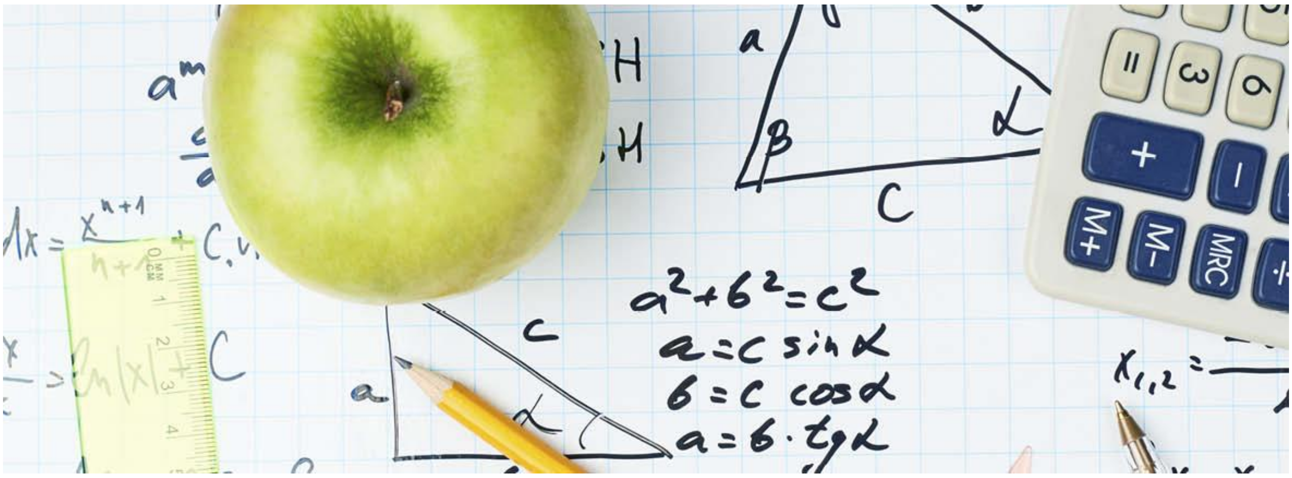
SDU

COHERE - Centre of Health Economics Research



Nonparametric Identification of a Time-Varying Frailty Model

By

Georgios Effaimidis

COHERE, Department of Business and Economics, University of Southern Denmark

COHERE discussion paper No. 6/2016

FURTHER INFORMATION
Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55,
DK-5230 Odense M Denmark
www.cohere.dk

Nonparametric Identification of a Time-Varying Frailty Model*

Georgios Effraimidis †

July 25, 2016

Abstract

In duration analysis, the Mixed Proportional Hazard model is the most common choice among practitioners for the specification of the underlying hazard rate. One major drawback of this model is that the value of the frailty term (i.e. unobserved factors) is time-invariant. This paper introduces a new model, the Mixed Random Hazard (MRH) model, which allows the frailty term to be time-varying. We provide sufficient conditions under which the new model is nonparametrically identified. Moreover, a theoretical framework is proposed for testing whether the true model is MRH. We conclude this paper with a discussion of how the arguments for the univariate MRH model can be extended to various multivariate problems.

Keywords: Competing Risks Model; Duration Analysis; Mixed Random Hazard; Time-varying frailty.

JEL Classification: C14, C31, C41

*I am extremely grateful to Geert Ridder for insightful comments and suggestions. I also thank Eileen Crimmins, Irene Botosaru, and Vanessa Smith as well as seminar participants at California Econometrics Conference (USC, 2015), Midwest Econometrics Group Conference (Federal Reserve Bank of Saint Louis, 2015), AHEW & AWEHE Workshop (University of Hawaii, 2015), Australasia Meeting of the Econometric Society (UTS, 2016), and an internal seminar at USC Center on Biodemography and Population Health. This research is based upon work supported by the Danish Council for Independent Research via the FP7 Marie Curie Actions COFUND (DFF: 1329-00074A). The warm hospitality of the USC Center on Biodemography and Population Health is gratefully acknowledged. Finally, I thank Lene Holbæk for promptly proofreading the manuscript. I bear full responsibility for the remaining errors.

†University of Southern Denmark, Department of Business and Economics. Email: georgiosnl@gmail.com.

1 Introduction

Duration analysis is emerging as a popular technique in diverse research fields such as economics, finance, medical science, marketing and political science (Van den Berg, 2001). It provides researchers with methods to analyze the distribution of the timing of an event of interest (e.g. transition from unemployment to employment, death, onset of a disease, corporate bankruptcy). Duration data are usually subject to random right censoring and/or left truncation. As a result, standard regression techniques cannot be applied to duration data. Duration models based on the underlying hazard rate are by far the most popular choice in empirical analysis for dealing with this type of incomplete information. In addition to that, the use of the hazard rate is flexible since it allows the inclusion of observed covariates that change over time- a pattern which is inherent to many problems due to their dynamic setting.

To be more concrete, let $T \in \mathcal{R}_+$ represent the duration of interest, $x \in \mathcal{X} \subset \mathcal{R}^d$ be a vector of observed covariates, and $V \in \mathcal{R}_+$ be a random variable which represents unobserved (to the researcher) heterogeneity and is commonly referred to as frailty. In the sequel, these terms will be used interchangeably. The hazard rate (also known as the transition rate) corresponding to the duration variable $T|x, V$ is defined as follows:

$$\theta(t|x, V) := \lim_{dt \to 0} (dt)^{-1} \mathbb{P}[t \leq T < t + dt | T \geq t, x, V]. \tag{1}$$

Heuristically speaking, the conditional on the individual observed as well as unobserved characteristics hazard rate $\theta(t|x, V)$ gives the speed at which the event of interest occurs at time t given that it has not occured at $t-$. A nice and appealing feature of hazard rate models is that there is one-to-one correspondence between the hazard rate and the underlying distribution function. In other words, as long as we have determined the hazard rate, we can compute the corresponding cumulative distribution function.

The Mixed Proportional Hazard (MPH) model (Van den Berg, 2001), which is the cornerstone of duration analysis, is overwhelmingly used for the specification of the hazard rate $\theta(t|x, V)$. In particular, introduce the positive-valued functions $\lambda : (0, \infty) \to (0, \infty)$ and $\phi : \mathcal{X} \to (0, \infty)$. The

MPH model has the following representation for all $t > 0$ and $x \in \mathcal{X}$:

$$\theta(t|x, V) = \lambda(t)\phi(x)V. \tag{2}$$

The function $\lambda(t)$ is called the baseline hazard and gives the change of the hazard rate over time and is identical across all units under consideration. The function $\phi(x)$ is called the regressor function and measures the effect of the observed covariates on the hazard rate. Elbers and Ridder (1982) show that the above model can be nonparametrically identified (up to scale normalizations) with the support of x being arbitrarily small (i.e. at least two support points) and statistically independent of V. That is, for an observed data distribution, the integral of λ, the function ϕ, and the distribution of V are uniquely pinned down. Moreover, the authors provide two different duration models that yield the same distribution of T in the absence of observed characteristics. The observational equivalence breaks in case the value of x takes on at least two values. That example essentially implies that observed heterogeneity is also a necessary condition for nonparametric identification of the MPH model. In another paper, Ridder and Woutersen (2003) derive an interesting result as they show that if the baseline hazard is bounded away from zero and infinity at zero, then the MPH model can be nonparametrically identified even in case the first moment of V is not finite. [1]

A common concern that arises when using the MPH model is that the stochastic variable V is time-invariant. [2] In other words, the value of the frailty term is determined at time zero and follows the unit under consideration until the occurence of the event of interest. However, in some cases it is somewhat problematic to claim that the value of the unobserved factors remains constant over time.

In the context of labor economics suppose that the transition rate from unemployement to

[1]In a related paper, Chiappori et al. (2015) discuss nonparametric identification and estimation of a nonlinear transformation model that includes the MPH model as a special case. In contrast to the paper by Elbers and Ridder (1982), they require at least one component of x to be (i) continuously distributed (with possibly bounded support) and (ii) stochastically independent of the unobserved terms conditional on the other (possibly endogenous) components of x.

[2]For notational simplicitly, we implicitly assume in the hazard rate equation (2) that the vector x is time-invariant. We should point out that the nonparametric identification result of Elbers and Ridder (1982) is true also when the components of x are stochastic processes with piecewise-constant trajectories.

employment is given by the MPH model, where V represents the (unobserved) motivation of the individual to find a job. One reason that the MPH model could be inappropriate for this case is the possible dependence of motivation on the elapsed duration in the unemployment state. In particular, we could expect that the pace of increase in the level of motivation will be higher for high ability individuals. Hence, we would need a more flexible specification that models the motivation as a function of time.

Another example where it is necessary to allow for time-varying frailty is the study of corporate defaults. It is well understood that the risk factors that affect the default probability are macroeconomic as well as firm-specific. Given that (i) the value of those factors will likely change over time and (ii) it is almost impossible to have information about all those factors, it would be useful to construct a duration model where the frailty can vary over time. Duffie et al. (2009), for instance, state that credit quality, which has a large impact on the probability of default, cannot be fully captured by credit rating (which is usually observed) and thus a time-varying frailty model is essential to capture this missing information.

Finally, a third motivating example for the use of duration models with time-varying frailty comes from mortality and epidemiological studies. In mortality studies, for instance, researchers often adopt a MPH specification for the individual age-dependent mortality rate, where V captures genetic as well as environmental influences. That is, it is implicitly assumed that those characteristics have the same numeric value over the entire lifespan. This assumption seems to be incorrect, as it is well known that the genome and environmental exposure change over the lifetime and consequently the frailty term should account for such changes over time.

In the past, a few papers have proposed duration models that allow the frailty term to change over time. Gjessing et al. (2003) and Botosaru (2013) consider a frailty model where the frailty is a weighted Lévy process with positive jumps (such a process is also known as a subordinator). Specifically, suppose that $Z = \{Z(t) : t \geq 0\}$ is a Lévy process with nonnegative jumps. The authors specify the hazard rate as follows:

$$\theta(t|Z) = \lambda(t)\phi(x) \int_0^t \alpha(u, t-u) dZ(W(u)). \tag{3}$$

The function α determines the effect of previous jumps of the process V on the hazard rate at time t, and W is a positive as well as nondecreasing function which transforms the time scale. Popular examples of such Lévy processes are the standard compound Poisson process, Gamma process, Stable process and Power Variance Function process.

In contrast to Gjessing et al. (2003), whose analysis is fully parametric, Botosaru (2013) focuses on semiparametric inference since she treats either the function λ or the function α as being completely unknown (but not both). [3] One criticism regarding model (3) is that the frailty can be only nondecreasing over time. Also, certain Lévy processes display piecewise constant paths, which implies that the hazard rate will be equal to zero on the corresponding time intervals. Another implication of the model is that all units have exactly the same value of frailty at time zero. In particular, $Z(0) = 0$, which implies that conditional on the observed heterogeneity all units are homogeneous at the beginning of the duration spells. Our belief is that this property does not make model (3) appropriate for mortality or epidemiological studies as it implicitly assumes that all individuals are identical at the moment they are conceived.

Some other studies (Enki et al., 2014; Paddy Farrington et al., 2012; Unkel et al., 2014) focus on shared frailty duration models with the frailty term to (continuously) vary over time.[4] There are two advantages of those models compared to (3). First, they allow the value of the frailty term to have an arbitrary time-dependence, that is, not only being a nondecreasing function of time. Second, at time $t = 0$ each unit has its own frailty value. On the other hand, those papers focus on fully parametric multivariate analysis where units that belong to the same cluster share the same (time-varying) unobserved characteristics.

To address the limitations of the above duration models, we construct a new univariate duration model and provide sufficient conditions under which the new model can be nonparametrically identified. Specifically, we build a new single hazard model that allows the presence of time-varying unobserved factors. The novelty of our approach lies in the development of a new model that generalizes the popular MPH model without imposing conditions on the behavior of the unobserved factors over time. For instance, the value of those unobserved factors can exhibit a

[3] Our conjecture is that her result could be extended probably in a fully nonparametric context by making use of a functional normalization as we do in our analysis below.

[4] Section 3 gives more details about the models developed in those papers.

decreasing pattern over time, which is not feasible in the hazard equation (3). The key findings of the paper are the following. First, sufficient conditions are provided for the nonparametric identification of the underlying model. In summary, by having at least one continuous covariate and fixing t we can recover first the baseline hazard by a functional normalization and subsequently the distribution of the unobserved terms for the particular value of t.

Second, we theoretically show that by following a two-step procedure it is possible under a set of (weak) conditions to infer whether or not the frailty term is time-varying. In the first step, the goal is to answer if the underlying hazard rate has the MPH specification. If the answer is negative, the second step tries to answer whether or not the non-MPH specification is due to the presence of time-varying characteristics.

Finally, we show how the arguments for the new univariate hazard model can be extended to various multivariate problems, including the competing risks model and a model with one treatment variable and multiple competing risks. One striking result of this paper is that the semiparametric test statistic for independece that has been developed by Van den Berg and Effraimidis (2015) for the competing risks model with time-invariant frailties is also applicable to our new competing risks model.

The rest of this paper is structured as follows. Section 2 introduces a new duration model, which allows the frailty term to change over time. Section 3 focuses on the identification of the model. We first list and discuss in detail the required assumptions and their implications from different perspectives. Subsequently, we provide the identification result under the set of those assumptions. Section 4 describes a theoretical framework for testing whether the duration data are from the new model we consider in this paper. Section 5 shows how the identification result for the univariate case can be extended, by appropriate modifications, to different multivariate problems. We are also concerned with the competing risks model, which is, unquestionably, the most complicated class of duration models. Section 6 demonstrates how the model of this paper can also serve as the building block in the evaluation of treatment effects when the outcome(s) as well as the treatment status are represented by duration variables and the unobserved factors change over time. Finally, section 7 concludes and gives directions for future research.

2 The Mixed Random Hazard Model

Consider the random variable U, which is standard uniformly distributed (i.e., U is uniformly distributed on [0,1]). Furthermore, introduce the map $r : (0, \infty) \times [0, 1] \to (0, \infty)$. To model the dependence of T on X and U, we specify the hazard rate of $T|x, U$ as follows:

$$\theta(t|x, U) = \lambda(t)\phi(x)r(t, U). \tag{4}$$

We will refer to the above model as the Mixed Random Hazard (MRH) model. The assumption that U follows a standard uniform distribution does not restrict the generality. In particular, any stochastic variable with a certain cumulative distribution function (CDF) P can be written as the inverse of that CDF, P^{-1}, which can be absorbed into the function r. To see this, assume that $\tilde{\theta}(t|x, V) = \lambda(t)\phi(x)\tilde{r}(t, V)$. By definition, $V = P^{-1}(U)$ and hence, $\tilde{\theta}(t|x, V) = \lambda(t)\phi(x)\tilde{r}(t, P^{-1}(U)) = \lambda(t)\phi(x)r(t, U) = \theta(t|x, U)$, where $r(t, U) := \tilde{r}(t, P^{-1}(U))$. Note that in case $r(t, U) = P^{-1}(U) = V$, where V is a random variable with nonnegative support, the above model reduces to the MPH model.

As in the MPH model, the positive functions $\lambda(t)$ and $\phi(x)$ have exactly the same role. The main difference between the MPH model and the MRH model is the time dependence of the unobserved heterogeneity. Particularly, in the MPH model it is constant over time from the beginning of the duration spell, while in the MRH specification it is modeled through the term $r(t, U)$, which implies that it generally varies over time. Moreover, as we will discuss later, the conditions required for the nonparametric identification do not impose any restriction on the effect of t on the function r. In other words, the frailty term has arbitrary time dependence, a property that does not hold true for model (3). More importantly, the formulation of the MRH model implies that at the beginning of the duration spells there is unobserved heterogeneity among individuals, which is not true for (3). *In summary, the MRH model has two key-properties that are not jointly present in other hazard rate models: Time-varying unobserved factors as well as the presence of those factors at the starting time of the underlying process.*

Note that we could also write the hazard rate as $\theta(t|x, U) = \phi(x)r(t, U)$, that is, have the function $\lambda(t)$ absorbed into $r(t, U)$. The reason why we include both $\lambda(t)$ and $r(t, U)$ in the model is that those two quantities have totally different interpretation. The function $\lambda(t)$ captures the evolution of the underlying hazard rate, whereas $r(t, U)$ refers to time-varying unobserved factors. In mortality studies, for example, $\lambda(t)$ describes the evolution of the mortality rate over the lifespan, which is common for all individuals. On the other hand, in the analysis of mortality data, the quantity $r(t, U)$ will capture individual-specific genetic and environmental factors that have an impact on the mortality rate and change over time. Later, we describe in more detail such an example related to mortality studies as well as an example from financial applications.

The new hazard rate model is somewhat related to the nice work of Brinch (2007), who breaks the conventional proportionality of the MPH model. Specifically, he shows that the inclusion of just one time-varying covariate with a single jump (multiple jumps are acceptable, too) allows the nonparametric identification of a hazard model where time and observed covariates do not neccessarily act multiplicatively on the hazard rate. Thus, the difference of our model from the model of Brinch (2007) is that we allow arbitrary interaction between time and unobserved risk factors, whereas he focuses on hazard models with arbitrary interaction between time and observed (time-varying) covariates.

Below, we present two examples as motivation for the construction of a hazard model, which will allow the presence of time-varying unobserved covariates.

Example 1 (Aging of Human Populations) *Life expectancy across the world has substantially increased as the individual mortality rate at every age has been considerably postponed due to economic and medical improvements (Vaupel, 2010). The shape of the individual mortality rate is determined by the aging rate which is defined as the growth rate of the individual mortality rate. Equivalently, the aging rate is defined as the derivative with respect to t of the logarithm of the mortality rate. The most popular specification for the mortality rate is the so-called Gompertz frailty model. The mortality rate at age t is given by*

$$\theta(t|x, U) = a \exp(bt)\phi(x)P^{-1}(U), \;\; a > 0, b > 0 \tag{5}$$

where x represents socioeconomic and demographic characteristics (e.g., gender and/or income *level) and $P^{-1}(U)$ refers to unobserved genetic and environmental influences.*

To capture the possible dependence on time of those unobserved factors, we can consider the hazard model (4) for some time-varying unobserved factors, $r(t, U)$. A major difference between models (4) and (5) in this context is that the aging rate in the former is b, whereas in the latter is $b + \partial_t \ln r(t, U)$, where ∂_t represents partial derivative with respect to t.[5] *Clearly, the second specification is more flexible, as variation (across individuals and over time) of genetic and environmental factors consistutes a strong reason to believe that the aging rate will be individual-specific and time-varying.*

A possible specification for r, useful for mortality studies, is the following: $r(t, U) = \frac{h(U)\exp(tP^{-1}(U))}{\exp(tP^{-1}(U))+1}$, where h is a positive real-valued function and P is a probability distribution with support on (a subset of) the real line. Therefore, an example of a mortality model which incorporates time-varying frailty can be written as follows:

$$\theta(t|x, U) = \exp(bt)\phi(x)\frac{h(U)\exp(tP^{-1}(U))}{\exp(tP^{-1}(U)) + 1}. \quad (6)$$[6]

Note that if $Var\left[P^{-1}(U)\right] = 0$, then $r(t, U)$ does not depend on t and thus we are back to the MPH model. The presence of the function $h(U)$ in the numerator has as consequence the existence of unobserved heterogeneity also at $t = 0$. Recall that this feature is not true for model (3). In the above example, the aging rate is time-varying; particularly, it is equal to $b+h(U)\partial_t \ln \frac{\exp(tP^{-1}(U))}{\exp(tP^{-1}(U))+1}$, where the second term is a positive function. In contrast to the Gompertz frailty model, which assumes a time-invariant and identical aging rate across all individuals, the mortality model (6) gives an individual-specific (unobserved) aging rate, which increases over time. In the appendix, we consider the MPH and MRH models for this case and we compare the two respective survival probability curves.

Another noteworthy point is that if we model the time-varying frailty as $r(t, U) = \frac{\exp(tP^{-1}(U))}{\exp(tP^{-1}(U))+1}$,

[5] For model (4) we get $\ln \theta(t|x, U) = bt + \ln \phi(x) + \ln r(t, U)$, while for model (5) we have $\ln \theta(t|x, U) = bt + \ln \phi(x) + \ln P^{-1}(U)$. By employing the definition of the aging rate, the corresponding values are obtained.

[6] Another candidate expression for the mortality rate is

$$\theta(t|x, U) = \exp(bt)\phi(x)\exp(P^{-1}(U)t).$$

In the next section, we discuss why such specification is not appropriate.

then all the individuals would be, ceteris paribus, exactly the same at time $t = 0$ as they are with model (3). More importantly, heterogeneity among survivors would converge to a degenerate distribution with single value equal to 1. In other words, this specification would be at the starting point of the process as well as in the limit equivalent to the standard Gompertz model without frailty. In view of the above discussion, it is pretty clear that the MRH model is quite flexible and allows for different time-dependence patterns of the unobserved heterogeneity.

Below, an example is presented to show the importance of time-varying frailty models in credit risk management.

Example 2 (Credit Risk Analysis) *Credit Value Adjustment (CVA) is regarded as the most appropriate quantity for measuring the counterparty credit risk. It is equal to the difference in the value of the portfolio before and after the adjustment for a possible default from the counterparty. Since the estimation of CVA requires knowledge of the density function of the default, we should model the underlying hazard rate (also known as stochastic intensity). Based on Hull and White (2012), we can use (4) for the hazard rate, where x is a vector of (possibly time-varying) covariates, such as (in case of corporate defaults) distance to default or 3-month Treasury bill rate and $r(t, U)$ is a positive-valued time-varying unobserved covariate.*

We should point out that our time-varying frailty model is an alternative way of modeling the hazard rate of default. Existing credit models specify the hazard rate as an affine function of a multidimensional Markov process. On the other hand, the specification $r(t, U)$ implies that the unobserved heterogeneity is not a stochastic process since U does not depend on t. In other words, conditional on the value of the unobserved characteristics at some time t, the evolution of those characteristics is nonstochastic along any time trajectory.

3 Identification

In this section, we turn our attention to the identification of the MRH model. We first state and discuss in detail the conditions that are needed to identify our model. Next, we state the main identification result of this paper.

The following four assumptions are required for the establishment of the main result.

Assumption 1 *The function* $\phi : \mathcal{X} \to (0, \infty)$ *is such that it attains all values in an open connected subset* $(0, \infty)$. *Additionally,* $\phi(x^\star) = 1$ *for some a priori chosen* $x^\star \in \mathcal{X}$.

Assumption 2 *The random variable* U *is independent of* x. *The limit* $\lim_{t\to 0} r(t, U)$ *exists and there is an integrable function* $Z : [0, 1] \to (0, \infty)$ *such that* $r(t, U) \le Z(U)$ *for every* $t > 0$ *and* $U \in [0, 1]$. *Also,* $0 < \int_0^1 \lim_{t\to 0} r(t, u)du < \infty$.

Assumption 3 *The map* $U \mapsto \int_0^t \lambda(\omega) r(\omega, U)d\omega$ *is strictly monotonic for all* $t > 0$.

Assumption 4 *For some a priori chosen* $u^\star$, *it holds that* $r(t, u^\star) = 1$.

Assumption 1 is stronger than the corresponding assumption in Elbers and Ridder (1982) as it requires x to have at least one element that takes on values in an interval. Note that no large support conditions are imposed on the support of that element. In other words, the support of that element can be any arbitrarily small open subset of the real line. The commonly used normalization $\phi(x^\star) = 1$ is innocuous, as we can always divide $\phi(x)$ by $\phi(x^\star)$ and multiply $\lambda(t)$ or $r(t, U)$ by $\phi(x^\star)$.

Assumption 2 is needed for the identification of ϕ and reduces to the standard finite mean assumption for the MPH model. We should point out that this assumption rules out, for instance, the following specification for the hazard rate: $\theta(t|x, U) = a \exp(bt)\phi(x) \exp(P^{-1}(U)t)$, where P^{-1} is the inverse CDF of a distribution (absolutely continuous with respect to the Lebesgue measure) with nonnegative support. This is due to the fact that the quantity $\exp(P^{-1}(U)t)$ is not bounded by an integrable function in U. Such specification could potentially be useful in mortality studies, where the quantity $\exp(P^{-1}(U)t)$ would capture either (i) time-varying unobserved characteristics or (ii) a mortality process such that the aging rate is an individual-specific random quantity. [7]

Assumption 3 is essential for the inversion of a certain map, and therefore identification of the model is not possible without its presence. Note that this assumption in the context of the MPH model would imply that the random variable U is absolutely continuous with respect to the

[7] The aging rate associated with that mortality rate either in (i) or (ii) is equal to $b + P^{-1}(U)$.

Lebesgue measure. That is, assumption 3 rules out distributions for the unobserved heterogeneity whose support have a finite number of points. However, Elbers and Ridder (1982) do not make use of such a condition in order to nonparametrically identify the MPH model. This further minor restriction, used for the identification of our model, is needed because of the nonseperability between t and U in the function r. Another implication of this assumption is that the function $r(t, U)$ cannot be used to model an unobserved random shock that occurs at some unknown point in time and affects the hazard rate from that point onwards. For instance, such random shock could be incorporated into the hazard rate by specifying the function $r(t, U)$ as follows: $r(t, U) = 1\{P^{-1}(U) > t\}$. Now, as we can see, the function $r(t, .)$ is nondecreasing for every t but not strictly increasing as the assumption 3 requires. In this example, $r(t, .)$ is a piecewise function with a single jump at t. Hence, jump discontinuities in the hazard rate are allowed only due to discontinuities of the baseline hazard. [8]

Finally, a few comments are in order regarding assumption 4. Note that this assumption is essentially a functional normalization, as we can always multiply $\lambda(t)$ and divide $r(t, U)$ by $r(t, u^\star)$. This assumption is closely related to two different parametric shared time varying frailty models that have been developed in previous studies. Paddy Farrington et al. (2012) construct their model as follows: $r(t, U) = 1 + h_1(U)h_2(t)$. On the other hand, Enki et al. (2014) suggest a model in which $r(t, U) = \exp(h_1(U)h_2(t))$. The functions h_1 and h_2 are known up to a finite dimensional parameter vector. In the first paper $h_1(U) = P^{-1}(U) - 1$, while in the second paper $h_1(U) = \log(P^{-1}(U))$, where $\mathbb{E}[P^{-1}(U))] = 1$ and P^{-1} is parametrically specified. For both of these two models, the function $r(t, U)$ is equal to 1 for $U = P(1)$. Our slightly different normalization stems from the fact that we follow a nonparametric approach. As a result, the CDF P is part of the function λ, which does not allow us to write down the normalization by explicitly using P. A similar functional normalization in a different duration analysis context is adopted by Evdokimov (2011) who nonparametrically identifies bivariate hazard rates where the

[8]Such discontinuities can also arise due to the presence of time-dependent covariates that jump either up or down. Generally speaking, the results of this paper are also true when we explicitly include such time-varying covariates in our model.

baseline hazard as well as the unobserved factors depend on observed covariates. [9]

It is worth noting that a special case of the MRH model can be viewed as a hazard rate with random individual-specific baseline hazard without unobserved heterogeneity. In particular, by letting $\lambda(t) = 1$ for every $t > 0$ we have $\theta(t|x, U) = \phi(x)r(t, U)$, where the baseline hazard rate is now represented by the quantity $r(t, U)$. That is, in this case the function $r(t, U)$ captures the evolution of the underlying hazard rate over time rather than (time-varying) unobserved heterogeneity. In mortality studies, we could have $\theta(t|x, U) = \exp(P^{-1}(U)t)\phi(x)$, where P^{-1} is the inverse of a probability measure with nonnegative support such as lognormal or gamma. For this specification of the mortality rate, the (unobserved) individual aging rate would be equal to $P^{-1}(U)$. If the support of the probability distribution of a random variable is concentrated on a single point, that is, P is a degenerate distribution, then this model reduces to the standard Gompertz model (without unobserved heterogeneity). Another application of a hazard rate with random baseline hazard can be found in labor economics. For instance, let the transition rate from unemployement to employment be expressed as $\theta(t|x, U) = t^{P^{-1}(U)}\phi(x)$, where again P^{-1} is the inverse of a probability measure with positive support. If P is a degenerate distribution, then that transition rate corresponds to a standard Weibull model.[10] The problem with the two above (parametric regarding the function $r(t, U)$) examples is that the function $r(t, U)$ does not satisfy assumption 2. In particular, the map $t \mapsto \exp(P^{-1}(U)t)$ cannot be bounded by an integrable function of U and also $\int_0^1 t^{P^{-1}(U)}du = \infty$ for $P^{-1}(U) > 0$. Thus our preference is to stick to the interpretation of the MRH model as a hazard model with deterministic baseline hazard and time-varying unobserved factors such that our work is closely related to previous studies (Enki et al., 2014; Paddy Farrington et al., 2012; Unkel et al., 2014)

Define $\Lambda(t) := \int_0^t \lambda(\omega)d\omega$ and $R(t, U) := \int_0^t r(\omega, U)d\omega$. We now state the first main result of this paper.

[9] Evdokimov (2011) studies the following bivariate shared-frailty duration model:

$$\theta_j(t|x, U) = \kappa_j(t, x)h(x, P^{-1}(U))$$

for some nonnegative functions κ and h, $j = A, B$. To nonparametrically identify his model, he assumes that $\kappa_j(1, x) = 1$.

[10] Analogously, we could also model the transition rate from employment to unemployment as $\theta(t|x, U) = t^{P^{-1}(U)}\phi(x)$, with P^{-1} the inverse of a probability measure with support on the interval [-1,0], which is a generalization of the Weibull model.

Proposition 1 *Let assumptions 1 − 4 hold. Then, the functions ϕ, Λ, and R are identified.*

Proof.

The survival function $\mathbb{P}[T > t|x]$ is given by the following mixture representation for all $t > 0$ and $x \in \mathcal{X}$:

$$\mathbb{P}[T > t|x] = \int_0^1 \exp\left(-\phi(x)\int_0^t \lambda(\omega)r(\omega, u)d\omega\right) du. \tag{7}$$

The corresponding density $f(t|x)$ is expressed as follows:

$$f(t|x) = \phi(x)\lambda(t) \int_0^1 r(t, u) \exp\left(-\phi(x)\int_0^t \lambda(\omega)r(\omega, u)d\omega\right) du. \tag{8}$$

By assumption 1, we have $\phi(x^\star) = 1$. Hence,

$$\frac{f(t|x)}{f(t|x^\star)} = \frac{\phi(x) \int_0^1 r(t, u) \exp\left(-\phi(x)\int_0^t \lambda(\omega)r(\omega, u)d\omega\right) du}{\int_0^1 r(t, u) \exp\left(-\int_0^t \lambda(\omega)r(\omega, u)d\omega\right) du}. \tag{9}$$

In view of assumption 2, we can conclude by applying the dominated convergence theorem for any $x \in \mathcal{X}$ that

$$\lim_{t\to 0} \int_0^1 r(t, u) \exp\left(-\phi(x)\int_0^t \lambda(\omega)r(\omega, u)d\omega\right) du = \int_0^1 \lim_{t\to 0} r(t, u)du, \tag{10}$$

where the quantity on the right-hand side is strictly positive and finite. Therefore, for any $x \in \mathcal{X}$

$$\lim_{t\to 0} \frac{f(t|x)}{f(t|x^\star)} = \phi(x). \tag{11}$$

Define as G_t the CDF of the random variable $\int_0^t \lambda(\omega)r(\omega, U)d\omega$ for fixed t. Let $\mathcal{L}_t$ be the Laplace transform of G_t. We can write for each $t > 0$ and $x \in \mathcal{X}$

$$S(t|x) = \int_0^\infty \exp(-u\phi(x))dG_t(u).$$

Hence, $S(t|x) = \mathcal{L}_t(\phi(x))$. By assumption 1, we can vary ϕ in an open interval, and thus we can trace out the Laplace Transform (LT) $\mathcal{L}_t$, as it is real analytic. Consequently, we can identify the CDF G_t for any $t > 0$.

Now, denote by $\tau(t, .)$ the inverse map of $U \mapsto \int_0^t \lambda(\omega) r(\omega, U) d\omega$. Note that the inverse exists due to assumption 3. One can write for each $t > 0$ and $u > 0$

$$\mathbb{P}\left[\int_0^t \lambda(\omega) r(\omega, U) d\omega \leq u\right] = G_t(u).$$

The above formula implies (assuming without loss of generality that $U \mapsto \int_0^t \lambda(\omega) r(\omega, U) d\omega$ is strictly increasing)

$$\mathbb{P}\left[U \leq \tau(t, u))\right] = G_t(u).$$

By using the fact that U follows a standard uniform distribution we obtain for each t

$$\tau(t, u) = G_t(u).$$

The right- hand side is known which yields identification of $\tau(t, .)$ for any t and consequently of $\int_0^t \lambda(\omega) r(\omega, U) d\omega$ for every t.

Therefore, we can trace out the product $\lambda(t) r(t, U)$ for almost any $t > 0$ and all $U \in [0, 1]$, which results in identification of the product $\lambda(t) r(t, u^\star)$ for almost all $t > 0$. By assumption 4, we know that $r(t, u^\star) = 1$. Consequently, we can identify the quantity $\lambda(t)$ for almost every $t > 0$, which in turn leads to the identification of Λ on $[0, \infty)$. Recall that for almost every $t > 0$ and all $U \in [0, 1]$ the value of $\lambda(t) r(t, U)$ is known. Since we have traced out $\lambda(t)$ for almost every $t > 0$, the identification of $r(t, U)$ for almost every $t > 0$ and all $U \in [0, 1]$ is straightforward. Hence, R is identified on $[0, \infty) \times [0, 1]$. ■

Proposition 1 establishes the identification of the MRH model. However, as in the case of most of univariate hazard rate models, the proof is not very contructive in the sense that we cannot directly use the results for the estimation of most of the underlying quantities. Only the regressor function ϕ can be estimated "directly" by using the observed hazard rates. Botosaru

(2013) discusses semiparametric estimation of (3) and suggests to use second derivatives of the corresponding survival functions (i.e., $S(t|x)$ and $S(t|x^\star)$) to estimate the regressor function. Similar idea could be applied to the MRH model. However, as we will discuss here, a different estimation method for the regressor function can be employed without requiring differentiation of nonparametrically estimated quantities. Generally speaking, it is good to avoid differentiation of estimators since it slows down the convergence rate of the underlying estimators.

To fix ideas, suppose that the duration data are randomly right censored and/or left truncated. Recall that

$$\lim_{t\to 0}\frac{f(t|x)}{f(t|x^\star)} = \phi(x).$$

Obviously, the densities $f(t|x)$ and $f(t|x^\star)$ cannot be estimated by using standard kernel methods due to the incomplete nature of the data. Botosaru (2013) suggests an estimator for the regressor function that is based on the following propery:

$$\lim_{t\to 0}\frac{\partial_{tt}\ln S(t|x)}{\partial_{tt}\ln S(t|x^\star)} = \phi(x),$$

where ∂_{tt} denotes the second derivative with respect to t. The estimator of ϕ is simply obtained by using appropriate estimators for the quantities on the left-hand side. As already mentioned, that estimator is somewhat problematic due to the need to estimate second derivatives by making use of kernel-based smoothing techniques. Furthermore, that estimator implicitly assumes that there is no random right censoring and/or left truncation provided that the popular (conditional) Kaplan-Meier estimator, which deals with incomplete duration data, cannot be implemented in this context due to its discontinuity, and thus nondifferentiability, with respect to t.

To overcome the aforementioned problem, we can use the hazard rate $q(t|x)$ of the duration variable $T|x$, which is nonparametrically identified and easily estimated from incomplete data. In mathematical notation,

$$q(t|x) := \lim_{dt\to 0}(dt)^{-1}\,\mathbb{P}[t \le T < t+dt|T \ge t, x].$$[11]

[11]Denote by C the censoring variable with $T \perp\!\!\!\perp C \mid x$, where the symbol $\perp\!\!\!\perp$ implies independence between the

Let $Q(t|x) := \int_0^t q(\omega|x)d\omega$ be the corresponding integrated hazard rate. Hence, it follows that $f(t|x) = q(t|x)\exp(-Q(t|x))$ and $S(t|x) = \exp(-Q(t|x))$. Consequently, for all $t > 0$ and $x \in \mathcal{X}$

$$\lim_{t\to 0}\frac{f(t|x)}{f(t|x^\star)} = \lim_{t\to 0}\frac{q(t|x)}{q(t|x^\star)}. \tag{12}$$

Consistent estimation of $q(t|x)$ can be performed by using the nonparametric kernel methods of Nielsen and Linton (1995). In contrast to Botosaru (2013), the proposed method in this paper succesfully deals with incomplete duration data and does not require estimation of any derivative.

Finally, in a similar spirit to Chiappori et al. (2015), the result of proposition 1 can be also extended to the following model

$$\theta(t|x,y) = \lambda(t)\phi(x,y)r(U),$$

where $y \in \mathcal{Y}$ is allowed to be dependent on U and the vectors U and $x \in \mathcal{X}$ are stochastically independent given y. For tracing out the function ϕ, the identification strategy parallels the steps of the proof of proposition 1. Next, by following similar steps, we can uniquely determine the functions Λ and R. More precisely, the following equality between the CDFs holds: $\mathbb{P}\left[\int_0^t \lambda(\omega)r(\omega,U)d\omega \leq u|x,y\right] = \mathbb{P}\left[\int_0^t \lambda(\omega)r(\omega,U)d\omega \leq u|y\right]$ for any $t > 0, x \in \mathcal{X}, y \in \mathcal{Y}$. The latter is true due to the conditional (on y) independence between U and x. Hence, the underlying LTs will be equal, namely, $\mathcal{L}_{t|y,x}(.) = \mathcal{L}_{t|y}(.)$. By changing the value of x we can trace out the conditional LT $\mathcal{L}_{t|y}$, which in turn allows us to uniquely detrmine the CDF $\mathbb{P}\left[\int_0^t \lambda(\omega)r(\omega,U)d\omega \leq u|y\right]$. Knowledge of that conditional CDF allows us to compute the unconditional CDF $\mathbb{P}\left[\int_0^t \lambda(\omega)r(\omega,U)d\omega \leq u\right]$ provided that we can integrate out y with respect to its density, which is observed from the data. Then, the identification of Λ and R is straightforward.

.

We choose to pursue our study with model (4), as this is a generalization of model (2), which

underlying random variables. By the independence condition,

$$q(t|x) = \lim_{dt\to 0}(dt)^{-1}\mathbb{P}[t \leq T < t+dt, T < C|T \geq t, C \geq t, x].$$

This example is concerned only with random right censoring. Similar arguments can be applied to the case with random left truncation.

is usually considered and imposes the condition that explanatory covariates are not dependent on the unobserved factors.

4 Nonparametric Hypothesis Testing

In this section, we provide a formal discussion regarding a two-step nonparametric procedure, which can be used to test for the MRH model. We first show how we can nonparametrically test whether the underlying hazard rate is MPH. In case we reject the MPH model, we can proceed with testing whether the underlying model is MRH. The first step is closely related to the work of Chiappori et al. (2015) and Lewbel et al. (2015), who consider nonparametric testing of additivity between the regressor function and the error term in nonlinear transformation models. Our procedure is more flexible since it allows the presence of random right censoring and left truncation. [12]

We will assume that x consists of two continuously distributed components, that is, $x = (x_1, x_2)$. We should mention that x_2 is needed only for the second step. That is, the first step is valid also in case x is a single continuous variable. Throughout this section, we will use the notation $\theta(t|x, U)$ without imposing any restriction on the type of interaction among t, x, and U.

4.1 Testing for the MPH model

Our first goal is to examine whether the underlying hazard rate has the MPH specification. We denote by $\mathcal{M}_{ALL}$ the set of all admissible structures corresponding to all possible models that can be employed to model the hazard rate $\theta(t|x, U)$. Furthermore, we consider the family $\mathcal{M}_{MPH}$, which includes all the admissible structures of the MPH model. Specifically, we have

$$\mathcal{M}_{MPH} := \{\text{all triplets } (\lambda, \phi, P^{-1}) : \theta(t|x, U) = \lambda(t)\phi(x)P^{-1}(U)\}. \tag{13}$$

[12] The derivation of the asymptotics is ongoing work.

We are interested in analyzing the problem of testing the hypothesis

$$\begin{aligned}
&\mathbb{H}_0 : \theta(t|x, U) \in \mathcal{M}_{MPH} \\
&\mathbb{H}_1 : \theta(t|x, U) \in \mathcal{M}_{ALL} \backslash \{\mathcal{M}_{MPH}\}.
\end{aligned} \tag{14}$$

Let

$$\rho_i(t, x) := \frac{f(t|x)}{\partial_{x_i} S(t|x)}, \tag{15}$$

where f and S denote the probability density function and survival function, respectively, and ∂_{x_i} signifies the partial derivative with respect to x_i ($i = 1, 2$). It is not difficult to see that for the MPH model, which is described by the family (13) it holds for all $t > 0$ and $x \in \mathcal{X}$ that

$$\rho_i(t, x) = -\frac{\phi(x)\lambda(t)}{\partial_{x_i}\phi(x)\Lambda(t)}. \tag{16}$$

Looking at the above equation, we note that the (nonparametrically identified) ratio $\rho_i(t, x)$ is multiplicatively seperable in t and x.

It can easily be shown that the ratio $\rho_i(t, x)$ does not generally possess that multiplicativity property in case the hazard rate is not MPH. For instance, in the presence of time-varying frailty we get for the MRH model

$$\rho_i(t, x) = -\frac{\phi(x)\lambda(t) \int_0^1 r(t, u) \exp\left(-\phi(x) \int_0^t \lambda(\omega) r(\omega, u) d\omega\right) du}{\partial_{x_i}\phi(x) \int_0^1 \left[\int_0^t \lambda(\omega) r(\omega, u) d\omega\right] \exp\left(-\phi(x) \int_0^t \lambda(\omega) r(\omega, u) d\omega\right) du}.$$

Of course, nonproportionality does not only result due to time-varying frailty. Another example is when the unobserved factors are constant over time but the interaction between t and x is not proportional. Specifically, if $\theta(t|x, U) = \kappa(t, x) P^{-1}(U)$ we have

$$\rho_i(t, x) = -\frac{\kappa(t, x) \int_0^1 P^{-1}(u) \exp\left(-\int_0^t \kappa(\omega, x) d\omega P^{-1}(u)\right) du}{\partial_{x_i} \int_0^t \kappa(\omega, x) d\omega \int_0^1 P^{-1}(u) \exp\left(-\int_0^t \kappa(\omega, x) d\omega P^{-1}(u)\right) du}.$$

Consequently, the hypothesis testing problem (14) is equivalent to

$$\begin{aligned} &\mathbb{H}_0 : \tfrac{\rho_i(t,x)\rho_i(\tilde{t},y)}{\rho_i(t,y)\rho_i(\tilde{t},x)} = 1 \text{ for almost every } x, y \in \mathcal{X} \text{ and } t, \tilde{t} > 0 \\ &\mathbb{H}_1 : \mathbb{H}_0 \text{ is false,} \end{aligned} \tag{17}$$

where the null hypothesis obviously corresponds to the MPH model. [13] Note that $\mathbb{H}_0$ is true for $i = 1$ if and only if $\mathbb{H}_0$ is true for $i = 2$ and thus, as we mentioned at the beginning of this section, the first step can be done by using variation only in x_1 (or x_2). Chiappori et al. (2015) and Lewbel et al. (2015) develop a nonparametric test for additivity in nonlinear transformation models. However, their asymptotic theory is applicable only to cases in which the duration variable is always observed.

In case the duration variable is randomly right censored and/or left truncated, the quantity $\rho_i(t, x)$ can be nonparametrically estimated by using the following property: $\rho_i(t, x) = -\frac{q(t|x)}{\partial_{x_i} Q(t|x)}$ Recall that $f(t|x) = q(t|x) \exp(-Q(t|x))$, $S(t|x) = \exp(-Q(t|x))$, and then it is easy to show that $\rho_i(t, x) = -\frac{q(t|x)}{\partial_{x_i} Q(t|x)}$. Linton et al. (2011) suggest a consistent and asymptotically normal estimator of $Q(t|x)$.

4.2 Testing for the MRH Model

As explained in the previous subsection, the negation of $\mathbb{H}_0$ corresponding to problem (14) does not necessarily imply that the hazard rate is MRH. Hence, once we reject $\mathbb{H}_0$ we should try further to investigate whether this is due to the presence of a time-varying frailty. However, without further assumptions it is not possible to design a hypothesis-testing problem where we will be able to make the distinction between the MRH model and all the other models with admissible strucures not included in $\mathcal{M}_{MPH}$.

[13] Alternatively,

$$\begin{aligned} &\mathbb{H}_0 : \partial_{tx} \ln \rho_i(t, x) = 0 \text{ for almost every } x \in \mathcal{X} \text{ and } t > 0 \\ &\mathbb{H}_1 : \mathbb{H}_0 \text{ is false,} \end{aligned}$$

where ∂_{tx} denotes the cross-derivative with respect to t and x.

We now define the following two families:

$$\mathcal{M}_{MRH} := \{ \text{ all triplets } (\lambda, \phi, r) : \theta(t|x, U) = \lambda(t)\phi(x)r(t, U)\}, \tag{18}$$

and

$$\mathcal{M}_{IN} := \{\text{all structures of hazard rate models} : \theta(t|x, U) = \theta(t|\phi(x), U).\} \tag{19}$$

Clearly, $\mathcal{M}_{MRH}$ gives the admissible structures of the MRH model. $\mathcal{M}_{IN}$ gives the admissible structures of all models in which x affects the hazard rate through the function ϕ without imposing any restriction on the interaction among t, $\phi(x)$ and U. As we will show later, the family $\mathcal{M}_{IN}$ rules out admissible structures of models that share a common feature with the MRH model.

We are now going to formulate the hypothesis-testing problem as follows:

$$\begin{aligned} &\mathbb{H}_0 : \theta(t|x, U) \in \mathcal{M}_{MRH} \\ &\mathbb{H}_1 : \theta(t|x, U) \in \mathcal{M}_{ALL} \backslash \{\mathcal{M}_{MPH} \cup \mathcal{M}_{MRH} \cup \mathcal{M}_{IN}\}. \end{aligned} \tag{20}$$

To continue, we also introduce the ratio

$$\pi(t, x) := \frac{\partial_{x_1} S(t|x)}{\partial_{x_2} S(t|x)}. \tag{21}$$

We can readily check that under the MRH model, we have for all $t > 0$ and $x \in \mathcal{X}$

$$\pi(t, x) = \frac{\partial_{x_1} \phi(x)}{\partial_{x_2} \phi(x)}. \tag{22}$$

In view of the above equation, we can reformulate the hypothesis testing problem (20) as

$$\begin{aligned} &\mathbb{H}_0 : \tfrac{\pi(t,x)}{\pi(\tilde{t},x)} = 1 \text{ for almost every } x \in \mathcal{X} \text{ and } t, \tilde{t} > 0 \\ &\mathbb{H}_1 : \mathbb{H}_0 \text{ is false.} \end{aligned} \tag{23}$$ [14]

Acceptance of the null hypothesis is equivalent to stating that the hazard rate has the MRH specification. Note that in case $\theta(t|x, U) \in \mathcal{M}_{IN}$, the ratio $\pi(t, x)$ does not depend on t, and this is the reason for not including those specifications in the alternative hypothesis. In particular, if $\theta(t|x,U) \in \mathcal{M}_{IN}$, it is straighforward to derive that $\pi(t,x) = \frac{\partial_{x_1}\phi(x)}{\partial_{x_2}\phi(x)}$, and therefore $\pi(t,x) = 1$ for every $t > 0$ and $x \in \mathcal{X}$. Examples of hazard rates which belong to $\mathcal{M}_{IN}$ are (assuming that $x = (x_1, x_2)$): $\theta(t|x,U) = \exp(tx_1)\exp(tx_2)U$ and $\theta(t|x,U) = \lambda(t)\exp(Ux_1)\exp(Ux_2)$.

5 Multivariate Duration Models

Our identification methodology in section 3 can be appropriately modified for the identification of different types of multivariate duration models under the presence of time-varying unobserved factors. For the sake of convenience, we will analyze only bivariate duration models where the two underlying duration variables are denoted by T_A and T_B. The analysis with more than two duration variables is identical and thus we omit it. Our focus is on (i) the bivariate duration model with parallel durations and (ii) the competing risks model.

We now introduce the maps $\lambda_j : (0,\infty) \to (0,\infty)$, $\phi_j : \mathcal{X} \to (0,\infty)$ and $r_j : (0,\infty) \times [0,1] \to (0,\infty)$ for $j = A, B$. The respective hazard rates of $T_A|x, U_A$ and $T_B|x, U_B$ are expressed as follows:

$$\begin{aligned} \theta_A(t|x,U_A) &= \lambda_A(t)\phi_A(x)r_A(t,U_A), \\ \theta_B(t|x,U_B) &= \lambda_B(t)\phi_B(x)r_B(t,U_B), \end{aligned} \tag{24}$$

where U_A and U_B follow the standard uniform distribution.

Throughout this section we will make use of assumptions $1 - 4$ for each risk. We only need to slightly strengthen the first assumption. In particular, we also require that the vector of the observed covariates x contains at least two continuously distributed elements such that the vector-valued function $(\phi_A(x), \phi_B(x))$ takes on values in an open subset of $\mathcal{R}_+ \times \mathcal{R}_+$. A necessary requirement for the latter to be true is that x contains at least two continuously distributed covariates, x_1 and x_2. The other three assumptions are the same and should hold for each hazard rate equation.

Let us introduce some extra notation employed in the next two subsections. Define $\Upsilon_j(t) :=$

$\int_0^t \lambda_j(\omega) r_j(\omega, U_j) d\omega$, $j = A, B$. For fixed t, $\Upsilon_j(t)$ is a random variable. The notation $\mathcal{L}_{t_A t_B}$ is employed to denote for fixed t_A, t_B the LT of the distribution of the stochastic vector $(\Upsilon_A(t_A), \Upsilon_B(t_B))$. We will also use the indicator functions $D_A := 1\{T_A < T_B\}$ and $D_B := 1\{T_B < T_A\}$ and the subdensity functions $f_j(t|x) := \lim_{dt \to 0} (dt)^{-1} \mathbb{P}[t \leq \min(T_A, T_B) < t + dt, D_j = 1|x]$.

5.1 Parallel Durations

We first focus on the case where the duration variables are parallel (e.g. age at onset of a disease in siblings. retirement age of couples). The proof for the identifcation of Λ_j, ϕ_j, R_j closely follows the steps of the proof of Proposition 1. It remains to recover the joint CDF of (U_A, U_B). Pick some $t_A > 0, t_B > 0$. For any $u_A > 0, u_B > 0$

$$\mathbb{P}[U_A \leq u_A, U_B < u_B] = \mathbb{P}\left[\Upsilon_A(t_A) \leq \int_0^{t_A} \lambda_A(\omega) r_A(\omega, u_A) d\omega, \Upsilon_B(t_B) \leq \int_0^{t_B} \lambda_B(\omega) r_B(\omega, u_B) d\omega\right] \tag{25}$$

provided that the functions $r_j(\omega, .)$ are strictly increasing. The identification of Λ_j, R_j implies that the quantity $\int_0^t \lambda_j(\omega) r_j(\omega, u) d\omega$ can be traced out for any $t > 0$ and $v > 0$. Thus, to identify the joint CDF of (U_A, U_B), we just need to look at the CDF of $(\Upsilon_A(t_A), \Upsilon_B(t_B))$ for some fixed t_A, t_B. Clearly,

$$\begin{aligned}\mathbb{P}[T_A > t_A, T_B > t_B|x] &= \int_0^1 \int_0^1 \prod_{j=A,B} \exp\left(-\phi_j(x) \int_0^{t_j} \lambda_j(\omega) r_j(\omega, u_j) d\omega\right) dG(u_A, u_B) \\ &= \mathcal{L}_{t_A t_B}\left(\phi_A(x), \phi_B(x)\right). \end{aligned} \tag{26}$$

Fixing t_A, t_B and varying ϕ_A and ϕ_B, we can uniquely identify $\mathcal{L}_{t_A t_B}$ and subsequently recover the joint CDF of $(\Upsilon_A(t_A), \Upsilon_B(t_B))$. The latter gives, by virtue of (25), the identification of the CDF of (U_A, U_B). In fact, the variation of $(\phi_A(x), \phi_B(x))$ in an open subset of the positive real plane is needed only to recover the joint CDF of (U_A, U_B) and not for the other functions of the model. Consequently, we obtain the following statement:

Proposition 2 *Let assumptions 1 − 4 hold for each risk. Furthermore, the vector* $(\phi_A(x), \phi_B(x))$

takes on values in an open subset of $(0,\infty)^2$*. Then the functions* $\phi_A, \phi_B, \Lambda_A, \Lambda_B$*,* R_A, R_B*, and the joint CDF of* (U_A, U_B) *for the bivariate duration model (24) with parallel durations are identified.*

Note that even if $U_A = U_B$ with probability one and $r_A(t,U) = r_B(t,U)$ for all $t > 0$, we still have to consider seperately each hazard equation, which implies that the presence of x is necessary for this case, too.

Honoré (1993) identifies the above bivariate duration model for $r_A(t,U_A) = r_B(t,U_B) = P^{-1}(U)$ without using variation in x (i.e. $\phi_A(x) = \phi_B(x) = 1$ for any $x \in \mathcal{X}$). To see this, we will make use of the subdensity functions $f_A(t)$ and $f_B(t)$ (we do not condition on x since for this particular case it does not affect the hazard rates). Honoré (1993) exploits the fact that $\frac{f_A(t_A)}{f_B(t_B)} = \frac{\lambda_A(t_A)}{\lambda_B(t_B)}$ for almost every $(t_A,t_B) \in (0,\infty)^2$. By integrating out t_B, the identification (up to scale normalization) of Λ_A is straightforward, and subsequently of Λ_B. On the contrary, our case is more complicated since the unobserved heterogeneity depends explicitly on time, and hence it can be easily shown that $\frac{f_A(t_A)}{f_B(t_B)} = \frac{\lambda_A(t_A)}{\lambda_B(t_B)}$ only for $t_A = t_B$, which in turn does not allow us to perform marginal integration. [15] [16]

5.2 The Competing Risks Model

In this subsection, we show how we can nonparametrically identify a competing risks model with time-varying frailties and also test for (i) MPH structure and (ii) dependence between the risks. In this model, one is assumed to observe only the minimum between T_A and T_B, where the corresponding risks are mutually exclusive (e.g., death due to cancer vs. death due to cardiovascular disease, full-time employment vs part-time employment, and bankruptcy vs acquisition of a corporate).

[15] Let $U_A = U_B = U$ with probability one. Also, $r_A(t,U) = r_B(t,U) = r(t,U)$ for any $t > 0$ and $U \in [0,1]$. We have

$$\frac{f_A(t_A)}{f_B(t_B)} = \frac{\lambda_A(t_A)\int_0^1 r(t_A,u)\prod_{j=A,B}\exp\left(-\int_0^{t_j}\lambda_j(\omega)r(\omega,u)d\omega\right)du}{\lambda_B(t_B)\int_0^1 r(t_B,u)\prod_{j=A,B}\exp\left(-\int_0^{t_j}\lambda_j(\omega)r(\omega,u)d\omega\right)du}.$$

[16] On a related note, the time-varying nature of the unobserved heterogeneity does not allow us to infer whether the frailties are the same for the two underlying units. On the other hand, for the model of Honoré (1993), we can test for the presence of identical time-varying frailty by studying if the ratio $\frac{f_A(t_A)}{f_B(t_B)}$ is proportional in t_A and t_B.

5.2.1 Identification of the MRH Competing Risks Model

Heckman and Honoré (1989) and Abbring and Van den Berg (2003a) provide an identification result for the MPH competing risks model. [17] Our goal is to establish a nonparametric identification of the MRH competing risks model.

Replacing f with f_A or f_B in the proof of Proposition 1, we can solve for ϕ_A, ϕ_B, respectively. Next, we notice that for $t > 0$

$$\begin{aligned}\mathbb{P}[T_A > t, T_B > t|x] &= \int_0^1 \int_0^1 \prod_{j=A,B} \exp\left(-\phi_j(x)\int_0^t \lambda_j(\omega) r_j(\omega, u_j) d\omega\right) dG(u_A, u_B) \\ &= \mathcal{L}_{tt}\left(\phi_A(x), \phi_B(x)\right). \end{aligned} \tag{27}$$

Following analogous steps to the proof of proposition 1, by varying the functions ϕ_A, ϕ_B appropriately, we can trace out for fixed t the LT $\mathcal{L}_{tt}$ and, subsequently the joint distribution of $(\Upsilon_A(t), \Upsilon_B(t))$. Therefore, we can identify the two marginal distributions for each t. This permits us to identify the functions $\Lambda_A, \Lambda_B, R_A$, and R_B by working analogously to the proof of our proposition 1. As shown in the previous subsection (formula (25)), knowledge of the distribution of $(\Upsilon_A(t), \Upsilon_B(t))$ gives us the distribution of (U_A, U_B). In view of the above discussion we can state the following proposition:

Proposition 3 *Let assumptions 1 − 4 hold for each risk. Furthermore, the vector $(\phi_A(x), \phi_B(x))$ takes on values in an open subset of $(0, \infty)^2$. Then the functions $\phi_A, \phi_B, \Lambda_A, \Lambda_B$, R_A, R_B and the joint CDF of (U_A, U_B) for the competing risks model (24) are identified.*

In fact, our proof is an alternative way to identify Λ_A, Λ_B and the distribution of V_A, V_B in the model of Abbring and Van den Berg (2003a) without using differential equations. Nevertheless, our methodology is valid only if the underlying frailties are absolutely continuous with respect to the Lebesgue measure, whereas their identification result does not need such a constraint.

[17] Lee (2006) discusses semiparametric identification of nonlinear transformation competing risks models by parameterizing the functions ϕ_j. On the other hand, Lee and Lewbel (2013) consider nonparametric identification for linear transformation models. Both papers are different from the papers of Heckman and Honoré (1989) and Abbring and Van den Berg (2003a) as they do not require $t \to 0$ (i.e., the so-called identification at zero) for the derivation of the main results.

5.2.2 Testing for MPH and MRH Structure in the Competing Risks Model

Concerning hypothesis testing in the competing risks setup- unless we are willing to make some rather strong assumptions about the underlying models- we cannot follow an approach similar to the one of section 4. Yet, we can test for the MRH model by just considering a one step procedure where the null hypothesis refers to the MPH model, while the alternative hypothesis corresponds to the MRH model such that the variable t has an effect on the function r. In mathematical notation, we now have

$$\begin{aligned} &\mathbb{H}_0 : \theta_j(t|x, U) \in \mathcal{M}_{MPH} \\ &\mathbb{H}_1 : \theta_j(t|x, U) \in \mathcal{M}_{MRH} \backslash \{\mathcal{M}_{MPH}\}, \end{aligned} \tag{28}$$

where we examine each j seperately .

Irrespective of which hypothesis is correct, the identification strategy for the functions ϕ_A, ϕ_B is exactly the same for both of MPH and MRH models by making use of the nonparametrically identified (from the data) subdensity functions $f_A(t|x)$ and $f_B(t|x)$. Since ϕ_A, ϕ_B are known, we can condition on these two quantities rather than on x. Now, in a similar manner as in section 4, we define $\rho_j(t, \phi_A(x), \phi_B(x)) := \frac{f_j(t|\phi_A(x),\phi_B(x))}{\partial_{\phi_j} S(t|\phi_A(x),\phi_B(x))}$, where $S(t|\phi_A(x), \phi_B(x))$ is the survival function of $\min(T_A, T_B)|(\phi_A(x), \phi_B(x))$. Identical calculations to the ones of section 4 reveal that under the MPH model for the risk j, the quantity $\rho_j(t, \phi_A(x), \phi_B(x))$ is factorized into the product of two functions of t and ϕ_j, respectively. More precisely, $\rho_j(t, \phi_A(x), \phi_B(x)) = -\phi_A(x)\frac{\lambda_A(t)}{\Lambda_A(t)}$ for the MPH model. Therefore, by a straightforward modification of (17), we can construct the two respective hypothesis-testing problems (one for each risk). In the presence of random right censoring and/or left truncation, we can utilize the following property for the development of the nonparametric test: $\rho_j(t, \phi_A(x), \phi_B(x)) = -\frac{\psi_j(t|\phi_A(x),\phi_B(x))}{\partial_{\phi_j} \Psi(t|\phi_A(x),\phi_B(x))}$, where $\psi_j(t|\phi_A(x), \phi_B(x))$ denotes the cause-specific hazard rate of $\min(T_A, T_B)|(\phi_A(x), \phi_B(x))$ and $\Psi(t|\phi_A(x), \phi_B(x)) := \int_0^t \psi(\omega|\phi_A(x), \phi_B(x))d\omega$. Formally,

$$\psi_j(t|\phi_A(x), \phi_B(x)) := \lim_{dt \to 0} (dt)^{-1} \mathbb{P}[t \leq \min(T_A, T_B) < t+dt, D_j = 1| \min(T_A, T_B) \geq t, \phi_A(x), \phi_B(x)]. \tag{29}$$

and

$$\psi(t|\phi_A(x), \phi_B(x)) := \lim_{dt \to 0} (dt)^{-1} \mathbb{P}[t \leq \min(T_A, T_B) < t + dt | \min(T_A, T_B) \geq t, \phi_A(x), \phi_B(x)]. \quad (30)$$

Notice that $\psi(t|\phi_A(x), \phi_B(x)) = \psi_A(t|\phi_A(x), \phi_B(x)) + \psi_B(t|\phi_A(x), \phi_B(x))$ and that the functions on the right as well as left-hand side can be nonparametrically identified by using the data without imposing any condition on the dependence between U_A and U_B once the functions ϕ_A and ϕ_B have been identified. To see that $\rho_j(t, \phi_A(x), \phi_B(x)) = \frac{\psi_j(t|\phi_A(x),\phi_B(x))}{\partial_{\phi_j} \Psi(t|\phi_A(x),\phi_B(x))}$, we use the facts that

$$f_j(t|\phi_A(x), \phi_B(x)) = \psi_j(t|\phi_A(x), \phi_B(x)) \exp(-\Psi(t|\phi_A(x), \phi_B(x)))$$

and

$$S(t|\phi_A(x), \phi_B(x)) = \exp(-\Psi(t|\phi_A(x), \phi_B(x))).$$

As we mentioned above, if we would like to develop a similar two-step procedure to the one of section 4, we would need further (rather strong) assumptions to detect whether the corresponding hazard rate is of the MRH type. An example of such an (difficult to justify) assumption could be that the vector x affects the hazard rate of each risk through two different indices. Thus, for each risk j, we would have

$$\begin{aligned} \theta_A(t|x, U_A) &= \lambda_A(t)\phi_{A_1}(x)\phi_{A_2}(x)r_A(t, U_A), \\ \theta_B(t|x, U_B) &= \lambda_B(t)\phi_{B_1}(x)\phi_{B_2}(x)r_B(t, U_B). \end{aligned} \quad (31)$$

In this case, we would need the extra condition that the vector-valued function $(\phi_{j_1}(x), \phi_{j_2}(x))$ varies in an open subset of $(0, \infty)^2$.

5.2.3 Testing for Independence in the MRH Competing Risks Model

One of the major challenges in the competing risks model is to examine whether $U_A \perp\!\!\!\perp U_B$, that is, whether the unobserved variables are independent. It is easy to show that independence between U_A and U_B occurs if and only if $T_A \perp\!\!\!\perp T_B | x$. Van den Berg and Effraimidis (2015) have derived a test statistic for the semiparametric mixed hazard model under a single-index restriction. [18]

In this subsection, we show that a similar procedure can be followed for investigating whether or not the underlying unobserved factors are independent. [19] The idea is based on the fact that for each $t > 0$ the survival function factorizes into the product of two functions, where the first function depends on $\phi_A(x)$ and the second function depends on $\phi_B(x)$, if and only if $U_A \perp\!\!\!\perp U_B$. First, note that for each $t > 0$ and $x \in \mathcal{X}$ $U_A \perp\!\!\!\perp U_B$ yields

$$\begin{aligned}\mathbb{P}[T_A > t, T_B > t|x] &= \int_0^1 \int_0^1 \prod_{j=A,B} \exp\left(-\phi_j(x)\int_0^t \lambda_j(\omega) r_j(\omega, u_j) d\omega\right) dG(u_A, u_B) \\ &= \mathcal{L}_t^A\left(\phi_A(x)\right)\mathcal{L}_t^B\left(\phi_B(x)\right),\end{aligned}$$

where $\mathcal{L}_t^j$ refers to the LT of the random variable $\Upsilon_j(t) = \int_0^t \lambda_j(\omega) r_j(\omega, u_j) d\omega$ (recall that we fix t).

On the other hand, suppose that

$$\mathbb{P}[T_A > t, T_B > t|x] = \mathcal{L}_{tt}\left(\phi_A(x), \phi_B(x)\right) = H_t^A\left(\phi_A(x)\right) H_t^B\left(\phi_B(x)\right)$$

for any fixed $t > 0$ and $x \in \mathcal{X}$. The unknown functions H_t^A and H_t^B are real analytic. Since $\mathcal{L}_{tt}\left(\omega_A, \omega_B\right) = H_t^A\left(\omega_A\right) H_t^B\left(\omega_B\right)$ for an open rectangle of $\mathcal{R}_+ \times \mathcal{R}_+$ (provided that $(\phi_A(x), \phi_B(x))$ takes on values in an open rectangle of $\mathcal{R}_+ \times \mathcal{R}_+$), we obtain that this equality holds true on the whole $\mathcal{R}_+ \times \mathcal{R}_+$ due to the real analyticity of the functions on the left- as well as the right-hand side. By standard arguments related to LT, we get $H_t^j(\omega) = \mathcal{L}_t^j(\omega)$ for $\omega \in \mathcal{R}_+$. The latter implies that $\mathcal{L}_{tt}\left(\omega_A, \omega_B\right) = \mathcal{L}_t^A(\omega_A)\mathcal{L}_t^B(\omega_B)$ on $\mathcal{R}_+ \times \mathcal{R}_+$. Equivalently, for fixed $t > 0$ the random variables $\Upsilon_A(t)$ and $\Upsilon_B(t)$ are independent. Hence, $U_A \perp\!\!\!\perp U_B$.

[18] By using the notation of the current paper, the authors model the hazard rate of $T_j|x, U_j$ as follows:

$$\theta_j(t|x, U_j) = \mu(t, \beta_j^{'} x) P_j^{-1}(U_j),$$

where μ is a positive function and $\beta_j \in \mathcal{R}^d$.

[19] To keep the notation simple throughout, we discuss the case where the minimum is always observed and is not randomly right censored/left truncated. The extension is rather straightforward.

6 Treatment Effects

In this section we show how our new time-varying frailty model can be used in the context of treatment effects. Specifically, following Drepper and Effraimidis (2016), at time $t_0 = 0$ the subject under consideration enters an initial state and faces two competing hazard rates θ_1, θ_2. The two exit hazards are affected by an endogenous treatment that occurs at time $S = s$ with hazard rate θ_S. In female mortality studies, for instance, the two competing risks could correspond to mutually exclusive causes of death such as heart attack vs. cancer, and the treatment could represent age at conception. Another example is the effect of smoking on different causes of death. [20]

The trivariate hazard model can be expressed as follows:

$$\begin{aligned}
\theta_1(t|S,x,U_A) &= \lambda_A(t)\ \phi_A(x)\ \delta_A(t|S,x)^{1(t>S)}\ r_A(t,U_A) \\
\theta_2(t|S,x,U_B) &= \lambda_B(t)\ \phi_B(x)\ \delta_B(t|S,x)^{1(t>S)}\ r_B(t,U_B) \\
\theta_S(s|x,U_S) &= \lambda_S(s)\ \phi_S(x) r_S(t,U_S),
\end{aligned} \tag{32}$$

where 1 is the indicator function. In the first step and working analogously to subsection 5.2, we can nonparametrically identify ϕ_j, Λ_j, and R_j $(j = A, B, S)$ as well as the distribution of the trivariate random vector (U_A, U_B, U_S). In particular, we have to use the subdensity functions $f_j(t|x) = \lim_{dt\to 0} (dt)^{-1} \mathbb{P}[t \leq \min(T_A, T_B, S) < t + dt, D_j = 1|x]$ and the survival function $\mathbb{P}[T_A > t, T_B > t, S > t|x]$. In the second step, by using similar arguments to the work of Drepper and Effraimidis (2016), we can nonparametrically identify the treatment effect functions $\int_0^t \delta_j(\omega|S,x)$ for $j = A, B$.

The above result implies that the time-varying frailty model of this paper can also be adopted for the construction of social networks where the outcomes represent the transition from an initial state to a final state. Drepper and Effraimidis (2015) construct such a model and analyze peer effects among siblings in the use of drugs.

[20] The case of one treatment and one risk can be viewed as a special case of the model in this section. Alternatively, its identification can be achieved by following the strategy of Abbring and Van den Berg (2003b) and steps described in subsection 5.2.

7 Conclusions

The contribution of this paper is the construction of a new duration model, the Mixed Random Hazard model, which can be used for the specification of a hazard rate where the value of the frailty (unobserved risk factors) changes over time. The presence of time-varying frailty can be justified for several types of problems. By exploiting continuous variation of the observed covariate(s), we can nonparametrically identify the underlying functions. Moreover, hazard rate models with random baseline hazard and without unobserved heterogeneity are a special case of the Mixed Random Hazard model. Two popular examples of such models are the Gompertz and Weibull hazard rates where the underlying parameters are random. It is shown that these two specifications do not satisfy one of the (sufficient) conditions needed for the nonparametric identification result.

We also describe a two-step nonparametric procedure that can form the basis of a nonparametric test for determining whether the hazard rate has a Mixed Random Hazard specification. Future research efforts need to develop a formal test statistic with asymptotic theory for this important and challenging problem. We also show that nonparametric identification is possible for (i) the multivariate duration model with parallel durations and (ii) the competing risks model. An interesting research finding is that the semiparametric test statistic for competing risks independence developed by Van den Berg and Effraimidis (2015) is applicable to the Mixed Random Hazard competing risks model if we model the regressor functions as linear indices.

Finally, future research may consider nonparametric identification of the multivariate duration model in which the duration variables are sequential, that is, a duration model with lagged duration dependence.

Appendix

In this appendix, we present and compare two different hazard rate models. The first one has the specification

$$\theta(t|x, V) = a \exp(bt)V$$

and the second one is expressed as follows:

$$\theta(t|x, V) = a \exp(bt) \frac{2V \exp(Vt)}{\exp(Vt) + 1}.$$

To simplify the analysis, we have omitted x. The first hazard rate corresponds to the MPH model, whereas the second hazard rate corresponds to the MRH model. The random variable V is gamma distributed with shape parameter and scale parameter equal to 1. Note that at the beginning of the process the distribution of the unobserved heterogeneity is the same for both models. We set $a = 0.0004$ and $b = 0.1$.

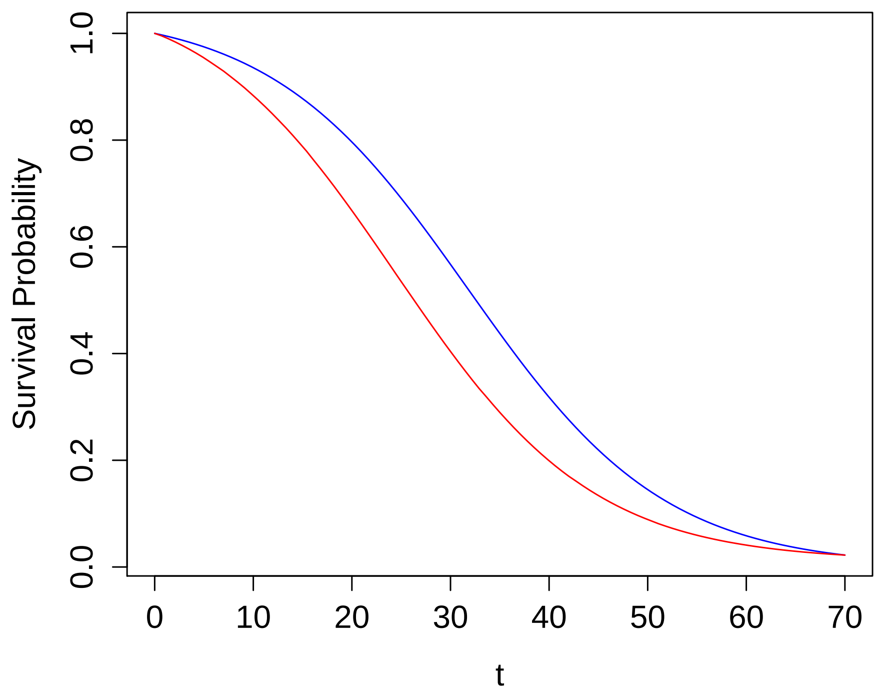


Figure 1: Plot of the survival curve for two different models: MPH model (blue line), MRH model (red line).

As we can observe from the above graph, the survival probability at any fixed $t > 0$ is larger for the MPH model. This is due to the fact that for any given V, the value of the MRH frailty is larger than the value of the MPH frailty. Consequently, for any given V and each $t > 0$, the MRH hazard rate has a larger value than the MPH hazard rate.

References

Abbring, J. H. and Van den Berg, G. J. (2003a), "The identifiability of the mixed proportional hazards competing risks model," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 701–710.

— (2003b), "The nonparametric identification of treatment effects in duration models," *Econometrica*, 71, 1491–1517.

Botosaru, I. (2013), "A Duration Model with Dynamic Unobserved Heterogeneity," *Working Paper*.

Brinch, C. N. (2007), "Nonparametric identification of the mixed hazards model with time-varying covariates," *Econometric Theory*, 23, 349–354.

Chiappori, P. A., Komunjer, I., and Kristensen, D. (2015), "Identification and Estimation of Nonparametric Transformation Models," *Journal of Econometrics*, 188, 22–39.

Drepper, B. and Effraimidis, G. (2015), "Social Networks in Duration Models," *Working Paper*.

— (2016), "Identification of the Timing-of-Events Model with Multiple Competing Exit Risks from Single-Spell Data," *Working Paper, Submitted*.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009), "Frailty correlated default," *The Journal of Finance*, 64, 2089–2123.

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.

Enki, D. G., Noufaily, A., Farrington, C. P., et al. (2014), "A time-varying shared frailty model with application to infectious diseases," *The Annals of Applied Statistics*, 8, 430–447.

Evdokimov, K. (2011), "Nonparametric identification of a nonlinear panel model with application to duration analysis with multiple spells," *Department of Economics, Princeton University*.

Gjessing, H. K., Aalen, O. O., Hjort, N. L., et al. (2003), "Frailty models based on Lévy processes," *Advances in Applied Probability*, 35, 532–550.

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

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

Hull, J. and White, A. (2012), "CVA and wrong-way risk," *Financial Analysts Journal*, 68, 58–69.

Lee, S. (2006), "Identification of a competing risks model with unknown transformations of latent failure times," *Biometrika*, 93, 996–1002.

Lee, S. and Lewbel, A. (2013), "Nonparametric identification of accelerated failure time competing risks models," *Econometric Theory*, 29, 905–919.

Lewbel, A., Lu, X., and Su, L. (2015), "Specification testing for transformation models with an application to generalized accelerated failure-time models," *Journal of Econometrics*, 184, 81–96.

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

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

Paddy Farrington, C., Unkel, S., and Anaya-Izquierdo, K. (2012), "The relative frailty variance and shared frailty models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 673–696.

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

Unkel, S., Farrington, C., Whitaker, H. J., and Pebody, R. (2014), "Time varying frailty models and the estimation of heterogeneities in transmission of infectious diseases," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63, 141–158.

Van den Berg, G. and Effraimidis, G. (2015), "A Semiparametric Test of Competing Risks Dependence Using Covariates," *Working Paper*.

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

Vaupel, J. W. (2010), "Biodemography of human ageing," *Nature*, 464, 536–542.