the triad t_s, t_p, and t_r, being these latter realizations of the underlying random variables T_s, T_p, and T_r. These market outcomes represent a natural experiment because whenever an event occurs it may affect the hazard rate of subsequent, yet not observed, events. Our data consists of 1562 realizations of these three duration variables measured in months after introduction. We will refer to this market phenomenon as the *big crash* because it consists of a permanent reduction in both sales and price levels.

We define the *sales crash* as the moment after the first sales peak when the sales decrease rate is closest to zero. That is, the timing of the sales crash is marked by the moment when sales stop falling dramatically and stabilize at a new level. Note that we do not define the sales crash based on the rates of the sales drop but our definition led us to identify sales drops that are deep (63% decrease on average) and that occurred at a very fast rate (within 3 to 6 months). Hence, the *sales crash* is very distinct from the *sales saddle* that is usually a gradual and not so deep decay in sales (Chandrasekaran and Tellis, 2011; Goldenberg et al., 2002) that occurs long after a product's introduction. Next, we define *sales recovery* as the sales peak that follows the sales crash and that is larger than 10% of the total sales. We measure the time to *sales recovery* with the time spanning in between the sales crash and the sales recovery. Finally, we define the *price crash* as the moment when the largest price cut occurs regardless of its depth. Our definitions closely follow the literature on saddles. For example, Goldenberg et al. (2002) define the saddle as a decline of 10% or 20% in sales that follows an initial sales peak and they define the start of the saddle as the moment just before the sales drop. Goldenberg et al. (2002) define the end of the saddle, or recovery, as the time when sales recover their initial peak level and they focus their study on saddles that have a duration of more than two years. Likewise, Chandrasekaran and Tellis (2011) define the start of the saddle as the first drop in sales that is greater than 10% and that lasts more than 2 years. Finally, Golder and Tellis (2004) define the start of the slowdown as the moment when sales fall for two consecutive years after takeoff. Like the saddle studies, we used heuristic rules and visual inspection to identify inflection points.[1]

The identification of events is done through the analysis of the monthly time series of market sales (in units) and average market prices of video-games sold in the US from 1995 to 2001. These data cover three major video-game platforms, nine product categories, eighty five firms and seven years (1995-2001). The coverage of products reaches approximately 65% of the retailers in the US market (sampled by NPD Group) and the products in the sample represent more than 400 million units and 14 billion dollar sales. In addition, we count with quality ratings for each of the products. We also count with 48 observed characteristics for each video game and these are its platform, publisher, quality, season and year of launch, and its genre. Given the nature of the data, our analysis will focus on market-level outcomes. Such causal analysis at the market level is common in the economics literature, for examples see Kaminsky and Reinhart (1999) and Schnabel (2004).

In Figure 1 we illustrate the identification of the sales crash, the sales recovery, and the price crash for three of the most popular video-games in our data. On the left side of the panel (a) we observe the unit sales of the title "Grand Turismo" and how its sales go from 600 hundredth thousand at introduction and then drop to less than 50 thousand by its fifth month in the market. It is the fifth month when sales stabilize around a new low level and that is where we mark the sales crash. The sales recovery is the first sales peak that comes next to the sales crash. We notice that the sales recovery may be due to a completely exogenous factor, like seasonality, and we control for it in our modeling approach. On the right side of the panel (a) we plot the price of "Grand Turismo" and we identify the price crash at the moment when the largest price cut happens. Finally, on the panels (b) and (c), we illustrate the definition of the sales-price crash for the titles "Star Wars" and "Crash Bandicoot", respectively.

[1]Our tests indicate that other alternative and simple rules of identification are linear transformations of our original measures and therefore our results remain without significant change.

The data reveals that the price crash occurs before the sales crash 73% of the time while it occurs after the sales crash 27% of the time (see Table 1). The price crash occurs before the sales recovery 90% of the time whereas it occurs after the recovery 10% of the time. The sales recovery is unobserved for 63% of the products in the sample and it is observed for the remaining 37%. Finally, the depth of the sales crash is close to 60% while the depth of the price crash is close to 23%. These last two figures remain fairly the same across firms, genres, and years (see Table 2).

3 Model

We introduce a continuous-time triple hazard model which permits three important sources of variation in the timing of events that may arise from i) observed covariates, ii) unobserved and correlated heterogeneity, and iii) causality between events. Our model is based on the bivariate hazard model of Abbring and Van den Berg (2003). In basic terms, the model developed by Abbring and Van den Berg (2003) concerns two parallel duration variables and the authors identify the distribution of the correlated unobserved heterogeneity between the corresponding events and the *unidirectional* causal effect of one event (*the treatment*) on a second duration outcome.

The model we study permits *bidirectional* causal effects between two pairs of parallel durations and it also incorporates causal dependence between sequential duration variables. The first pair of parallel durations are the sales crash timing and the price crash timing while the second pair consists of the sales recovery timing and the price crash timing. The sequential duration variables are the sales crash and the sales recovery. Figure 2 illustrates how our empirical application involves these two pairs of parallel durations and one pair of sequential events.

Our model assumes that events cannot be perfectly anticipated and hence we analyze the effect of a focal event on others only after the focal event is realized. Abbring and Van den Berg (2003) also need this assumption. Assuming no anticipation means that market participants cannot predict and expect with extreme accuracy the timing when the market events occur. Van Heerde et al. (2005) argue that no anticipation is plausible in fast-moving consumer goods and they also discuss guidelines for evaluating when such assumption may be reasonable. In our empirical application we are concerned with an industry that had 345 new product introductions per year, that is almost one introduction per day. Thus, it is likely that market events across such large array of products cannot be fully anticipated even though market participants may be aware that price cuts or sales crashes may occur. In spite of the no anticipation assumption, we control for the correlation among the timing of market events and hence any endogenous relationship between the price crash, the sales crash, and sales recovery arising from unobservables (like the strategic interaction of consumers and firms) is accounted for.

The model consists of three stochastic duration variables: i) $T_p|t_s, t_r, x, v_p$, ii) $T_s|t_p, x, v_s$ and iii) $T_r|t_s, t_p, x, v_r$; where the notation $\mathcal{T}|\mathcal{F}$ denotes the stochastic variable $\mathcal{T}$ conditional on the information set $\mathcal{F}$. The marginal distributions of these stochastic durations are characterized by the corresponding *structural hazard rates* $\ddot{\lambda}_p(t|t_s, t_r, x, v_p)$, $\ddot{\lambda}_s(t|t_p, x, v_s)$ and $\ddot{\lambda}_r(t|t_p, t_s, x, v_r)$. The structural hazard rate $\ddot{\lambda}_j(t|.)$, for $j \in \{p, s, r\}$, represents the speed at which the event of interest may occur at time t given that it has not occurred before, and also given other information which are associated with the occurrence of this specific event. That is, the three relevant hazard rates can be mathematically expressed as follows

$$\begin{aligned} \ddot{\lambda}_p(t|t_s, t_r, x, v_p) &= (dt)^{-1}\,\mathbf{P}(t \le T_p < t + dt|T_p \ge t, t_s, t_r, x, v_p) \\ \ddot{\lambda}_s(t|t_p, x, v_s) &= (dt)^{-1}\,\mathbf{P}(t \le T_s < t + dt|T_s \ge t, t_p, x, v_s) \\ \ddot{\lambda}_r(t|t_p, t_s, x, v_r) &= (dt)^{-1}\,\mathbf{P}(t \le T_r < t + dt|T_r \ge t, t_s, t_r, x, v_r) \end{aligned} \tag{1}$$

for all $t > 0$, where $\mathbf{P}(\cdot)$ denotes a generic probability measure and dt is an infinitesimal time increment. Thus, $\ddot{\lambda}_p(t|t_s, t_r, x, v_p)$ is the probability of observing a price crash at time t given that is has not occurred yet and given the information set t_s, t_r, x, and v_p. Similarly, $\ddot{\lambda}_s(t|t_p, x, v_s)$ and $\ddot{\lambda}_r(t|t_p, t_s, x, v_r)$ denote the probability of observing a sales crash or a recovery at time t given they have not occurred before. Note that each product i is characterized by a vector of observable characteristics x being the realization of $X \subset \mathbf{R}^d$ (in our empirical example $d = 48$), and by a vector of unobserved correlated factors $(v_p, v_s, v_r) \in \mathbf{R}^3_+$ which represent time-invariant unobserved or non-measurable explanatory variables and they are the realization of the random variables V_p, V_s, and V_r. The sub-index i is omitted from all variables and functions for exposition purposes.

Next we describe how we model the relationship among T_p, T_s, and T_r and how we model their structural hazards. The aim of the model is to identify and to measure how the hazard rate of each duration variable is affected by the occurrence of two possibly related events. Specifically, our model incorporates four time-varying causal effects and one time-invariant causal effect. The time-varying effects are: i) the effect of the realization of the sales crash on the hazard rate of the price crash, ii) the effect of the price crash on the hazard rate of the sales crash, iii) the effect of the price crash realization on the hazard rate of the sales recovery, and iv) the effect of sales recovery on the hazard rate of the price crash. The time-invariant effect measures the effect of the realization of the sales crash on the sales recovery.

The causal effects are incorporated as follows. We assume that the hazard rate of the event of interest changes deterministically after the realization of one and/or the two other events. Specifically, we assume that the hazard rate of the sales crash timing T_s at t increases/decreases by a factor of $\delta_s^p(t|t_p)$ after $t > t_p$ that is after the price crash occurs. In a similar way, we assume that the hazard rate of the price crash timing T_p increases/decreases at t by a factor of $\delta_p^s(t|t_s)$ after the sales crash occurs that is when $t > t_s$, and by a factor of $\delta_p^r(t|t_s, t_r)$ after the sales recovery occurs that is when $t > t_s + t_r$. In addition, we assume that the hazard rate of the sales recovery timing T_r increases/decreases at t by a factor of $\delta_r^p(t|t_p, t_s)$ when the price crash occurs after the occurrence of the sales crash but before the recovery, that is when $t_s < t_p < t_s + t_r$. Finally, we assume that if the price crash occurs before the sales crash, then the price crash affects the sales recovery indirectly through its effect on the sales crash. Note that hazard rates are strictly positive and hence the functions δ must also be positive. If $0 < \delta < 1$ the hazard rates decrease, $\delta > 1$ implies an increase in the hazard rates, whereas $\delta = 1$ implies that the hazard rates is not affected by the occurrence of the other events. In the model, the functions δ consists of time-varying coefficients and hence we measure the time-varying effect of the "treatment" event on the hazard rate of the two remaining events. We discuss all the parametrization details and the intuition behind these functions in Section 3.1.

Formally, we adopt the following specification for the structural hazard rates

$$\begin{aligned}
\ddot{\lambda}_p(t|t_s, t_r, x, v_p) &= \lambda_p(t)\phi_p(x)\delta_p^s(t|t_s)^{1\{t>t_s\}}\delta_p^r(t|t_s, t_r)^{1\{t>t_s+t_r\}}v_p \\
\ddot{\lambda}_s(t|t_p, x, v_s) &= \lambda_s(t)\phi_s(x)\delta_s^p(t|t_p)^{1\{t>t_p\}}v_s \\
\ddot{\lambda}_r(t|t_s, t_p, x, v_r) &= \lambda_r(t)\phi_r(x)\delta_r^s(t_s)\delta_r^p(t|t_p, t_s)^{1\{t_s<t_p<t_s+t\}}v_r,
\end{aligned} \tag{2}$$

where $1\{.\}$ is an indicator function that is equal to one if the event within the brackets occurs and zero otherwise. The λ_j is the baseline hazard for $j \in \{p, s, r\}$ and it captures the duration dependence, that is, the effect of time on the likelihood of the occurrence of the event of interest. The quantity ϕ_j is called regressor function and it captures the effect of observed covariates in the overall level of the hazard rates. The "causal effects functions" δ are commonly known in the literature as "treatments" and, as described in the previous paragraph, they capture the causal effects among the duration variables. The variables v_s, v_p and v_r refer to unmeasurable or unobserved (to the researcher) characteristics. Note that the hazard rate corresponding to the sales recovery is directly affected by the timing of the sales crash by way of the function δ_r^s. This type of dependence is known as lagged causal dependence (Van den Berg, 2001; Heckman and Borjas, 1980). Notice that lagged dependence cannot exist between the duration variables T_p and T_r as they are parallel to each other and as a consequence, information about their realization in not available at the beginning of their respective duration process.

3.1 Model Parametrization

In this section we discuss the parametrization of the baseline hazard, the regressor functions, and the causal effect functions introduced which are the main elements of the model in Equation (2).

We assume that the baseline hazards follow an Expo-power specification. Namely, we have for $t > 0$, $\alpha_j > 0$ and $\gamma_j \in \mathbf{R}$, where $j \in \{p, s, r\}$,

$$\begin{aligned}
\lambda_p(t; \alpha_p, \gamma_p) &= \alpha_p t^{\alpha_p - 1} e^{\gamma_p t}, \\
\lambda_s(t; \alpha_s, \gamma_s) &= \alpha_s t^{\alpha_s - 1} e^{\gamma_s t}, \\
\lambda_r(t; \alpha_r, \gamma_r) &= \alpha_r t^{\alpha_r - 1} e^{\gamma_r t}.
\end{aligned} \tag{3}$$

We choose an Expo-power specification because it permits several shapes for the baseline hazard and because it is the most flexible function among other alternatives (Seetharaman and Chintagunta, 2003; Saha and Hilton, 1997). In particular, the Expo-power allows for U-shape, inverted U-shape, increasing, and decreasing hazard functions and it nests the Gompertz (when $\alpha_j = 1$) and the Weibull (when $\gamma_j = 0$), two parametrization which have gained great popularity in empirical analysis. See for example Franses and Paap (2004, Chapter 8) or Seetharaman and Chintagunta (2003). Note also that although the Expo-power does not include as a special case the log-logistic, it does allow for inverted U-shape which is the main attractive feature of the latter. In contrast with the log-logistic, the Expo-power has a closed form expression for its integrated hazard function and this greatly facilitates model estimation.

The regressor functions follow an exponential specification which is widely used in empirical analysis. In particular, for $\beta_r^s \in \mathbf{R}$, $\beta_j \in \mathbf{R}^d$, $x \in \mathbf{R}^d$, $j \in \{p, s, r\}$, and $t > 0$ we have that

$$\begin{aligned} \phi_p(x; \beta_p) &= \exp\left(x^{'}\beta_p\right), \\ \phi_s(x; \beta_s) &= \exp\left(x^{'}\beta_s\right), \\ \phi_r(x; \beta_r) &= \exp\left(x^{'}\beta_r\right). \end{aligned} \tag{4}$$

In view of the above definitions, the parameter vector β_j for $j = p, s, r$ measures the effect of the observed covariates on the occurrence time of the respective event.

Next, we define the time-varying causal effects as piece-wise constant functions with K intervals $[Y_l, Y_{l+1}]$, where $l = 0, ..., K-1$, $Y_0 = 0$, and $Y_K = \infty$. A piece-wise specification is highly flexible and it is suitable for our study because we have no prior knowledge about the shape of the causal effects. Notice that as the number of intervals increases, the more "non-parametric" that this specification becomes. Moreover, such piece-wise specifications are commonly applied to measure causal dependence among parallel durations (Van den Berg, 2001) and they were first introduced by Freund (1961). Below we provide further intuition about how these causal functions work.

The causal effect function δ_p^s quantitatively specifies the effect of the sales crash on the price crash and it is parametrized for $t > t_s > 0$ as

$$\delta_p^s(t|t_s;\theta_p^s) = \sum_{l=0}^{K-1} 1\{Y_l \leq t - t_s < Y_{l+1}\} \exp(\theta_{pl}^s). \tag{5}$$

where we make use of an exponential specification to ensure that the causal effect is strictly positive.

The role of the causal effect function in Equation (5) is to select the relevant coefficient l out of K possible coefficients. In this case the function selects the relevant $\exp(\theta_{pl}^s)$ coefficient that affects the structural hazard rate when the treatment duration is equal to $t - t_s$ and when this treatment duration lays in the interval $[Y_l, Y_{l+1})$. In the model implementation we define seven intervals each of two months of length. These are $[0, 2)$, $[2, 4)$,..., and $[12, \infty)$. Notice that the last causal effect is selected whenever the treatment duration is longer than 12 months. For example, in terms of the first equation in (2), if $t < t_s$, then the structural hazard rate of the price crash $\ddot{\lambda}_p(t|\dots)$ would be equal to $\lambda_p(t)\phi_p(x)v_p$. That is, before t_s, the structural hazard rate of the price crash is not affected by the sales crash. In contrast, if we had t laying after the sales crash but before the sales recovery, that is when $t_s < t < t_s + t_r$, then the structural hazard of the price crash $\ddot{\lambda}_p(t|\dots)$ would be equal to $\lambda_p(t)\phi_p(x)\delta_p^s(t|t_s;\theta_p^s)v_p$. Moreover, if the treatment duration $t - t_s$ would be smaller than 2 months, for example, then the structural hazard rate would be equal to $\lambda_p(t)\phi_p(t)\exp(\theta_{pl}^s)v_p$ with $l = 1$. Notice that the sales crash affects the hazard of the price crash only when $t > t_s$. The structural hazard is defined in a similar way when $t > t_s + t_r$, that is when both the sales crash and the sales recovery are observed. In this latter case, two causal functions would affect the structural hazard as defined in Equation (2).

Next, the function δ_p^r corresponds to the effect of the sales recovery on the price crash and it is modeled for $t > t_s + t_d > 0$ as

$$\delta_p^r(t|t_s, t_r;\theta_p^r) = \sum_{l=0}^{K-1} 1\{Y_l \leq t - t_s - t_d < Y_{l+1}\} \exp(\theta_{pl}^r), \tag{6}$$

The function δ_s^p gives the effect of a price crash on the event of sales crash. We parametrize it for $t > t_p > 0$ as

$$\delta_s^p(t|t_p;\theta_s^p) = \sum_{l=0}^{K-1} 1\left\{Y_l \leq t - t_p < Y_{l+1}\right\} \exp\left(\theta_{sl}^p\right). \tag{7}$$

Furthermore, δ_r^p measures the effect of a price crash on the propensity of a sales recovery. We adopt the next parametrization where

$$\delta_r^p(t|t_p, t_s;\theta_r^p) = \sum_{l=0}^{K-1} 1\left\{Y_l \leq t - t_p + t_s < Y_{l+1}\right\} \exp\left(\theta_{rl}^p\right) \tag{8}$$

for $t > t_p - t_s > 0$.

Later we refer to the causal effects as vectors that consist of the all coefficients of the K intervals. That is, $\theta_p^s = (\theta_{p1}^s, \theta_{p2}^s, ..., \theta_{pK}^s)$, $\theta_p^r = (\theta_{p1}^r, \theta_{p2}^r, ..., \theta_{pK}^r)$, $\theta_s^p = (\theta_{s1}^p, \theta_{s2}^p, ..., \theta_{sK}^p)$, $\theta_r^p = (\theta_{r1}^p, \theta_{r2}^p, ..., \theta_{rK}^p)$ to denote the collection of the four sets of causal effects.

The causal effect of the sales crash on the sales recovery is operationalized as

$$\delta_r^s(t_s) \quad = \quad \exp\left(t_s\beta_r^s\right). \tag{9}$$

The parameter β_r^s quantitatively specifies the effect of the sales crash t_s on the hazard rate of the sales recovery. This latter specification was first introduced by Heckman and Borjas (1980) as a way to measure the causal effect of a past event on a second event that always proceeds the first. In our application, the sales recovery can only occur after the sales crash. That is the reason why this causal effect is operationalized different from the others.

Finally, we assume that the random vector (V_p, V_s, V_r) follows a trivariate log-normal distribution, that is,

$$\mathbf{P}\Big(\log(v_s, v_p, v_r)\Big) = (2\pi)^{-3/2}|\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}(v_s, v_p, v_r)\Sigma^{-1}(v_s, v_p, v_r)'\Big), \tag{10}$$

where Σ is the variance-covariance matrix and whose off-diagonal elements give the correlation between the logarithm of the corresponding random variables.

3.2 Model Identification

The identification strategy of the model can be roughly described as follows. First, we identify the variation coming from regressor functions, the baseline hazards, and the distribution of the unobserved terms by considering the whole population regardless of the sequence of the realized events. Secondly, we identify the causal effects. To achieve this, we analyze all sub-populations of market outcomes where the events of interest occur in the same sequential order. For instance, we identify the effect of the price crash on the sales crash based on the sub-population where the price crash precedes the sales crash. In this latter sub-population of cases, we observe the price crash at several timings t_p and we can untangle how the variation in t_p affects the observed variance in the occurrence timing of the sales crash that follows at time t_s. A similar identification strategy applies for all other causal effects. All details about identification are provided in the Appendix A.

4 Results

Table 3 presents the four sets of causal effects coefficients (in Equations (5) to (8)). These are: i) The causal effect coefficients (θ_s^p) of the price crash on the sales crash (upper panel), ii) the causal effect coefficients (θ_r^p) of the price crash on the sales recovery (second panel from the top), iii) the causal effects coefficients (θ_p^s) of the sales crash on the price crash (third panel), and iv) the causal effects coefficients (θ_p^r) of the sales recovery on the price crash (bottom panel). As we can see in the upper panel of Table 3, the signs of the causal effects of the price crash on the sales crash are negative and this implies that the realization of a price crash decreases significantly the hazard rate of the sales crash for all subsequent periods following the price crash. In contrast, on the third panel we notice that the realization of a sales crash increases the hazard rate of the price crash in all following periods after the sales crash time. Notice how the effect between the sales crash and the price crash are asymmetric and that the stronger effect goes from the sales crash to the price crash. In Figure 4 we plot the four sets of causal effects and their time variation. In the upper right panel, we notice how the realization of a price crash increases the hazard rate of a sales crash after the third month that follows the price crash and it stays relatively constant (around 1.3) up to the twelfth month after the price crash. The reverse effect of the recovery on the price crash has no effect on the price crash for most periods as the confidence bounds for most of these last coefficients contain zero. Finally, we report the coefficient of the causal effect of the sales crash on the sales recovery in Table 4 (bottom of third column) together with the other regressor functions coefficients. The -0.213 coefficient for t_s is significant and it implies that the longer it takes the saddle to occur, the less likely the sales recovery becomes.

In Table 4 we present the coefficients of the regressor functions in Equation (4). Our results indicate that there is substantial heterogeneity driven by seasons, platforms, products' genres, publishers, and by quality. Quality has a significant -1.197 coefficient in the sales crash regressor function and this implies that the hazard rate of a sales crash decreases as quality increases. In the same fashion, an increase in quality decreases the hazard rate of a price crash given the -0.863 and -0.205 significant coefficients in the price crash regressor function. A higher quality also predicts a decrease in the hazard rate of the sales recovery given the significant -0.929 and -0.309 coefficients for quality in the regressor function of the sales recovery. This last results is surprising but notice that Goldenberg et al. (2002) also find that the sales of more successful products may fall at faster than less successful ones. Seasons are also important, for example, we find that a sales crash is more likely during the first four months of the year than during the last four months. See how the coefficients first four months (January to February) are either not significantly different from zero or they are smaller than coefficients of the last four months. In contrast, the price crash is more likely during the shopping season of December than in the two months preceding it; see the non-significant -0.084 coefficient for December and the significant -0.098 and -0.162 coefficients for October and November (mid column of Table 4). Several of the platforms, publisher, and genre effects are also significant and they explain observed heterogeneity in the timing of the sales crash, the price crash, and the sales recovery.

In Table 5 we present the coefficient estimates of the baseline hazards (Equation (3)). All coefficients are significantly different from zero with the exception of the γ_r of the sales recovery hazard. This implies that the baseline hazard of the sales recovery follows a Weibull distribution while the baseline hazards of the sales crash and the price crash follow an Expo-power function. Note that the Expo-power hazard reduces to a Gompertz when $\alpha = 1$ and to a Weibull when $\gamma = 0$.

In Figure 5 we present the shape of the three baselines hazards given the parameter estimates. The baseline hazard of the sales crash (top panel) shows an inverted U-shape with a peak at the 10th month after introduction. In contrast, the baseline hazard of the price crash (mid panel) has a decreasing shape and it reaches zero by the second year (24 months) after introduction. This last result implies that that the likelihood of a price crash increases only due to the realization of a sales crash because without such realization the price crash hazard rate would decline with time. Finally, we find that the baseline hazard of the sales recovery (bottom panel) is initially high and that it decays but very mildly and it stays relatively constant after the tenth month.

In Table 6 we report the covariance matrix of the distribution of the random effects (Equation (10)). Our results indicate that there is a strong positive dependence among the random effects and this implies that an unobserved shock would change the hazard rate of all events by a similar amount and by the same direction. In other words, the realization of a random shock delays or advances evenly all events.

5 Discussion

This paper proposes a new causal triple hazard model to estimate the inter-dependence among three events. The model can statistically identify three sources of variation in the timing of events that may arise from i) observed covariates, ii) unobserved correlated covariates, and iii) causality between events. Our identification results prove that the model can uniquely identify among these three sources of variation in the timing of events. Hence, we contribute to the marketing literature by proposing a method for incorporating correlation among the three events and simultaneously for testing whether an event is causally linked to others. This section summarizes the main findings, presents managerial implications, and lists limitations and topics for further research.

5.1 Summary of Findings

- The price crash occurs before the sales crash 73% of the time while it occurs after the sales crash 27% of the time. The depth of the price crash is close to 23% while the depth of the sales crash is close to 60%.

- The occurrence of price crash causally and significantly lowers the hazard rate of a sales crash whereas the occurrence of a sales crash causally and significantly increases the hazard of a price crash. The latter effect is stronger than the former.

- The sales crash is the only determining factor that causally and significantly increases the hazard rate of price crashes given that the baseline hazard of the latter is decreasing in time.

- The price crash causally and significantly increases the hazard rate of a sales recovery whereas the sales recovery has a positive but insignificant effect on the occurrence of a price crash. Products that face a late sales crash are less likely to face a sales recovery.

- The results are valid after controlling for observed heterogeneity that is significantly driven by seasons, quality, platforms, publishers and genres. We also control for unobserved heterogeneity and we find that this latter factor delays or advances evenly all events in time.

5.2 Managerial Implications

Managers could use our method to analyze whether the co-occurrence of events is due to causality among them or due to exogenous marketing variables. For example, the model can provide managers with new insights about the purchase timing of multiple durable goods, the click-stream across multiple websites, or the adoption timing of multiple technologies. Moreover, the data and empirical analyzes seem to indicate that firms will inevitably face sales crashes and most of the time firms wait for the crash before cutting prices. Surprisingly, we find that a sales crash is not completely inevitable and that managers may benefit from analyzing when to cut prices in order to delay such dramatic sales drop. Fast moving consumer goods like mobile phones, fiction novels, and designer's clothes usually sell high at introduction or during their relevant season but a crash may follow the high-season. Thus, managers in these industries may benefit from analyzing whether such dramatic sales drop can be avoided or at least delayed by modifying their marketing policies.

5.3 Limitations and Further Research

Our empirical investigation can lead to further research topics due to its limitations. First, except for the time-varying causal effects among events, we do not control for any time-varying covariates of the marketing mix like advertising or promotion. A model that includes such variables would represent an important and challenging development in causal hazard models. Second, our analyzes focus on two main moments of the sales data (the sales crash and recovery) and a single moment of the price series (the price crash) and the model leaves out all other information about the sales price data. It could be that market participants dynamically update their beliefs about the realization of events as new information is available and we consider that the formation of expectations is an interesting avenue for further research. Third, we use aggregate data and we do not model individual level behavior. We consider that modeling the individual level behavior of consumers and firms facing price-sales crashes is also an interesting avenue for further research. Finally, our results are valid only when market expectations are assumed to be incorporated through the unobserved time-invariant shocks. All these topics are interesting avenues for further research.

6 Tables and Figures

	No. of Products
Total	1725
Total in Sample	1562
No Sales Crash	1
Sales Crash - Unobserved Recovery	996
Sales Crash - Observed Recovery	565
No Price Crash	87
Price Crash	1475
Price Crash Before Sales Crash	417
Price Crash After Sales Crash	1145
Price Crash Before Sales Recovery	1407
Price Crash After Sales Recovery	155

Table 1: Study Sample

	Price Crash	Sales Crash	% Depth of Sales Crash	% Depth of Price Crash
Average by Firm				
ACCLAIM	9.45	3.61	-58.57%	-22.25%
ACTIVISION	7.35	3.77	-61.70%	-24.39%
CAPCOM	7.09	4.27	-59.64%	-25.08%
EIDOS INTERACTIVE	5.61	4.44	-63.77%	-22.24%
ELECTRONIC ARTS	5.82	4.63	-66.15%	-26.71%
HASBRO	7.47	4.04	-66.19%	-25.38%
INFOGRAMES	14.65	4.09	-60.55%	-21.27%
INTERPLAY	13.18	3.37	-65.36%	-20.04%
KONAMI	7.18	3.51	-61.96%	-28.02%
MIDWAY	5.33	4.62	-58.43%	-23.83%
NAMCO	6.91	4.17	-57.18%	-23.11%
NINTENDO	7.67	4.13	-63.70%	-22.56%
SEGA	5.18	4.38	-57.20%	-24.65%
SONY	6.82	3.95	-58.16%	-21.99%
THQ	7.22	3.93	-62.90%	-22.09%
Average by Genre				
ACTION	7.25	3.84	-62.69%	-24.28%
ADVENTURE	9.50	3.20	-55.98%	-22.84%
DRIVING	6.06	4.52	-57.56%	-23.79%
FAMILY	8.95	4.32	-55.11%	-26.29%
FIGHTING	7.84	4.60	-64.01%	-22.66%
SHOOTER	6.92	4.49	-63.83%	-23.02%
SIMULATIONS	7.78	3.69	-58.52%	-24.25%
SPORTS	6.93	4.10	-63.04%	-23.50%
STRATEGY	8.31	3.74	-61.06%	-24.08%
Average by Year				
1995	8.59	4.00	-60.00%	-24.93%
1996	6.86	4.47	-59.59%	-25.06%
1997	7.20	4.04	-61.42%	-23.15%
1998	7.69	3.80	-63.67%	-23.65%
1999	6.89	3.94	-64.01%	-23.81%
2000	7.84	4.12	-59.08%	-23.54%
2001	9.13	4.29	-40.38%	-12.89%

Table 2: Descriptive Statistics of the Timing of Sales and Price Crash

Price Crash → Sales Crash

Index	$[\mathbf{Y_l}, \mathbf{Y_{l+1}})$	θ^p_{sl}	HPDR 95%
$l = 1$	[0-2)	-0.418 **	(-0.543, -0.261)
$l = 2$	[2-4)	-0.529 **	(-0.695, -0.306)
$l = 3$	[4-6)	-0.974 **	(-1.279, -0.678)
$l = 4$	[6-8)	-1.402 **	(-1.82, -0.995)
$l = 5$	[8-10)	-0.811 **	(-1.259, -0.434)
$l = 6$	[10-12)	-0.987 **	(-1.488, -0.390)
$l = 7$	[12-∞)	-0.362	(-1.085, 0.420)

Price Crash → Sales Recovery

Index	$[\mathbf{Y_l}, \mathbf{Y_{l+1}})$	θ^p_{rl}	HPDR 95%
$l = 1$	[0-2)	-0.116	(-0.372, 0.102)
$l = 2$	[2-4)	0.866 **	(0.66, 1.061)
$l = 3$	[4,6)	1.129 **	(0.939, 1.321)
$l = 4$	[6,8)	1.147 **	(0.926, 1.373)
$l = 5$	[8-10)	1.111 **	(0.815, 1.362)
$l = 6$	[10,12)	1.410 **	(1.119, 1.706)
$l = 7$	[12,∞)	-0.247 *	(-0.573, 0.041)

Sales Crash → Price Crash

Index	$[\mathbf{Y_l}, \mathbf{Y_{l+1}})$	θ^s_{pl}	HPDR 95%
$l = 1$	[0,2)	1.08 **	(0.677, 1.236)
$l = 2$	[2,4)	1.774 **	(1.269, 1.931)
$l = 3$	[4,6)	2.202 **	(1.699, 2.423)
$l = 4$	[6,8)	2.741 **	(2.218, 3.010)
$l = 5$	[8-10)	2.769 **	(2.271, 3.107)
$l = 6$	[10,12)	3.003 **	(2.301, 3.461)
$l = 7$	[12,∞)	3.380 **	(2.784, 3.862)

Sales Recovery → Price Crash

Index	$[\mathbf{Y_l}, \mathbf{Y_{l+1}})$	θ^r_{pl}	HPDR 95%
$l = 1$	[0,2)	-0.057	(-0.342, 0.195)
$l = 2$	[2,4)	-0.084	(-0.453, 0.274)
$l = 3$	[4,6)	0.554 **	(0.172, 0.944)
$l = 4$	[6,8)	0.850 **	(0.401, 1.302)
$l = 5$	[8-10)	0.288	(-0.494, 1.033)
$l = 6$	[10,12)	0.025	(-0.961, 1.129)
$l = 7$	[12,∞)	1.004 **	(0.204, 1.643)

Notes: The interval limits $[Y_l, Y_{l+1})$ are defined in months.

Table 3: Estimates of Causal Effects

Coefficient	Sales Crash β_s	Price Crash β_p	Sales Recovery β_r
Intercept	0.000	0.000	0.000
PLATFORM			
Saturn	-0.159 **	-0.139 *	0.005
Nintendo 64	0.013	-0.058	-0.15 **
Multiplatform (PS + Saturn)	-0.177 **	-0.017	-0.12 **
Multiplatform (PS + N64)	-0.144 **	-0.039	-0.158 **
GAME GENRE			
Action	-0.213 **	-0.095 **	-0.128 *
Adventure	-0.055	-0.176 **	-0.098
Driving	-0.311 **	-0.066	-0.102 **
Family	-0.295 **	-0.115 *	-0.085
Fighting	-0.127 *	-0.211 **	-0.215 **
Shooter	-0.319 **	-0.042	-0.131 **
Simulations	-0.029	-0.303 **	-0.233 **
Strategy	-0.173 **	-0.128 *	-0.010
FIRM			
ACCLAIM	-0.350 **	-0.047	-0.154 *
ACTIVISION	-0.069	0.051	0.088
CAPCOM	0.030	-0.072	-0.037
EIDOS INTERACTIVE	-0.144 *	0.058	-0.085
ELECTRONIC ARTS	-0.078	-0.048	-0.099
HASBRO	-0.133 *	-0.008	-0.044
INFOGRAMES	-0.227 **	0.035	0.030
INTERPLAY	-0.136 *	-0.053	0.018
KONAMI	-0.116	-0.117 *	-0.194 **
MIDWAY	-0.140 **	-0.126 *	-0.135 *
NAMCO	-0.063	-0.239 **	-0.102
NINTENDO	0.031	-0.257 **	-0.014
SEGA	-0.012	-0.063	0.001
Small Publisher 1	-0.293 **	-0.203 **	-0.222 **
Small Publisher 2	-0.157 **	-0.031	-0.106 *
Small Publisher 3	-0.195 **	-0.113 *	-0.191 **
THQ	-0.109	-0.082	-0.134 *
YEAR			
1996	-0.480 **	-0.167 **	-0.189 **
1997	-0.368 **	-0.184 **	-0.244 **
1998	-0.254 **	-0.175 *	-0.357 **
1999	-0.332 **	-0.064	-0.331 **
2000	-0.480 **	-0.324 **	-0.400 **
2001	-0.219 **	-0.223 *	-0.032
LAUNCH MONTH			
February	-0.075	-0.221 **	-0.378 **
March	-0.167 **	-0.19 **	0.027
April	-0.064	-0.004	-0.045
May	-0.203 **	-0.186 **	-0.01
June	-0.044	-0.176 **	-0.529 **
July	-0.296 **	-0.104	-0.089
August	-0.142 **	-0.18 **	-0.049
September	-0.262 **	-0.094	-0.263 **
October	-0.431 **	-0.098 *	-0.321 **
November	-0.282 **	-0.162 **	-0.098 *
December	-0.211 **	-0.084	-0.049
QUALITY			
Quality	-1.197 **	-0.863 **	-0.929 **
Quality^2	0.043	-0.205 **	-0.309 **
t_s	–	–	-0.213 **

*,** mean that 0 is not included in the 90% and 95% Highest Posterior Density Regions (HPDR), respectively. $^+$ The base platform is the PlayStation, the base genre is Sports, the base firm is Sony, the base year is 1995 and the base month is January. Intercept is equal to zero for identification.

Table 4: Coefficients in Regressor Functions

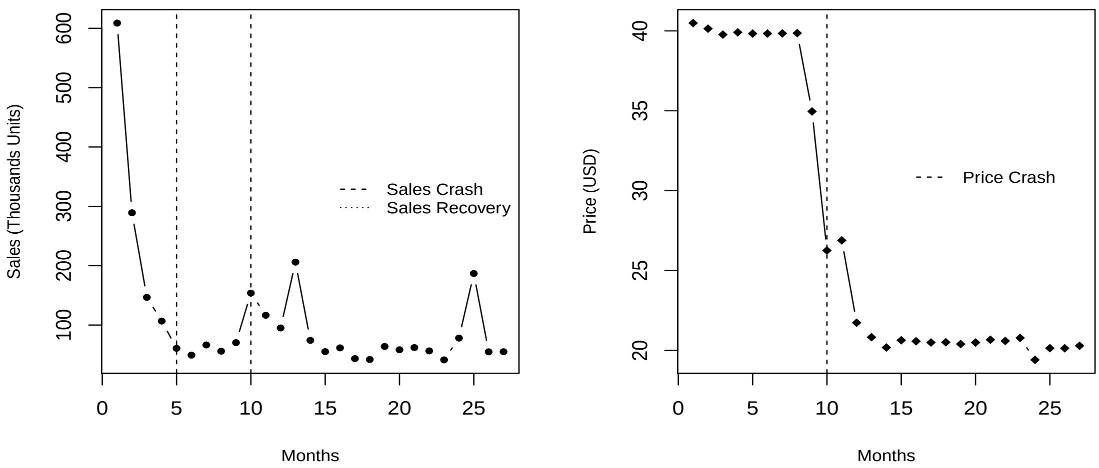
	Coefficient	HPDR 95%
Sales Crash		
α_s	1.916 **	(1.490, 1.967)
γ_s	-0.007 **	(-0.010, -0.004)
Price Crash		
α_p	0.917 **	(0.870, 1.356)
γ_p	-0.241 **	(-0.272, -0.07)
Sales Recovery		
α_r	0.649 **	(0.589, 0.733)
γ_r	0.008	(-0.035, 0.064)

Table 5: Coefficients of Expo-Power Baseline Hazard Functions

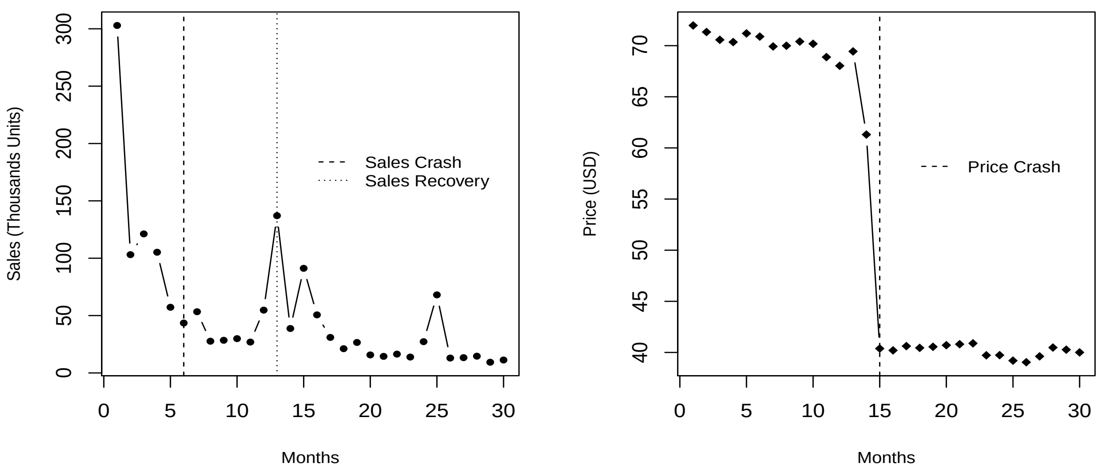
Covariance Matrix			
	v_s	v_p	v_r
v_s	0.0642 **	0.0785 **	0.0731 **
v_p		0.1365 **	0.1116 **
v_r			0.1023 **
Correlation Matrix			
	v_s	v_p	v_r
v_s	1	0.8958 **	0.9439 **
v_p		1	0.9727 **
v_r			1

*,** mean that 0 is not included in the 90% and 95% Highest Posterior Density Regions (HPDR), respectively.

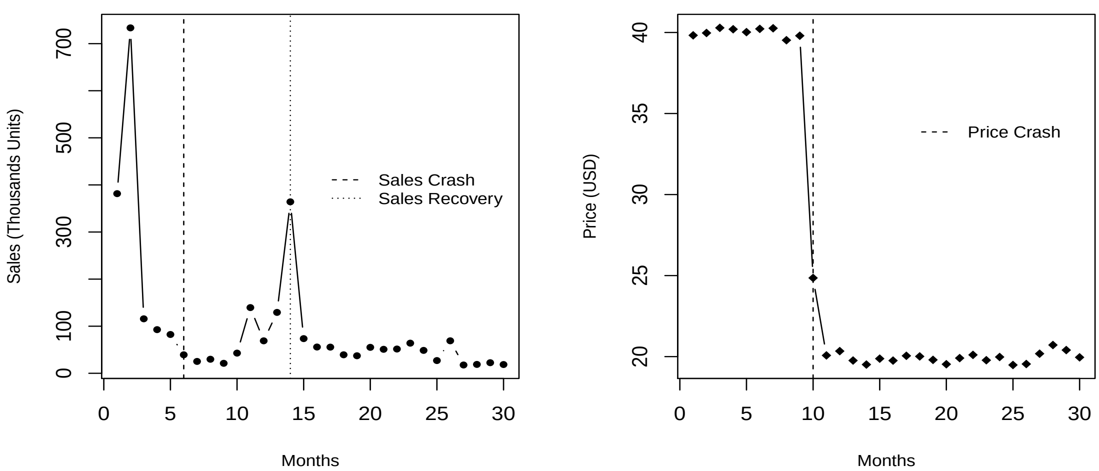
Table 6: Covariance and Correlation Matrix (Distribution Random Effects)



(a) **Events for Grand Turismo**



(b) **Events for Star Wars**



(c) **Events for Crash Bandicoot**

Figure 1: Events Definitions

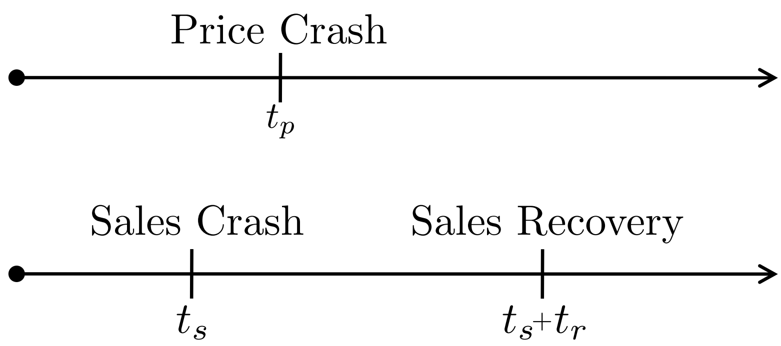


Figure 2: Triple Durations

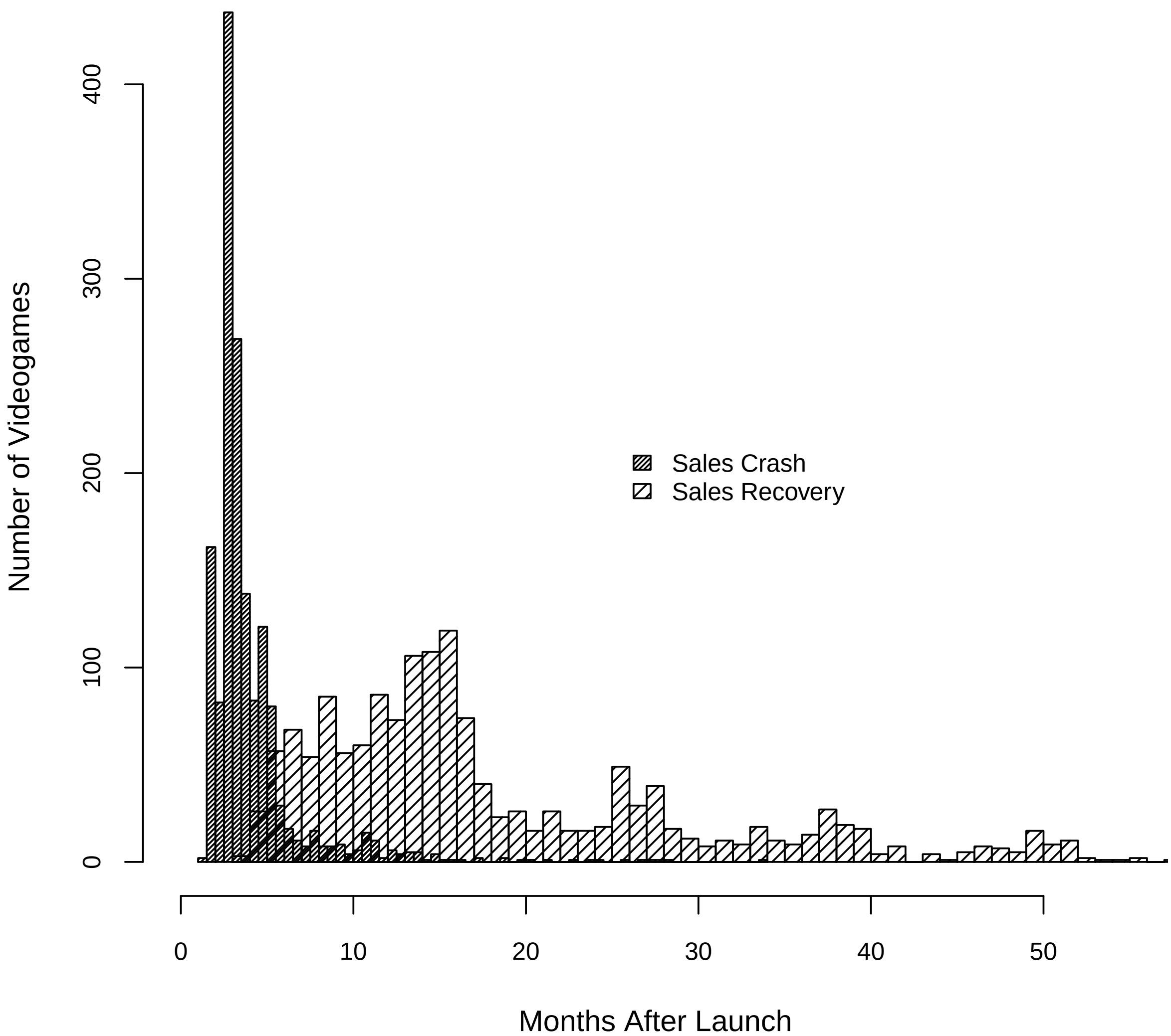


Figure 3: Histograms of the Sales Crash and Sales Recovery Times

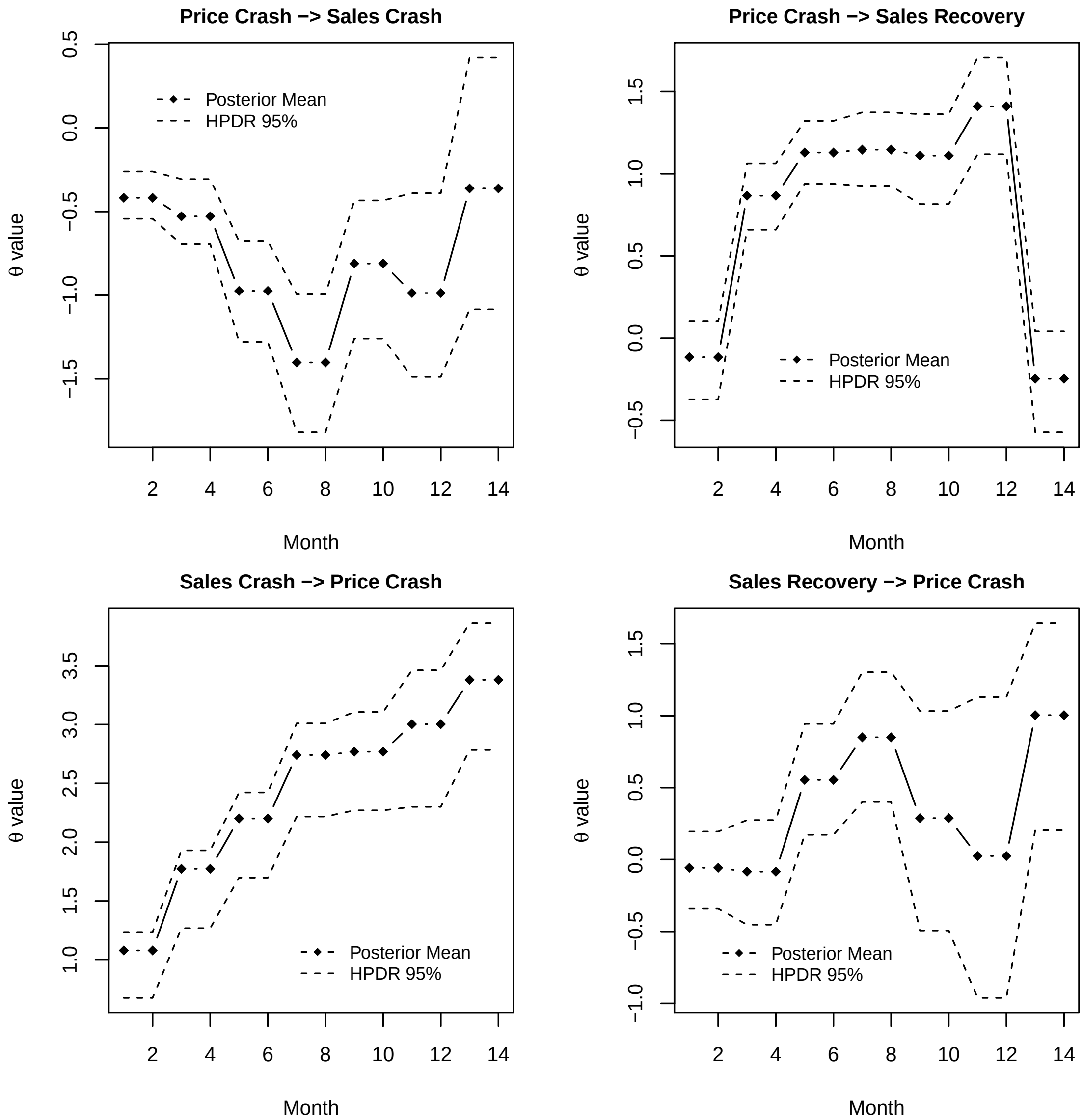


Figure 4: Estimates of the Time-varying Causal Effects

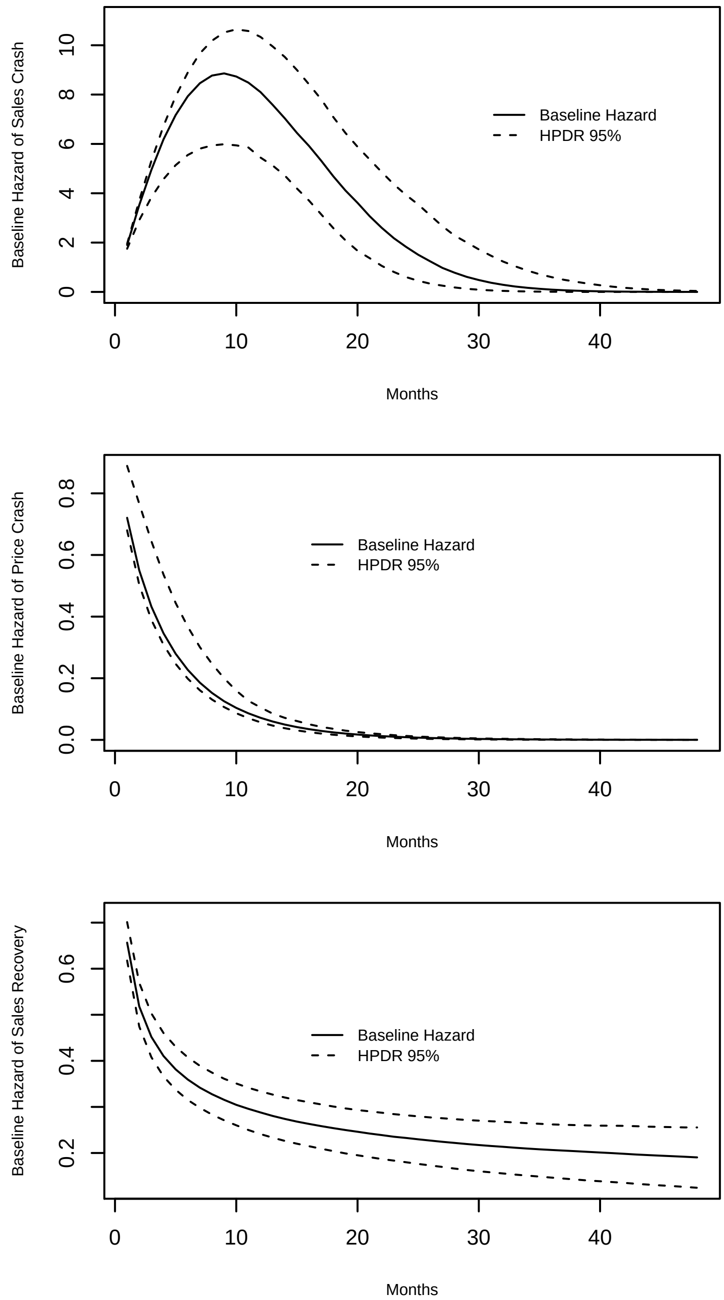


Figure 5: Baseline Hazards

A Identification Proof

The proof is organized as follows. First, we introduce some notation that is used throughout the proof and the appendix. Second, we describe the assumptions that the model needs for identification. Finally, we prove seven propositions that identify all of the model structural elements. These elements are i) the regressor functions and ii) the integrated hazard functions (Proposition 1, 2, and 4), iii) the distribution of the unobserved heterogeneity (Proposition 3), and finally iv) the integrated causal effects functions (Proposition 1, 4, 5, 6, and 7).

A.1 Definitions and notation

The integrated baseline hazards are defined as

$$\begin{aligned}
\Lambda_p(t) :=&\ \ \textstyle\int_0^t \lambda_p(\omega)d\omega \\
\Lambda_s(t) :=&\ \ \textstyle\int_0^t \lambda_s(\omega)d\omega \\
\Lambda_r(t) :=&\ \ \textstyle\int_0^t \lambda_r(\omega)d\omega.
\end{aligned} \tag{A-1}$$

for $t > 0$. Next, we define the sub-survival functions

$$\begin{aligned}
Q_p(t_p, t_s, t_r|x) =&\ \ \mathbf{P}(T_p > t_p, T_s > t_s, T_r > t_r, T_s > T_p|x) \\
Q_{sp}(t_p, t_s, t_r|x) =&\ \ \mathbf{P}(T_p > t_p, T_s > t_s, T_r > t_r, T_s < T_p < T_s + T_r|x) \\
Q_s(t_p, t_s, t_r|x) =&\ \ \mathbf{P}(T_p > t_p, T_s > t_s, T_r > t_r, T_s < T_p|x) \\
Q_r(t_p, t_s, t_r|x) =&\ \ \mathbf{P}(T_p > t_p, T_s > t_s, T_r > t_r, T_s + T_r < T_p|x)
\end{aligned} \tag{A-2}$$

for all $t_p, t_s, t_r > 0$ and any $x \in \mathcal{X}$. The first sub-survival function refers to the sub-population where the price crash occurs before the sales crash. The second sub-survival function concerns the sub-population where the price crash is realized after the sales crash but before the sales recovery. The third sub-survival function refers to the sub-population in which the sales crash precedes the price crash. Finally, the last sub-survival function is about the sub-population where the price crash occurs after the realization of the sales recovery and consequently after the sales crash.

We define the integrated version of the product between the structural hazard functions and the causal effects functions for $t_p, t_s, t_r > 0$ as

$$\begin{array}{lll} \Upsilon_s^p(t_s|t_p) & := \int_{t_p}^{t_s} \lambda_j(\omega)\delta_s^p(\omega|t_p)d\omega, & \text{for } t_s > t_p, \\ \Upsilon_p^s(t_p|t_s) & := \int_{t_s}^{t_p} \lambda_j(\omega)\delta_p^s(\omega|t_s)d\omega, & \text{for } t_p > t_s, \\ \Upsilon_p^r(t_p|t_s,t_r) & := \int_{t_s+t_r}^{t_p} \lambda_p(\omega)\delta_p^s(\omega|t_s)\delta_p^r(\omega|t_s,t_r)d\omega, & \text{for } t_p > t_s+t_r, \\ \Upsilon_r^p(t_r|t_p,t_s) & := \int_{t_p-t_s}^{t_r} \lambda_r(\omega)\delta_r^p(\omega|t_p,t_s)d\omega, & \text{for } t_r > t_p-t_s > 0. \end{array} \tag{A-3}$$

The integrated causal effects are defined for $t_p, t_s, t_r > 0$ as

$$\begin{array}{lll} \Delta_s^p(t_s|t_p) := & \int_{t_a}^{t_b} \delta_s^p(\omega|t_p)d\omega, & \text{for } t_s > t_p, \\ \Delta_p^s(t_p|t_s) := & \int_{t_a}^{t_b} \delta_p^s(\omega|t_s)d\omega, & \text{for } t_p > t_s, \\ \Delta_p^r(t_p|t_s,t_r) := & \int_{t_s+t_r}^{t_p} \delta_p^r(\omega|t_s,t_r)d\omega, & \text{for } t_p > t_s+t_r, \\ \Delta_r^p(t_r|t_p,t_s) := & \int_{t_p-t_s}^{t_r} \delta_r^p(\omega|t_p,t_s)d\omega. & \text{for } t_r > t_p-t_s > 0. \end{array} \tag{A-4}$$

Finally, we make use of the Laplace Transform $\mathcal{L}_G$ of the trivariate random vector (V_p, V_s, V_r) which is defined as

$$\mathcal{L}_G(\omega_1, \omega_2, \omega_3) := \int_{\mathbf{R}_+^3} \exp(-\omega_1 v_1 - \omega_2 v_2 - \omega_3 v_3) dG(v_1, v_2, v_3) \tag{A-5}$$

for $(\omega_1, \omega_2, \omega_3) \in \mathbf{R}_+^3$.

A.2 Assumptions

Assumption 1 *The functions $\phi_p : \mathcal{X} \rightarrow (0,\infty), \phi_s : \mathcal{X} \rightarrow (0,\infty), \phi_r : \mathcal{X} \rightarrow (0,\infty), \delta_r^s : \mathbf{R}_+ \rightarrow (0,\infty)$ are continuous with $\phi_p(x^*) = \phi_s(x^*) = \phi_r(x^*) = \delta_r^s(t^*) = 1$ for some priory chosen $x^* \in \mathcal{X}$ and $t^* > 0$.*

Assumption 2 *It holds $\Lambda_p(t^*) = \Lambda_s(t^*) = \Lambda_r(t^*) = 1$ for some priory chosen $t^* > 0$, and $\Lambda_r(t_r)\delta_r^s(t_s) \rightarrow 0$ as $t_r \rightarrow 0$ and $t_s \rightarrow 0$.*

Assumption 3 *The vector of the regressor functions $(\phi_p(x), \phi_s(x); x \in \mathcal{X})$ attain all values in a nonempty open subset of $(0,\infty)^2$. There exists $\tilde{x} \in \mathcal{X}$, with $\tilde{x} \neq x^*$, such that $\phi_r(\tilde{x}) \neq \phi_r(x^*)$.*

Assumption 4 *The distribution of the trivariate random vector (V_p, V_s, V_r) is G. It also holds $\mathbf{E}(V_p) < \infty$, $\mathbf{E}(V_s) < \infty$, $\mathbf{E}(V_r) < \infty$.*

Assumption 5 *The causal effects functions $\delta_s^p : \mathbf{R}_+^2 \to (0, \infty)$, $\delta_p^s : \mathbf{R}_+^2 \to (0, \infty)$, $\delta_p^r : \mathbf{R}_+^3 \to (0, \infty)$, and $\delta_r^p : \mathbf{R}_+^3 \to (0, \infty)$ are such that the quantities defined in (A-3) and (A-4) exist and are finite.*

Assumption 1 and 2 impose mild smoothness conditions and normalizations about the functions ϕ_j for $j \in \{p, s, r\}$. It also states normalizations about the function δ_r^s and the integrated baseline hazard rates $\Lambda_p, \Lambda_s, \Lambda_r$. These normalizations imply that the model is identified up to scale normalizations. The statement $\Lambda_r(t_r)\delta_r^s(t_s) = 0$ for $t_r \to 0$ and $t_s \to 0$ is needed for identification of ϕ_r and it is satisfied if conventional parametric approaches are adopted such as exponential specification for the δ_r^s and expo-power specification for the λ_r. Note that Assumption 2 allows $\lim_{t\to\infty} \Lambda_j(t) < \infty$ for $j \in \{p, s, r\}$, which permits the underlying distribution to be defective, namely, the probability that the event of interest is not realized is strictly positive.

Assumption 3 ensures that we can independently vary ϕ_p, ϕ_s when x ranges over $\mathcal{X}$. This assumption is needed to identify the baseline hazards and the regressor functions of the hazard rates corresponding to the price crash as well as the sales crash. On the other hand, we do not make such an assumption about ϕ_r as we can exploit continuous variation in t_s which enters the hazard rate of the sales recovery. A necessary condition for Assumption 3 to hold is at least two components of the x to have continuous variation. For ease of the presentation, we assume that all the components of x are continuous. However, our identification result is valid even in case some of the components of x have discrete variation such that Assumption 3 remains true.

Assumption 4 guarantees that the expectation of the non-negative random variables V_p, V_s, V_r are finite. This is a standard assumption in duration analysis and is necessary for identification of this type of models. Assumption 5 imposes the condition that all the corresponding defined quantities are finite.

A.3 Propositions

Proposition 1 *Let Assumptions 1-5 hold. Then, the regressor functions ϕ_p, ϕ_s, the integrated baseline hazards Λ_p, Λ_s, and the integrated causal effect Δ_s^p are identified.*

This proposition postulates identification of the regressor functions and the integrated baseline hazards which correspond to the price crash and the sales crash. It also deals with the identification of the integrated version of the function which describes the effect of the price crash on the sales crash. Its proof is next.

Proof. This result is a direct consequence of Proposition 2 of Abbring and Van den Berg (2003). In particular, T_p, T_s are two parallel duration variables and as soon as the former is realized, the hazard of the latter is affected which is the same model setup of Abbring and Van den Berg (2003). ∎

Proposition 2 *Let Assumptions 1-5 hold. Then, the regressor function ϕ_r is identified.*

This proposition concerns the identification of the regressor function of the structural hazard rate corresponding to the sales recovery. Its proof is next.

Proof. For almost all (t_p, t_s, t_r), with $0 < t_p < t_s$, $t_r > 0$, and $x \in \mathcal{X}$ we have

$$\begin{aligned} &- \frac{\partial^3}{\partial t_p \partial t_s \partial t_r} Q_p(t_p, t_s, t_r|x) \\ &= \lambda_p(t_p)\phi_p(x)\lambda_s(t_s)\phi_s(x)\delta_s^p(t_s|t_p)\lambda_r(t_r)\phi_r(x)\delta_r^s(t_s) \\ &\times \mathcal{L}_G^{(psr)}(\phi_p(x)\Lambda_p(t_p), \phi_s(x)\left(\Lambda_s(t_s) + \Upsilon_s^p(t_p|t_s)\right), \phi_r(x)\Lambda_r(t_r)\delta_r^s(t_s)), \end{aligned} \tag{A-6}$$

where $\mathcal{L}_G^{(psr)}(\omega_1, \omega_2, \omega_3) = \partial^3 \mathcal{L}_G(\omega_1, \omega_2, , \omega_3)/\partial\omega_1\partial\omega_2\partial\omega_3$ for $(\omega_1, \omega_2, \omega_3) \in \mathbf{R}_+^3$.

Given that $\phi_p(x^*) = \phi_s(x^*) = \phi_r(x^*) = 1$ (Assumption 1), we can show that

$$\lim_{t_p \to 0, t_s \to 0, t_r \to 0} \frac{\frac{\partial^3}{\partial t_p \partial t_s \partial t_r} Q_p(t_p, t_s, t_r|x)}{\frac{\partial^3}{\partial t_p \partial t_s \partial t_r} Q_p(t_p, t_s, t_r|x^*)} = \phi_p(x)\phi_s(x)\phi_r(x). \tag{A-7}$$

By definition of the corresponding quantities (see the paragraph following the discussion of the Assumptions), the left hand side of the above equation is non-parametrically identified from the data. Moreover, the ϕ_p, ϕ_s have been identified (by Proposition 1) and therefore the ϕ_r is identified on its entire support $\mathcal{X}$. ∎

Proposition 3 *Let Assumptions 1-5 hold. Then, the trivariate distribution G is identified.*

This proposition deals with the identification of the distribution of the unobserved heterogeneity which generates correlated effects among the underlying duration variables. To prove the result, we use a similar methodology to the proof of Theorem 3 of Honoré (1993) for identifying the functions of interest. In particular, Honoré (1993) focuses on the identification of bivariate duration models in which the duration variables are successive. Hence, given that T_r always follows T_s we adopt the same approach.

Proof. Define the open connected sets $(a_1, a_2) \subset \mathbf{R}_+$ and $(b_1, b_2) \subset \mathbf{R}_+$, with $t^* \in (b_1, b_2)$ and $a_2 < b_1$. Define also the function $\chi : (a_1, a_2) \times (b_1, b_2) \to \mathcal{X}_{\mathcal{S}} \subset \mathcal{X}$ such that the quantities $\phi_p(\chi(t_p, t_s))\Lambda_p(t_p)$ and $\phi_s(\chi(t_p, t_s))(\Lambda_s(t_p)+\Upsilon_s^p(t_s|t_p))$ remain constant for any $(t_p, t_s) \in (a_1, a_2) \times (b_1, b_2)$. Thus, for almost all $(t_p, t_s) \in (a_1, a_2) \times (b_1, b_2)$, $(\hat{t}_p, t^*) \in (a_1, a_2) \times (b_1, b_2)$ and $x \in \mathcal{X}$, we get

$$\lim_{t_r \to 0} \frac{\frac{\partial^3}{\partial t_p \partial t_s \partial t_d} Q_p(t_p, t_s, t_r|\chi(t_p, t_s))}{\frac{\partial^3}{\partial \hat{t}_p \partial t^* \partial t_d} Q_p(\hat{t}_p, t^*, t_r|\chi(\hat{t}_p, t^*))} = \frac{\phi_p(\chi(t_p, t_s))\lambda_s(t_s)\phi_s(\chi(t_p, t_s))\delta_s^p(t_s|t_p)\delta_r^s(t_s)}{\phi_p(\chi(\hat{t}_p, t^*))\lambda_s(t^*)\phi_s(\chi(\hat{t}_p, t^*))\delta_s^p(t^*|\hat{t}_p)\delta_r^s(t^*)}. \tag{A-8}$$

Recall that we have already identified $\phi_p, \phi_s, \Lambda_s$, and Δ_s^p (Proposition 1). Identification of Λ_s and Δ_s^p yields almost everywhere identification of λ_s and δ_s^p, respectively. Hence, by using the normalization $\delta_r^s(t^*) = 1$ (Assumption 1), we can show identification of δ_r^s on the open set (b_1, b_2).

Additionally, for almost all t_p, t_r such that $0 < t_p < t_s$, all $t_r > 0$, and $x \in \mathcal{X}$, we have that

$$\begin{aligned} \frac{\partial^2}{\partial t_p \partial t_s} Q_p(t_p, t_s, t_r|x) &= \lambda_p(t_p)\phi_p(x)\lambda_s(t_s)\phi_s(x)\delta_s^p(t_s|t_p) \\ &\times \mathcal{L}_G^{(ps)}(\phi_p(x)\Lambda_p(t_p), \phi_s(x)\left(\Lambda_s(t_s) + \Upsilon_s^p(t_p|t_s)\right), \phi_r(x)\Lambda_r(t_r)\delta_r^s(t_s)), \end{aligned} \tag{A-9}$$

where $\mathcal{L}_G^{(ps)}(\omega_1, \omega_2, \omega_3) = \partial^2 \mathcal{L}_G(\omega_1, \omega_2, , \omega_3)/\partial\omega_1\partial\omega_2$ for $(\omega_1, \omega_2, \omega_3) \in \mathbf{R}_+^3$. Recall that $\Lambda_r(t^*) = 1$ and thus

$$\begin{aligned} \frac{\partial^2}{\partial t_p \partial t_s} Q_p(t_p, t_s, t^*|x) &= \lambda_p(t_p)\phi_p(x)\lambda_s(t_s)\phi_s(x)\delta_r^s(t_s|t_p) \\ &\times \mathcal{L}_G^{(ps)}(\phi_p(x)\Lambda_p(t_p), \phi_s(x)\left(\Lambda_s(t_p) + \Upsilon_s^p(t_p|t_s)\right), \phi_r(x)\delta_r^s(t_s)). \end{aligned} \tag{A-10}$$

The left hand side of the above equation is directly identified from the data. The quantity $\mathcal{L}_G^{(ps)}$ on the right hand side is unknown. By Propositions 1 and 2, $\phi_p, \Lambda_p, \phi_s, \Lambda_s, \Delta_s^p$, and ϕ_r have been identified. Identification of $\Lambda_p, \Lambda_s, \Delta_s^p$ implies almost everywhere identification of $\lambda_p, \lambda_s, \delta_s^p$, respectively. Also, δ_r^s is known only on (b_1, b_2). By varying appropriately (t_p, t_s) over $(a_1, a_2) \times (b_1, b_2)$ we can then identify the function $\mathcal{L}_G^{(ps)}$ on an nonempty open subset of $\mathbf{R}_+^3$ and consequently on the whole $\mathbf{R}_+^3$ due to its real analytical property. Recall that $\mathcal{L}_G^{(ps)}$ is the corresponding mixed partial derivative of $\mathcal{L}_G$ .Hence, we can also identify the $\mathcal{L}_G$. Identification of the latter yields identification of G due to the one to one relationship between the Laplace Transform and some certain cumulative distribution function. ■

Proposition 4 *Let Assumptions 1-5 hold. Then, the integrated baseline hazard Λ_r and the function δ_r^s are identified. Its proof is next.*

The above proposition establishes the identification of the integrated baseline hazard which corresponds to the structural hazard rate of the sales recovery. It also gives the identification of the effect of the timing of the sales crash on the structural hazard rate of the sales recovery.

Proof. By the proof of Proposition 3, we know that for almost all t_p, t_s such that $0 < t_p < t_s$, all $t_r > 0$, and $x \in \mathcal{X}$,

$$\frac{\partial^2}{\partial t_p \partial t_s} Q_p(t_p, t_s, t_r|x) = \lambda_p(t_p)\phi_p(x)\lambda_s(t_s)\phi_s(x)\delta_s^p(t_s|t_p)$$

$$\times \mathcal{L}_G^{(ps)}(\phi_p(x)\Lambda_p(t_p), \phi_s(x)\left(\Lambda_s(t_s) + \Upsilon_s^p(t_p|t_s)\right), \phi_r(x)\Lambda_r(t_r)\delta_r^s(t_s)), \quad \text{(A-11)}$$

where $\mathcal{L}_G^{(ps)}$ is defined in the proof of Proposition 3. The left hand side is observed from the data. We set $t_s = t^*$ and recall that $\delta_r^s(t^*) = 1$ as Assumption 1 states. By Propositions $1, 2, 3$, the quantities $\phi_p, \Lambda_p, \phi_s, \Lambda_s, \Delta_s^p, \phi_r$, and G have been identified. Identification of $\Lambda_p, \Lambda_s, \Delta_s^p$ yields almost everywhere identification of $\lambda_p, \lambda_s, \delta_s^p$, respectively, and identification of G gives identification of $\mathcal{L}_G^{(ps)}$. Hence, for $t_s = t^*$ all the terms on the right hand side of the above equations are known except for Λ_r. Given the fact that the $\mathcal{L}_G^{(ps)}$ is strictly monotonic in its arguments, we can identify Λ_r from the above equation. By using the latter result and analogous arguments to the identification of Λ_r, we can identify for δ_r^s from (A-11) on its whole support. This completes the proof. ■

Proposition 5 *Let Assumptions 1-5 hold. Then, the integrated causal effect of the sales crash on the price crash Δ_p^s is identified.*

Proof. For almost all $t_s > 0$, every $t > 0$ and $x \in \mathcal{X}$ we obtain,

$$\begin{aligned}&\frac{\partial}{\partial t_s} Q_s(t + t_s, t_s, t|x) \\ &= \lambda_s(t_s)\phi_s(x)\mathcal{L}_G^{(s)}(\phi_p(x)(\Lambda_p(t_s) + \Upsilon_p^s(t + t_s|t_s)), \phi_s(x)\Lambda_s(t_s), \phi_r(x)\delta_r^s(t_s)\Lambda_r(t))),\end{aligned} \tag{A-12}$$

where $\mathcal{L}_G^{(s)}(\omega_1, \omega_2, \omega_3) = \partial\mathcal{L}_G(\omega_1, \omega_2, \omega_3)/\partial\omega_2$ for $(\omega_1, \omega_2, \omega_3) \in \mathbf{R}_+^3$. The left hand side is directly observed from the data. Moreover, $\phi_p, \Lambda_p, \phi_s, \Lambda_s, \phi_r, \Lambda_r, \delta_r^s$,and G have been identified (Propositions $1, 2, 3, 4$). Identification of $\Lambda_p, \Lambda_s, \Lambda_r$ yields almost everywhere identification of $\lambda_p, \lambda_s, \lambda_r$, respectively. Also, we know $\mathcal{L}_G$ and consequently $\mathcal{L}_G^{(s)}$ as we have identified G (i.e., use the one to one relationship between a cumulative distribution function and a Laplace Transform). Given that $\mathcal{L}_G^{(s)}$ is strictly monotonic function in its first argument the identification of Υ_p^s follows. Recall the definition $\Delta_p^s(t|t_s) = \int_0^t \frac{\partial\Upsilon_p^s(t|t_s)}{\partial\omega} [\lambda_p(\omega)]^{-1} d\omega$. Given that we have identified Λ_p and consequently λ_p almost everywhere, the identification of Δ_p^s directly follows. ■

Proposition 6 *Let Assumptions 1-5 hold. Then, the integrated causal effect of the sales recovery on the price crash Δ_p^r is identified.*

Proof. For almost all $t_s, t_r > 0$, every $t_p > 0$, with $t_s + t_r < t_p$, and any $x \in \mathcal{X}$ we have,

$$\begin{aligned}&\frac{\partial^2}{\partial t_s \partial t_r} Q_r(t_p, t_s, t_r|x) \\ &= \lambda_s(t_s)\phi_s(x)\lambda_r(t_r)\phi_r(x)\delta_r^s(t_s) \\ &\times \mathcal{L}_G^{(sr)}(\phi_p(x)(\Lambda_p(t_s) + \Upsilon_p^s(t_r + t_s|t_s) + \Upsilon_p^r(t_p|t_s, t_r)), \phi_s(x)\Lambda_s(t_s), \phi_r(x)\delta_r^s(t_s)\Lambda_r(t))),\end{aligned} \tag{A-13}$$

where $\mathcal{L}_G^{(sr)}(\omega_1, \omega_2, \omega_3) = \partial^2\mathcal{L}_G(\omega_1, \omega_2, , \omega_3)/\partial\omega_2\partial\omega_3$ for $(\omega_1, \omega_2, \omega_3) \in \mathbf{R}_+^3$. The left hand side is non-parametrically described by the data. Following analogous arguments with the Proposition 4 and using also the result of the latter for the right hand side, the identification of Υ_p^r follows. But, $\Delta_p^r(t|t_r) = \int_0^t \frac{\partial\Upsilon_p^r(t|t_r)}{\partial\omega} [\lambda_p(\omega)]^{-1} d\omega$. Given that we have identified Λ_p, which implies almost everywhere identification of λ_p, we obtain identification of Δ_p^r. ■

Proposition 7 *Let Assumptions 1-5 hold. Then, the integrated causal effect of the price crash on the sales recovery Δ_r^p is identified.*

Proof. We consider for almost all $t_p, t_s > 0$, every $t_r > 0$, with $t_s < t_p < t_p + t_r$, and any $x \in \mathcal{X}$, the following quantity

$$\begin{aligned}
&\frac{\partial^2}{\partial t_p \partial t_s} Q_{sp}(t_p, t_s, t_r|x) \\
&= \lambda_p(t_p)\phi_p(x)\delta_p^s(t_p|t_s)\lambda_s(t_s)\phi_s(x) \\
&\times \mathcal{L}_G^{(ps)}(\phi_p(x)(\Lambda_p(t_s) + \Upsilon_p^s(t_p|t_s)), \phi_s(x)\Lambda_s(t_s), \phi_r(x)\delta_r^s(t_s)(\Lambda_r(t_p - t_s) + \Upsilon_r^p(t_r|t_p, t_s))). \quad \text{(A-14)}
\end{aligned}$$

The left hand side of the above equation is described by the data. Moreover, making use of identical steps to Proposition 4 for the right hand side, we can identify the function Υ_r^p. By using the relationship $\Delta_r^p(t|t_p) = \int_0^t \frac{\partial \Upsilon_r^p(t|t_p)}{\partial \omega} \left[\lambda_r(\omega)\right]^{-1} d\omega$ and the identification of Λ_r, which yields almost everywhere identification of λ_r, we uniquely determine Δ_r^p. ■

B Model Likelihood

In the present section we describe the product-level likelihood contributions for each of the three duration variables. We first introduce some extra notation that we apply throughout this section. The data consist of N realizations of $(\tilde{T}_p, \tilde{T}_s, \tilde{T}_r, \gamma_p, \gamma_s, \gamma_r, X)$, where $\tilde{T}_j = \min(C_j, T_j)$, $\gamma_j = 1\{T_j \leq C_j\}$ for $j \in \{p, s, r\}$. Throughout this section we denote the realization of $\tilde{T}_j$ and X by $\tilde{t}_j$ and x, respectively. The C_j is a censoring variable that is equal to the number of months for which we observe data for each product. For example, if we have 24 months of data and we do not observe an event being realized within these 24 months, then $\tilde{t}_j$ will be equal to $C_j = 24$ for $j \in \{p, s\}$. Note that we suppress the sub-index i from all expressions for exposition purposes.

In addition, the data are such that when neither T_p nor T_s has been realized and one of them is censored the other duration variable is censored at the same point of time as well. Similarly, when $T_s < T_p$ and none of T_p and T_r has been realized, censoring of one of the latter automatically results in censoring at the same point of time of the other variable as well. Finally, by construction of the model, censoring of T_s implies that we have $\tilde{T}_r = 0$.

Let v_p, v_s and v_r denote realization of V_p, V_s, V_r, respectively. For ease of notation we do not place a subscript on these three variables as we do not observe them. We first form the likelihood contribution

$$l(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r | x, v_p, v_s, v_r) := l_p(\widetilde{t}_p | \widetilde{t}_s, \widetilde{t}_r, x, v_p) l_s(\widetilde{t}_s | \widetilde{t}_p, \widetilde{t}_r, x, v_p) l_r(\widetilde{t}_r | \widetilde{t}_p, \widetilde{t}_s, x, v_p), \tag{B-1}$$

where the likelihoods on the right hand side of the above display refers to $\widetilde{t}_p, \widetilde{t}_s$ and $\widetilde{t}_r$, respectively. Next, we integrate out the unobserved terms v_p, v_s, v_r with respect to the probability measure G in order to obtain the likelihood contribution $l(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r | x)$, that is,

$$l(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r | x) := \int_{\mathbf{R}^3_+} l(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_{di} | x, v_p, v_s, v_r) dG(v_p, v_s, v_r). \tag{B-2}$$

The derivation of the form of the likelihoods is based on the following general idea. If we do not have uncensored observation the contribution equals the density which in turn equal the product between the respective instantaneous hazard rate and survival function. On the other hand, in case an observation is censored its likelihood contribution equals the survival function.

Let $\mathcal{T}$ denote a positive duration variable with structural hazard rate $\ddot{\lambda}(.)$. Denote by $\tilde{\mathcal{T}} = \min(\mathcal{T}, \mathcal{T}^c)$ and by $\tilde{\tau}$ the realization of $\tilde{\mathcal{T}}$. Then, the likelihood contribution for $\tilde{\tau}$ is given by

$$\begin{aligned} l(\tilde{\tau}) &= 1\{\mathcal{T} \leq \mathcal{T}^c\} \ddot{\lambda}(\tilde{\tau}) \exp\left(- \int_0^{\tilde{\tau}} \ddot{\lambda}(\omega) d\omega \right) \\ +1\{\mathcal{T} &> \mathcal{T}^c\} \exp\left(- \int_0^{\tilde{\tau}} \ddot{\lambda}(\omega) d\omega \right). \end{aligned}$$

The likelihood corresponding for $\widetilde{t}_p$ is

$$
\begin{aligned}
l_p(\widetilde{t}_p|\widetilde{t}_s,\widetilde{t}_r,x,v_p) \;=\;& \gamma_p 1\left\{\widetilde{t}_p<\widetilde{t}_s\right\} v_p\phi_p(x)\lambda_p(\widetilde{t}_p)\exp\left(-\phi_p(x)\Lambda_p(\widetilde{t}_p)v_p\right)\\
&+\gamma_s 1\left\{\widetilde{t}_s<\widetilde{t}_p\le\widetilde{t}_s+\widetilde{t}_r\right\}\left(\phi_p(x)\lambda_p(\widetilde{t}_p)\delta_p^s(\widetilde{t}_p|\widetilde{t}_s)v_p\right)^{\gamma_p}\times\\
&\exp\left(-\phi_p(x)\left(\Lambda_p(\widetilde{t}_s)+\Upsilon_p^s(\widetilde{t}_p|\widetilde{t}_s)v_p\right)\right)\\
&+\gamma_s\gamma_r 1\left\{\widetilde{t}_s+\widetilde{t}_r<\widetilde{t}_p\right\}\left(\phi_p(x)\lambda_p(\widetilde{t}_p)\delta_p^s(\widetilde{t}_p|\widetilde{t}_s)\delta_p^r(\widetilde{t}_p|\widetilde{t}_s,\widetilde{t}_r)v_p\right)^{\gamma_p}\times\\
&\exp\left(-\phi_p(x)\left(\Lambda_p(\widetilde{t}_s)+\Upsilon_p^s(\widetilde{t}_p|\widetilde{t}_s)+\Upsilon_p^r(\widetilde{t}_p|\widetilde{t}_s,\widetilde{t}_r)\right)v_p\right)\\
&+(1-\gamma_p)(1-\gamma_s)\exp\left(-\phi_p(x)\Lambda_p\left(\widetilde{t}_p\right)v_p\right). \qquad \text{(B-3)}
\end{aligned}
$$

We now provide some intuition for the expression above. The first term is the likelihood contribution in case the price crash occurs before the sales crash and consequently the hazard rate is affected by neither the sales crash nor the sales recovery. The second term describes the likelihood contribution in case the price crash happens after the sales crash but before the sales recovery. Note that it is at this latter case that the causal effect function δ_p^s enters the model and affects the instantaneous hazard rate. The third term equals the likelihood contribution in case the sales crash and sales recovery precede the price crash. In this latter case, two causal effects, δ_p^s and δ_p^r, affect the instantaneous hazard rate. Finally, the last term is equal to the likelihood contribution in case the timing of price crash and sales crash is censored. The term $\Upsilon_p^s(\widetilde{t}_p|\widetilde{t}_s,x)$ is defined as follows,

$$
\begin{aligned}
\Upsilon_p^s(\widetilde{t}_p|\widetilde{t}_s) \;=\;& \sum_{l=0}^{K-1} 1\left\{Y_l\le\widetilde{t}_p-\widetilde{t}_s<Y_{l+1}\right\}\\
&\times\sum_{\eta=0}^{l}\exp\left(\theta_{p\eta}^s\right)\left[\left(\Lambda_p(\widetilde{t}_p)\right)^{1\{\eta=l\}}\left(\Lambda_p(\widetilde{t}_s+Y_{\eta+1})\right)^{1\{0\le\eta<l\}}\right.\\
&\left.-\left(\Lambda_p(\widetilde{t}_s+Y_\eta)\right)^{1\{0<\eta\le l\}}\left(\Lambda_p(\widetilde{t}_s)\right)^{1\{\eta=0\}}\right], \qquad \text{(B-4)}
\end{aligned}
$$

where we use the definition of Υ_p^s and the fact that

$$\delta_p^s(\widetilde{t}_p|\widetilde{t}_s;\theta_p^s)=\sum_{l=0}^{K-1}1\left\{Y_l\le\widetilde{t}_p-\widetilde{t}_s<Y_{l+1})\right\}\exp(\theta_{pl}^s).$$

Concerning the term $\Upsilon_p^r(\widetilde{t}_p|\widetilde{t}_s, \widetilde{t}_r)$, by using the corresponding definition and also that

$$\delta_p^r(\widetilde{t}_p|\widetilde{t}_s, \widetilde{t}_r; \theta_p^r) = \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_p - \widetilde{t}_s - \widetilde{t}_r < Y_{l+1}\right\} \exp(\theta_{pl}^r),$$

we get

$$\begin{aligned}
\Upsilon_p^r(\widetilde{t}_p|\widetilde{t}_s, \widetilde{t}_r) &= \int_{\widetilde{t}_s+\widetilde{t}_r}^{\widetilde{t}_p} \lambda_p(w)\delta_p^s(w|\widetilde{t}_s)\delta_p^r(w|\widetilde{t}_s, \widetilde{t}_r)dw \\
&= \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_p - \widetilde{t}_s - \widetilde{t}_r < Y_{l+1}\right\} \sum_{\eta=0}^{l} \exp\left(\theta_{p\eta}^r\right) \\
&\quad \times \int_{\widetilde{t}_s+\widetilde{t}_r+Y_\eta}^{1\{\eta=l\}\widetilde{t}_p+1\{0\leq\eta<l\}\left[Y_{\eta+1}+\widetilde{t}_s+\widetilde{t}_r\right]} \lambda_p(w)\delta_p^s(w|\widetilde{t}_s)dw. \qquad \text{(B-5)}
\end{aligned}$$

Recall that

$$\delta_p^s(\widetilde{t}_p|\widetilde{t}_s; \theta_p^s) = \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_p - \widetilde{t}_s < Y_{l+1}\right\} \exp(\theta_{pl}^s).$$

Then, the last line of (B-5) becomes

$$\begin{aligned}
&\sum_{\rho=0}^{K-1} 1\left\{Y_\rho \leq \widetilde{t}_r + Y_\eta < Y_{\rho+1}\right\} \sum_{\omega=\rho}^{K-1} 1\left\{Y_\omega \leq 1\{\eta=l\}\widetilde{t}_p + 1\left\{0\leq\eta<l\right\}\left[Y_{\eta+1}+\widetilde{t}_s+\widetilde{t}_r\right] - \widetilde{t}_s < Y_{\omega+1}\right\} \\
&\sum_{\iota=\rho}^{\omega} \exp\left(\theta_{p\iota}^s\right) \left[\left[\Lambda_p(1\{\eta=l\}\widetilde{t}_p + 1\left\{0\leq\eta<l\right\}\left[Y_{\eta+1}+\widetilde{t}_s+\widetilde{t}_r\right])\right]^{1\{\iota=\omega\}} \left[\Lambda_p(\widetilde{t}_s+Y_{\rho+1})\right]^{1\{\rho\leq\iota<\omega\}}\right. \\
&\left. - \left[\Lambda_p(\widetilde{t}_s+Y_\iota)\right]^{1\{\rho<\iota\leq\omega\}} \left[\Lambda_p(\widetilde{t}_s+\widetilde{t}_r+Y_\eta)\right]^{1\{\iota=\rho\}}\right]. \qquad \text{(B-6)}
\end{aligned}$$

The likelihood for $\widetilde{t}_s$ is expressed as

$$\begin{aligned}
l_s(\widetilde{t}_s|\widetilde{t}_p, x, v_s) &= \gamma_s 1\left\{\widetilde{t}_s < \widetilde{t}_p\right\} \phi_s(x)\lambda_s(\widetilde{t}_s)v_s \exp\left(-\phi_s(x)\Lambda_s(\widetilde{t}_s)v_s\right) \\
&\quad + \gamma_p 1\left\{\widetilde{t}_p < \widetilde{t}_s\right\} \left(\phi_s(x)\lambda_s(\widetilde{t}_s)\delta_s^p(\widetilde{t}_s|\widetilde{t}_p)v_s\right)^{\gamma_s} \\
&\quad \times \exp\left(-\phi_s(x)\Lambda_s(\widetilde{t}_p) + \Upsilon_s^p(\widetilde{t}_s|\widetilde{t}_p)v_s\right) \\
&\quad + (1-\gamma_s)(1-\gamma_p)\exp\left(-\phi_s(x)\Lambda_s(\widetilde{t}_s)v_s\right) \qquad \text{(B-7)}
\end{aligned}$$

The first term is the likelihood contribution if the sales crash is realized before the price crash and therefore the hazard rate is not affected by the price crash. The second term describes the likelihood contribution in case the price crash is realized before the price crash. As a consequence, the instantaneous hazard rate is multiplied by the function δ_s^p which captures the causal effect of the price crash on the sales crash. Finally, the last term represents the likelihood contribution in case both durations corresponding to the price crash and sales crash are censored. By making use of the definition of Υ_s^p and the fact that

$$\delta_s^p(\widetilde{t}_s|\widetilde{t}_p;\theta_s^p) = \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_s - \widetilde{t}_p < Y_{l+1}\right\} \exp\left(\theta_{sl}^p\right),$$

it follows

$$\begin{aligned}
\Upsilon_s^p(\widetilde{t}_s|\widetilde{t}_p) \;=\; & \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_s - \widetilde{t}_p < Y_{l+1}\right\} \\
& \times \sum_{\eta=0}^{l} \exp\left(\theta_{s\eta}^p\right)\left[\left(\Lambda_s(\widetilde{t}_s)\right)^{1\{\eta=l\}}\left(\Lambda_s(\widetilde{t}_p+Y_{\eta+1})\right)^{1\{0\leq\eta<l\}}\right. \\
& \left. - \left(\Lambda_p(\widetilde{t}_p+Y_\eta)\right)^{1\{0<\eta\leq l\}}\left(\Lambda_p(\widetilde{t}_p)\right)^{1\{\eta=0\}}\right]. \qquad \text{(B-8)}
\end{aligned}$$

Finally, the likelihood for $\widetilde{t}_r$ is given by

$$\begin{aligned}
l_r(\widetilde{t}_r|\widetilde{t}_p,\widetilde{t}_s,x,v_r) \;=\; & \gamma_s\gamma_p 1\left\{\widetilde{t}_p < \widetilde{t}_s\right\} \times \exp\left(-\phi_r(x)\Lambda_r(\widetilde{t}_r)\delta_r^s(\widetilde{t}_s)v_r\right) \\
& \times \left(\phi_r(x)\lambda_r(\widetilde{t}_r)\delta_r^s(\widetilde{t}_s)v_r\right)^{\gamma_r} \\
& +\gamma_s\gamma_p 1\left\{\widetilde{t}_s < \widetilde{t}_p < \widetilde{t}_s+\widetilde{t}_r\right\} \\
& \times \exp\left(-\phi_r(x)\delta_r^s(\widetilde{t}_s)\left(\Lambda_r(\widetilde{t}_p-\widetilde{t}_s)+\Upsilon_r^p(\widetilde{t}_r|\widetilde{t}_p,\widetilde{t}_s)\right)v_r\right) \\
& \times \left(\phi_r(x)\lambda_r(\widetilde{t}_r)\delta_r^p(\widetilde{t}_r|\widetilde{t}_p,\widetilde{t}_s)\delta_r^s(\widetilde{t}_s)v_r\right)^{\gamma_r} \\
& +\gamma_s\gamma_p 1\left\{\widetilde{t}_s+\widetilde{t}_r<\widetilde{t}_p\right\}\exp\left(-\phi_r(x)\delta_r^s(\widetilde{t}_s)\Lambda_r(\widetilde{t}_r)v_r\right) \\
& +\gamma_s(1-\gamma_p)\times\exp\left(-\phi_r(x)\Lambda_r(\widetilde{t}_r)\delta_r^s(\widetilde{t}_s)v_r\right) \\
& \times\left(\phi_r(x)\lambda_r(\widetilde{t}_r)\delta_r^s(\widetilde{t}_s)v_r\right)^{\gamma_r}+(1-\gamma_s). \qquad \text{(B-9)}
\end{aligned}$$

The first term is the likelihood contribution if the price crash is realized before the sales recovery (note that by definition the sales crash always precedes the sales recovery) and therefore the hazard rate of the latter is not directly affected by the occurrence of the former. The second term describes the likelihood contribution in case the price crash is realized after the price crash but before the sales recovery. Hence, the instantaneous hazard rate is multiplied by δ_r^p. The third term describes the likelihood contribution for the case in which the price crash occurs after the realization of the sales recovery and therefore the hazard rate of the latter is not affected by the former. The fourth term gives the likelihood contribution if the price crash occurs after the sales crash but we do not observe the time to the occurrence of the latter. Finally the last term specifies the likelihood contribution for the case in which the timing of the sales crash is censored and therefore we do not observe at all the duration which corresponds to the sales recovery. As a consequence, the instantaneous hazard rate is multiplied by the function δ_s^p which captures the causal effect of the price crash on the sales crash. By making use of the definition of Υ_r^p and the fact that

$$\delta_r^p(\widetilde{t}_r|\widetilde{t}_p, \widetilde{t}_s; \theta_r^p) = \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_r - \widetilde{t}_p + \widetilde{t}_s < Y_{l+1}\right\} \exp\left(\theta_{rl}^p\right),$$

we get

$$\begin{aligned}
\Upsilon_r^p(\widetilde{t}_r|\widetilde{t}_p, \widetilde{t}_s) &= \sum_{l=0}^{K-1} 1\left\{Y_l \leq \widetilde{t}_r - \widetilde{t}_p + \widetilde{t}_s < Y_{l+1}\right\} \\
&\quad \times \sum_{\eta=0}^{l} \exp\left(\theta_{r\eta}^p\right) \Big[\left(\Lambda_r(\widetilde{t}_r)\right)^{1\{\eta=l\}} \left(\Lambda_d(\widetilde{t}_p - \widetilde{t}_s + Y_{\eta+1})\right)^{1\{0\leq\eta<l\}} \\
&\quad - \left(\Lambda_d(\widetilde{t}_p - \widetilde{t}_s)\right)^{1\{\eta=0\}} \left(\Lambda_d(\widetilde{t}_p - \widetilde{t}_s + Y_\eta)\right)^{1\{0<\eta\leq l\}}\Big]. \qquad \text{(B-10)}
\end{aligned}$$

Then $l(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r|x)$ is then obtained by applying the formula (B-1).

C Prior Specification and Model Estimation

We use a Bayesian model specification and complete the model with the following priors. We assume that

$$\begin{aligned}
p(\alpha_j) &\propto 1 && \text{for } j = p, s, r \\
p(\beta_j) &\propto |\Omega|^{-1/2} \exp(-\tfrac{1}{2}\beta_j \Omega^{-1} \beta_j') && \text{for } j = p, s, r \qquad \text{(C-1)}
\end{aligned}$$

where Ω is equal to an identity matrix times 10. And we assume that

$$
\begin{array}{ll}
p(\theta_{j1}) \propto (1/\tau_0^2)\exp(-\frac{1}{2\tau_0^2}\theta_{j1}^2) & \text{for } j = p, s, r \\
p(\theta_{j\eta} - \theta_{j(\eta-1)}) \propto (1/\tau_1^2)\exp(-\frac{1}{2\tau_1^2}(\theta_{j\eta} - \theta_{j(\eta-1)})^2) & \text{for } j = p, s, r \text{ and } \eta = 2, \ldots, K.
\end{array}
\tag{C-2}
$$

where τ_0^2 and τ_1^2 are set equal to 100. We set no prior for Σ and we define

$$
p(\vartheta) = \prod_j \prod_\eta p(\beta_j)p(\alpha_j)p(\theta_{j1})p(\theta_{j\eta} - \theta_{j(\eta-1)}) \tag{C-3}
$$

where ϑ is a vector containing α_j for $j = s, p, r$, $\theta_{j\eta}$ for $j = s, p, r$ and $\eta = 1, \ldots, K$, β_j for $j = s, p, r$, and Σ.

The posterior distribution of ϑ is given by is then

$$
p(\vartheta|\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r, x) = p(\vartheta) \prod_i l(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r|x) \tag{C-4}
$$

where we use $(\widetilde{t}_p, \widetilde{t}_s, \widetilde{t}_r, x)$ to refer to the duration variables and covariates for all observations.

We sample each of the elements of ϑ using a Random Walk Metropolis step within a Gibbs sampler that updates each of the elements of ϑ at a time. We use a block update for β_j and a block update for $\theta_{j\eta}$. In addition, we use an auxiliary 3×3 matrix H with each r of its row equal to $[\epsilon_{r1}, \epsilon_{r2}, \epsilon_{r3}]$ and we sample each ϵ_{rc} from a random walk and use $H'H$ as the proposal for Σ. This specification allow us to avoid an informative prior specification for Σ, like the Wishart. The MCMC chain ran for 10,000 iterations and we discarded the first 5,000 draws and then applied a thinning factor of 3.

References

Abbring, Jaap H., Gerard J. Van den Berg. 2003. The nonparametric identification of treatment effects in duration models. *Econometrica* **71**(5) 1491–1517.

Abbring, Jaap H., Gerard J. Van den Berg, Jan C. Van Ours. 2005. The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *The Economic Journal* **115**(505) 602–630.

Agarwal, Rajshree, Barry L. Bayus. 2002. The market evolution and sales takeoff of product innovations. *Management Science* **48**(8) 1024–1041.

Braun, Michael, David A. Schweidel. 2011. Modeling customer lifetimes with multiple causes of churn. *Marketing Science* **30**(5) 881–902.

BusinessWeek Online. 2007. Apple averts a "Fanboy Rebellion" . September 7th. URL `http://tiny.cc/72rdr`.

Chandrasekaran, Deepa, Gerard Tellis. 2008. Global takeoff of new products: Culture, wealth, or vanishing differences. *Marketing Science* **27**(5) 844–860.

Chandrasekaran, Deepa, Gerard J. Tellis. 2011. Getting a grip on the daddle: Chasms or cycles? *Journal of Marketing* **75**(4) 21–34.

Chintagunta, Pradeep K. 1998. Inertia and variety seeking in a model of Brand-Purchase timing. *Marketing Science* **17**(3) 253–270.

Chintagunta, Pradeep K., Xiaojing Dong. 2006. Hazard/survival models in marketing. Rajiv Grover, Marco Vriens, eds., *The handbook of marketing research: Uses, Misuses, and Future Advances*, 1st ed., chap. 21. UK: Sage Publications Ltd, 441–454.

Chintagunta, Pradeep K., Sudeep Haldar. 1998. Investigating purchase timing behavior in two related product categories. *Journal of Marketing Research* **35**(1) 43–53.

Danaher, Peter J., Michael S. Smith. 2011. Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science* **30** 4–21.

Franses, Philip Hans, Richard Paap. 2004. *Quantitative Models in Marketing Research*. Cambridge University Press, Cambridge, U.K.

Freund, John. E. 1961. A bivariate extension of the exponential distribution. *Journal of The American Statistical Association* **56**(296) 971–977.

Goldenberg, Jacob, Barak Libai, Eitan Muller. 2002. Riding the saddle: How cross-market communications can create a major slump in sales. *Journal of Marketing* **66**(2) 1–16.

Golder, Peter N, Gerard J Tellis. 1997. Will it ever fly? Modeling the takeoff of really new consumer durables. *Marketing Science* **16**(3) 256–270.

Golder, Peter N., Gerard J. Tellis. 2004. Growing, growing, gone: Cascades, diffusion, and turning points in the product life cycle. *Marketing Science* **23**(2) 207–218.

Hauser, John, Gerard J. Tellis, Abbie Griffin. 2006. Research on innovation: A review and agenda for Marketing Science. *Marketing Science* **25**(6) 687.

Heckman, James J., George J. Borjas. 1980. Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica* **47**(187) 247–283.

Honoré, Bo E. 1993. Identification results for duration models with multiple spells. *Review of Economic Studies* **60**(1) 241–246.

Kaminsky, Graciela L., Carmen M. Reinhart. 1999. The Twin Crises: The causes of banking and Balance-Of-Payments problems. *The American Economic Review* **89**(3) 473–500.

Park, Young-Hoon, Peter S. Fader. 2004. Modeling browsing behavior at multiple websites. *Marketing Science* **23**(3) 280–303.

Peres, Renana, Eitan Muller, Vijay Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* **27**(2) 91–106.

Saha, Atanu, Lynette Hilton. 1997. Expo-power: A flexible hazard function for duration data models. *Economics Letters* **54**(3) 227–233.

Schnabel, Isabel. 2004. The German Twin Crisis of 1931. *The Journal of Economic History* **64**(3) 822–871.

Schweidel, David A., Peter S. Fader, Erick T. Bradlow. 2008. A bivariate timing model of customer acquisition and retention. *Marketing Science* **27**(5) 829–843.

Seetharaman, P.B., Pradeep K. Chintagunta. 2003. The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business and Economic Statistics* **21**(3) 368–382.

Talukdar, Deby, Karunakaran Sudhir, Andrew Ainslie. 2001. Identifying similarities in diffusion patterns across countries and products: A Bayesian variance components approach. *Marketing Science* **21**(1) 97–114.

Tellis, Gerard J., Stefan Stremersch, Eden Yin. 2003. The international takeoff of new products: The role of economics, culture, and country innovativeness. *Marketing Science* **22**(2) 188–208.

Vakratsas, Demetrios, Ceren Kolsarici. 2008. A dual-market diffusion model for a new prescription pharmaceutical. *International Journal of Research in Marketing* **25**(4) 282–293.

Van den Berg, Gerard, Bettina Drepper. 2011. A unique bond: Does losing your co-twin affect your remaining life-span? *Unpublished Manuscript, University of Mannheim* .

Van den Berg, Gerard J. 2001. Duration models: Specification, identification and multiple durations. *Handbook of Econometrics* **5** 3381–3460.

Van den Bulte, Christophe. 2000. New product diffusion acceleration: Measurement and analysis. *Marketing Science* **19**(4) 366–380.

Van Everdingen, Y.M., Dennis Fok, Stefan Stremersch. 2009. Modeling global spillover of new product takeoff. *Journal of Marketing Research* **46**(5) 637–652.

Van Heerde, Herald J., Marnik G. Dekimpe, William P. Putsis Jr. 2005. Marketing models and the Lucas critique. *Journal of Marketing Research* **42**(1) 15–21.

Van Ours, Jan C. 2003. Is cannabis a stepping-stone for cocaine? *Journal of Health Economics* **22**(4) 539–554.

Van Ours, Jan C., Jenny Williams. 2009. Why parents worry: Initiation into cannabis use by youth and their educational attainment. *Journal of Health Economics* **28**(1) 132–142.

Wired Magazine. 2007. Four mistakes Apple made with the iPhone price drop. URL `http://tiny.cc/5m889`.

Asymptotic Theory for Nonparametric Estimation of Cumulative Incidence Functions

1 Introduction

Competing risks analysis is a very important field of applied statistics. The competing risks model encompasses a large class of statistical models where the subject is exposed to several risks at the same point in time and fails due to only one of these risks. The risk that triggers the failure is referred to as the cause of failure. The elapsed time between the starting point of time and the failure is called failure time. Competing risks models are relevant for many applications in demography, economics, medical research and reliability analysis. In labor economics, competing risks for an unemployed person could be employment and leaving the labor force. In finance, competing risks for a limit order are cancellation and execution. In studies for the failure of firms, competing risks could be acquisition and bankruptcy. In cancer studies, competing risks are usually recurrence of the original cancer and successful treatment of the original tumor. Kalbfleisch and Prentice (1991) and Van den Berg (2001) provide a detailed overview on this important topic.

Tsiatis (1975) has shown that the complete identification of the competing risks model is not feasible. In particular, for any joint distribution of the failure time and the cause of failure we can always construct a competing risks model with independent risks that will generate this particular joint distribution. However, it is possible to achieve nonparametric identification of two functions without claiming anything about the dependence structure of the underlying risks. The first one is the cumulative incidence function and the second one the cause specific hazard rate. The former gives the probability of failure by a certain point of time due to a particular risk in the presence of all other risks, whereas the latter gives the instantaneous rate of failure due to a specific risk in the presence of all other risks. The cumulative incidence function and the cause specific hazard rate are important tools in competing risks analysis for several reasons. First, they can be employed in determining which risk is greater than the others, namely, which risk is more likely to trigger the failure. Second, the two quantities can be used to compare two different groups (e.g., different treatment, different race) with respect to a particular risk, that is, to investigate for which group the risk of interest is greater. Aly et al. (1994) propose relevant test statistics for the first case, whereas Gray (1988), Lin et al. (1997), Bajorunaite and Klein (2007), andSchaubel and Wei (2011) build different test statistics for the second case. Lee and Whang (2009) propose a class of nonparametric tests for testing the significance of treatment effects that can be directly applied by using cumulative incidence functions if the failure time and the cause of failure are always observed. Finally, the knowledge of the shape of the cumulative incidence function and the cause specific hazard rate is important in its own right as we can get useful insights about the occurrence of failures as well as the effects of covariates on failures.

The goal of this paper is to derive asymptotic results for the nonparametric estimator of the cumulative incidence function in cases in which continuous covariates affect the realization of the failure time, and the cause of failure is Missing At Random for some observations. More precisely, we provide results on the uniform convergence rate and pointwise asymptotic normality of our estimator. We will rely on counting process theory as well as kernel smoothing theory to derive the asymptotic results. Counting processes techniques are very successful in dealing with censored data that commonly arise in survival analysis problems. Andersen et al. (1993) and Fleming and Harrington (1991) provide an excellent review of this important field. On the other hand, Nielsen and Linton (1995), Mammen and Nielsen (2007), and Linton et al. (2011) address problems within the context of survival analysis as well as of conventional regression analysis by combining the classical counting process theory with kernel smoothing methods. Nonparametric techniques have recently received substantial attention because they do not require functional form assumptions regarding the underlying model and consequently they do not suffer from the problem of model misspecification which can potentially give erroneous statistical results. Peng and Fine (2008), for instance, develop a two sample nonparametric test statistic for the significance of a particular covariate on the survival probability of prostate cancer patients. Their findings contrast the results of the parametric analysis which adopts a Cox proportional hazard form for the cause specific hazard rates.

The proposed nonparametric estimator is complementary to *i*) the developed (semi)-parametric procedures for the estimation of the cumulative incidence function with right-censored data, *ii*) the suggested parametric methods for right-censored data where missing observations, concerning the cause of failure, occur randomly. Regarding the first category, there are two main approaches. The first one is to use either a Cox specification (Andersen et al., 1993) or an additive specification (Shen and Cheng, 1999) for the cause specific hazard rate and then, based on the estimation results of the first step, to estimate the cumulative incidence function. The second one is to directly model the cumulative incidence function. Fine and Gray (1999) adopt a fully parametric approach and make the assumption that the complementary log-log of the cumulative incidence function is on the proportional hazards form. Jeong and Fine (2007) use a Gombertz distribution to parameterize the cumulative incidence function. Moreover, Scheike et al. (2008) adopt a semiparametric approach that allows for time-varying effects of the covariates and includes as special cases the Cox's model, the Aalen's additive model and the combination of these two. On the other hand, Lu and Liang (2008) develop parametric estimation methods with randomly missing cause of failure. They adopt (augmented) inverse probability weighting scheme to deal with missing observations, and additive hazard specification for the underlying cause specific hazard rates.

Our focus on the concept of cumulative incidence function rather than cause specific hazard rate is driven by the fact that the former is the most commonly employed quantity in the empirical analysis of competing risks data. In particular, one drawback of the cause specific hazard rate is that it is not linked to some probability distribution, if the failure time is continuously distributed and the risks are not independent of each other. On the other hand, cumulative incidence function has an easy probabilistic interpretation as it equals the proportion of failures due to a particular risk by a certain point of time. Additionally, the use of cause specific hazard rates for the comparison of two different groups with respect to a particular risk may not give satisfactory answers. In particular, Gray (1988) describes an example in which the cause specific hazard rates of a certain risk for different groups follow a particular ordering; however, this ordering is not preserved under the comparison of the respective cumulative incidence functions.

The outline of the paper is as follows. Section 2 discusses the construction of the nonparametric estimator. In Section 3, asymptotic results for the nonparametric estimator are presented. Section 4 concludes. The technical proofs of the main results are relegated to the Appendix.

2 Nonparametric estimator

For expositional convenience only, we will focus on two possible risks. Let $L_1 > 0$ and $L_2 > 0$ represent the latent failure times that correspond to each risk, $L = \min(L_1, L_2)$ be the (actual) failure time and γ be a failure type indicator, that is, $\gamma = 1$ if $L_1 < L_2$ and $\gamma = 2$ if $L_2 < L_1$. Moreover, let X be a vector of observed covariates with support $\mathbf{X} \subset \mathbf{R}^d$ which are associated with the latent failure outcomes L_1 and L_2.

Let $x \in \mathbf{X}$ denote the realization of X. Define for each risk $j = 1, 2$ and $(t, x) \in \mathbf{R}_+ \times \mathbf{X}$ the cumulative incidence function

$$F_j(t|x) := \mathbf{P}(L \leq t, \gamma = j|x). \tag{1}$$

Namely, the value of $F_j(t|x)$ equals the probability of failure by time t due to risk j for a subject with characteristics x. Note that $\lim_{t\to\infty} F_j(t|x) = \mathbf{P}(\gamma = j|x) < 1$ if $\mathbf{P}(L_j < L_\xi|x) < 1$ for $\xi = 1, 2$ with $\xi \neq j$, that is, $F_j(t|x)$ is a subdistribution function. To obtain an explicit expression for the cumulative incidence function we also introduce the notion of the cause specific hazard rate. We will first suppose that the stochastic variables L_1 and L_2 are absolutely continuous (with respect to $\mathbf{R}_+$) and consequently the stochastic variable L as well. Then, the cause specific hazard rate function $\lambda_j(t, x)$ is defined as follows

$$\lambda_j(t, x)dt = \mathbf{P}(t \leq L < t + dt, \gamma = j|L \geq t, x) + o(dt) \tag{2}$$

for any $(t, x) \in \mathbf{R}_+ \times \mathbf{X}$, where the notation dt denotes throughout the paper an infinitesimal time increment. The $\lambda_j(t, x)$ quantitatively describes for a subject with characteristics x the instantaneous rate of failure at t due to risk j given survival up to $t-$. We also introduce the cumulative cause specific hazard rate $\Lambda_j(t, x) = \int_0^t \lambda_j(u, x)du$. Next, consider also the overall hazard rate $\lambda(t, x)$ with $\lambda(t, x) = \lambda_1(t, x) + \lambda_2(t, x)$ for all $(t, x) \in \mathbf{R}_+ \times \mathbf{X}$ and the corresponding cumulative overall hazard rate $\Lambda(t,x) = \int_0^t \lambda(u,x)du$. Note that the $\lambda(t,x)$ gives the instantaneous rate of failure at t due to either risk 1 or risk 2 given survival up to $t-$. By using (1) and (2) we get for $j = 1, 2$,

$$F_j(t|x) \simeq \int_0^t S(u - |x)d\Lambda_j(u,x), \tag{3}$$

where $S(t - |x) = \mathbf{P}(L \geq t|x)$. The latter survival function can be explicitly calculated by the product-integral formula as follows

$$S(t|x) = \prod_{u \leq t} \{1 - d\Lambda(u,x)\}. \tag{4}$$

We impose the condition that our study consists of n independent subjects. If there were not Missing At Random (MAR) observations we would observe n independently and identically distributed (i.i.d.) copies $(T_i, X_i, \tilde{\gamma}_i)$ $(i = 1, ..., n)$ of $(T, X, \tilde{\gamma})$, where $T = \min(L, Z)$, $\tilde{\gamma} = \gamma 1\{L \leq Z\}$, and $1\{B\}$ is the indicator function that is equal to one if the event B occurs and zero otherwise. The stochastic variable Z is called censoring variable and satisfies the property $Z \perp L_1, L_2 \mid X$. Namely, we would be encountered with the conventional right-censored competing risks model.

In this paper, we will nonparametrically estimate the cumulative incidence function, when the cause of failure is missing for some uncensored observations and the missing data mechanism satisfies the MAR assumption (Rubin, 1976; Little and Rubin, 1987). In particular, introduce the missing indicator variable R_i for the cause of failure. The value of R_i equals 0 if $T_i = L_i$ and the cause of failure is not observed. On the other hand, the stochastic variable R_i is equal to 1 if $T_i = L_i$ and the cause of failure is observed or if $T_i = Z_i$. The MAR scheme implies that

$$\mathbf{P}(R = 1|\tilde{\gamma}, \tilde{\gamma} > 0, T, x) = \mathbf{P}(R = 1|\tilde{\gamma} > 0, x) =: \pi(x). \tag{5}$$

Hence, $\pi(x)$ specifies the probability that the cause of failure is missing given the subject-level characteristics x, and the fact that $\tilde{\gamma}$ is strictly positive. In view of the above discussion, the dataset consist of n i.i.d. tuples $(T_i, X_i, R_i, R_i\tilde{\gamma}_i, 1\{\tilde{\gamma}_i > 0\})$. We also assume that

$$R \perp Z, L_1, L_2 | X. \tag{6}$$

In the above discussion we implicitly assume that the covariates are time-invariant. The same assumption is also made for the missing indicator R. Under this scenario, information on whether the failure cause is missing will be available at the beginning of the risk process. However, the time-invariant setup is adopted only for notational convenience. In particular, all the results in the sequel are true if both X and/or R are predictable. Predictability is ensured by imposing the condition that X and R are either caglad (i.e., left continuous with right limits) or predictable cadlag (i.e., right continuous with left limits). Predictability for X and R is roughly equivalent to saying that the value of the underlying processes at t is known just prior to this point of time. This implies for the missing mechanism that at every point of time we know whether the cause of failure will be missing in case the subject fails at the next moment of time.

We will study two estimators for the cumulative incidence function. In particular,

$$\hat{F}_j^C(t|x) = \int_0^t \hat{S}^C(u-|x)d\hat{\Lambda}_j^C(u,x), \qquad j = 1, 2, \tag{7}$$

and

$$\hat{F}_j^L(t|x) = \int_0^t \hat{S}^L(u-|x)d\hat{\Lambda}_j^L(u,x), \qquad j = 1, 2. \tag{8}$$

where the superscripts C and L refer to the type of smoothing with respect to the vector x. In particular, C is used for the local constant smoothing, whereas L is used for the local linear smoothing. To obtain these estimators, we need introduce some extra notation. Consider the counting processes $N_{ji}(t) = 1\{T_i \le t, \tilde{\gamma}_i = j, R_i = 1\}$, $N_{oi}(t) = 1\{T_i \le t, \tilde{\gamma}_i > 0, R_i = 0\}$. The first process describes whether the subject i has failed due to risk j at the time interval $[0,t]$ and the cause of failure is not missing, and the second process represents whether the subject i has failed due to either risk 1 or risk 2 at the time interval $[0,t]$ and the cause of failure is missing. We also introduce the counting process $\bar{N}_i(t) = 1\{T_i \le t, \tilde{\gamma}_i > 0\}$ which specifies whether the subject i has failed due to either risk 1 or risk 2 at the time interval $[0,t]$. It is straightforward to see that $\bar{N}_i(t) = N_{1i}(t) + N_{2i}(t) + N_{oi}(t)$. Moreover, consider the "at risk" predictable process $Y_i(t) = 1\{T_i \ge t\}$ which describes whether the subject i has survived and not be censored up to $t-$.

Let $\omega = (\omega_1, ..., \omega_d) \in \mathbf{R}^d$ and consider the quantity $\mathcal{K}_h(\omega) = \frac{1}{h^d}\prod_{p=1}^{d} K\left(\frac{\omega_p}{h}\right)$, where K is a probability density function with symmetric around zero compact support $\mathbf{K}$, and h is a sequence of non-negative numbers such that $h = o(n)$. Next, introduce the quantity $\mathcal{L}_h(\omega) = \frac{\mathcal{K}_h(\omega) - \mathcal{K}_h(\omega)\omega^{\mathrm{T}}\bar{D}^{-1}\bar{c}_1}{\bar{c}_0 - c_1^{\mathrm{T}}\bar{D}^{-1}\bar{c}_1}$, with

$$\bar{c}_0 = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h(x - X_i),$$

$$\bar{c}_{1\rho} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h(x - X_i)(x_\rho - X_{i\rho}),$$

$$\bar{d}_{\rho\kappa} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h(x - X_i)(x_\rho - X_{i\rho})(x_\kappa - X_{i\kappa}),$$

and $\bar{c}_1 = (\bar{c}_{1\rho})_{\rho=1}^{d}$ and $\bar{D} = (\bar{d}_{\rho\kappa})_{\rho,k=1}^{d}$. The notation x_ρ and $X_{i\rho}$ refer to the $\rho-$th element of the corresponding row vector. To proceed with the formal description of the two estimators we also consider the (local) weights $w_i^{\nu}(x), b_i^{\nu}(x)$ for $\nu = C, L$ (suppressing dependence on n for notational simplicity)

$$w_i^C(x) = \frac{R_i}{\hat{\pi}_*^C(X_i, \tilde{\gamma}_i)}\mathcal{K}_h(x - X_i)\Big/\sum_{i=1}^{n} \frac{R_i}{\hat{\pi}_*^C(X_i, \tilde{\gamma}_i)}\mathcal{K}_h(x - X_i),$$
$$w_i^L(x) = \frac{R_i}{\hat{\pi}_*^L(X_i, \tilde{\gamma}_i)}\mathcal{L}_h(x - X_i)\Big/\sum_{i=1}^{n} \frac{R_i}{\hat{\pi}_*^L(X_i, \tilde{\gamma}_i)}\mathcal{L}_h(x - X_i),$$

and

$$b_i^C(x) = \mathcal{K}_h(x - X_i)\Big/\sum_{i=1}^{n} \mathcal{K}_h(x - X_i),$$
$$b_i^L(x) = \mathcal{L}_h(x - X_i)\Big/\sum_{i=1}^{n} \mathcal{L}_h(x - X_i),$$

with $\hat{\pi}_*^{\nu}(X_i, \tilde{\gamma}_i) = 1\{\tilde{\gamma}_i \neq 0\}\hat{\pi}^{\nu}(X_i) + 1\{\tilde{\gamma}_i = 0\}$ and $\hat{\pi}^{\nu}(X_i) = \sum_{i=1}^{n} b_i^{\nu}(x)R_i$. Note that $\sum_{i=1}^{n} w_i^{\nu}(x) = \sum_{i=1}^{n} b_i^{\nu}(x) = 1$ for all $x \in \mathbf{X}$.

Introduce also the conditional distribution function $H(t|x) := \mathbf{P}(T > t|x)$. The corresponding estimators of the cumulative hazard rates $\Lambda_j(t, x)$ and $\Lambda(t, x)$ are given by

$$\hat{\Lambda}_j^{\nu}(t,x) = \sum_{i=1}^{n} \int_0^t \frac{w_i^{\nu}(x)}{\hat{\pi}_*^{\nu}(X_i,\tilde{\gamma}_i)\hat{H}^{\nu}(u-|x)} dN_{ji}(u), \quad \hat{\Lambda}^{\nu}(t,x) = \sum_{i=1}^{n} \int_0^t \frac{b_i^{\nu}(x)}{\breve{H}^{\nu}(t-|x)} d\bar{N}_i(u), \tag{9}$$

where the above integrals are written in Riemann–Stieltjes form, $\breve{H}^{\nu}(t-|x) = \sum_{i=1}^{n} w_i^{\nu}(x)1(T \geq t)$ and $\hat{H}^{\nu}(t-|x) = \sum_{i=1}^{n} b_i^{\nu}(x)1(T \geq t)$ for $\nu = C, L$ are estimators for the quantity $H(t-|x) = \lim_{u\uparrow t} H(u|x)$. Note that for the estimation of the $\Lambda_j(t,x)$ we employ inverse probability weighting (either local constant or local linear smoothing) scheme. Similar approach is popular in the standard regression context for dealing with MAR observations (Hu et al., 2010). On the other hand, we adopt the conventional smoothing (either local constant or local linear) techniques for the estimation of the cumulative hazard rate $\Lambda(t,x)$. The reason that we do not need inverse probability weights for its estimation is that we always observe the variable T_i and the stochastic variable $1\{\tilde{\gamma}_i > 0\}$ regardless of whether we observe the cause of failure.

The respective estimator of $S(t|x)$ is given by

$$\hat{S}^{\nu}(t|x) = \prod_{u\leq t} \left\{1 - d\hat{\Lambda}^{\nu}(u,x)\right\}, \tag{10}$$

where the product is taken over the discontinuity points of the estimator $\hat{\Lambda}^{\nu}(t,x)$. The estimators $\hat{\Lambda}_j^{\nu}(t,x)$ and $\hat{\Lambda}^{\nu}(t,x)$ are generalization of the Nelson-Aalen estimator, and $\hat{S}^{\nu}(t|x)$ is generalization of the Kaplan-Meier estimator in the sense that continuous covariates, which affect the realization of L_1, L_2 (and therefore L), are included in the model in a fully nonparametric way.

3 Asymptotic results

In this section, we will investigate the large sample properties of the estimator (7) and (8). We give some extra definitions that will be used in the sequel. We first assume that the support of X is of the form $\mathbf{X}=\bigotimes_{p=1}^{p=d}[x_{lp}, x_{up}] \subset\mathbf{R}^d$, with $x_{lp} < x_{up}$ for any $p = 1, ..., d$. We also define the internal region

$$\mathbf{X}_h = \left\{x \in \mathbf{X} : \left\{x - h\omega : \omega \in \mathbf{K}^d\right\} \subset \mathbf{X}\right\}. \tag{11}$$

The set $\mathbf{X}_h$ is a compact subset of the interior of $\mathbf{X}$ that contains all points that are sufficiently far away from the boundary in which the local constant smoother suffers from the so-called edge effects. To give an example, if $\mathbf{K} = [-1, 1]$ we would have $\mathbf{X}_h = \bigotimes_{p=1}^{p=d} [x_{lp} + h, x_{up} - h]$. For each $x \in \mathbf{X}_h$, let $\tau(x)$ be some real positive number such that $\tau(x) < \sup\{t \in \mathbf{R}_+ : H(t|x) > 0\}$ and introduce also the positive real number $\tau \leq \min\{\tau(x) : x \in \mathbf{X}_h\}$. Finally, $H(t, x) := \mathbf{P}(T > t|x)f(x)$, where $f(x)$ is the probability density function of the stochastic vector X, $H(t, x, \tilde{\gamma} > 0) := H(t, x|\tilde{\gamma} > 0)\mathbf{P}(\tilde{\gamma} > 0)$, $H(t, x, \tilde{\gamma} = 0) := H(t, x|\tilde{\gamma} = 0)\mathbf{P}(\tilde{\gamma} = 0)$ and $||K||_2^2 := \int K^2(u)du$. We will employ the following assumptions to derive the rate of uniform consistency rate and asymptotic normality of the proposed estimator. All the results are proved in the Appendix.

Assumption 1 *The derivatives of $\lambda_j(t, x)$ $(j = 1, 2)$ and $H(t|x)$ with respect to x are continuously differentiable up to order 2 on the interior of $[0, \tau(x)]$ for any $x \in \mathbf{X}_h$, and the corresponding derivatives are uniformly bounded. Moreover, the probability density function $f(x)$ is strictly positive on $\mathbf{X}_h$.*

Assumption 2 *The probability of missing cause of failure $\pi(x)$ is bounded away from zero, that is, $\pi(x) \geq \epsilon > 0$ for any $x \in \mathbf{X}$.*

Assumption 3 *The univariate kernel K (i) is a probability density function with compact support $\mathbf{K}$=$[-\mathcal{S}_k, \mathcal{S}_k]$, where $0 < \mathcal{S}_k < \infty$, (ii) is of order 2, that is,*

$$\int_{-\mathcal{S}_k}^{\mathcal{S}_k} K(u)u^\eta du = \begin{cases} 1, \ \eta = 0, \\ 0, \ \eta = 1, \\ \mu_2(K) < \infty, \ \eta = 2. \end{cases}$$

Assumption 4 *For the bandwidth sequence it holds $\ln n = o(nh^d)$.*

The above conditions are standard and very weak in the context of nonparametric estimation. Assumption 1 gives the smoothness conditions about the underlying functions. Assumption 2, which describes the properties of the kernel that can be used for the implementation of the nonparametric estimator, allows for most commonly used kernels such as Epanechnikov, Quartic and Triweight. Note that we work with product kernels just for ease of notation. All the results of the paper carry over to nonparametric estimators, which employ other multivariate kernels. Finally, Assumption 3 is standard in nonparametric estimation and is required to ensure uniform convergence for the proposed nonparametric estimation method.

Let $\alpha_n \equiv \left(\frac{\ln n}{nh^d}\right)^{\frac{1}{2}} + h^2$, $\Xi \equiv [0,\tau] \times \mathbf{X}_h$. The first main result presents the uniform consistency rate for the nonparametric estimator of the cumulative incidence functions.

Theorem 1 *Suppose Assumptions 1-4, hold. Then, for $j = 1, 2$ and $\nu = C, L$, we have, as $n \to \infty$*

$$\sup_{(t,x)\in\Xi} \left| \hat{F}_j^{\nu}(t|x) - F_j(t|x) \right| = O_p\left(\alpha_n\right).$$

The first part of the above convergence rate corresponds to the stochastic part, whereas the second to the bias part. Bordes and Gneyou (2011) have obtained a similar result for an estimator which is related to the cumulative incidence function without considering the possibility of MAR observations. Also, their estimator is based on local constant smoothing. It is also worth mentioning that for the calculation of the convergence rate we focus, for the variable t, on the compact set $[0, \tau]$ where the survival function $H(t|x)$ is bounded away from zero for all $x \in \mathbf{X}_h$. The same approach is also adopted by Dabrowska (1989) who studies uniform convergence of the conditional Kaplan-Meier estimator. This assumption is essential for deriving uniform consistency of the estimated quantities which are employed in the estimation of the cumulative incidence function.

To continue with the asymptotic distribution, we give some extra definitions. For any $t > 0$, introduce the filtration

$$\mathcal{F}_t = \sigma(N_{1i}(u), N_{2i}(u), N_{oi}(u), X_i, R_i, Y_i(u) : 0 \leq u \leq t, 1 \leq i \leq n),$$

where the notation σ specifies the sigma algebra generated by the events within the parenthesis. With these definitions, for each $j = 1, 2$, the counting processes $N_{ji}(t)$ and $N_{oi}(t)$, have stochastic intensity $\lambda_j(t, X_i)Y_i(t)R_i$ and $\lambda(t, X_i)Y_i(t)(1 - R_i)$, respectively, that is,

$$\lambda_j(t, X_i)Y_i(t)R_i dt = \mathbf{E}((N_{ji}(t + dt)_-) - N_{ji}(t)_- | \mathcal{F}_{t^-}),$$

$$\lambda(t, X_i)Y_i(t)(1 - R_i)dt = \mathbf{E}((N_{oi}(t + dt)_-) - N_{oi}(t)_- | \mathcal{F}_{t^-}),$$

where $N_{ji}(t)_- = \lim_{u \uparrow t} N_{ji}(u)$ and $N_{oi}(t)_- = \lim_{u \uparrow t} N_{oi}(u)$. Namely, the $\lambda_j(t, X_i)Y_i(t)R_i$ and $\lambda(t, X_i)(1 - R_i)Y_i(t)$ give the conditional average rate of change in the counting processes $N_{ji}(t)$ and $N_{oi}(t)$, respectively, over the interval $[t, t + dt)$. The stochastic intensities $\lambda_j(t, X_i)Y_i(t)R_i$ and $\lambda(t, X_i)Y_i(1 - R_i)$ are predictable with respect to $\mathcal{F}_t$. We will also assume that the $n-$variate process $\bar{\mathbf{N}}(t) = \{\bar{N}_1(t), ..., \bar{N}_n(t) : t \geq 0\}$ is a multivariate counting process, that is, it is not possible for two processes $\bar{N}_i(t), \bar{N}_\eta(t)$, with $i, \eta = 1, ..., n$ and $i \neq \eta$, to simultaneously jump. We should point out here that as $\bar{N}_i(t)$ jumps only if either $N_{1i}(t)$ or $N_{2i}(t)$ or $N_{oi}(t)$ jumps, the counting processes $N_{1i}(t)$, $N_{2i}(t)$ and $N_{oi}(t)$ will not jump simultaneously either. The latter facts are used for the derivation of the asymptotic variance of the nonparametric estimator.

For each $t > 0$, $j = 1, 2$, and $i = 1, ..., n$, consider the $\mathcal{F}_t-$measurable processes

$$M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u, X_i)Y_i(u)R_i du, \qquad M_{oi}(u) = N_{oi}(t) - \int_0^t \lambda(u, X_i)Y_i(u)(1 - R_i)du.$$

It is not difficult to verify that $\mathbf{E}M_{ji}(t) = 0$. Working analogously to the proof of Theorem 1 of Shorack and Wellner (2009), we note that for any $t > s > 0$, $\mathbf{E}\left[M_{ji}(t)|\mathcal{F}_s\right] = N_{ji}(s) - \int_0^s \lambda_j(u, X_i)Y_i(u)R_i du = M_{ji}(s)$ and $\mathbf{E}\left[M_{oi}(t)|\mathcal{F}_s\right] = N_{oi}(t) - \int_0^s \lambda_j(u, X_i)Y_i(u)(1 - R_i)du = M_{oi}(s)$ for all $t > s > 0$. Again, following the steps of the proof of Theorem 1 of Shorack and Wellner (2009) we get $\mathbf{E}M_{ji}^2(t) \leq 1$ and $\mathbf{E}\left[M_{oi}(t)\right]^2 \leq 1$ for each $t > 0$. The above discussion shows that $M_{ji}(t)$ and $M_{oi}(t)$ are zero-mean (local) square integrable martingales with respect to the filtration $\mathcal{F}_t$.

Next, our focus is on the asymptotic distribution of the nonparametric estimators (7) and (8) for any fixed value $x \in \mathbf{X}_h$. Let $\mathcal{D}\left[0, \tau(x)\right]$ denote the space of cadlag functions endowed with the Skorohod topology. Additionally, the symbol $\Longrightarrow$ will imply weak convergence. Define for

$j = 1, 2,$

$$d_j(t,u,x) = \frac{S(u|x)}{H(u,x)}, \qquad \rho_j(t,u,x) = -\frac{\int_u^t S(\epsilon|x)\lambda_j(\epsilon,x)d\epsilon}{H(u,x)}.$$

Additionally, for $\xi = 1, 2$, with $\xi \neq j$,

$$b_{jA}^C(t,x) = \sum_{l=0}^{1} \frac{\mu_2(K)h^r}{(2-l)!l!} \sum_{p=1}^{d} \int_0^t \frac{\partial^{2-l}\lambda_j(u,x)}{\partial x_p^{2-l}} \frac{\partial^l H(u,x)}{\partial x_p^l} \left[d_j(t,u,x) + \rho_j(t,u,x)\right] du$$
$$b_{jB}^C(t,x) = \sum_{l=0}^{1} \frac{\mu_2(K)h^2}{(2-l)!l!} \sum_{p=1}^{d} \int_0^t \frac{\partial^{2-l}\lambda_\xi(u,x)}{\partial x_p^{2-l}} \frac{\partial^l H(u,x)}{\partial x_p^l} \rho_j(t,u,x)du,$$

and

$$b_{jA}^L(t,x) = \frac{1}{2} \sum_{p=1}^{d} \int_0^t \frac{\partial^2 \lambda_j(u,x)}{\partial x_p^2} H(u,x) \left[d_j(t,u,x) + \rho_j(t,u,x)\right] du,$$
$$b_{jB}^L(t,x) = \frac{1}{2} \sum_{p=1}^{d} \int_0^t \frac{\partial^2 \lambda_\xi(u,x)}{\partial x_p^2} H(u,x)\rho_j(t,u,x)du,$$

Moreover,

$$v_{jA}(t,x) = ||K||_2^2 \int_0^t \left[\frac{1}{\pi(x)} H(u,x,\tilde{\gamma} > 0) + H(u,x,\tilde{\gamma} = 0)\right] d_j^2(t,u,x)\lambda_j(u,x)du,$$
$$v_{jB}(t,x) = ||K||_2^2 \int_0^t H(u,x)[\lambda_j(u,x) + \lambda_\xi(u,x)]\rho_j^2(t,u,x)du,$$
$$v_{jAB}(t,x) = 2\,||K||_2^2 \int_0^t H(u,x)d_j(t,u,x)\rho_j(t,u,x)\lambda_j(u,x)du,$$
$$\varsigma_1(t,x) = ||K||_2^2 \int_0^t H(u,x) \left[d_1(t,u,x)\rho_2(t,u,x) + \rho_1(t,u,x)\rho_2(t,u,x)\right] \lambda_1(u,x)du,$$
$$\varsigma_2(t,x) = ||K||_2^2 \int_0^t H(u,x) \left[d_2(t,u,x)\rho_1(t,u,x) + \rho_1(t,u,x)\rho_2(t,u,x)\right] \lambda_2(u,x)du.$$

Before stating Theorem 2, which establishes, pointwise in x, results for the weak convergence on a Gaussian process of our nonparametric estimator, we replace Assumption 4 by Assumption 5.

Assumption 5 *For the bandwidth sequence it holds i)* $\ln n = o(nh^d)$, *ii)* $nh^{d+4} = O(1)$.

Clearly, Assumption 5 is stronger than Assumption 4. The extra condition that appears in Assumption 5 has as consequence that the bandwidth parameter is optimally chosen in the sense that the bias squared and variance of the nonparametric estimator are of the same order.

We are now in a position to state Theorem 2.

Theorem 2 *Suppose Assumptions 1-3,5 hold. Then, for each $x \in \mathbf{X}_h$, we have, as $n \to \infty$*

$$\sqrt{nh^d}\begin{bmatrix} \hat{F}_1^{\nu}(t|x) - F_1(t|x) - b_{1A}^{\nu}(t,x) - b_{1B}^{\nu}(t,x) \\ \\ \hat{F}_2^{\nu}(t|x) - F_2(t|x) - b_{2A}^{\nu}(t,x) - b_{2B}^{\nu}(t,x) \end{bmatrix} \Longrightarrow \mathcal{N}(0, V(t,x))$$

over $\mathcal{D}\left[0, \tau(x)\right]^2$, where

$$V(t,x) = \begin{bmatrix} v_{1A}(t,x) + v_{1B}(t,x) + v_{1AB}(t,x) & \varsigma_1(t,x) + \varsigma_2(t,x) \\ \\ \varsigma_1(t,x) + \varsigma_2(t,x) & v_{2A}(t,x) + v_{2B}(t,x) + v_{2AB}(t,x) \end{bmatrix}$$

is a positive semidefinite matrix on $[0, \tau(x)]$ for each $x \in \mathbf{X}_h$.

To digest the above result about the asymptotic distribution, we make some comments. Recall that

$$\hat{F}_j^{\nu}(t|x) = \int_0^t \hat{S}^{\nu}(u - |x) d\hat{\Lambda}_j^{\nu}(u,x).$$

For the estimation of $F_j(t,x)$ we use the estimators $\hat{\Lambda}_j^{\nu}(t,x)$ and $\hat{S}^{\nu}(t,x)$. The bias and variance due to $\hat{\Lambda}_j^{\nu}(t,x)$ is captured by the terms $b_{jA}^{\nu}(t,x)$ and $v_{jA}(t,x)$. On the other hand, the bias and variance due to $\hat{S}^{\nu}(t,x)$ is captured by the terms $b_{jB}^{\nu}(t,x)$ and $v_{jB}(t,x)$. Moreover, the term $v_{jAB}(t,x)$ refers to the covariance due to the simultaneous estimation of $\hat{\Lambda}_j^{\nu}(t,x)$ and $\hat{S}^{\nu}(t,x)$. To gain some more intuition, assume that we know the survival function $S(t|x)$. Then, we obtain the estimator $\tilde{F}_j^{\nu}(t|x) = \int_0^t S(u - |x) d\hat{\Lambda}_j^{\nu}(u,x)$ for the cumulative incidence function. The asymptotic distribution of $\tilde{F}_j(t|x)$ will be then expressed as follows

$$\sqrt{nh^d}\left[\tilde{F}_j^{\nu}(t|x) - F_j(t|x) - b_{jA}^{\nu}(t,x)\right] \Longrightarrow \mathcal{N}(0, v_{jA}(t,x)).$$

Regarding the off-diagonal terms of the matrix $V(t,x)$, they capture the covariance between the estimators $\hat{F}_1^{\nu}(t|x)$ and $\hat{F}_2^{\nu}(t|x)$. By definition, the quantity $\varsigma_j(t,x)$ is expressed as the sum of two terms. The reason for the presence of these two terms is that in order to estimate $S(t|x)$, we have to estimate $\Lambda_1(t,x)$ and $\Lambda_2(t,x)$. Hence, for the estimation of $F_j(t|x)$ we need estimate "twice" the quantity $\Lambda_j(t,x)$ and "once" the quantity $\Lambda_\xi(t,x)$, with $\xi \neq j$. Finally, note that the correlation between the estimators $\hat{\Lambda}_j^\nu(t,x)$ and $\hat{\Lambda}_\xi^\nu(t,x)$ is asymptotically negligible as the correlation between the underlying counting process martingales which are employed for these two estimators are asymptotically negligible (see Appendix for the details).

Corollary 1 *Suppose Assumptions 1,3,4,5 hold and* $\pi(x) = 1$ *for all* $x \in \mathbf{X}_h$. *Then, for each* $x \in \mathbf{X}_h$, *we have, as* $n \to \infty$

$$\sqrt{nh^d}\begin{bmatrix} \hat{F}_1^\nu(t|x) - F_1(t|x) - b_{1A}^\nu(t,x) - b_{1B}^\nu(t,x) \\ \\ \hat{F}_2^\nu(t|x) - F_2(t|x) - b_{2A}^\nu(t,x) - b_{2B}^\nu(t,x) \end{bmatrix} \Longrightarrow \mathcal{N}(0, \ddot{V}(t,x))$$

over $\mathcal{D}\left[0,\tau(x)\right]^2$, *where*

$$\ddot{V}(t,x) = \begin{bmatrix} \dot{v}_{jA}(t,x) + v_{1B}(t,x) + v_{1AB}(t,x) & \varsigma_1(t,x) + \varsigma_2(t,x) \\ \\ \varsigma_1(t,x) + \varsigma_2(t,x) & \dot{v}_{jA} + v_{2B}(t,x) + v_{2AB}(t,x) \end{bmatrix},$$

with $\ddot{v}_{jA}(t,x) = ||K||_2^2 \int_0^t H(u,x) d_j^2(t,u,x)\lambda_j(u,x)du$.

To the best of our knowledge, the above result is the first one which characterizes the asymptotic distribution of the nonparametric estimator of cumulative incidence functions under the assumption that continuous covariates affect the latent failure times and the cause of failure is always observed. By abusing a bit our notation, Pepe (1991) shows that for a homogeneous population the following stochastic expansion holds uniformly in $t > 0$,

$$\hat{F}_j(t) - F_j(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t [d_j(t,u) + \rho_j(t,u)]dM_{ji}(u) + \frac{1}{n}\sum_{i=1}^{n}\int_0^t \rho_j(t,u)dM_{\xi i}(u) + o_p\left(n^{-\frac{1}{2}}\right).$$

Application of the martingale central limit theorem leads exactly to the variance-covariance matrix of Corollary 1 (of course by suppressing dependence on x).

To verify the validity of the statement of Corollary 1, we also consider the nonparametric estimator of the cumulative incidence function if there is no censoring (i.e., $Z = \infty$) and missing observations. For simplicity, we focus on the local constant estimator. Clearly, $N_{ji}(t) = 1\{T_i \leq t, \gamma_i = j\}$ and consequently,

$$\hat{F}_j^C(t|x) = \frac{1}{\sum_{i=1}^n \mathcal{K}_h(x-X_i)} \sum_{i=1}^n \int_0^t \mathcal{K}_h(x-X_i)dN_{ji}(u),$$
$$\hat{S}^C(t-|x) = \frac{1}{\sum_{i=1}^n \mathcal{K}_h(x-X_i)} \sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(t),$$

The absence of censoring implies $S(t|x) = H(t|x)$ for each $(t,x) \in \mathbf{R}_+ \times \mathbf{X}$ and we thus get

$$\begin{aligned}
\hat{F}_j^C(t|x) - F_j(t|x) &= \frac{1}{\sum_{i=1}^n \mathcal{K}_h(x-X_i)} \sum_{i=1}^n \int_0^t \mathcal{K}_h(x-X_i)dN_{ji}(u) - \int_0^t S(u-|x)\lambda_j(u,x)du \\
&= \frac{1}{\sum_{i=1}^n \mathcal{K}_h(x-X_i)} \sum_{i=1}^n \int_0^t \mathcal{K}_h(x-X_i)dM_{ji}(u) \\
&+ \int_0^t \left[\hat{S}^C(u-|x) - S(u-|x)\right] \lambda_j(u,x)du \\
&+ \frac{1}{\sum_{i=1}^n \mathcal{K}_h(x-X_i)} \sum_{i=1}^n \int_0^t \mathcal{K}_h(x-X_i)Y_i(u)\left[\lambda_j(u,X_i) - \lambda_j(u,x)\right]du \\
&= \left[\mathcal{V}_{jj}(t,x) + \mathcal{V}_{j\xi}(t,x) + \mathcal{B}_{jj}(t,x) + \mathcal{B}_{j\xi}(t,x)\right]\left[1 + o_p(1)\right],
\end{aligned}$$

where the quantities $\mathcal{V}_{jj}(t,x), \mathcal{V}_{j\xi}(t,x), \mathcal{B}_{jj}(t,x), \mathcal{B}_{j\xi}(t,x)$ are defined in the Appendix by setting $\pi_*(X_i, \tilde{\gamma}_i) = 1$ for this case. The second equality follows by $M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u, X_i)Y_i(u)du$, and the third equality is obtained by using the Duhamel equation (see proof of Theorem 2) for the expansion of the term $\hat{S}^C(u-|x) - S(u-|x)$, and the fact that $\inf_{x \in \mathbf{X}_h} \sum_{i=1}^n \mathcal{K}_h(x-X_i)/n \geq \epsilon + o_p(1)$ for large n, which is obtained by combining standard results in nonparametric density estimation (Gine and Guillou, 2002) and Assumption 1. Using arguments similar to the ones applied in the proof of Theorem 2, we get the distribution of Corollary 1.

4 Conclusions

This paper proposes a nonparametric method for estimating for each risk the corresponding cumulative incidence function in competing risks models, if continuous covariates affect the latent failure outcomes and the cause of failure is Missing At Random for some observations. The rate of uniform consistency on compact sets and pointwise asymptotic normality of the proposed estimator are derived. Existing estimation procedures, which account for covariates, are either fully parametric or semiparametric. In contrast to these estimation methods, the proposed estimator does not make any functional assumptions and thus it is robust under any specification for the underlying model. There are several topics for further research. First, the asymptotic distribution of the nonparametric estimator is characterized by unknown quantities that we need to estimate for statistical inference (e.g., construction of confidence intervals) that is not very appealing. Thus, we plan to derive the asymptotic distribution of the bootstrap estimator that will have the advantage of not needing to estimate unknown quantities. Second, interesting topic for future research is the use of the proposed nonparametric estimator for testing the significance of treatment effects in competing risks models and to extend the work of previous relevant articles that were cited in the introduction. For instance, the test statistic for the significance of distributional treatment effects which is developed by Lee and Whang (2009) is not directly applicable in cases in which the outcome is censored for some of the observations. Finally, it is worthwhile studying nonparametric estimation of the cumulative incidence functions in the multivariate competing risks model (Cheng et al., 2007) which is extension of the conventional competing risks model. In this setup, either the subjects are clustered or there are multiple observations for each subject. Multivariate competing risks models are useful, for instance, in the analysis of data for familial diseases or unemployment data where multiple unemployment spells are observed for each individual.

A Appendix

We begin with two lemmas, whose statements will be employed to prove the main results in the main text.

Define $F_{\tilde{\gamma}=j}(t|x) = \mathbf{P}(T \leq t, \tilde{\gamma} = j|x)$ for $j = 1, 2$. Consider the respective estimators of $F_{\tilde{\gamma}=j}(t|x)$ and $H(t-|x)$

$$\hat{F}^{\nu}_{\tilde{\gamma}=j}(t|x) = \sum_{i=1}^{n} w_i^{\nu}(x) N_{ji}(t), \qquad \breve{F}^{\nu}_{\tilde{\gamma}=j}(t|x) = \sum_{i=1}^{n} b_i^{\nu}(x) N_{ji}(t),$$

and

$$\hat{H}^{\nu}(t-|x) = \sum_{i=1}^{n} w_i^{\nu}(x) Y_i(t), \qquad \breve{H}^{\nu}(t-|x) = \sum_{i=1}^{n} b_i^{\nu}(x) Y_i(t).$$

It is clear that the local weight for the observation i of the estimators $\hat{F}^{\nu}_{\tilde{\gamma}=j}(t|x)$ and $\hat{H}^{\nu}(t-|x)$ is equal to the local weight of the estimators $\breve{F}^{\nu}_{\tilde{\gamma}=j}(t|x)$ and $\breve{H}^{\nu}(t|x)$, multiplied by $\frac{R_i}{\hat{\pi}^{\nu}_*(X_i)}$ for $\nu \in C, L$. Lemma 1 states that multiplication results in a stochastic error of order $O_p(\alpha_n)$.

Lemma 1 *Suppose Assumptions 1-4 hold. Then, we have for $\nu \in \{C, L\}$ and $j = 1, 2$, as $n \to \infty$,*

$$\sup_{(t,x)\in\Xi} \left| \hat{F}^{\nu}_{\tilde{\gamma}=j}(t|x) - \breve{F}^{\nu}_{\tilde{\gamma}=j}(t|x) \right| = O_p(\alpha_n),$$
$$\sup_{(t,x)\in\Xi} \left| \hat{H}^{\nu}(t-|x) - \breve{H}^{\nu}(t-|x) \right| = O_p(\alpha_n).$$

Proof. We restrict our attention to the local constant estimator. Similar algebraic calculations can be carried out for the local linear estimator and therefore we will skip this part.

To keep the notation simple, we omit the superscript C. Let

$$\hat{F}_{\tilde{\gamma}=j}(t,x) = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\hat{\pi}_*(X_i, \tilde{\gamma}_i)} \mathcal{K}_h(x - X_i) N_{ji}(t), \qquad \breve{F}_{\tilde{\gamma}=j}(t|x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) N_{ji}(t),$$

and

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\hat{\pi}_*(X_i, \tilde{\gamma}_i)} \mathcal{K}_h(x - X_i), \qquad \breve{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i).$$

Clearly,

$$\hat{F}_{\tilde{\gamma}=j}(t|x) = \frac{\hat{F}_{\tilde{\gamma}=j}(t,x)}{\hat{f}(x)}, \qquad \breve{F}_{\tilde{\gamma}=j}(t|x) = \frac{\breve{F}_{\tilde{\gamma}=j}(t,x)}{\breve{f}(x)}.$$

Define $\mathcal{F}_j(t,x) = |\hat{F}_{\tilde{\gamma}=j}(t,x) - \breve{F}_{\tilde{\gamma}=j}(t,x)|$. In the following we show that $\sup_{(t,x)\in\Xi} \mathcal{F}_j(t,x) = O_p(\alpha_n)$. It is straightforward to see that

$$\begin{aligned}
\sup_{(t,x)\in\Xi} \mathcal{F}_j(t,x) &\leq \sup_{(t,x)\in\Xi} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) N_{ji}(t) \left[\frac{R_i \left(\pi_*(X_i,\tilde{\gamma}_i) - \hat{\pi}_*(X_i,\tilde{\gamma}_i)\right)}{\hat{\pi}_*(X_i,\tilde{\gamma}_i)\pi_*(X_i,\tilde{\gamma}_i)} \right] \right| \\
&+ \sup_{(t,x)\in\Xi} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) N_{ji}(t) \left[\frac{R_i - \pi_*(X_i,\tilde{\gamma}_i)}{\pi_*(X_i,\tilde{\gamma}_i)} \right] \right| \\
&=: \sup_{(t,x)\in\Xi} \mathcal{F}_{jA}(t,x) + \sup_{(t,x)\in\Xi} \mathcal{F}_{jB}(t,x). \qquad \text{(A-1)}
\end{aligned}$$

Let $\beta_n \equiv \left(\frac{\ln n}{nh^d}\right)^{\frac{1}{2}} + h$. By definition, $\pi_*(X_i,\tilde{\gamma}_i) = 1\{\tilde{\gamma}_i \neq 0\}\pi(X_i) + 1\{\tilde{\gamma}_i = 0\}$ and hence, $\hat{\pi}_*(X_i,\tilde{\gamma}_i) = 1\{\tilde{\gamma} \neq 0\}\hat{\pi}(X_i) + 1\{\tilde{\gamma}_i = 0\}$. By Hansen (2008), $\hat{\pi}(X_i) - \pi(X_i) = O_p(\alpha_n)$ uniformly in $i = 1,...,n$, if $X_i \in \mathbf{X}_h$, and $\hat{\pi}(X_i) - \pi(X_i) = O_p(\beta_n)$ uniformly in $i = 1,...,n$, if $X_i \in \mathbf{X}\backslash\mathbf{X}_h$. Also, $\pi(X_i)$ is bounded away from zero with probability one, by making use of Assumption 2. Hence, $\min_{i\leq i\leq n} \hat{\pi}_*(X_i,\tilde{\gamma}_i)\pi_*(X_i,\tilde{\gamma}_i) \geq \epsilon + o_p(1)$ for $\epsilon > 0$ and consequently we get for positive constants $\mathcal{C}_1$ and $\mathcal{C}_2$, and with large probability

$$\begin{aligned}
\sup_{(t,x)\in\Xi} \mathcal{F}_{JA}(t,x) &\leq \mathcal{C}_1 O_p(\alpha_n) \sup_{x\in\mathbf{X}_h} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) 1\{X_i \in \mathbf{X}_h\} \right| \\
&+ \mathcal{C}_2 O_p(\beta_n) \sup_{x\in\mathbf{X}_h} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) 1\{X_i \in \mathbf{X}\backslash\mathbf{X}_h\} \right|. \qquad \text{(A-2)}
\end{aligned}$$

By applying results of Hansen (2008) we have with probability one

$$\sup_{x\in\mathbf{X}_h} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) 1\{X_i \in \mathbf{X}_h\} \right| = O(1), \qquad \text{(A-3)}$$

$$\sup_{x\in\mathbf{X}_h} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(x - X_i) 1\{X_i \in \mathbf{X}\backslash\mathbf{X}_h\} \right| = O(h). \qquad \text{(A-4)}$$

The above arguments imply that

$$\sup_{(t,x)\in\Xi} \mathcal{F}_{jA}(t,x) = O_p(\alpha_n). \tag{A-5}$$

To derive the stochastic order of $\sup_{(t,x)\in\Xi} \mathcal{F}_{jB}(t,x)$, we first note

$$\mathbf{E}\left\{\mathcal{K}_h(x - X_i)N_{ji}(t)\left[\frac{R_i - \pi_*(X_i,\tilde{\gamma}_i)}{\pi_*(X_i,\tilde{\gamma}_i)}\right]\right\} = 0. \tag{A-6}$$

Hence, employing analogous arguments to the ones of Hansen (2008), we can show that with probability one

$$\sup_{(t,x)\in\Xi} \mathcal{F}_{jB}(t,x) = O(\alpha_n). \tag{A-7}$$

Similarly, one can show that $\sup_{x\in\mathbf{X}_h} |\hat{f}(x) - \breve{f}(x)| = O_p(\alpha_n)$ which completes the proof for the first result of the Lemma. Also, the result $\sup_{(t,x)\in\Xi}\left|\hat{H}^{\nu}(t-|x) - \breve{H}^{\nu}(t-|x)\right| = O_p(\alpha_n)$ can be derived by following similar arguments. The proof is complete. ■

The following lemma deals with the uniform convergence rates of $\hat{\Lambda}_j(t,x)$ and $\hat{S}(t|x)$ over the compact set Ξ.

Lemma 2 *Suppose Assumptions 1-4 hold. Then, it holds for $\nu \in \{C, L\}$ and $j = 1, 2$, as $n \to \infty$,*

$$\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}^{\nu}_j(t,x) - \Lambda_j(t,x)\right| = O_p(\alpha_n),$$

and with probability one

$$\sup_{(t,x)\in\Xi}\left|\hat{S}^{\nu}(t|x) - S(t|x)\right| = O(\alpha_n).$$

Proof. We first argue that $\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}_j(t,x) - \Lambda_j(t,x)\right| = O_p(\alpha_n)$. Denote by $\breve{\Lambda}^{\nu}_j(t,x)$ the estimator of $\Lambda_j(t,x)$ if all observations were complete, that is, if $R_i = 1$ for each $i = 1, ..., n$. Clearly,

$$\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}^{\nu}_j(t,x) - \Lambda_j(t,x)\right| \le \sup_{(t,x)\in\Xi}\left|\hat{\Lambda}^{\nu}_j(t,x) - \breve{\Lambda}_j(t,x)\right| + \sup_{(t,x)\in\Xi}\left|\breve{\Lambda}^{\nu}_j(t,x) - \Lambda_j(t,x)\right|. \tag{A-8}$$

By employing results of González-Manteiga and Cadarso-Suarez (1994), we have

$$\sup_{(t,x)\in\Xi}\left|\breve{\Lambda}^{\nu}_j(t,x) - \Lambda_j(t,x)\right| = O(\alpha_n) \tag{A-9}$$

with probability one. Hence, we need to establish that $\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}_j^\nu(t,x)-\breve{\Lambda}_j^\nu(t,x)\right| = O_p(\alpha_n)$.

By definition, it is straightforward to see that

$$\hat{\Lambda}_j^\nu(t,x) = \int_0^t \frac{d\hat{F}^\nu_{\tilde{\gamma}=j}(u|x)}{\hat{H}^\nu(u-|x)}, \qquad \breve{\Lambda}_j^\nu(t,x) = \int_0^t \frac{d\breve{F}^\nu_{\tilde{\gamma}=j}(u|x)}{\breve{H}^\nu(u-|x)}.$$

By applying partial integration and triangle inequality we obtain

$$\begin{aligned}\left|\hat{\Lambda}_j^\nu(t,x)-\breve{\Lambda}_j^\nu(t,x)\right| &\le \left|\int_0^t \left[\frac{1}{\hat{H}^\nu(u-|x)}-\frac{1}{\breve{H}^\nu(u-|x)}\right] d\hat{F}^\nu_{\tilde{\gamma}=j}(u|x)\right| \\ &+ \left|\left[\hat{F}^\nu_{\tilde{\gamma}=j}(t|x)-\breve{F}^\nu_{\tilde{\gamma}=j}(t|x)\right]\frac{1}{\breve{H}^\nu(t-|x)}\right| \\ &+ \left|\int_0^t \left[\hat{F}^\nu_{\tilde{\gamma}=j}(u|x)-\breve{F}^\nu_{\tilde{\gamma}=j}(u|x)\right] d\left[\frac{1}{\breve{H}^\nu(u-|x)}\right]\right| \\ &=: \mathcal{D}_1(t,x)+\mathcal{D}_2(t,x)+\mathcal{D}_3(t,x). \end{aligned} \tag{A-10}$$

Note that

$$\frac{1}{\hat{H}^\nu(t-|x)}-\frac{1}{\breve{H}^\nu(t-|x)} = \frac{\breve{H}^\nu(t-|x)-\hat{H}^\nu(t-|x)}{\hat{H}^\nu(t-|x)\breve{H}^\nu(t-|x)}.$$

Hence,

$$\sup_{(t,x)\in\Xi}\mathcal{D}_1(t,x) \le \sup_{(t,x)\in\Xi}\left|\frac{\breve{H}^\nu(t-|x)-\hat{H}^\nu(t-|x)}{\hat{H}^\nu(t-|x)\breve{H}^\nu(t-|x)}\right| \sup_{(t,x)\in\Xi}\left|\hat{F}^\nu_{\tilde{\gamma}=j}(t|x)\right| \tag{A-11}$$

Lemma 1 implies that $\hat{H}^\nu(t-|x) = \breve{H}^\nu(t-|x)+O_p(\alpha_n)$ uniformly over Ξ. Also, by González-Manteiga and Cadarso-Suarez (1994), $\breve{H}^\nu(t-|x) = H^\nu(t-|x)+O_p(\alpha_n)$ uniformly over Ξ. By Assumption 1, we have $\inf_{(t,x)\in\Xi} H(t-|x) > 0$. Hence, $\inf_{(t,x)\in\Xi}\hat{H}(t-|x)\breve{H}(t-|x) \ge \epsilon + o_p(1)$. The latter entails

$$\begin{aligned}\sup_{(t,x)\in\Xi}\left|\frac{\breve{H}^\nu(t-|x)-\hat{H}^\nu(t-|x)}{\hat{H}^\nu(t-|x)\breve{H}^\nu(t-|x)}\right| &\le \inf_{(t,x)\in\Xi}\hat{H}(t-|x)\breve{H}(t-|x)\sup_{(t,x)\in\Xi}\left|\breve{H}^\nu(t-|x)-\hat{H}^\nu(t-|x)\right| \\ &= O_p(1)\cdot O_p(\alpha_n) = O_p(\alpha_n). \end{aligned} \tag{A-12}$$

Also, note that

$$\sup_{(t,x)\in\Xi}\left|\hat{F}^{\nu}_{\tilde{\gamma}=j}(t|x)\right| \le \sup_{(t,x)\in\Xi}\left|\hat{F}^{\nu}_{\tilde{\gamma}=j}(t|x) - \breve{F}^{\nu}_{\tilde{\gamma}=j}(t,x)\right| + \sup_{(t,x)\in\Xi}\left|\breve{F}^{\nu}_{\tilde{\gamma}=j}(t|x)\right| = O_p(1), \qquad \text{(A-13)}$$

where the equality makes use of Lemma 1, and results of González-Manteiga and Cadarso-Suarez (1994) for the second term of the right-hand side. Consequently, combining (A-11) (A-12) and (A-13), it follows $\sup_{(t,x)\in\Xi}\mathcal{D}_1(t,x) = O_p(\alpha_n)$. In a similar fashion, we can deduce that $\sup_{(t,x)\in\Xi}\mathcal{D}_2(t,x) = O_p(\alpha_n)$ and $\sup_{(t,x)\in\Xi}\mathcal{D}_3(t,x) = O_p(\alpha_n)$. The aforementioned results, combined with (A-8), (A-9), (A-10), lead to the conclusion that

$$\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}^{\nu}_j(t,x) - \Lambda_j(t,x)\right| = O_p(\alpha_n). \qquad \text{(A-14)}$$

For the second claim of the lemma we can directly infer that

$$\sup_{(t,x)\in\Xi}\left|\hat{S}^{\nu}_j(t|x) - S(t|x)\right| = O(\alpha_n) \qquad \text{(A-15)}$$

with probability one by making use of the results of González-Manteiga and Cadarso-Suarez (1994). This concludes the proof. ■

In view of Lemma 2, we proceed with the proof of Theorem 1.

Proof of Theorem 1. Clearly,

$$\begin{aligned}\hat{F}^{\nu}_j(t|x) - F_j(t|x) &= \int_0^t \hat{S}^{\nu}(u-|x)d\left[\hat{\Lambda}^{\nu}_j(u,x) - \Lambda_j(u,x)\right] + \int_0^t \left[\hat{S}^{\nu}(u-|x) - S(u-|x)\right]\lambda_j(u,x)du \\ &=: \hat{\Upsilon}^{\nu}_{jA}(t,x) + \hat{\Upsilon}^{\nu}_{jB}(t,x), \qquad \text{(A-16)}\end{aligned}$$

Triangle inequality entails

$$\sup_{(t,x)\in\Xi}\left|\hat{F}^{\nu}_j(t|x) - F_j(t|x)\right| \le \sup_{(t,x)\in\Xi}\left|\hat{\Upsilon}^{\nu}_{jA}(t,x)\right| + \sup_{(t,x)\in\Xi}\left|\hat{\Upsilon}^{\nu}_{jB}(t,x)\right|. \qquad \text{(A-17)}$$

Partial integration, triangle inequality and use of Lemma 2 yields for the first term

$$\sup_{(t,x)\in\Xi}\left|\hat{\Upsilon}^{\nu}_{jA}(t,x)\right| \le \sup_{(t,x)\in\Xi}\left|\hat{\Lambda}^{\nu}_{j}(t,x)-\Lambda_j(t,x)\right| \sup_{(t,x)\in\Xi}\left|\hat{S}(t|x)\right|$$
$$+\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}^{\nu}_{j}(t,x)-\Lambda_j(t,x)\right| \sup_{(t,x)\in\Xi}\left|\int_0^t d\hat{S}^{\nu}(u-|x)\right| = O_p(\alpha_n), \quad \text{(A-18)}$$

as $\sup_{(t,x)\in\Xi}\left|\hat{S}^{\nu}(t|x)\right| = O_p(1)$. Regarding the second term, it is straightforward to check that

$$\sup_{(t,x)\in\Xi}\left|\hat{\Upsilon}^{\nu}_{jB}(t,x)\right| \le \sup_{(t,x)\in\Xi}\left|\hat{S}^{\nu}(t|x)-S(t|x)\right| \sup_{(t,x)\in\Xi}\left|\Lambda_j(t,x)\right| = O(\alpha_n), \quad \text{(A-19)}$$

where we apply Lemma 2 and use the fact that $\sup_{(t,x)\in\Xi}|\Lambda_j(t,x)| = O(1)$. Combining (A-17)-(A-19) completes the proof. ■

Before proving Theorem 2 we state the martingale central limit theorem (Andersen et al., 1993; Kalbfleisch and Prentice, 1991).

Proposition 1 (Martingale Central Limit Theorem) *Consider the filtration $\mathcal{F}_t$ for any $t>0$ and the $\mathcal{F}_t-$martingales $\mathcal{M}^{(n)}_1(t),...,\mathcal{M}^{(n)}_n(t)$ for $n\to\infty$. Moreover, introduce for $j=1,2$ and $n\to\infty$ the $\mathcal{F}_{t-}-$measurable processes $g^{(n)}_{j1}(t),...,g^{(n)}_{jn}(t)$ (i.e., the processes are predictable with respect to $\mathcal{F}_t$), the $\mathbf{R}^2$-valued martingale processes $\mathcal{U}^{(n)}=(\mathcal{U}^{(n)}_1,\mathcal{U}^{(n)}_2)$, where*

$$\mathcal{U}^{(n)}_j(t) = \sum_{i=1}^{n}\int_0^t g^{(n)}_{ji}(u)d\mathcal{M}^{(n)}_i(u), \;\; j=1,2$$

and $\mathcal{U}^{(n)}_{\varepsilon}=(\mathcal{U}^{(n)}_{1\varepsilon},\mathcal{U}^{(n)}_{2\varepsilon})$, where

$$\mathcal{U}^{(n)}_{j\varepsilon}(t) = \sum_{i=1}^{n}\int_0^t g^{(n)}_{ji}(u)1(\left|g^2_{ji}(u)>\varepsilon\right|)d\mathcal{M}^{(n)}_i(u), \;\; j=1,2.$$

Suppose that $p\lim_{n\to\infty}\left[\mathcal{U}^{(n)}<t>\right]=\Sigma(t)$, where Σ is a positive semidefinite matrix, and also $p\lim_{n\to\infty}\left[\mathcal{U}^{(n)}_{\varepsilon}<t>\right]=0$ for any $\varepsilon>0$. Then,

$$\mathcal{U}^{(n)}(t) \Longrightarrow \mathcal{N}(0,\Sigma(t)).$$

For $\xi = 1, 2$, with $\xi \neq j$, introduce

$$\mathcal{V}_{jj}(t,x) = \sum_{i=1}^{n} \int_0^t \frac{\mathcal{K}_h(x - X_i)}{\pi_*(X_i, \tilde{\gamma}_i)} \left[d_j(t,u,x) + \pi_*(X_i, \tilde{\gamma}_i)\rho_j(t,u,x)\right] dM_{ji}(u)$$

$$\mathcal{V}_{j\xi}(t,x) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \mathcal{K}_h(x - X_i)\rho_j(t,u,x) dM_{\xi i}(u)$$

$$\mathcal{V}_{jo}(t,x) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \mathcal{K}_h(x - X_i)\rho_j(t,u,x) dM_{oi}(u)$$

and

$$\begin{aligned}\mathcal{B}_{jj}(t,x) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t & \left[d_j(t,u,x) \frac{R_i \mathcal{K}_h(x - X_i) Y_i(u)}{\pi_*(X_i, \tilde{\gamma}_i)} + \rho_j(t,u,x)\right] \\ & \times \left[\lambda_j(u, X_i) - \lambda_j(u,x)\right] du, \\ \mathcal{B}_{j\xi}(t,x) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t & \rho_j(t,u,x)\mathcal{K}_h(x - X_i) Y_i(u) \left[\lambda_\xi(u, X_i) - \lambda_\xi(u,x)\right] du.\end{aligned}$$

We now proceed with the proof of Theorem 2.

Proof of Theorem 2. We will show the asymptotic normality for the estimator $\hat{F}_j^C(t|x)$. The asymptotic distribution for $\hat{F}_j^L(t|x)$ can be derived by following similar arguments. For ease of notation we skip the superscript C.

As in the proof of Theorem 1 we write

$$\begin{aligned}\hat{F}_j(t|x) - F_j(t|x) &= \int_0^t \hat{S}(u - |x) d\left[\hat{\Lambda}_j(u,x) - \Lambda_j(u,x)\right] \qquad \text{(A-20)} \\ &\quad + \int_0^t \left[\hat{S}(u - |x) - S(u - |x)\right] \lambda_j(u,x) du \\ &= \hat{\Upsilon}_{jA}(t,x) + \hat{\Upsilon}_{jB}(t,x). \qquad \text{(A-21)}\end{aligned}$$

By the definition of $\hat{\Lambda}_j(t,x)$ and the property $M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u, X_i) Y_i(u) R_i du$, it follows

$$\begin{aligned}\hat{\Upsilon}_{jA}(t,x) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \hat{S}(u - |x) \frac{R_i}{\hat{\pi}_*(X_i, \tilde{\gamma}_i)} \frac{\mathcal{K}_h(x - X_i) dM_{ji}(u)}{\frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\hat{\pi}_*(X_i, \tilde{\gamma}_i)} \mathcal{K}_h(x - X_i) Y_i(u)} \\ + \frac{1}{n} \sum_{i=1}^{n} \int_0^t \hat{S}(u - |x) \frac{R_i}{\hat{\pi}_*(X_i, \tilde{\gamma}_i)} \frac{\mathcal{K}_h(x - X_i) Y_i(u) \left[\lambda_j(u, X_i) - \lambda_j(u,x)\right]}{\frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\hat{\pi}_*(X_i, \tilde{\gamma}_i)} \mathcal{K}_h(x - X_i) Y_i(u)} du. \qquad \text{(A-22)}\end{aligned}$$

Next, we work on $\hat{\Upsilon}_{jB}(t,x)$ by making use of the Duhamel equation

$$\hat{S}(t|x) - S(t|x) = -S(t|x)\int_0^t \frac{\hat{S}(u-|x)}{S(u|x)} d\left[\hat{\Lambda}(u,x) - \Lambda(u,x)\right].$$

In particular, by construction of $\hat{\Lambda}(t,x)$, the equalities $\bar{N}_i(t) = N_{1i}(t) + N_{2i}(t) + N_{oi}(t)$, $M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u,X_i)Y_i(u)R_i du$ and $N_{io}(t) - \int_0^t \lambda(u,X_i)Y_i(u)(1-R_i)du$, the Duhamel formula, the fact that the mapping $t \mapsto S(t|x)$ is continuous for all $x \in \mathbf{X}_h$ (recall that L is absolutely continuous), and doing some algebra we obtain

$$\begin{aligned}
\hat{\Upsilon}_{jB}(t,x) = &-\frac{1}{n}\sum_{i=1}^n \int_0^t \left[\int_u^t S(\epsilon|x)\lambda_j(\epsilon,x)d\epsilon\right] \frac{\hat{S}(u-|x)}{S(u|x)} \\
&\times \frac{\mathcal{K}_h(x-X_i)}{\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(u)} d(M_{1i}(u) + M_{2i}(u) + M_{oi}(u)) \\
&- \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{\hat{S}(u-|x)}{S(u|x)} \frac{\mathcal{K}_h(x-X_i)Y_i(u)\left[\lambda(u,X_i) - \lambda(u,x)\right]}{\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(u)} \\
&\times \left[\int_u^t S(\epsilon|x)\lambda_j(\epsilon,x)d\epsilon\right] du.
\end{aligned} \tag{A-23}$$

By Lemma 1 and continuity of the mapping $t \mapsto S(t|x)$, $\frac{\hat{S}(t-|x)}{S(t|x)} = 1 + o_p(1)$ uniformly over $t \in [0,\tau(x)]$ for each $x \in \mathbf{X}_h$. Making use of similar arguments with the case of standard nonparametric regression Hansen (2008) or working analogously to González-Manteiga and Cadarso-Suarez (1994) we can show that it holds, pointwise in x,

$$\sup_{t\in[0,\tau(x)]} \left|\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(t)\right| = H(t-,x).$$

Recall also that $\hat{\pi}_*(X_i,\tilde{\gamma}_i) = \pi_*(X_i,\tilde{\gamma}_i) + o_p(1)$ uniformly over $= 1,...,n$ with $\pi_*(X_i,\tilde{\gamma}_i)$ to be bounded away from zero. Furthermore, $\lambda(t,x) = \lambda_1(t,x) + \lambda_2(t,x)$ for any $(t,x) \in \mathbf{R}_+ \times \mathbf{X}_h$. Combining (A-20), (A-22) and (A-23) we get

$$\hat{F}_j(t|x) - F_j(t|x) = \left[\mathcal{V}_{j1}(t,x) + \mathcal{V}_{j\xi}(t,x) + \mathcal{V}_{oj}(t,x) + \mathcal{B}_{jj}(t,x) + \mathcal{B}_{j\xi}(t,x)\right]\left[1 + o_p(1)\right]. \tag{A-24}$$

uniformly over $[0,\tau(x)]$ for each $x \in \mathbf{X}_h$.

Based on the previous lines, we rewrite the difference $\hat{F}_j(t|x) - F_j(t|x)$ as follows

$$\begin{aligned}\hat{F}_j(t|x) - F_j(t|x) = \{&\mathcal{V}_{j1}(t,x) + \mathcal{V}_{j2}(t,x) + \mathcal{V}_{jo}(t,x) + \mathbf{E}\mathcal{B}_{jj}(t,x) + \mathbf{E}\mathcal{B}_{j\xi}(t,x) \\ &+ [\mathcal{B}_{jj}(t,x) - \mathbf{E}\mathcal{B}_{jj}(t,x)] + [\mathcal{B}_{j\xi}(t,x) - \mathbf{E}\mathcal{B}_{j\xi}(t,x)]\}\,[1 + o_p(1)] \qquad \text{(A-25)}\end{aligned}$$

uniformly over $[0, \tau(x)]$ for each $x \in \mathbf{X}_h$. In virtue of the above equation it is clear that for deriving the asymptotic distribution of $\hat{F}_j(t|x) - F_j(t|x)$ it suffices to consider the term

$$\begin{aligned}&\mathcal{V}_{j1}(t,x) + \mathcal{V}_{j2}(t,x) + \mathcal{V}_{jo}(t,x) + \mathbf{E}\mathcal{B}_{jj}(t,x) + \mathbf{E}\mathcal{B}_{j\xi}(t,x) \\ &+ [\mathcal{B}_{jj}(t,x) - \mathbf{E}\mathcal{B}_{jj}(t,x)] + [\mathcal{B}_{j\xi}(t,x) - \mathbf{E}\mathcal{B}_{j\xi}(t,x)]\,.\end{aligned}$$

In the sequel, we show that the three terms in the above display give the variance of the estimator, the fourth and the fifth term give the bias of the estimator, whereas the last two terms are asymptotically negligible. Application of the martingale central limit theorem yields $\sqrt{nh^d}\mathcal{V}_{j1}(t,x) \Longrightarrow \mathcal{N}(0, v_{jj}(t,x))$, $\sqrt{nh^d}\mathcal{V}_{j\xi}(t,x) \Longrightarrow \mathcal{N}(0, v_{j\xi}(t,x))$, and $\sqrt{nh^d}\mathcal{V}_{jo}(t,x) \Longrightarrow \mathcal{N}(0, v_{jo}(t,x))$, where

$$\begin{aligned}v_{jj}(t,x) &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h^2(x - X_i)}{\pi_*^2(X_i,\tilde{\gamma}_i)}\left[d_j(t,u,x) + \pi_*(X_i,\tilde{\gamma}_i)\rho_j(t,u,x)\right]^2 d < M_{ji} > (u)\right] \\ &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h^2(x - X_i)}{\pi_*^2(X_i,\tilde{\gamma}_i)}\left[d_j(t,u,x) + \pi_*(X_i,\tilde{\gamma}_i)\rho_j(t,u,x)\right]^2 Y_i(u)R_i\lambda_j(u,X_i)du\right] \\ &= \frac{||K||_2^2}{\pi(x)}\int_0^t H(u,x,\tilde{\gamma} > 0)\left[d_j(t,u,x) + \pi(x)\rho_j(t,u,x)\right]^2\lambda_j(u,x)du \\ &+ ||K||_2^2\left[\int_0^t H(u,x,\tilde{\gamma} = 0)\left[d_j(t,u,x) + \rho_j(t,u,x)\right]^2\lambda_j(u,x)du\right] \qquad \text{(A-26)}\end{aligned}$$

$$\begin{aligned}
v_{j\xi}(t,x) &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)d<M_{\xi i}>(u)\right]\\
&= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)Y_i(u)R_i\lambda_\xi(u,X_i)du\right]\\
&= ||K||_2^2\,\pi(x)\int_0^t H(u,x,\tilde{\gamma}>0)\rho_j^2(t,u,x)\lambda_\xi(u,x)du\\
&+ ||K||_2^2\int_0^t H(u,x,\tilde{\gamma}=0)\rho_j^2(t,u,x)\lambda_\xi(u,x)du, \qquad \text{(A-27)}
\end{aligned}$$

and

$$\begin{aligned}
v_{jo}(t,x) &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)d<M_{oi}>(u)\right]\\
&= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)Y_i(u)\lambda(u,X_i)(1-R_i)du\right]\\
&+ ||K||_2^2\,[1-\pi(x)]\int_0^t H(u,x,\tilde{\gamma}>0)\rho_j^2(t,u,x)\lambda(u,x)du. \qquad \text{(A-28)}
\end{aligned}$$

where the last equalities in (A-26), (A-27) and (A-28) follow by using the definition of $\mathcal{K}_h(x-X_i)$, change of variables and dominated convergence theorem. For deriving the above expressions we have also used the fact that it is not possible for two components of the process $\bar{\mathbf{N}}(t) = \{\bar{N}_1(t),...,\bar{N}_n(t) : t \geq 0\}$ to simultaneously jump, which in turn implies the same for the process $\mathbf{N}_j(t) = \{N_{j1}(t),...,N_{jn}(t) : t \geq 0\}$, where $j = 1,2$. Therefore, by carrying out simple algebraic calculations we get $d < M_{ji}, M_{j\eta} > (t) = \text{Cov}(dM_{ji}(t), dM_{ji}(t)|\mathcal{F}_{t-}) = o(dt)$ and $d < M_{oi}, M_{o\eta} > (t) = \text{Cov}(dM_{oi}(t), dM_{o\eta}(t)|\mathcal{F}_{t-}) = o(dt)$ for any $t > 0$, $j = 1,2$, $i,\eta = 1,...,n$, with $i \neq \eta$. Furthermore, by construction the counting processes $N_{1i}(t)$, $N_{2i}(t), N_{oi}(t)$ cannot simultaneously jump (in fact, as soon as the one jumps, the other is not possible to jump) and we thus get $d < M_{1i}, M_{2i} > (t) = \text{Cov}(dM_{1i}(t), dM_{2i}(t)|\mathcal{F}_{t-}) = o(dt)$ and $d < M_{ji}, M_{oi} > (t) = \text{Cov}(dM_{ji}(t), dM_{oi}(t)|\mathcal{F}_{t-}) = o(dt)$. The latter facts, combined with the equality $\lambda(t,x) = \lambda_1(t,x) + \lambda_2(t,x)$ for any $(t,x) \in \mathbf{R}_+ \times \mathbf{X}_h$, and some algebra imply

$$\sqrt{nh^d}\,[\mathcal{V}_{jj}(t,x) + \mathcal{V}_{j\xi}(t,x) + \mathcal{V}_{jo}(t,x)] \Longrightarrow \mathcal{N}(0, v_{jA}(t,x) + v_{jB}(t,x) + v_{jAB}(t,x)). \qquad \text{(A-29)}$$

Next, we proceed with the stable parts in a similar manner to the approach of Nielsen and Linton (1995). Let $\omega = (\omega_1, \omega_2, ..., \omega_d)$. It is easy to check that

$$\begin{aligned}
\mathbf{E}\mathcal{B}_{jj}(t,x) &= \int_{\mathbf{R}^d}\int_0^t \prod_{p=1}^{d} K(\omega_p)\left[\lambda_j(u,x-h\omega)-\lambda_j(u,x)\right]H(u,x-h\omega)\left[d_j(t,u,x)+\rho_j(t,u,x)\right]dud\omega \\
&= \sum_{l=0}^{1}\frac{\mu_2(K)h^2}{(2-l)!l!}\sum_{p=1}^{d}\int_0^t \frac{\partial^{2-l}\lambda_j(u,x)}{\partial x_p^{2-l}}\frac{\partial^l H(u,x)}{\partial x_p^l}\left[d_j(t,u,x)+\rho_j(t,u,x)\right]du + o_p(h^2)
\end{aligned} \tag{A-30}$$

and

$$\begin{aligned}
\mathbf{E}\mathcal{B}_{j\xi}(t,x) &= \int_{\mathbf{R}^d}\int_0^t \prod_{p=1}^{d} K(\omega_p)\left[\lambda_\xi(u,x-h\omega)-\lambda_\xi(u,x)\right]H(u,x-h\omega)\rho_j(t,u,x)dud\omega \\
&= \sum_{l=0}^{1}\frac{\mu_2(K)h^2}{(2-l)!l!}\sum_{p=1}^{d}\int_0^t \frac{\partial^{2-l}\lambda_\xi(u,x)}{\partial x_p^{2-l}}\frac{\partial^l H(u,x)}{\partial x_p^l}\rho_j(t,u,x)du + o_p(h^2),
\end{aligned} \tag{A-31}$$

where the first equalities in the two above equations are obtained by the definition of $\mathcal{K}_h(x - X_i)$ and the second equalities by applying $r-$th Taylor series expansion with Lagrange remainder for the difference $\lambda_j(u, x - h\omega) - \lambda_j(u, x)$ $(j = 1, 2)$ and the quantity $H(u, x - h\omega)$, along with the fact that the K is of order r. Also, by working in a completely analogous way it is straightforward to show that $\mathbf{E}\mathcal{B}^2_{j\iota}(t, x) = O(nh^{d-2})^{-1}$ for $\iota = j, \xi$, which in turn gives $\mathbf{E}(\mathcal{B}_{j\iota}(t, x) - \mathbf{E}\mathcal{B}_{j\iota}(t, x))^2 = O(nh^{d-2})^{-1} = o(nh^d)^{-1}$ and consequently, by Chebyshev's inequality,

$$|\mathcal{B}_{j\iota}(t,x) - \mathbf{E}\mathcal{B}_{j\iota}(t,x)| = o_p(nh^d)^{-\frac{1}{2}}. \tag{A-32}$$

To calculate the off-diagonal terms of the matrix $V(t, x)$ we exploit analogous steps as for the derivation of $v_{jA}(t, x), v_{jB}(t, x)$ and $v_{jAB}(t, x)$. As previously discussed, we note that $d < M_{ji}, M_{j\eta} > (t) = o(dt)$, $d < M_{oi}, M_{o\eta} > (t)$, $d < M_{1i}, M_{2i} > (t) = o(dt)$, and $d < M_{ji}, M_{oi} > (t)$ for all $t > 0$, any $j = 1, 2$ and $i, \eta = 1, ..., n$, with $i \neq \eta$. Hence, application of martingale central limit theorem for the covariance terms yields

$$\varsigma_{11}(t,x) = p \lim_{n\to\infty} \left[\frac{h^d}{n} \sum_{i=1}^{n} \int_0^t \frac{\mathcal{K}_h^2(x-X_i)}{\pi_*(X_i,\tilde{\gamma}_i)} [d_1(t,u,x) + \pi_*(X_i,\tilde{\gamma}_i)\rho_1(t,u,x)]\rho_2(t,u,x) d < M_{1i} > (u) \right]$$
$$= ||K||_2^2 \int_0^t H(u,x,\gamma > 0)[d_1(t,u,x) + \pi(x)\rho_1(t,u,x)]\rho_2(t,u,x)\lambda_1(u,x)du$$
$$+ ||K||_2^2 \int_0^t H(u,x,\gamma = 0)[d_1(t,u,x) + \rho_1(t,u,x)]\rho_2(t,u,x)\lambda_1(u,x)du, \quad \text{(A-33)}$$

$$\varsigma_{22}(t,x) = p \lim_{n\to\infty} \left[\frac{h^d}{n} \sum_{i=1}^{n} \int_0^t \frac{\mathcal{K}_h^2(x-X_i)}{\pi_*(X_i,\tilde{\gamma}_i)} [d_2(t,u,x) + \pi_*(X_i,\tilde{\gamma}_i)\rho_2(t,u,x)]\rho_1(t,u,x) d < M_{2i} > (u) \right]$$
$$= ||K||_2^2 \int_0^t H(u,x,\gamma > 0)[d_2(t,u,x) + \pi(x)\rho_2(t,u,x)]\rho_1(t,u,x)\lambda_2(u,x)du$$
$$+ ||K||_2^2 \int_0^t H(u,x,\gamma = 0)[d_2(t,u,x) + \rho_2(t,u,x)]\rho_1(t,u,x)\lambda_2(u,x)du, \quad \text{(A-34)}$$

and

$$\varsigma_o(t,x) = p \lim_{n\to\infty} \left[\frac{h^d}{n} \sum_{i=1}^{n} \int_0^t \mathcal{K}_h^2(x-X_i)\rho_1(t,u,x)\rho_2(t,u,x) d < M_{oi} > (u) \right]$$
$$= ||K||_2^2 \left[1 - \pi(x)\right] \int_0^t H(u,x,\gamma > 0)\rho_1(t,u,x)\rho_2(t,u,x)\lambda(u,x)du, \quad \text{(A-35)}$$

where we make use of the definition of $\mathcal{K}_h(x-X_i)$, change of variables and dominated convergence theorem. By simple algebra, we deduce the desired result. ∎

References

Aly, E. E. A. A., Kochar, S. C., and McKeague, I. W. (1994), "Some tests for comparing cumulative incidence functions and cause-specific hazard rates," *Journal of the American Statistical Association*, 994–999.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993), *Statistical models based on counting processes*, Springer Verlag.

Bajorunaite, R. and Klein, J. (2007), "Two-sample tests of the equality of two cumulative incidence functions," *Computational statistics & data analysis*, 51, 4269–4281.

Bordes, L. and Gneyou, K. (2011), "Uniform convergence of nonparametric regressions in competing risks model with right censoring," *Statistics & Probability Letters*.

Cheng, Y., Fine, J. P., and Kosorok, M. R. (2007), "Nonparametric association analysis of bivariate competing-risks data," *Journal of the American Statistical Association*, 102, 1407–1415.

Dabrowska, D. M. (1989), "Uniform consistency of the kernel conditional Kaplan-Meier estimate," *The Annals of Statistics*, 17, 1157–1167.

Fine, J. P. and Gray, R. (1999), "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, 496–509.

Fleming, T. and Harrington, D. (1991), *Counting processes and survival analysis*, vol. 8, Wiley Online Library.

González-Manteiga, W. and Cadarso-Suarez, C. (1994), "Asymptotic properties of a generalized Kaplan-Meier estimator with some applications," *Communications in Statistics-Theory and Methods*, 4, 65–78.

Gray, R. J. (1988), "A class of K-sample tests for comparing the cumulative incidence of a competing risk," *The annals of statistics*, 1141–1154.

Hansen, B. E. (2008), "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24, 726.

Hu, Z., Follmann, D. A., and Qin, J. (2010), "Semiparametric dimension reduction estimation for mean response with missing data," *Biometrika*, 97, 305–319.

Jeong, J. and Fine, J. (2007), "Parametric regression on cumulative incidence function," *Biostatistics*, 8, 184–196.

Kalbfleisch, J. and Prentice, L. (1991), *The Statistical analysis of failure time data*, Wiley New York.

Lee, S. and Whang, Y. (2009), "Nonparametric tests of conditional treatment effects," .

Lin, D. et al. (1997), "Non-parametric inference for cumulative incidence functions in competing risks studies," *Statistics in medicine*, 16, 901–910.

Linton, O., Mammen, E., Nielsen, J. P., and Van Keilegom, I. (2011), "Nonparametric regression with filtered data," *Bernoulli*, 17, 60–87.

Little, R. J. A. and Rubin, D. B. (1987), *Statistical analysis with missing data*, vol. 4, Wiley New York.

Lu, W. and Liang, Y. (2008), "Analysis of competing risks data with missing cause of failure under addtive hazards model," *Statistica Sinica*, 19, 219–234.

Mammen, E. and Nielsen, J. (2007), "A general approach to the predictability issue in survival analysis with applications," *Biometrika*, 94, 873–892.

Nielsen, J. and Linton, O. (1995), "Kernel estimation in a nonparametric marker dependent hazard model," *The Annals of Statistics*, 1735–1748.

Peng, L. and Fine, J. P. (2008), "Nonparametric tests for continuous covariate effects with multi-state survival data," *Biometrics*, 64, 1080–1089.

Pepe, M. S. (1991), "Inference for events with dependent risks in multiple endpoint studies," *Journal of the American Statistical Association*, 770–778.

Rubin, D. (1976), "Inference and missing data," *Biometrika*, 63, 581–592.

Schaubel, D. and Wei, G. (2011), "Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring," *Biometrics*, 67, 29–38.

Scheike, T., Zhang, M., and Gerds, T. (2008), "Predicting cumulative incidence probability by direct binomial regression," *Biometrika*, 95, 205–220.

Shen, Y. and Cheng, S. C. (1999), "Confidence bands for cumulative incidence curves under the additive risk model," *Biometrics*, 55, 1093–1100.

Shorack, G. R. and Wellner, J. A. (2009), *Empirical processes with applications to statistics*, vol. 59, Society for Industrial Mathematics.

Tsiatis, A. (1975), "A nonidentifiability aspect of the problem of competing risks," *Proceedings of the National Academy of Sciences*, 72, 20.

Van den Berg, G. J. (2001), "Duration models: specification, identification and multiple durations," *Handbook of econometrics*, 5, 3381–3460.

The present dissertation was completed and written independently and without outside assistance and all auxiliary means were declared.

Georgios Effraimidis

Georgios Effraimidis

FIELDS

Duration Analysis, Applied Econometrics, Nonparametrics, Demography

EMPLOYMENT

Since September 2012	Postdoc at University of Southern Denmark

EDUCATION

July 2012	PhD in Economics (summa cum laude), University of Mannheim, VU University Amsterdam, *Essays in Multivariate Duration Models*, Advisor: Gerard J. van den Berg.
Sep. 2007	MPhil in Economics, Erasmus University of Rotterdam.
June 2005	MSc in Econometrics and Management Science, Erasmus University of Rotterdam.
June 2004	"Ptychion" in Industrial Management and Technology, University of Piraeus.

ACADEMIC VISITS

Spring 2011	University of Southern California, Department of Economics.
Spring 2010	University of Southern California, Department of Economics.

RESEARCH PAPERS

[1] *Asymptotic Theory for Nonparametric Estimation of Cumulative Incidence Functions*, working paper.

[2] *Association Measures in Bivariate Gamma Frailty Models*, with Gerard J. van den Berg, working paper.

[3] *Nonparametric Identification of Dynamic Treatment Effects in Competing Risks Models*, with Ina Drepper, working paper.

[4] *Social Interaction Effects in Duration Models*, with Ina Drepper, working paper.

[5] *A Semiparametric Test of Competing Risks Dependence Using Covariates*, with Gerard J. van den Berg, research in progress.

[6] *Stemming the Big Crash: A Triple Hazard Analysis of Price and Sales Crashes of New Products*, with Carlos Hernandez-Mireles and Gerard Tellis, working paper.

TEACHING EXPERIENCE

Spring 2011	Econometrics (graduate), University of Southern California.
Spring 2010	Econometrics (graduate), University of Southern California.
Fall 2008	Microeconomics (graduate), VU University Amsterdam.

CONFERENCE PRESENTATIONS

May 2011	Schaeffer center, Los Angeles.
June 2010	European Meeting of Statisticians, Piraeus.
Dec. 2010	University of Mannheim.
Sep. 2009	Royal Statistical Society, Edinburgh.
June 2009	Stochastic Processes and Applications, Berlin.
May 2009	Tinbergen Institute, Amsterdam.
Apr. 2009	Erasmus University Rotterdam.
Jan. 2009	VU University Amsterdam.
Aug. 2008	Econometric Society European Meeting, Milan.
Dec. 2007	RTN Meeting, Amsterdam.

OTHER SKILLS

Computer:	Stata, Latex, Microsoft Office.
Programming:	Gauss, Matlab, R.
Languages:	Greek (Native), English (Fluent), German (Intermediate), Dutch (Beginner).

SCHOLARSHIPS

2008-2009	George and Victoria Karelias Foundation Scholarship (20.000 EURO).
2005-2008	Bakalas Foundation Scholarship (36.000 EURO).