Universität St.Gallen

# Nonparametric identification in nonseparable duration models with unobserved heterogeneity

Petyo Bonev

March 2020 Discussion Paper no. 2020-05

School of Economics and Political Science, Department of Economics

University of St.Gallen

| | |
|---|---|
| Editor: | Vanessa Pischulti<br>University of St.Gallen<br>School of Economics and Political Science<br>Department of Economics<br>Müller-Friedberg-Strasse 6/8<br>CH-9000 St.Gallen<br>Phone +41 71 224 23 07<br>Email seps@unisg.ch |
| Publisher: | School of Economics and Political Science<br>Department of Economics<br>University of St.Gallen<br>Müller-Friedberg-Strasse 6/8<br>CH-9000 St.Gallen<br>Phone +41 71 224 23 07 |
| Electronic Publication: | http://www.seps.unisg.ch |

# Nonparametric identification in nonseparable duration models with unobserved heterogeneity[1]

Petyo Bonev

Author's address: Prof. Dr. Petyo Bonev
Swiss Institute for Empirical Economic Research (SEW)
University of St.Gallen
Varnbüelstrasse 14
CH-9000 St.Gallen
Email petyo.bonev@unisg.ch
Website https://sites.google.com/site/petyobbonev/

[1] I thank Gerard J. van den Berg, Richard W. Blundell, Christoph Breunig, Christian Brinch, Andrew Chesher, Dennis Kristensen, Michael Lechner and participants at seminars at the University College of London, ETH Zürich and the Humboldt University of Berlin for their helpful and critical comments. Special thanks to Georgios Effraimidis for his helpful ideas and suggestions.

## Abstract

We study nonparametric identification of nonseparable duration models with unobserved heterogeneity. Our models are nonseparable in two ways. First, genuine duration dependence is allowed to depend on observed covariates. Second, observed and unobserved characteristics may interact in an arbitrary way. Our study develops novel identification strategies for a comprehensive account of typical duration model settings. In particular, we show identification in single-spell models with and without time-varying covariates, in multiple models with shared frailty and lagged duration dependence, in single-spell and multiple-spell competing risks models, and in treatment effects models where treatment is assigned during the individual spell in the state of interest.



## Keywords

Duration models, identification, unobserved treatment heterogeneity, nonseparable models, competing risks, treatment effect, job search, unemployment

## JEL Classification

C14, C41, J64

# 1 Introduction

A major strategy for identification in duration models is to impose that the hazard function of the duration variable is multiplicatively separable with respect to genuine duration dependence, observed and unobserved covariates. The Mixed Proportional Hazard (MPH) model, which incorporates this assumption, is the most commonly used duration model in the econometric literature. Separability is a powerful source of identifying variation. It can be used to distinguish between genuine duration dependence and spurious duration dependence caused by dynamic selection. In models with competing risks, separability has been used to break the nonidentification result of Tsiatis (1975).

Despite these appealing properties, multiplicative separability has two major drawbacks. First, separating covariates from duration dependence is difficult to justify with economic theory. In job search models, for example, this is only justified with myopic agents and under very particular parametric assumptions, see Van den Berg (2001). These theoretical arguments are reinforced by recent experimental studies that find that duration dependence varies in a complex, nonproportional way with observed covariates, see Eriksson and Rooth (2014), Farber et al. (2016) and Kroft et al. (2013). Second, separating observed and unobserved characteristics implies that the effect of an observed covariate is the same for all individuals with identical observed characteristics. This assumption, however, is often too restrictive. While unobserved effect heterogeneity has been addressed by numerous studies in the context of regression models (see e.g. Matzkin (2007)), it has remained largely unaddressed in duration context.

The objective of this paper is to develop methods that address the drawbacks of fully separable models. I study identification of hazard models of the type

$$\theta(t \mid X, V) = \lambda(t, X) r(X, V),$$

where $\theta$ is the structural hazard function of a duration random variable and $X$ and $V$ represent observed and unobserved individual characteristics, respectively. These models are nonseparable in the sense that they allow for (i) arbitrary observed heterogeneity in the "true" duration dependence (through the function $\lambda$) and (ii) arbitrary interaction between observed and unobserved covariates (through the function $r$). The functions $\lambda$ and $r$ are unknown to the econometrician and are not assumed to belong to a parametric family.

This paper achieves identification in a comprehensive account of empirically relevant settings. In particular, I develop identification strategies for single-spell models with and without time-varying covariates, in multiple-spell models with shared frailty and lagged duration dependence, in single-spell and multiple-spell competing risks models, as well as in treatment effects models where treatment is assigned during the individual spell in the state of interest.

Identification in this study relies on two main strategies. At the core of my single-spell approaches is the mild assumption that there is one "special" regressor, whose effect on the individual hazard is homogeneous w.r.t. unobserved characteristics. No assumption is imposed on the genesis of this regressor. Furthermore, this regressor is allowed to impact the "weeding out" process (i.e. the change of the conditional distribution of $V$ given $X$ over time) in an unrestricted way. This is the first paper to show identification in single-spell hazard models with interaction between observed and unobserved characteristics. Other single-spell nonparametric studies rely either on the fully separable structure of the MPH model, Elbers and Ridder (1982), Heckman and Singer (1984), Ridder (1986), Horowitz (1999), Chiappori et al. (2015), or at least on the multiplicative separability of the unobserved heterogeneity as in the Mixed Hazard (MH) model, Brinch (2007), McCall (1994).

My multiple-spell models rely on a mild fixed-effects assumption and do not require a "special" separable regressor. Multiple-spell methods are often used in labor market studies, in particular when two or more unemployment spells per individual are available, or when there is information on previous employment spells. I allow previous spells to impact the hazard of subsequent spells (lagged duration dependence). My multiple-spell model can thus be interpreted as a hazard version of a dynamic panel model with fixed-effects. In the context of unemployment duration, lagged duration dependence, sometimes also referred to as scarring effects, is an object of interest on its own and its analysis has a long tradition, Heckman and Borjas (1980). Related lagged duration studies either rely on fully separable variation in the observed heterogeneity, Honoré (1993), or at least impose separability of the unobservables, Frijters (2002) and Picchio (2012).[1,2] I show that a fixed-effects assumption is sufficient to avoid these restrictive assumptions. Importantly, I allow the lagged duration effect to be arbitrary heterogeneous w.r.t. observed characteristics. This assumption is motivated by recent empirical findings in the unemployment duration literature, see e.g. Cockx and Picchio (2013). The models in this paper also nest a nonseparable shared frailty model, which is widely used in demographics and appropriate when multiple individuals from the same group share unobserved characteristics, Hougaard (2000). Finally, in addition to my identification results, I derive a novel property of nonseparable multi-spell hazard models that can be used for graphical model diagnostics and testing.

Next, I study identification in nonseparable competing risks models. Competing risks models arise naturally in unemployment duration context when more than one exit destination is possible, see e.g. Kyyrä and Ollikainen (2008) for an empirical example. Competing risks models based on the separable MPH specification have been studied by Heckman and Honoré (1989), Abbring and Van Den Berg (2003a) and Horny and Picchio (2010). I apply the single-spell and multiple-spell results presented in this paper to achieve identification in nonseparable competing risks models. I also give a general characterization of the single-spell strategies that *can* lead to identification in competing risks models. In particular, the special regressor assumption is necessary.

Finally, I study nonseparable duration models which allow treatments to be assigned during the spell of observation. Typical examples for such treatments are Active Labor Market Policies (e.g. job search trainings), or punitive reductions of unemployment benefits. Such a setup was first studied by the seminal paper of Abbring and Van Den Berg (2003b). They achieve identification in an augmented bivariate MPH model. This paper generalizes their results to a nonseparble context.

In all presented models, identification of the function $r$ draws on insights from the literature on nonaddtive random functions, see Matzkin (2003) and Chesher (2007). In particular, $r$ is assumed to be strictly increasing in the unobserved component. This is a natural gener-

[1] Frijters (2002) imposes in addition that the baseline hazard is the same for all spells.

[2] A nonseparable multiple-spell model is also studied in Evdokimov (2010). Identification there rules out lagged duration dependence and relies on having at least three spells per individual.

alization of the MPH assumption $r(x, v) = id(v)$. Important MPH features such as spurious negative duration dependence in the empirical hazard translate to the general nonseparable setup in a straightforward way. Thus, this paper provides a link between the literature on identification in duration mixture models and the literature on identification in nonseparable regression models.

The paper is structured as follows. In section 2, we present and motivate our model. Our identification results in single- and multiple spell single risk models are presented in sections 3. Section 4 presents applications to competing risks and duration treatment effect models. Section 5 concludes. All proofs are in the appendix.

# 2 Model and motivation

For illustrative purposes, we build our exposition on a labor market example. Suppose that $n$ unemployed individuals are searching for a job. Denote by $T_i, i = 1, \dots, n$ the duration of unemployment of individual $i$, with $T_1, \dots, T_n$ assumed to be independent and drawn from the same stochastic law. Let $\theta(t)$ be the unconditional hazard of $T_i$ at elapsed length of unemployment $t \geq 0$, $\theta(t) = \lim_{h \to 0} \Big( P\{T \in [t, t+h) | T \geq t\}/h \Big)$ (we omit the index $i$ whenever this is possible). Let the random vector $X_i$ represent observed characteristics of individual $i$ that impact the duration $T_i$. $X_i$ is assumed to have realized (just) prior entry into unemployment, so that its value is determined at $t = 0$. Typical examples are wage and experience in the preceding job spell, highest degree of education obtained until the moment of inflow into unemployment and gender. The realizations of $X_i$, denoted by small letters $x$, are assumed to be elements of a set $\mathbb{X} \subset \mathbb{R}^k$, where $k$ is a positive integer. By way of definition, $X_i$ is time-constant. We consider time-varying covariates in section 3.2. Further, let $V_i$ be a one-dimensional nonnegative unobserved random variable. $V_i$ represents a single index of all unobserved individual characteristics that impact $T_i$. A typical example for factors contained in $V_i$ is noncognitive skills. Analogously to the definition of $X_i$, $V_i$ is required to be time-constant and determined before the individual spell start.

Conditionally on $X_i$ and $V_i$, the individual hazard is assumed fully specified. Henceforth, we write $\theta(t|X, V)$ to denote the individual hazard with notation in analogy to conditional probabilities. The empirical or observed hazard is denoted by $\theta(t|X)$. Note that an important property of mixture hazard models (i.e. of hazard models allowing for unobserved heterogeneity $V$) is that $T$ is a random variable even conditionally on realizations of $X$ and $V$. The residual randomness, referred to as "the effect of luck" by Lancaster (1979), has been only scarcely discussed in the literature and has no established meaning. Lancaster (1990) interprets it as some intrinsic individual uncertainty, whereas Heckman (1991) distinguishes between characteristics known to the individual which are captured by $V$, and characteristics that are unknown to the individual and subsumed by the residual randomness. Decisions of agents are therefore based solely on the value of $X, V$. Heckman (1991) acknowledges, however, that this distinction has an arbitrary character. For the rest of the paper, we follow Heckman (1991) and assume that the residual randomness is due to idiosyncratic factors (noise) which are independent of the factors determining the decisions of the agents. This interpretation is compatible with the well-known fact that the transformed duration $\Theta(T|X, V) := \int_0^T \theta(t|X, V)$ is independent of $X$ and $V$, see e.g. Lancaster (1990).

With these preliminaries, consider the following model, which we refer to as the **Generalized Mixed Hazard (GMH)** model:

$$\theta(t|X,V) = \lambda(t,X)r(X,V) \tag{1}$$

The functions $\lambda$ and $r$ are unknown to the econometrician. Two important special cases are the Mixed Proportional Hazard (MPH) model,

$$\theta(t|X,V) = \lambda_{MPH}(t)\theta_0(X)V \tag{2}$$

and the Mixed Hazard (MH) model,

$$\theta(t|X,V) = \lambda_{MH}(t,X)V. \tag{3}$$

Both the MPH and MH models impose multiplicative separability of $V$. As observed by Lancaster (1985) and Chesher (2002), multiplicative separability of $V$ effectively reduces the number of sources of stochastic variation from 2 to 1. The GMH model, on the contrary, allows $X$ and $V$ to interact through the function $r$ in an arbitrary way. Furthermore, similarly to the MH model, the GMH model allows the duration dependence $\lambda$ to depend on observed covariates.

In appendix section A, we briefly discuss the relation of the GMH model to other existing models. In a nutshell, the Accelerated Time Failure (AFT) model and its generalizations the GAFT and the EGAFT models are not nested in the GMH model. Conversely, the GMH model is also not nested in those models.

We now motivate the GMH model with several empirical examples.

**Example 1: unobserved treatment heterogeneity.** Van den Berg and Van der Klaauw (2006) evaluate the effect of counseling and monitoring (CM) on the re-employment chances of unemployed workers within a social experiment in the Netherlands. At inflow into unemployment, workers are either assigned at random to CM ($Z = 1$) or to a control group with no such services ($Z = 0$). The model estimated in their paper is the MPH model

$$\theta(t|X,Z,V) = \lambda(t)\exp\{X\beta + \delta Z + \ln V\}$$

where $X$ denotes a list of controls. Separability of $Z$ and $V$ precludes the effect of the training from depending on unobserved noncognitive abilities such as locus of control. Recent empirical and theoretical evidence suggests however that individuals with higher levels of locus of control benefit more from the CM process through more active participation, Caliendo et al. (2015). The simplest model that can capture this relationship is

$$\theta(t|X,Z,V) = \lambda(t)\exp\{X\beta + \delta Z + \gamma Z\ln V + \ln V\}. \tag{4}$$

This is a hazard version of the location-scale model in quantile regression, see He (1997). The coefficient $\gamma$ captures the unobserved effect heterogeneity of $Z$.

**Example 2: heterogeneous duration dependence.** The baseline hazard function $\lambda_{MPH}$ in model (2) represents the (genuine) duration dependence of the hazard. A commonly discussed reason for negative duration dependence of the unemployment hazard is stigma. Stigma occurs when the willingness of employers to hire unemployed individuals decreases with increasing spell of unemployment. For example, screening models predict that employers use the length of the current unemployment spell as a (hidden) productivity signal, see e.g. Lockwood (1991). Recent experimental studies use fictitious job applications to actual job postings to analyze the effect of unemployment duration on the call-back rate. These studies provide evidence that the stigma-driven duration dependence might depend in a complex way on individual characteristics, in particular on the level of education and experience, as well on the type of occupation. As an example, Eriksson and Rooth (2014) find large drops of call-back rates over time for low- and medium-skilled occupations but little to no effect for high skilled jobs. Similar effects are found by Weber (2014). Farber et al. (2016) find that longer employment histories compensate for long ongoing unemployment spells, thus reducing or even eliminating stigma effects. Kroft et al. (2013) find significant heterogeneity of the duration dependence across levels of tightness of the local labor market. These empircial findings cannot be incorporated in a separable model. In particular, separability of $\lambda$ and $X$ implies that individual characteristics only shift the level of duration dependence. We thus complement the economic theory arguments against separability outlined in Van den Berg (2001).

The nonseparable duration dependence $\lambda(t, X)$ in model (1), on the contrary, allows for a flexible interaction between time and observed characteristics. As an example, consider the Weibull specification

$$\lambda(t, x) = \lambda_0 \lambda_1(x) t^{\lambda_1(x)-1} \tag{5}$$

with a scale parameter $\lambda_0$ and a shape parameter $\lambda_1$ whose value is now allowed to depend on the value of the observed characteristics.[3] The findings of Farber et al. (2016) can now be modeled as setting $\psi_1 < 1$ for shorter employment histories and $\psi_1 = 1$ for long employment histories. Alternatively, flexibility can be achieved in a piecewise-constant bazeline hazard, in which the coefficients depend on $x$.

**Example 3: heterogeneous measurment error.** Consider a case in which the duration variable is measured with error, see e.g. Abrevaya and Hausman (1999). Common reasons for measurement errors in duration context are time aggregation, Bergström and Edin (1992), as well as under-reporting in retrospective surveys, Mathiowetz and Duncan (1988) and Aït-Sahalia (2007). Lancaster (1985) shows that a multiplicative measurement error in the duration variable, together with a Weibull specification of the latent baseline hazard, lead to a MPH model, with the multiplicative $V$ being a result of the measurement error. His model imposes that the measurement error does not depend on observed covariates. In general, however, the measurement error will depend on demographic characteristics, Bound et al. (1989). Accounting for that possibility naturally leads to a nonseparable model, as we now demonstrate. Denote by $T^*$ the latent duration variable, which might be measured imprecisely. Assume that the observed duration $T$ satisfies

$$\begin{aligned} T &= g(T^*, \eta) = T^*(1/\eta), \\ \eta &= k(X, V). \end{aligned}$$

[3] The Weibull specification, in which $\lambda_1$ does not depend on $x$, is commonly used in empirical studies in the context of an MPH model, see Lancaster (1979) for an early paper and Van den Berg (2001) for further examples.

where $\eta$ is a measurement error, $k$ is some unknown function and $V \perp\!\!\!\perp (X, T^*)$. Let the conditional distribution of the latent variable follow a Weibull specification, $P\{T^* \leq t \mid X = x\} = 1 - exp\{-t^\alpha \psi(x)\}$ with $\alpha$ being a (unknown) scalar and $\psi$ an unknown function. Then the individual hazard of $T$ satisfies $\theta(t \mid x, v) = \alpha t^{\alpha-1}\psi(x)k(x,v)^\alpha$, which is a special case of model (1) with $\lambda(t,x) = \alpha t^{\alpha-1}\psi(x)$ and $r(x,v) = k(x,v)^\alpha$.

# 3 Identification

## 3.1 Two basic assumptions

Suppose first that $T$ is fully observed. This assumption is relaxed in section 4.1. Denote by $F_{T,X}$ and $F_{X,V}$ the distributions of $(T,X)$ and $(X,V)$, respectively, and by $G$ the marginal distribution of $V$. Following a convention in the analysis of identification, we assume that the distribution of the observables $F_{T,X}$ is known to the econometrician, see e.g. Lewbel (2019). We call a tripple $\mathscr{S} = (\lambda, r, F_{X,V})$ that satisfies (1) a model structure. Each structure implies exactly one distribution $F_{T,X}$. Two structures are observationally equivalent if they imply the same distribution $F_{T,X}$. A feature of a structure, say $\lambda$, is identified under a set of assumptions, if, under these assumptions, the value of the feature does not vary among any set of observationally equivalent structures, Chesher (2003). We also say that model (1) is identified under a given set of assumptions, if no two structures $\mathscr{S}, \mathscr{S}', \mathscr{S} \neq \mathscr{S}'$ are observationally equivalent.

The following two assumptions are adopted in all identification results in this paper.

**Assumption A1:** the function $r : \mathbb{X} \times \mathbb{R}^+ \to \mathbb{R}$ is (i) nonnegative and (ii) strictly increasing in its second component.

A1 is trivially fulfilled for the MPH and MH models. A1(i) ensures that the hazard function is nonnegative. A1(ii) is borrowed from the literature on nonseparable regression models, Matzkin (2003), where it simply implies a monotonic relationship between unobservables and outcomes. In the case of duration models, however, it has an important *additional* implication. Good risks (individuals with high values of $V$) have a higher exit rate out of unemployment for every fixed $t$ and $X$ than bad risks (low values of $V$). Thus, the distribution of $V$ changes over time $t$ spent in unemployment, with the proportion of bad risks increasing. This process, called weeding out, Lancaster (1979), or dynamic selection, Van den Berg (2001), creates a spurious negative duration dependence of the hazard. Formally, the semi-elasticity of the observed hazard of the GMH model w.r.t. time fulfills

$$\frac{\partial \ln \theta(t|x)}{\partial t} = \frac{\partial \ln \lambda(t,x)}{\partial t} - \frac{Var[V_x|T \geq t, X = x]}{\mathbb{E}[V_x|T \geq t, X = x]}\lambda(t,x),$$

where we set $V_x := r(x,V)$ and $Var$ denotes the variance of a random variable. The above equality implies that $\frac{\partial \ln \theta(t|x)}{\partial t} < \frac{\partial \ln \lambda(t,x)}{\partial t}$. Put in words, for each subgroup of individuals characterized by a given value $x$, observed duration dependence is more negative than the true duration dependence.

To describe the set of observationally equivalent models under A1, assume that the distribution $G$ is strictly increasing and denote by $G^{-1}$ its inverse. Define the function $\bar{r} : \mathbb{X} \times \mathbb{R}^+ \to \mathbb{R}, \bar{r}(x,y) = r(x, G^{-1}(y))$. Furthermore, let $\bar{V} = G(V)$. Then $\bar{V}$ is uniformly distributed on $[0, 1]$, $\bar{r}$ is nonnegative and strictly increasing in its second argument, and it holds

$$\bar{r}(x, \bar{V}) = r(x, V) \tag{6}$$

for each $x \in \mathbb{X}$. (6) implies that identification requires either normalization of $G$ or additional restrictions on $r$. In this paper, we follow Matzkin (2003) and normalize $G$, see assumption A2(i) below. Further normalization approaches discussed in Matzkin (2003) are not natural to the context of duration models and we do not pursue them.

**Assumption A2:** the unobserved random variable $V$ is (i) uniformly distributed on the interval $[0, 1]$ and (ii) is independent of $X$.

Assumption A2 (ii) is a standard assumption in the literature on hazard mixture models. The identification of single-spell models typically depends crucially on it. Honoré (1993) shows that multiple-spell MH models are identified without imposing A2 (ii). We come back to this point in section 3.3.

**Example 1, continued.** The scale-location model (4) satisfies assumption A1 under the restriction $\gamma Z \geq 0$.

Under A1 and A2, identification boils down to retrieving $\lambda$ and $r$ from the data. It is clear, however, that this cannot be achieved without further assumptions. There are two reasons for this. First, without further restrictions on $\lambda$ and $r$, it is possible to shift separable components depending on $X$ between $\lambda$ and $r$ without changing the DGP. As an example, the structures $(\lambda_1, r_1)$, $(\lambda_2, r_2)$ with

$$\lambda_1(t, x) = \psi(t) \exp\{\beta_1 x_1 + \beta_2 x_2\}, \qquad r_1(x, v) = \exp \beta_3 x_3 + v \tag{7}$$

$$\lambda_2(t, x) = \psi(t) \exp\{\beta_1 x_1\}, \qquad r_2(x, v) = \exp \beta_2 x_2 + \beta_3 x_3 + v \tag{8}$$

are observationally equivalent. Second, assume for the moment that $r$ does not depend on $x$, so that the problem of shifting components of $x$ does not exist. In the following lemma, we state that the MH model is not identified.

**Lemma 1** *Assume that $r(x_1, V) = r(x_2, V) = r(V)$ for any two $x_1, x_2 \in \mathbb{X}$ and define $\bar{V} = r(V)$. Denote by $\bar{G}$ the distribution of $\bar{V}$. Then, there exists a generalized hazard $\tilde{\lambda}$ and a distribution $\tilde{G}$ of a nonnegative random variable $\tilde{V}$, such that the structures $(\lambda, \bar{G})$ and $(\tilde{\lambda}, \tilde{G})$ are observationally equivalent.*

Thus, identification is hampered by two distinct problems. The first one arises from the interplay of $\lambda$ and $r$ and the second one from the MH model. Before we present solutions to these problems in the next sections, we briefly study the relative importance of $r$ and $\lambda$ for identification. We can state the following result.

**Proposition 1** *Suppose that model* (1) *holds.*

*(i) If $\lambda$ is known, then under assumptions A1 and A2, the function $r$ is identified.*

*(ii) If $r$ is known, then under assumptions A1 and A2, the function $\lambda$ is identified.*

To interpret part (i) of proposition 1, transform model (1) in the following way

$$\ln \int_0^T \theta(t|X, V)dt = \ln \int_0^T \lambda(t, X)dt + \ln r(X, V). \tag{9}$$

By rearranging, we obtain

$$Y = k(X, V) + \varepsilon, \tag{10}$$

where $Y = \ln \int_0^T \lambda(t, X)dt$ is observed, $k(X, V) = -\ln r(X, V)$ and $\varepsilon = \ln \epsilon = \ln \int_0^T \theta(t|X, V)dt$. The error term $\epsilon$ is a unit exponential r.v. and is independent of $X$ and $V$. Note that the identification result remains valid under any known distribution of $\epsilon$. Thus, proposition 1 implies that the identification result of Matzkin (2003) holds even if we add a second source of unobserved stochastic variation, as long as its distribution is known. Hence, the doubly-stochastic regression model (10) with a known transformation behaves as a standard nonseparable regression model. The assumptions needed for identification are identical in those two cases.

## 3.2 Nonparametric identification in single-spell models

In this section, we assume that the econometrican observes exactly one spell of unemployment for each individual in the sample. At the core of our approach is the assumption that at least one regressor, say $X_1$, can be be excluded from the function $r$.

**Assumption A3.** There exists a random subvector $X_1$ with dimension $d_1 \geq 1$ and domain $\mathbb{X}_1$, such that $X = (X_1, X_2)$ has realizations in $\mathbb{X} = \mathbb{X}_1 \times \mathbb{X}_2 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and

$$r(x_1, x_2, v) = r(x_1^*, x_2, v) \tag{11}$$

for all $x_1, x_1^* \in \mathbb{X}_1, x_2 \in \mathbb{X}_2$ and all $v \in \mathbb{R}^+$.

A3 justifies the notation $r(x_2, V)$. Unlike the special regressor assumption of Lewbel (1998), the special treatment here relates to the effect heterogeneity and not to the joint distribution of observables and unobservables. Therefore, knowledge of the selection process alone is not sufficient for motivating A3. One approach to discipline the model choice is to rely on evidence on treatment effect heterogeneity provided by studies with particularly rich numbers of covariates. The following example illustrates this point.

**Example 1, continued.** In the context of example 1, suppose the researcher assumes that the major unobservable characteristics behind $V$ are some characteristics of the case worker (CW) assigned to the individual. Evidence for interaction of labor market treatments and characteristics of the CW can be found for example in Knaus et al. (2017). They use flexible estimators and a rich administrative dataset with Swiss unemployed to study the effect heterogeneity of job search programs. The study finds no heterogeneity of the effect w.r.t. most of the case worker characteristics. For example, the estimated effects are homogeneous w.r.t. tenure, age and gender of the CW, as well as w.r.t. whether CW and the unemployed have the same gender. These results could be used to make an informed model choice when CW characteristics are unobserved.

An important aspect of the search for separable covariates, highlighted by the above example, is to specify in a first step what is in the error term. Having knowledge about the nature of the error term has been often used in the literature as a source of identifying variation, see Heckman (2008). In the following example, knowledge about the error term together with economic theory inform the choice of the separable covariate.

**Example 4: purchase of durable goods.** This example builds on the study of Boizot et al. (2001). Consider a household that consumes a certain durable product. Let $T$ represent the duration to a purchase and $X_1$ the relative price of that product with respect to a substitute. If $V$ captures unobserved taste for this product, then individuals with high values of $V$ are likely to respond differently to an increase in $X_1$ than individuals with low values of $V$. Now suppose that the product is a regularly purchased product, such as noodles. Boizot et al. (2001) argue that in such cases, the taste can be assumed homogeneous among households. Then $V$ could capture instead (costly) storage space. Assume that rice is the only substitute of noodles. Since packages of noodles and rice have a similar size, the effect of an increase in the relative price of noodles can be safely assumed to have a similar effect on the time to purchase for low and high values of $V$. In addition, it can be assumed that the hazard of $T$ is monotonic in available space.

We now complement assumption A3 with two "competing" assumptions, A4 and A4', on the separable regressor.

**Assumption A4.** (i) The separable regressor $X_1$ from assumption A3 is also separable from the genuine duration dependence, that is

$$\lambda(t, X) = \mu(t, X_2)\phi(X_1, X_2) \tag{12}$$

(ii) For every $x_2 \in \mathbb{X}_2$, the set $\{\phi(x_1, x_2) : x_1 \in \mathbb{X}_1\}$ contains a non-empty open subset of $(0, \infty)$.

Under assumptions A3 and A4 (i), model (1) can be written as

$$\theta(t|X, V) = \theta(t|X_1, X_2, V) = \mu(t, X_2)\phi(X_1, X_2)r(X_2, V). \tag{13}$$

For a fixed $X_2 = x_2$, (13) reduces to an MPH model. (13) can be therefore interpreted as a generalized MPH model. Part (ii) requires that there is sufficient variation in the separable regressor $X_1$. This is a modification of a standard assumption in the MPH context, see e.g. assumption 3 in Elbers and Ridder (1982) and assumption 6b in Van den Berg (2001). The variation in $\phi$ is required for every element $x_2$. This requires that there is variation in $X_1$ for each value of $X_2$. In addition, A4 (ii) implies that (at least one element of) $X_1$ is continuous. A4 (ii) can be replaced with the following weaker assumption: the set $\mathbb{X}_1$ contains at least two elements $x_1, x_1'$ such that for every $x_2 \in \mathbb{X}_2$ it holds $\phi(x_1, x_2) \neq \phi(x_1', x_2)$. The elements $x_1, x_1'$ can depend on $x_2$. See also the remark in the proof of proposition 2.

The alternative assumption A4' requires the separable covariate $X_1$ to be time-varying, i.e. the value of $X_1$ depends on time, and $X_1 = (X_1(t))_{t \in \mathbb{R}_+}$ is a stochastic process. $X_2$ and $V$ are as before time-constant random variables. $x_1$ denotes a path of $X_1$ and is a deterministic function of time.

**Assumption A4'.** (i) $X_1$ is a predictable process. (ii) The hazard of $T$ at each $t$ depends only on the value $X_1(t)$. (iii) For each $x_2$, there are two paths of $X_1$, $z_1, z_2$ with $z_1(t) = z_2(t)$ for all $t$ in some fixed open interval $(t_0, t_1)$ and $S\{t_0 \mid z_1, x_2\} \neq S\{t_0 \mid z_2, x_2\}$, where $S(t \mid x_1, x_2)$ is the observed survival function.

The predictability of $X_1$ is commonly invoked in proportional models, see e.g. Kalbfleisch and Prentice (1980) and the discussion in section 4.2. in Van den Berg (2001). The value of $X_1(t)$ must be known just before $t$. This precludes anticipation by the individual which is not observed by the econometrician. A typical example is when an individual expects a child, which might affect the current job search. If this is not observed by the econometrician, then the process is not predictable. Assumption A2 (ii) now means that $V \perp\!\!\!\perp (X_1(t), X_2)$ for each $t$. Together with predictability, this implies that "only depends on past and outside random variation", see Van den Berg (2001) p. 3399 as well as Andersen et al. (1996). "Outside variation" means that $X_1$ is not determined within the model. A4' (iii) generalizes Condition 1 in Brinch (2007). Here, both the paths $z_1$ and $z_2$, as well as the interval $(t_0, t_1)$, are allowed to depend on $x_2$. A4' (iii) excludes processes that are constant over time (such as initial endowments), or do not change value over individuals (such calendar time), or are deterministic function of time (such as years of experience or age, measured throughout the spell). These requirements make it possible to distinguish between the genuine duration dependence and the time-varying effect of $X_1$. In addition, processes $X_1$ that are fully determined by the value of $X_2$ are also excluded. With A4', the hazard can be written as

$$\theta(t \mid X, V) = \theta(t \mid X_1(t), X_2, V) = \lambda(t, X_1(t), X_2) r(X_2, V). \tag{14}$$

There are two important implications of (14). First, at some elapsed duration $t_0$, past variation of $X_1$ (i.e. variation at $t < t_0$) impacts the hazard only through dynamic selection, that is, through the distribution of $r(x_2, V)$ at $t_0$. This insight provides a source of identifying variation. Second, information about future values $X_1(t)$, $t > t_0$ are allowed to impact the hazard at $t_0$ only through $X_1(t_0)$. This can be viewed as a no-anticipation assumption, see Abbring and Van Den Berg (2003b), although strictly speaking anticipation is allowed as long as it is symmetric between the individual and the econometrician.

**Example 5: Job search on the job.** To identify the effect of having children on labor market outcomes, Lundborg et al. (2017) use exogenous variation in the number of children caused by the randomness of an In Vitro Fertilization (IVF) process. In their setting, let $T$ be the duration until a new job is found while searching on the job and $X_1(t)$ the number of children. It holds $X_1(t) \perp\!\!\!\perp V$ since the IVF process is idiosyncratic and not dependent on factors influencing the job market history. In addition, it is plausible to assume that only current values of $X_1$ impact the search outcome. Past number of children impact the current search outcome only through the current number of children. Moreover, both predictability and no anticipation can be defended on the grounds that the IVF outcome is equally uncertain for the individual and the econometrician. The intentions of the individual to have additional children are known to the econometrician through inflow into the IVF register.

In addition to assumption A4 (A4'), we need the following regularities and normalization assumptions.

**Assumption A5.** The function $\lambda$ obtains only nonnegative values. For each $t \in [0, \infty)$ and each $x \in \mathbb{X}$, $\Lambda(t, x) := \int_0^t \lambda(w, x) dw$ exists and is finite.

**Assumption A6.** For every $x_2 \in \mathbb{X}_2$, there is a known $x_1^* = x_1^*(x_2) \in \mathbb{X}_1$ and a known $t^* = t^*(x_2) \in [0, \infty)$ such that $\phi(x_1^*, x_2) = 1$ and $\Lambda(t^*, x_2) = 1$.

**Assumption A6'.** For every $x_2 \in \mathbb{X}_2$, there is a known $t^* = t^*(x_2)$ and $x_1^* = x_1^*(x_2)$, such that $\lambda(t^*, x_1^*(t^*), x_2) = 1$.

**Assumption A7.** For every $x \in \mathbb{X}$, the random variable $r(x, V)$ has a finite mean.

Assumption A5 is an innocuous regularity assumption. It generalizes a standard assumption in mixture hazard models, see e.g. assumption 2 in Elbers and Ridder (1982). In the context of assumption A4, it requires $\mu$ to be integrable for each $x_2 \in \mathbb{X}_2$. In the context of assumption A4', it requires the integral $\int_0^t \lambda(s, x_1(s), x_2)dt$ to exist for every finite positive $t$.

A6 and A6' are scale normalization assumptions needed under A4 and A4', respectively. Consider first A6. For each $X_2 = x_2$, the conditional model is a standard MPH model. A6 reduces to a standard normalization assumption, see e.g. assumption 7 in Van den Berg (2001). Without A6, $\mu$ and $\phi$ would be identified only up to a scale for each value $x_2$. Depending on the context, the researcher might be willing to choose $t^*, x_1^*$ independently of $x_2$, for example $t^* = x^* = 0$. If A6' is dropped, the class of observationally equivalent structures $(\Lambda, r)$ can be described in the following way. For a fixed $x_2$, denote by $\Lambda_{x_2}$ and $V_{x_2}$ the expressions $\Lambda(., ., x_2)$ and $r(x_2, V)$, respectively. Further, let $G_{x_2}$ and $\mathscr{L}_{x_2}$ be the distribution of $V_{x_2}$ and the corresponding Laplace transform. For any $c > 0$, it holds

$$S(t \mid x_1, x_2) = L_{x_2}(\Lambda_{x_2}(t, x_1)) = \mathscr{L}_{x_2}(\frac{1}{c}(c\Lambda_{x_2}(t, x_1))).$$

Therefore, for any value $x_2$, the strata MH models $(\Lambda_{x_2}, L_{x_2})$ and $(\tilde{\Lambda}_{x_2}, \tilde{\mathscr{L}}_{x_2})$ are observationally equivalent if there exist a constant $c$, such that $\tilde{\Lambda}_{x_2} = c\Lambda_{x_2}$ and $\tilde{\mathscr{L}}_{x_2}(s) = \mathscr{L}_{x_2}(\frac{1}{c}s)$ for every $s \in [0, \infty)$. Assumption A6' normalizes the strata MH models corresponding to different values $x_2$.

Finally, A7 is a nontestable normalization assumption. It can be replaced by the assumption of Heckman and Singer (1984) on the tail of the distribution of $r(x, V)$. See Ridder (1990) for an extensive discussion. In some cases, A7 can be justified with economic theory, see the discussion in section 5.5 of Van den Berg (2001).

With these assumptions, we can state the following result.

**Proposition 2** *Under assumptions A1-A3, A4, A5, A6 and A7, the GMH model* (1) *is identified.*

**Proposition 3** *Under assumptions A1-A3, A4', A5, A6' and A7, the GMH model* (1) *is identified.*

While detailed proofs are provided in the appendix, let us give some intuition on these results. For a given $x_2$, define $\theta_{x_2}(t|x_1, v_{x_2}) := \theta(t|x_1, x_2, v)$ and $\lambda_{x_2}$ analogously. Each of the assumptions A4 and A4' ensure that for each $x_2$, the corresponding $x_2$-strata MH model $\theta_{x_2}(t|x_1, v_{x_2}) = \lambda_{x_2}(t|x_1)v_{x_2}$ is identified (that is, that the pair $(\lambda_{x_2}, G_{x_2})$ is identified). A4 uses the separability of $x_1$ as in the MPH model, while A4' the time variation of the covariates as in Brinch (2007). The function $r$ is identified over quantiles of the identified distributions $(G_{x_2})_{x_2 \in \mathbb{X}_2}$. This intuition leads naturally to the following result:

**Proposition 4** *Let assumptions A1-A3 hold. Then the GMH model* (1) *is identified with single spells if and only if for each $x_2 \in \mathbb{X}_2$ the corresponding single-spell $x_2$-strata MH model structure $(\lambda_{x_2}, G_{x_2})$ is identified.*

Thus, any strategy to identify the single-spell MH model can be used to identify the GMH model (1) under assumptions A1-A3. We presented two such strategies, A4 and A4'. A third one is the time-varying covariates strategy developed in Ruf and Wolter (2019). Their approach relies on martingale properties of a correctly specified MH model.

**Remark 1 (Overidentifying restrictions)** *Melino and Sueyoshi (1990) derive overidentifying restrictions for the MPH model. These restrictions are helpful, as they can be used to test the model. It is easy to see that their result caries over to the GMH model under assumptions A3-A4, since for each value $x_2$, the GMH model reduces to a MPH model. The overidentification result of Melino and Sueyoshi (1990), however, does not carry over to the model under restrictions A3-A4', because $\lambda$ is not multiplicative in $X_1$. In the next section, we derive a novel property for the GMH model, which can be used to derive a test-statistics.*

## 3.3 Nonparametric identification in multiple-spells models

Let $T_1$ and $T_2$ be two duration variables. In this subsection, we study identification of the following bivariate model:

$$\theta_1(t|X,V) = \lambda_1(t,X)r(X,V) \tag{15}$$

$$\theta_2(t|x,T_1=t_1,v) = \lambda_2(t,X)\psi(t_1,X)r(x,V) \tag{16}$$

The hazard model for the first duration is the GMH model (1). The hazard model for the second duration is an augmented GMH model which allows $T_1$ to have an effect on the hazard of $T_2$. This effect may depend on $X$ in an arbitrary way. We discuss two distinct setups of interest which differ w.r.t to the a priori assumption on $\psi$. In both setups, both $T_1$ and $T_2$ are assumed to be fully observed.

**Setup 1.** Suppose that $T_1, T_2$ describe the random length of two spells of the same individual. The first spell is finished before the begin of the second spell. As an example, $T_1, T_2$ might represent the durations of two (consecutive) unemployment spells. Alternatively, $T_1$ might be the length of the last employment spell and $T_2$ the length of the current unemployment spell. The spells are thus not required to be of the same type. The notation implies that $X$ and $V$ realize prior to the begin of the first spell. The function $\psi$ captures the so called lagged duration effect. Because of the consecutive character of the spells, $T_1$ is fully known to the individual (and the potential employer) throughout the second spell and is thus fully "anticipated". As a result, $T_1$ has an impact on $\theta_2$ right from the beginning. This model will be contrasted to the model in section 4.2 where $T_1$ arrives as a surprise *during* the spell of $T_2$. Notably, $\lambda_1, \lambda_2$ are allowed to be different. The generalized error $r(x,v)$, on the contrary, is restricted to be the same in both spells. Thus, given the dependence of $r$ on $X$, model (15), (16) can be interpreted as a hazard version of a dynamic fixed-effects panel data model (the analogy is not entirely correct though, since here $T_1$ and $T_2$ could be outcomes of different types).

Identification of lagged duration models has been first considered by Heckman and Borjas (1980), Honoré (1993), and more recently by Horny and Picchio (2010) and Picchio (2012). Recent empirical studies of lagged duration dependence can be found in Doiron and Gørgens (2008), Cockx and Picchio (2013), Dorsett and Lucchino (2018), among others.

**Setup 2.** Suppose now that $T_1, T_2$ describe duration variables of two distinct individuals. The individuals are assumed to share (observed and) unobserved characteristics. Such a context may arise when $T_1, T_2$ describe duration outcomes of twins, of employees in the same firm, or individuals from some common background, see Hougaard (2000) for an overview. Importantly, since the spells are assumed "parallel", i.e. one does not require sequential realizations, it must be also assumed that there are no cross-effects, $\psi = 1$. In section 4.2, we relax this assumption. To distinguish between the two setups, following the conventions in the literature, we refer to them as lagged duration dependence model (setup 1) and shared frailty model (setup 2).

**Remark 2 (A third setup)** *The following combination of the two setups is also considered. As in Setup 1, let $T_1, T_2$ represent two unemployment spells of the same individual. The experimental study of Eriksson and Rooth (2014) suggests that in certain cases employment experienced after $T_1$ might offset the lagged duration dependence between $T_1$ and $T_2$. In such cases, $\psi = 1$ can be assumed. Similarly, Doiron and Gørgens (2008) find the length of that past employment spells $T_1$ does not matter for subsequent unemployment spells $T_2$, as long as one conditions on the dummy variable "being employed". In our model, this can be done automatically by the sample choice when only individuals with $T_1 > 0$ are considered.*

Consider the following assumptions.

**Assumption A5".** (i) For each $j = 1, 2$, the function $\lambda_j$ obtains nonnegative values. $\lambda_j(t, x)$ is continuous in $t$ for each $x \in \mathbb{X}$. $\Lambda_j(t, x)$ is finite for each $x$. (ii) The function $\psi$ takes only positive values and is differentiable in $t$ for each $x \in \mathbb{X}$.

**Assumption A6".** For each $x \in \mathbb{X}$ and $j = 1, 2$ there exists known $t_j^* = t_j^*(x) \in \mathbb{R}^+$ such that (i.) $\lambda_1(t_1^*, x), \Lambda_2(t_2^*, x)$ are known and (ii) $\psi(t_1^*, x), \partial_t \psi(t_1^*, x)$ are also known, where $\partial_t$ denotes the derivative w.r.t. $t$..

Assumption A5" (i) is slightly stronger than A5. It can be relaxed by requiring that $t$ can be varied such that $\Lambda_2(t, x)$ obtains all values in an open interval. A5" (ii) requires smoothness of $\psi$. This is a mild assumption. In particular, consider the case in which the lagged duration dependence (that is, $T_1$) is treated as any other covariate in the hazard function of $T_2$. Then, under the standard MPH specification $\lambda(t, x) = h(t) \exp x\beta$, assumption A5" (ii) is fulfilled. A6" (i) has a role equivalent to those of A6 and A6'. Assumption A6" (ii) is a scale normalization assumption similar to the one made in Honoré (1993). A convenient normalization is $\psi(0, x) = 1$ for all $x$, for which case no normalization of $\Lambda_2$ is needed. Note that A6" (ii) requires the same $t^*$ for both $\lambda_1$ and $\psi$. This somewhat stronger assumption can be relaxed under additional assumptions on the distribution of $V$, see section B in the appendix.

We can now state the main result of this section.

**Proposition 5** *(i) Under assumptions A1, A5" and A6", the functions $\lambda_1, \lambda_2, \psi$ are identified. (ii) If in addition A2 is satisfied, then also the function $r$ is identified.*

We now discuss the key aspects of proposition 5 and relate them to the literature. First, the only other study allows for interaction in observed and unobserved factors is Evdokimov (2010). His model, however, rules out lagged duration dependence.

Second, part (i) of proposition 5 states that independence of $X$ and $V$ is not necessary for identifying $\lambda_1, \lambda_2, \psi$. This result is of particular importance since most studies focus either on

Table 1: Comparison of assumptions in the literature

| Paper | (a) | (b) | (c) | (d) | (e) | (f) | (g) |
|---|---|---|---|---|---|---|---|
| Evdokimov (2010) | Yes | Yes | Yes | No | No | No | Yes |
| Model 1, Honoré (1993) | No | Yes | Yes | No | No | Yes | No |
| Model 3, Honoré (1993) | No | No | No | Yes | No | No | No |
| Picchio (2012) | No | Yes | Yes | Yes | Yes | Yes | Yes |
| This paper | Yes | Yes | Yes | Yes | Yes | Yes | No |

Each column represents one assumption. (a) Interaction of $X$ and $V$ allowed. (b) Independence of $X$ and $V$ not needed to identify model structure (apart from $r$). (c) No finite mean assumption on the unobservables needed. (d) Model allows for lagged duration dependence. (e) Lagged duration dependence depends on $x$. (f) Fixed-effects assumption needed. (g) More than two spells needed.

the genuine duration dependence or on scarring effects, and not on the effect of a particular covariate on the hazard. In addition, independence is often hard to defend. Similarly to our result, independence is not required in Evdokimov (2010), in the shared frailty model (model 1) in Honoré (1993), and in Picchio (2012). However, model 1 in Honoré (1993) does not allowed for lagged duration dependence, the approach in Evdokimov (2010) requires at least three spells per individual, and Picchio (2012) assumes that there are recurrent data for each outcome (i.e. at least two observations for each $T_i$, $i = 1, 2$).

Third, we do not impose a finite mean assumption on $V_x$. The lagged duration model (model 3) in Honoré (1993) needs a finite mean assumption, while Evdokimov (2010), Picchio (2012) and model 1 in Honoré (1993) do not.

Fourth, as in Picchio (2012), we allow the lagged duration dependence $\psi$ to depend in an arbitrary way on $X$. A brief comparison to proposition 2 is due. If we treat $t_1$ in model (16) as a regular covariate, then we obtain a model equivalent to the model under assumption A4. In particular, $\psi = \phi$ and $\mu = \lambda$. The underlying sources of identification in both models are however fundamentally different. In the single-spell model, separable variation triggered by $\phi$ is used to identify $\mathcal{L}_G$ and eventually $\mu$. In the multiple-spell model, identification of $\psi$ and variation in $t_1$ are not necessary for identification of $\lambda$.

Fifth, the price for the generality of result 5 is the mild fixed-effect assumption $r_1 = r_2 = r, V_1 = V_2 = V$. Similar assumptions are adopted in Picchio (2012) and in model 1 in Honoré (1993), while Evdokimov (2010) allows the function $r$ to vary across periods.

Table 1 provides a summary of this discussion.

**Remark 3 (Alternative proof)** *We provide a second identification proof in the appendix that requires independence for the identification of $\lambda_1, \lambda_2, \psi$, see the remark at the end of the proof of proposition 5 in the appendix. The identification proof switches the order of identification: $r$ is identified before $\lambda_1, \lambda_2, \psi$ are identified. This distinction is important when independence can be credibly maintained and when $r$ is not treated as a nuisance parameter but as a parameter of interest.*

**Towards testing the multiple-spell model.** We now introduce a general property of the multiple-spell nonseparable model that can be used in graphical model diagnostics and is potentially testable. For $t_1, t_2 \in \mathbb{R}^+$, define

$$\rho(t_1, t_2, x) := \frac{\partial_{t_1} S(t_1, t_2|x)}{\partial_{t_2} S(t_1, t_2|x)}, \tag{17}$$

where $S$ is the joint survival function

$$S(t_1, t_2|x) := \mathbf{P}(T_1 > t_1, T_2 > t_2|x). \tag{18}$$

$\rho$ is nonparametrically identified under mild regularity conditions (see e.g. Kalbfleisch and Prentice (1980)), provided the denominator is not zero. It holds the following proposition.

**Proposition 6** *Suppose that for $t_1, t_2, y_1, y_2 \in \mathbb{R}^+$ the quantities $\rho(t_1, t_2, x), \rho(y_1, y_2, x), \rho(t_1, y_2, x)$ and $\rho(y_1, t_2, x)$ are well defined. In addition, suppose that the hazards of $T_1$ and $T_2$ satisfy the model assumptions* (15) *and* (16)*, respectively. Then:*

*(i) if $\psi = 1$, it holds*

$$\rho(t_1, t_2, x)\rho(y_1, y_2, x) - \rho(t_1, y_2, x)\rho(y_1, t_2, x) = 0 \ a.e \ ., \tag{19}$$

*(ii) if $\psi$ is allowed to vary with $t$ and $x$, then property* (19) *holds only if either (a) $\Lambda_2(t, x)$ does not depend on $t$ given $x$, i.e. $\Lambda_2(t, x) = f(x)$ where $f$ is a function of $x$, or (b) $\lambda_1(t, x) = g(x)$ and $\psi'(t, x) = k(x)$, where $g$ and $k$ are functions of $x$, or (c) $\psi$ and $\lambda_1$ are proportional in $t$ and $x$ and the parts depending on $t$ are equal, that is*

$$\lambda_1(t, x) = \eta(t)\phi_1(x) \tag{20}$$

$$\psi'(t, x) = \eta(t)\phi_2(x), \tag{21}$$

*where $\eta$ is common and $\phi_1$ and $\phi_2$ might be different.*

The choice of $t_1, t_2, y_1, y_2$ is arbitrary as long as all quantities are well-defined. Thus, proposition (6) states a property that can be used to test the model under both setups 1 and 2. Deriving the explicit test statistic and its properties is not in the scope of this paper.

# 4 Applications to competing risks and treatment effects models

## 4.1 Identification in nonseparable competing risks models

An individual might consider several destinations of the transition out of unemployment. As an example, elderly unemployed might choose to search for a new job or to withdraw from the labor force, Kyyrä and Ollikainen (2008). In addition, the researcher might want to distinguish between full-time employment, part-time employment, employment on a short-term/long-term contract, and other forms of employment, see e.g. Portugal and Addison (2008). These destinations are typically modeled as competing risks. The main feature of a competing-risks model is that the duration under each risk is latent and only the duration until the first cause of exit is observed. Without further structure, the competing risks model is nonparametrically nonidentified, Tsiatis (1975). We now apply the identification strategies discussed in the previous section to achieve identification of the GMH competing risks model.

**Single-spell setting.** Consider a single-spell setting with only two risks, $A$ and $B$. A generalization to a case with more than two risks is straightforward. Denote by $T_i$ the latent duration under risk $i = A, B$. For each individual in the sample, the researcher observes $\tilde{T} := \min\{T_A, T_B\}$ and the indicator function $\mathbb{1}\{T_A < T_B\}$ which informs about the cause of exit. Thus, instead of the joint survival function $P\{T_A > t_A, T_B > t_B\}$, the econometrician "knows" the crude survival functions $P\{T_i > t, T_j > T_i\}, i, j = A, B, i \neq j$. Denote by $V_A$ and $V_B$ the risk-specific unobserved characteristics that impact $T_A$ and $T_B$, respectively. We assume that $(X, V_i)$ fully determine the distribution of $T_i$ and that $T_A$ and $T_B$ are conditionally independent given $(X, V_A, V_B)$. $V_A$ and $V_B$ are allowed to be dependent, while $X \perp\!\!\!\perp (V_A, V_B)$. This is the standard setup in competing risks models with covariates, see Abbring and Van Den Berg (2003a), Heckman and Honoré (1989), Horny and Picchio (2010). We assume the bivariate GMH model

$$\theta_A(t|X, V_A, V_B) = \theta_A(t|X, V_A) = \lambda_A(t, X) r_A(X, V_A) \tag{22}$$
$$\theta_B(t|X, V_A, V_B) = \theta_B(t|X, V_B) = \lambda_B(t, X) r_B(X, V_B). \tag{23}$$

$r_i$ and $\lambda_i$ are allowed to be risk-specific. We can now state the following result.

**Proposition 7** *Suppose that $V_A, V_B$ fulfill assumptions A2 and A7, and let each of the hazards $\theta_A, \theta_B$ satisfy assumptions A1, A3-A6, where A4 (ii) is replaced by the condition*

***A4 (ii)':*** *for each $x_2 \in \mathbb{X}_2$, the set $\{(\phi_A(x, x_2), \phi_B(x, x_2)) : x \in \mathbb{X}_1\}$ contains a nonempty open subset of $\mathbb{R}^+ \times \mathbb{R}^+$.*

*Then, $\mu_i, \theta_i, r_i, i = A, B$ are identified from the data.*

Proposition 7 is a generalization of proposition 2 in Abbring and Van Den Berg (2003a). Note that identification of $\phi_A, \phi_B$ does not require independence of $X$ and $V_A, V_B$ and the variation condition A4 (ii)'.

**Multiple-spells setting.** Consider now a setting with competing risks and multiple spells. In particular, for $i = A, B$, we assume the model

$$\theta_{i1}(t|X, V_i) = \lambda_{i1}(t, X) r_i(X, V_i) \tag{24}$$
$$\theta_{i2}(t|X, T_{i1} = t^*_{i1}, V_i) = \lambda_{i2}(t, X) \psi_i(t^*_{i1}, X) r_i(X, V_i). \tag{25}$$

The researcher observes $(X, \tilde{T}_k = \min\{T_{Ak}, T_{Bk}\}, \mathbb{1}\{T_{Ak} < T_{Bk}\}), k = 1, 2$. $(T_{A1}, T_{A2})$ are conditionally independent from $(T_{B1}, T_{B2})$ given $X, V_A, V_B$. However, conditionally on $(X, V_A, V_B)$, $T_{i1}$ may impact $T_{i2}$. We can state the following result.

**Proposition 8** *Assume that both $(T_{A1}, T_{A2}, X, V_A)$ and $(T_{B1}, T_{B2}, X, V_B)$ satisfy the conditions of proposition 5. (i) Then for $i = A, B$, $k = 1, 2$, $\lambda_{ik}$ and $\psi_i$ are identified from the data. (ii) If in addition the following condition holds:*

***A5" (iii):*** *for each $x \in \mathbb{X}$, the set*

$$\{(\Lambda_{A1}(t, x) + \psi_A(t, x)\Lambda_{A2}(\tilde{t}, x), \Lambda_{B1}(t, x) + \psi_B(t, x)\Lambda_{B2}(\tilde{t}, x)) : t, \tilde{t} \in \mathbb{R}^+\}$$

*contains a nonempty open subset of $\mathbb{R}^+ \times \mathbb{R}^+$,*

*then also $r_A, r_B$ are identified.*

A comparison to existing results should provide some intuition. The closest setup to the one considered in proposition 8 is the one studied in Horny and Picchio (2010). They prove identification in a competing risk model with multiple spells using the MPH model structure, thus generalizing the single-risk results by Honoré (1993) (see his proposition 5 and our discussion in section 3.3). In the models of Honoré (1993) and Horny and Picchio (2010), it is crucial to identify the joint distribution of the unobservables in order to distinguish between the lagged duration dependence $\psi_i$ and the genuine duration dependence in the first spell $\lambda_{i1}$. To identify the distribution of the unobservables in those models, separable variation in $x$ is necessary. This separable variation is provided by the structure of the MPH model. In the GMH model studied in this section, we do not dispose of such variation. Instead, we restrict the functions $r_{ik}$ and the unobservables $V_{ik}$ to be the same for both spells, i.e. $r_{i1} = r_{i2}$ and $V_{i1} = V_{i2}$ for each $i = A, B$. This leads to identification of $\lambda_{i1}, \psi_i$. Identification of $r_i$ is achieved *subsequently*. Thus, the fixed-effects assumption is the price for the higher flexibility. Similar source of identifying variation is employed by Abbring and Van Den Berg (2003a) in their multiple-spell model. Our model, however, is substantially richer. We allow $X$ to interact with $V$ and we allow for lagged duration dependence. Both these aspects are not allowed by Abbring and Van Den Berg (2003a). Intuitively, distinguishing between $\lambda_{i1}$ and $\psi_i$ is possible, because the ratio of the subsurvival functions of the two spells depends on $\psi_i$ in a way that varies with the elapsed duration in the second spell, which is not the case for $\lambda_{i1}$.

**Discussion on (non)identification with time-varying covariates.** We showed thus far that both our strategy under assumption A4 and having multiple spells as in section 3.3 lead to identification in the GMH competing risks model. This is not the case for our strategy under assumption A4'. The following discussion should provide the intuition. Identification with risk-specific unobservables requires variation on an open subset of $\mathbb{R}^+ \times \mathbb{R}^+$. This variation is necessary to identify the bivariate Laplace transform of the joint distribution of the unobservables. However, due to the censoring problem, the subsurvival function restricts the variation to depend on a single time argument $t$. Thus, without separable variation in $x$, no identification of the distribution of the unobservables is possible. Hence, the bivariate version of the MH model considered by Brinch (2007) (and in general, of any single-spell MH model with risk-specific unobservables but without separable $x$-variation) is not identified. Since the MH model is nested in the GMH model, the single-spell GMH model is not identified with competing risks under assumption A4'.

## 4.2 Nonparametric identification of treatment effects when the treatment is assigned during the spell

Thus far, we considered effects of variables $X, V$ that realize prior to or at begin of the spell in the state of interest. However, a factor that impacts the hazard might materialize *during* the individual spell. The effect of such variable (in the following referred to as treatment) would typically depend on the precise *timing*, i.e. on the elapsed duration until the exposure begins and on the duration of the exposure.

**Example 6: Job Creation Schemes (JCS)** Bergemann et al. (2017) study the effect of JCS on job search outcomes in East Germany shortly after the German reunification.

The typical JCS consisted of temporary work opportunities for the unemployed in the public and nonprofit sector. Bergemann et al. (2017) find that the elapsed unemployment duration until program matters: the longer it takes until the program starts, the more negative are the associated lock-in effects of the JCS. In addition, the effect varies with the length of the exposure, with negative effects in the first 11 months followed by positive but insignificant effect for later months.

Formally, let the r.v. $S_i$ denote the time since entry into the state of interest (e.g. unemployment) until exposure to a treatment for individual $i$ begins. $S_i$ is a duration variable that might be censored by the duration in the state of interest $T$. In the example above, $S$ is the random time until an individual starts temporary work as part of a JCS, and $T$ is the duration of unemployment. Denote the hazard of $S$ by $\theta_S$ and define the vector $V = (V_T, V_S)$, where $V_T, V_S$ are two unobserved scalar r.v.. We consider the following bivariate duration model:

$$\theta_T(t|S = s, X = x, V) = \begin{cases} \lambda_T(t, x) r_T(x, V_T) & \text{if } t < s \\ \lambda_T(t, x)\delta(t, s, x) r_T(x, V_T) & \text{if } t \geq s. \end{cases} \tag{26}$$

$$\theta_S(t|X = x, V) = \lambda_S(t, x) r_S(x, V_S). \tag{27}$$

Equation (27) specifies the hazard of the treatment variable $S$. This is the nonseparble GMH model (1) studied in the previous sections. It depends on $V$ only through the $S$-specific unobservables $V_S$. Equation (26) specifies the hazard of $T$. It is an augmented GMH model. Before the treatment has been assigned $(t < s)$, this is simply the GMH model (1). After the begin of the exposure, the hazard function is modified by the component $\delta$, which can be interpreted as a treatment effect. Conditionally on $X$ and $V$, $S$ can impact $T$ only through $\delta$. The effect $\delta$ is allowed to depend on the elapsed time $s$ until exposure begins and on the elapsed exposure time that is inferred from $(t, s)$. $\delta$ is also allowed to depend on observed heterogeneity $x$. Since for $(t < s)$ there is no treatment effect, model (26) can be interpreted as satisfying the "No Anticipation" assumption: the treatment is not anticipated by individuals or individuals cannot act upon their information about its timing, see Abbring and Van Den Berg (2003b) for a discussion. $V_S$ and $V_T$ are allowed to be dependent, so that the model allows for endogenous selection into the treatment.

Model (26), (27) is a generalization of the widely-used bivariate MPH model considered in the seminal paper of Abbring and Van Den Berg (2003b) (see their model 1A on page 1503). In particular, unlike Abbring and Van Den Berg (2003b), I allow for (i) interaction of observed covariates $X$ and genuine duration dependence $t$ through the functions $\lambda_j$ and (ii) interaction of observed and unobserved heterogeneity, $X$ and $V$, through the functions $r_j$, $j = T, S$.

Note the difference to the lagged duration model (15), (16). The treatment effect $\delta$ appears only for values $t > s$, while the lagged duration effect $\psi$ appears at any elapsed duration of the second duration variable. This difference reflects the conceptual difference in the timing of the two models. While in the treatment effect model the treatment arrives during the spell of unemployment, in the lagged duration model, the treatment has realized prior to the begin of the spell of unemployment.

The main result of this section is stated in the following proposition.

**Proposition 9** *Assume that the functions* $\lambda_j, r_j, j = S, T$ *and the joint distribution of* $r_T(X, V_T)$ *and* $r_S(X, V_S)$ *are known. Then, the treatment effect* $\delta$ *is identified from the data.*

The proof of this proposition is novel and does not follow the steps of the proof in Abbring and Van Den Berg (2003b). The strategy for identifying the full model, however, does lean on the insights in Abbring and Van Den Berg (2003b) and consists of two steps. In a first step, consider model (26), (27) for values $t < s$. For such values, the model represents a GMH competing risks model. Under the assumptions from section 4.1, we can identify this model, i.e. we can identify $\lambda_j, r_j, j = T, S$ and the joint distribution of $r_T(X, V_T), r_S(X, V_S)$. In a second step, we can identify the treatment effect $\delta$ due to proposition (9). The following proposition states identification of the full model.

**Proposition 10** *Suppose that for* $t < s$*, the bivariate GMH competing-risks model* (26), (27) *is characterized by the assumptions of either proposition 7 or proposition 8. Then,* $\lambda_j, r_j, j = S, T$*, the joint distribution of* $r_T(X, V_T)$ *and* $r_S(X, V_S)$*, as well as* $\delta$ *are identified from the data.*

# 5 Discussion

In this paper, we provided identification results for nonseparable duration models. Our models are nonseparable in two ways. First, genuine duration dependence is allowed to depend on observed covariates. Second, observed and unobserved characteristics may interact in an arbitrary way. The models considered in the paper constitute a comprehensive set of settings considered in theoretical and applied duration studies. In particular, we showed identification in single-spell models with and without time-varying covariates, in multiple models with shared frailty and lagged duration dependence, in single-spell and multiple-spell competing risks models, and in treatment effects models where treatment is assigned during the individual spell in the state of interest. A natural follow-up of our results would be to additionally allow unobserved heterogeneity to interact with duration dependence. This remains a question for future research.

# A Appendix: Relation to existing models

In this subsection, we study the relation of the GMH model to the following alternative models:

1. The accelerated failure time (AFT) model imposes the nonparametric regression relation
$$\ln T = -\ln \theta_0(X) + \ln T_0, \tag{28}$$
with $T_0$ being some unobserved baseline duration variable with an unspecified distribution and $\theta_0$ an unknown function, see Kalbfleisch and Prentice (1980).

2. The generalized AFT (GAFT) model, developed by Ridder (1990) is characterized by the regression

$$\ln \int_0^T \psi(s)ds = -\ln \theta_0(X) + U, \tag{29}$$

where $U$ is unobserved stochastic variation and $\psi$ is a positive unknown function. This model nests the AFT ($\psi = 1$) and the MPH ($U = -\ln V + \ln \int_0^T \theta(s \mid x, v)ds$) models. It was developed because of the observation, that although the MPH has a more general regression form than the AFT, it does not nest it because of implicit restrictions imposed on the generalized error term $U$.[4] The GAFT model is closely related to the transformation models studied in Horowitz (1996),

$$\Lambda(T) = X\beta + U, \tag{30}$$

with $\Lambda$ an unknown (monotone) transformation.

3. The extended GAFT (EGAFT) model, studied by Brinch (2011) imposes

$$P\{T \geq t \mid X = x\} = \mathbb{T}(\Lambda(t, x)), \tag{31}$$

where $\mathbb{T}$ is any positive, strictly decreasing, continuously differentiable function with $\mathbb{T}(0)$ and $\Lambda(t, x) = \int_0^t \lambda(s, x)ds$. The EGAFT model nests the GAFT and the MH models.

For the analysis of the relations between the GMH and these models, it is useful to write the GMH in a regression form with transformed duration. Denote by $\varepsilon$ the transformed duration $\ln \int_0^T \theta(s \mid x, v)ds$. As noted above, it has a type-I extreme-value distribution. Moreover, $\varepsilon$ is independent of $(X, V)$ per construction. The GMH model can be written as

$$\ln \Lambda(T, X) = -\ln r(X, V) + \varepsilon. \tag{32}$$

Our first observation is that the AFT model is not nested in the GMH model. The line of reasoning is analogous to the result in Ridder (1990). In particular, assume that $\ln T_0$ in (28) has a standard normal distribution. Then model (32) cannot be written in the form (28), precisely due to the distribution of $\varepsilon$. In obvious notation, with AFT $\subset$ GAFT $\subset$ EGAFT, we obtain that the GAFT, EGAFT models are also not nested in the GMH model.

Second, the GMH model is not nested in the EGAFT model. To see this, note that under model (1) we can write for the observed survival function

$$\begin{aligned} P\{T \geq t \mid X = x\} &= \int_{\Omega_V} P\{T \geq t \mid X = x, V = v\} dG_V(v) = \\ &= \int_{\Omega_V} (1 - \exp\{-\Lambda(t, x) r(x, v)\}) dG_V(v). \end{aligned} \tag{33}$$

If for some $x_1 \neq x_2$, the values $\Lambda(t, x_1)$ and $\Lambda(t, x_2)$ are equal, then the values of (33) at $(t, x_1)$ and $(t, x_2)$ will in general be different due to the presence of $r(x, v)$ (unless a very specific form of $r$ is chosen). Thus, in the context of the GMH model, there exists no map $\mathbb{T} : \mathbb{R}_+ \to \mathbb{R}_+$ which satisfies the conditions of the EGAFT model.

[4] In the MPH model, $U$ cannot be a normal distribution because $\ln \int_0^T \theta(s \mid x, v)ds$ is distributed as type-I extreme-value distribution.

# B An alternative identification approach for the multiple spell lagged duration model

In the following, we state an assumption on the distribution of $V$ which helps identify the model without the normalization requirement A6" (ii).

**A8** For each $t_1$ and $x \in \mathbb{X}$, it holds

$$\mathbb{E}[V|T_2 \geq t, T_1 = t_1, X = x] = \mathbb{E}[V|T_1 \geq t, X = x] \tag{34}$$

A8 contains two parts. First, the conditional distribution of $V$ given $T_2$ and $X$ does not depend on $X$. $T_1$ impacts the hazard of $T_2$ through $\psi$ though (direct effect). It is excluded from determining the dynamic selection, once $X$ is accounted for. In this sense, $T_1$ can be viewed as an exclusion restriction with respect to the structure determining $G|T_2 \geq t, X$.

Define the difference $\mathcal{D}(t_1, t_2, t*, x) := \ln \theta_2(t_2|x, T_1 = t*, x) - \ln \theta_1(t_1|x)$. $\mathcal{D}$ is nonparametrically identified from the data. It holds the following proposition.

**Proposition 11** *(i) Under assumptions A1, A5" (ii) and A8, it holds:*

$$\psi(t', x) = \exp\{\mathcal{D}(t, t, t', x) + w(x)\}, \tag{35}$$

*where $t$ is arbitrary and $w : \mathbb{X} \to \mathbb{R}$ is a function of $x$.*

*(ii) Furthermore, if for any $x$ there is a $\tilde{t} = \tilde{t}(x)$ such that $\psi(\tilde{t}, x)$ is known, then under A1, A5" and A6" (i), the functions $\lambda_1, \lambda_2, \psi$ are identified.*

*(iii) If in addition to the assumptions in (ii) also assumption A2 holds, then also $r$ is identified.*

A discussion of (11) is due. Result (i) is closely related to the first-difference approach in panel data. Under A8 the fixed effect (the averaged $r(x, V)$) cancels out. Result (i) also implies that the model is overidentified. In particular,

$$\partial_t \ln \psi(t', x) = \partial_{t'} \ln \mathcal{D}(t, t, t', x) \tag{36}$$

does not depend on $t$. This property can be useful for testing the hypothesis

$$H_0 : \theta_2(t|x, t_1, v) = \lambda(t, x)\psi(t, x)r(x, V).$$

Consider namely the more general model

$$\theta_1(t|x, v) = \lambda_1(t, x)r(x, V) \tag{37}$$

$$\theta_2(t|x, t_1, v) = \lambda_2(t, x, t_1)r(x, V) \tag{38}$$

**Corollary 1** *Within the class of models described by equations* (37) *and* (38)*, under A8, $H_0$ is true iff $\partial_{t'} \ln \mathcal{D}(t, t, t', x)$ does not depend on $t$.*

# C Proofs

## C.1 Proofs of results in section 3.1

**Proof of lemma 1.** Assume that $r(x, V)$ does not depend on $x$, $r(x, V) = r(V)$ and set $\tilde{V} = r(V)$. The model transforms into the standard MH model,

$$\theta(t \mid X, \tilde{V}) = \lambda(t, X).\tilde{V}, \tag{39}$$

where now $\tilde{V}$ has an unknown distribution $G$. Let $S(t \mid X)$ be the (observed) survival function of $T$ given $X$, $S(t \mid X) = P\{T \geq t \mid X\}$, and $\mathcal{L}_G$ the Laplace transform of $\tilde{V}$, defined as

$$\mathcal{L}_G(s) = \int_0^\infty e^{-sv} dG(v).$$

Define $\Lambda(t, x) = \int_0^t \lambda(s, x) ds$. Further, let $\tilde{G}$ be some other distribution. We obtain

$$S(t \mid X) = \mathcal{L}_G(\Lambda(t, X)) = \mathcal{L}_{\tilde{G}}(\mathcal{L}_{\tilde{G}}^{-1}(\mathcal{L}_G(\Lambda(t, X)))). \tag{40}$$

(40) shows, that $(\Lambda, \mathcal{L}_G)$ and $(\mathcal{L}_{\tilde{G}}^{-1} \circ \mathcal{L}_G \circ \Lambda, \mathcal{L}_{\tilde{G}})$ are observationally equivalent. ■

**Proof of proposition 1.** (i) For $X = x$, set $Y_x = \Lambda(T, x)$ and $V_x = k_x(V) := k(x, V)$. Because $\varepsilon$ is independent of $V$, it is also independent of $V_x$ for each $x$. Moreover, the moment-generating functions $M_{Y_x}$, $M_{V_x}$ and $M_\varepsilon$ of respectively $Y_x$, $V_x$ and $\varepsilon$ satisfy

$$M_{Y_x}(s) = M_{V_x+\varepsilon}(s) = M_{V_x}(s) M_\varepsilon(s) \tag{41}$$

for all admissible $s$. It holds thus $M_{V_x}(s) = M_{Y_x}(s)/M_\varepsilon(s)$, and therefore $M_{V_x}$ is identified (the distributions of both $Y_x$ and $\varepsilon$ are known: the first per assumption and the second per construction). Since a moment-generating function uniquely determines the distribution function, the distribution function $F_{V_x}$ of $V_x$ is identified. Finally, for some $q$ and a fixed $x$, it holds

$$F_{V_x}(q) = P\{V_x \leq q\} = P\{k(x, V) \leq q\} = P\{V \geq k^{-1}(x, q)\} = 1 - k^{-1}(x, q), \tag{42}$$

where the inverse is taken with respect to the second element. The third equality utilizes the strict monotonicity of $k$ and the last equality is due to the uniform distribution of $V$. This identifies $k$ and therefore $r$.

(ii) The proof of $(ii)$ is trivial and follows from $S(t \mid X = x) = \mathcal{L}_{G_{V^x}}(\Lambda(t, x))$, where $G_{V^x}$ is the distribution function of $r(x, V)$. Since $\mathcal{L}_{G_{V^x}}$ is known, varying $t$ identifies $\Lambda(, x)$ for each $x$, see Elbers and Ridder (1982). ■

## C.2 Proofs of results in section 3.2

**Proof of proposition 2.** The proof proceeds in three steps.

**Step 1.** Let $S(t|x) := \mathbb{P}(T > t|x)$ and $\Psi(t, x_2) := \int_0^t \mu(s, x_2) ds$. We have for all $t > 0$ and $x \in \mathbb{X}$

$$S(t|x) = \int_0^\infty \exp(-\phi(x_1, x_2)\Psi(t, x_2) r(x_2, v)) dv, \tag{43}$$

(the conditional density of $V$ is equal to 1 because of assumption A2). Let $x_2 \in \mathbb{X}_2$ be arbitrary element and choose $x_1^*$ such that $\phi(x_1^*, x_2) = 1$ (A6 (i)). It is straightforward to show that for almost any $x \in \mathbb{X}, t \in \mathbb{R}^+$

$$\frac{\partial t S(t|x)}{\partial t S(t|x_1^*, x_2)} = \frac{\phi(x_1, x_2) \int_0^\infty r(x_2, v) \exp(-\phi(x_1, x_2)\Psi(t, x_2) r(x_2, v)) dv}{1 \int_0^\infty r(x_2, v) \exp(-1\Psi(t, x_2) r(x_2, v)) dv}, \tag{44}$$

where the notation $\partial t$ is used for the partial derivative with respect to $t$. Letting $t \to 0$ under A7 we identify the function $\phi$.

**Step 2.** For any $x_2 \in \mathbb{X}_2$, denote by $G_{x_2}$ the distribution of the quantity $r(x_2, V)$ and let $V_{x_2} := r(x_2, V)$. In view of the above definitions we can write

$$S(t|x) = \int_0^\infty \exp(-\phi(x_1, x_2)\Psi(t, x_2) v_{x_2}) dG_{x_2}(v_{x_2}) \tag{45}$$

Let $\mathscr{L}_{x_2}$ denote the Laplace Transfrom of $G_{x_2}$. It holds

$$S(t|x) = S(t|x_1, x_2) = \mathscr{L}_{x_2}(\phi(x_1, x_2)\Psi(t, x_2)) \tag{46}$$

By choosing $t^*$ such that $\Psi(t^*, x_2) = 1$ (A6) and varying $x_1$ on an open interval (A4 (ii)), we can identify the $\mathscr{L}_{x_2}$ and consequently the cumulative distribution function $G_{x_2}$ for all $x_2 \in \mathbb{X}_2$. Then, by making use of (46) the identification of $\Psi$ follows.

**Step 3.** Next, denote by $\tau(x_2, .)$ the inverse map of $r(x_2, .)$ (A1). One can write for each $\omega \in \mathbb{R}_+$

$$\mathbb{P}(r(x_2, V) \le \omega) = G_{x_2}(\omega) \tag{47}$$

The above implies

$$\mathbb{P}(V \le \tau(x_2, \omega)) = G_{x_2}(\omega). \tag{48}$$

Exploiting the fact that $V$ follows the standard uniform distribution (A2) we obtain

$$\tau(x_2, \omega) = G_{x_2}(\omega) \tag{49}$$

The right hand side is known which yields identification of $\tau$ and consequently of the function $r$.

**Remark.** An alternative step 2 would require the following weaker assumption about the variation of $X_1$: the set $\mathbb{X}_1$ contains at least two elements $x_1, x_1'$ such that for every $x_2 \in \mathbb{X}_2$ it holds $\phi(x_1, x_2) \neq \phi(x_1', x_2)$. The elements $x_1, x_1'$ can depend on $x_2$. The proof in step 2 proceeds then as in the appendix of Elbers and Ridder (1982). ∎

**Proof of proposition 3.** Let $x_2$ be a chosen fixed value of $X_2$. Furthermore, let $\mathscr{S}$ be the set of all feasible tuples $(\Lambda_{x_2}, \mathscr{L}_{x_2})$ in the corresponding $x_2$-strata-MH model. Define the relation $\sim$ on $\mathscr{S}$ in the following way: $(\Lambda_{x_2}, \mathscr{L}_{x_2}) \sim (\tilde{\Lambda}_{x_2}, \tilde{\mathscr{L}}_{x_2})$ if there exists a positive number $c$, such that $\tilde{\Lambda}_{x_2} = c\Lambda_{x_2}$ and $\tilde{\mathscr{L}}_{x_2}(s) = \mathscr{L}_{x_2}(\frac{1}{c}s)$ for every $s \in [0, \infty)$. Thus, $\sim$ is an equivalence relation on $\mathscr{S}$ and it induces a partition $\mathscr{P}$ of $\mathscr{S}$. Further, let $\mathscr{F}$ be the set of all possible conditional distributions of $T$ given $X$. Theorem 1 in Brinch (2007) states that under assumptions A1-A3, A4', A5, A6' and A7 the map $\zeta : \mathscr{P} \to \mathscr{F}$ is injective. Under the normalization assumption A6', each equivalence class in $\mathscr{P}$ consists of a single element.

This identifies $\Lambda_{x_2}$ and $G_{x_2}$ for each $x_2$. Following step 3 from the proof of proposition 2, identification of $r$ follows under assumptions A1 and A2. ■

**Proof of proposition 4.** Suppose that for each $x_2$, the $x_2$-strata MH model structure $(\lambda_{x_2}, G_{x_2})$ is identified. This implies that $\lambda$ is identified. Thus, with proposition 1 (i), $r$ is identified too. Alternative proof follows the lines of Step 3 in the proof of proposition 2.

Suppose that for one $\bar{x}_2$, the corresponding $(\lambda_{\bar{x}_2}, G_{\bar{x}_2})$ are not identified. Then there are functions $\tilde{\lambda}_{\bar{x}_2}, \tilde{G}_{\bar{x}_2}$, such that the two $\bar{x}_2$-strata MH models are observationally equivalent. Define $\hat{\lambda} := \lambda$ for all $x_2 \neq \bar{x}_2$ and $\hat{\lambda}(t, x_1, \bar{x}_2) = \tilde{\lambda}_{\bar{x}_2}(t, x_1)$ for all $t, x_1$. Define analogously $\hat{r}$. Then the structures $(\lambda, r)$ and $(\hat{\lambda}, \hat{r})$ are observationally equivalent. ■

## C.3 Proofs of results in section 3.3

**Proof of proposition 5.** The proof proceeds in 3 steps.

**Step 1: identifying** $\Lambda_2$ Define the quantities

$$\Theta_1(t|x, V_1) := \int_0^t \theta_1(\omega|x, V)d\omega.$$

$$\Theta_2(t|x, t_1, V_1) := \int_0^t \theta_1(\omega|x, t_1, V)d\omega.$$

$$S(t_1, t_2|x) = \mathbf{E}_V \left[\exp(-\Theta_1(t_1|x, V) - \Theta_2(t_2|x, t_1, V))\right]. \tag{50}$$

Under the model assumptions (15), (16), we have

$$S(t_1, t_2|x) = \mathbf{E}_V \left[\exp(-\Lambda_1(t_1, x)r(x, V) - \Lambda_2(t_2, x)\psi(t_1, x)r(x, V))\right]. \tag{51}$$

For $t_1, t_2 \in \mathbb{R}^+$, define

$$\rho(t_1, t_2, x) := \frac{\partial_{t_1} S(t_1, t_2|x)}{\partial_{t_2} S(t_1, t_2|x)} \tag{52}$$

$\rho$ is nonparametrically identified under mild regularity conditions (see e.g. Kalbfleisch and Prentice (1980)), provided the denominator is not zero. Simple algebra gives

$$\rho(t_1, t_2, x) = \frac{\mathbf{E}_V \left[(\lambda_1(t_1, x) + \partial_t\psi(t_1, x)\Lambda_2(t_2, x))r(x, V)\exp(-\Theta_1(t_1|x, V) - \Theta_2(t_2|x, t_2, V))\right]}{\mathbf{E}_V \left[\psi(t_1, x)\lambda_2(t_2, x)r(x, V)\exp(-\Theta_1(t_1|x, V) - \Theta_2(t_2|x, t_1, V))\right]}. \tag{53}$$

which simplifies to

$$\rho(t_1, t_2, x) = \frac{\lambda_1(t_1, x) + \partial_t\psi(t_1, x)\Lambda_2(t_2, x)}{\psi(t_1, x)\lambda_2(t_2, x)}. \tag{54}$$

For a given $x$, choose $t_1^*$ according to A6". Denote by $a(x), b(x)$ and $c(x)$ the known values $\lambda_1(t^*(x), x), \partial_t\psi(t^*(x), x)$ and $\psi(t^*(x), x)$, respectively, and by $g(t_2, x), \tilde{g}(t_2, , x)$ the functions $\rho(t_1^*(x), t_2, x)$ and $\frac{c}{b}\rho(t_1^*(x), t_2, x)$. Ignoring the dependence on $x$, (54) can be written as

$$g(t_2) = \frac{a + b\Lambda_2(t_2)}{c\lambda_2(t_2)}. \tag{55}$$

If $b = 0$, then $\lambda_2(t_2) = \frac{a}{cg(t_2)}$. If $b \neq 0$, then by multiplying both sides of (55) by $c/b$, we obtain

$$\tilde{g}(t_2) = \frac{a + b\Lambda_2(t_2)}{b\lambda_2(t_2)}. \tag{56}$$

It follows from (56) that we can write

$$\tilde{g}(t_2) = \frac{\partial ln(a + b\Lambda_2(t_2))}{\partial t}. \tag{57}$$

so that $ln(a+b\Lambda_2(t_2)) = \int_0^{t_2} \tilde{g}(t)dt + d$ for some unknown constant $d$. Using the normalization on $\Lambda_2$ we get the constant $d$ by choosing $t_2$ appropriately. This identifies $\Lambda_2$ and thus $\lambda_2$.

**Step 2: Identifying $\lambda_1$ and $\psi$** Ignoring the dependence on $x$, define $k(t, \tilde{t}) := \rho(t, \tilde{t})\lambda_2(\tilde{t})$. Since $\lambda_2$ is identified, $k$ is a known function. Multiplying (54) on both sides with $\lambda_2(\tilde{t})$, we obtain

$$k(t, \tilde{t}) = \frac{\lambda_1(t) + \partial_t\psi(t)\Lambda_2(\tilde{t})}{\psi(t)}. \tag{58}$$

Letting $\tilde{t} \to 0$ and varying $t$ identifies the ratio $\frac{\lambda_1(t)}{\psi(t)} = k(t, 0)$. Observing that

$$\frac{k(t, \tilde{t}) - k(t, 0)}{\Lambda_2(\tilde{t})} = \frac{\partial_t\psi(t)}{\psi(t)} = \frac{\partial \ln \psi(t)}{\partial t},$$

and using the normalization restrictions, we identify $\psi$. Then, $\lambda_1$ is identified from (79). Analogous steps lead to identification of $\psi$ and $\lambda_1$.

**Step 3: Identifying $r$** For a given $x$, it follows from (51) that

$$S(t_1, t_2|x) = \mathcal{L}_{G_x}(\Lambda_1(t_1, x) + \Lambda_2(t_2, x)\psi(t_1, x)), \tag{59}$$

which identifies $\mathcal{L}_{G_x}$ since the functions $\Lambda_1(t_1, x), \Lambda_2(t_2, x), \psi(t_1, x)$ are identified. Identification of $r(x, V)$ follows the lines of Step 3 in the proof of proposition 2. The proof is complete.

**Remark 4** *Here are alternative steps 2, 3 and 4 that switch the order of identification and use the stronger independence assumption*

***Alternative step 2: identifying*** $r$. *Consider the marginal survival function* $S_2(t|x) := \mathbb{P}(T_2 > t|x)$ *and denote by* $G_x$ *the cumulative distribution function of* $r(x, V)$ *for each* $x \in \mathbb{X}$. *The Laplace Transform of the latter is denoted by* $\mathcal{L}_x$. *It holds*

$$S_2(t|x, t_1) = \mathcal{L}_x(\Lambda_2(t, x)\psi(t_1, x)). \tag{60}$$

*Choosing* $t_1^*$ *according to assumption A6" (ii) and varying* $t$ *on an open interval traces out* $\mathcal{L}_x$ *and consequently* $G_x$ *for any* $x \in \mathbb{X}$. *By using analogous arguments to step 3 in the proof of proposition 2, we identify the function* $r$.

***Alternative step 3: identifying*** $\psi$. *Since* $\mathcal{L}_x$ *and* $\Lambda_2$ *are now identified,* $\psi$ *is now identified from equality* (60).

***Alternative step 4: identifying*** $\lambda_1$. *Using* (54), *we identify* $\lambda_1$. ***Remark.*** *An alternative approach to the identification of* $\lambda_1$, *that does not require identification of* $\psi$, *uses the equality*

$$S_1(t|x) = \mathcal{L}_x(\Lambda_1(t, x)). \tag{61}$$

■

**Proof of proposition 6.** Consider equality (54) in the proof of proposition 5. If $\psi = 1$, then

$$\rho(t_1, t_2, x) = \frac{\lambda_1(t_1, x)}{\lambda_2(t_2, x)}. \tag{62}$$

and the validity of (19) is evident.

Now let $\psi$ be allowed to vary with $t$ and $x$. In the following, we ignore the dependence on $x$. Define

$$D = \rho(t_1, t_2)\rho(y_1, y_2) - \rho(t_1, y_2)\rho(y_1, t_2). \tag{63}$$

We consider solutions of $D = 0$. The denominators of $\rho(t_1, t_2, x)\rho(y_1, y_2)$ and $\rho(t_1, y_2)\rho(y_1, t_2)$ are equal, so setting $D = 0$ cancels them out. Consider the nominator of $D$. Canceling out equal parts and rearranging, we obtain that $D = 0$ is equivalent to

$$(\lambda_1(t_1)\psi'(y_1) - \lambda_1(y_1)\psi'(t_1))(\Lambda_2(y_2) - \Lambda_2(t_2)) = 0. \tag{64}$$

This is trivially fulfilled if $\Lambda_2(t, x) = f(x)$, i.e. if it is constant for each $x$. It is also trivially fulfilled if $\lambda_1(t, x) = g(x)$ and $\psi'(t, x) = k(x)$. If this is not the case, we obtain

$$\frac{\lambda_1(t_1)}{\lambda_1(y_1)} = \frac{\psi'(t_1)}{\psi'(y_1)}, \tag{65}$$

where by $\psi'$ we denote the derivative of $\psi$ w.r.t. $t$. Denote for a fixed $y^*$ the values $\lambda_1(y^*), \psi'(y^*)$ by $a$ and $b$, respectively. Then $\lambda_1(t) = c\psi'(t)$, where $c = a/b$. Thus, the only solution that satisfies (65) consists of the functions

$$\lambda_1(t, x) = \eta(t)\phi_1(x) \tag{66}$$

$$\psi'(t, x) = \eta(t)\phi_2(x), \tag{67}$$

where $\eta$ is common and $\phi_1$ and $\phi_2$ might be different. ■

## C.4 Proofs of results in section 4.1

**Proof of proposition 7.** For $i, j = A, B, j \neq i$, define the subsurvival functions

$$\partial_t Q_i(t|x) = P\{T_i > t, T_j > T_i\}.$$

$Q_i$ is nonparametrically identified from the data. We will use the equality

$$\partial_t Q_i(t|x) = \partial_{t_i} S(t, t) \qquad \text{for } i = A, B, \tag{68}$$

see Tsiatis (1975). Furthermore, it holds

$$\begin{aligned} S(t_A, t_B|x) = \int_{\mathbb{R}^+ \times \mathbb{R}^+} & \exp\{-\Psi_A(t_A, x_2)\phi_A(x_1, x_2)r_A(x_2, v_A) \\ & -\Psi_B(t_B, x_2)\phi_B(x_1, x_2)r_B(x_2, v_B)\} dG_{V_A, V_B}(v_A, v_B). \end{aligned} \tag{69}$$

For a given $x$ and any $v$, define $V_i^x := r_i(x, V_i), G_i^x(v) := P\{V_i^x < v\}, i = A, B$, $G_{v_a,v_b}^x := P\{V_A^x < v_a, V_B^x < v_b\}$ and $\mathcal{L}_x$ the bivariate Laplace transform of $G_{v_a,v_b}^x$. With this notation, (69) can be rewritten as

$$S(t_A, t_B|x) = \mathcal{L}_x(\Psi_A(t_A, x_2)\phi_A(x_1, x_2), \Psi_B(t_B, x_2)\phi_B(x_1, x_2)), \tag{70}$$

and with (68), we obtain

$$\partial_t Q_i(t|x) = \mu_i(t, x_2)\phi_i(x_1, x_2)\mathcal{D}_i\mathcal{L}_x(\Psi_A(t, x_2)\phi_A(x_1, x_2), \Psi_B(t, x_2)\phi_B(x_1, x_2)), \tag{71}$$

where with $\mathcal{D}_i$ the derivative w.r.t. to the i-th component, $i = A, B$. Choose $x_1^* = x_1^*(x_2)$ such that $\phi_A(x_1^*, x_2) = 1$. It follows that

$$\frac{\partial_t Q_A(t|x_1, x_2)}{\partial_t Q_A(t|x_1^*, x_2)} = \tag{72}$$

$$\frac{\phi_A(x_1, x_2)\mathcal{D}_A\mathcal{L}_x(\Psi_A(t, x_2)\phi_A(x_1, x_2), \Psi_B(t, x_2)\phi_B(x_1, x_2))}{\mathcal{D}_A\mathcal{L}_x(\Psi_A(t, x_2), \Psi_B(t, x_2)\phi_B(x_1^*, x_2))}. \tag{73}$$

Letting $t \to 0$ leads to identification of $\phi_A$. The identification of $\phi_B$ follows analogously. Repeating steps 2 and 3 of the proof of proposition 2 for each of the components of the bivariate Laplace $\mathcal{L}_x$ leads to identification of $\mathcal{L}_x$, the joint distribution $G_{v_a,v_b}^x$ and the functions $r_A$ and $r_B$.

Finally, pick an arbitrary $\bar{x}_2 \in \mathbb{X}_2$ and hold it fest. The biivariate GMH model reduces to a bivariate MPH model (that depends on the value $\bar{x}_2$). For this model, it remains to identify $\Psi_A^{\bar{x}_2}(t), \Psi_B^{\bar{x}_2}(t)$ with $\Psi_i^{\bar{x}_2}(t) := \Psi_i(t, \bar{x}_2)$, $i = A, B$. Identification of $\Psi_i^{\bar{x}_2}(t)$ follows from part (c) of the proof of proposition (2 ) in Abbring and Van Den Berg (2003a). Since the choice of $\bar{x}$ is arbitrary, this leads to identification of $\Psi_A, \Psi_B$. The proof is complete. ■

**Proof of proposition 8.** (i) It holds for the conditional survival function

$$S(t_1, t_2, t_3, t_4|x) := P\{T_{A1} > t_1, T_{B1} > t_2, T_{A2} > t_3, T_{B2} > t_4|x\}$$
$$= \mathbb{E}_{V_A,V_B}[\exp -\Theta_{A1}(t_1|x, V_A) - \Theta_{B1}(t_2|x, V_B) - \Theta_{A2}(t_3|x, V_A) - \Theta_{B2}(t_4|x, V_B)].$$

Define the "two-period" subsurvival functions

$$Q_{A1}(t_{A1}; t_3, t_4|x) := P\{T_{A1} > t_{A1}, T_{B1} > T_{A1}, T_{A2} > t_3, T_{B2} > t_4\}$$
$$Q_{A2}(t_{A2}; t_1, t_2|x) := P\{T_{A1} > t_1, T_{B1} > t_2, T_{A2} > t_{A2}, T_{B2} > T_{A2}\}.$$

$Q_{B1}, Q_{B2}$ are defined equivalently. It is trivial to show that

$$S(t, t, t_3, t_4|x) = Q_{A1}(t; t_3, t_4|x) + Q_{B1}(t; t_3, t_4|x) \tag{74}$$
$$S(t_1, t_2, t, t|x) = Q_{A2}(t; t_1, t_2|x) + Q_{B2}(t; t_1, t_2|x) \tag{75}$$

and consequently

$$\partial_t Q_{A1}(t; t_3, t_4|x) = \partial_{t_1} S(t, t, t_3, t_4|x)$$
$$\partial_t Q_{A2}(t; t_1, t_2|x) = \partial_{t_3} S(t_1, t_2, t, t|x),$$

with analogous equalities for $B$. Furthermore, for a given $x$, we use the notation $V_i^x, G_i^x \dots$, etc., introduced in the previous proof. Thus, it holds

$$S(t_1,t_2,t_3,t_4|x) = \mathcal{L}_x\{\Lambda_{A1}(t_1,x)+\psi_A(t_1,x)\Lambda_{A2}(t_3,x), \Lambda_{B1}(t_2,x)+\psi_B(t_2,x)\Lambda_{B2}(t_4,x)\}, \quad (76)$$

and hence

$$\begin{aligned}
&\partial_t Q_{A1}(t;t_3,t_4|x) = (\lambda_{A1}(t,x)+\partial_t\psi_A(t,x)\Lambda_{A2}(t_3,x))\mathcal{D}_A\mathcal{L}_x\{\Lambda_{A1}(t,x)\\
&+\psi_A(t,x)\Lambda_{A2}(t_3,x), \Lambda_{B1}(t,x)+\psi_B(t,x)\Lambda_{B2}(t_4,x)\}\\
&\partial_t Q_{A2}(t;t_1,t_2|x) = \psi_A(t_1,x)\lambda_{A2}(t,x)\mathcal{D}_A\mathcal{L}_x\{\Lambda_{A1}(t_1,x)\\
&+\psi_A(t_1,x)\Lambda_{A2}(t,x), \Lambda_{B1}(t_2,x)+\psi_B(t_2,x)\Lambda_{B2}(t,x)\}
\end{aligned}$$

and as a result

$$\frac{\partial_t Q_{A1}(t;\tilde{t},\tilde{t}|x)}{\partial_t Q_{A2}(\tilde{t};t,t|x)} = \frac{\lambda_{A1}(t,x)+\partial_t\psi_A(t,x)\Lambda_{A2}(\tilde{t},x)}{\psi_A(t,x)\lambda_{A2}(\tilde{t},x)} \quad (77)$$

and from (54) in the proof of proposition 5, we obtain

$$\frac{\partial_t Q_{A1}(t;\tilde{t},\tilde{t}|x)}{\partial_t Q_{A2}(\tilde{t};t,t|x)} = \rho(t,\tilde{t},x). \quad (78)$$

Since $Q_{A1}(t;\tilde{t},\tilde{t}|x)$ and $Q_{A2}(\tilde{t};t,t|x)$ are identified from the data, (78) implies that $\rho$ is also identified. As a result, identification of $\Lambda_{A2}$ can be derived in an equivalent way to identification of $\Lambda_2$ in step 1 in the proof of proposition 5. Identification of $\Lambda_{B2}$ is achieved analogously.

Next, ignoring the dependence on $x$, define $k(t,\tilde{t}) := \frac{\partial_t Q_{A1}(t;\tilde{t},\tilde{t}|x)}{\partial_t Q_{A2}(\tilde{t};t,t|x)}\lambda_{A2}(\tilde{t},x)$. Since $\lambda_{A2}(\tilde{t},x)$ is identified, $k$ is a known function. By multiplying (77) on both sides with $\lambda_{A2}(\tilde{t},x)$, we obtain

$$k(t,\tilde{t}) = \frac{\lambda_{A1}(t)+\partial_t\psi_A(t)\Lambda_{A2}(\tilde{t})}{\psi_A(t)}. \quad (79)$$

Letting $\tilde{t}\to 0$ and varying $t$ identifies the ratio $\frac{\lambda_{A1}(t)}{\psi_A(t)} = k(t,0)$. Observing that

$$\frac{k(t,\tilde{t})-k(t,0)}{\Lambda_{A2}(\tilde{t})} = \frac{\partial_t\psi_A(t)}{\psi_A(t)} = \frac{\partial \ln\psi_A(t)}{\partial t},$$

and using the normalization restrictions, we identify $\psi_A$. Then, $\lambda_{A1}(t,x)$ is identified from (79). Analogous steps lead to identification of $\psi_B$ and $\lambda_{B1}(t,x)$.

(ii) Under condition A5" (iii), $\mathcal{L}_x$ can be identified from

$$S(t,t,\tilde{t},\tilde{t}|x) = \mathcal{L}_x\{\Lambda_{A1}(t,x)+\psi_A(t,x)\Lambda_{A2}(\tilde{t},x), \Lambda_{B1}(t,x)+\psi_B(t,x)\Lambda_{B2}(\tilde{t},x)\}. \quad (80)$$

by varying $t$ and $\tilde{t}$. This leads to identification of $G^x_{v_a,v_b}$. The identification of $r_A, r_B$ follows as in step 3 of the proof of proposition 2. The proof is complete. ■

**Proof of proposition 11.** First, it holds per definition

$$
\begin{aligned}
&\mathcal{D}(t,t,t^*,x) = \ln\theta_2(t|x,T_1 = t*,x) - \ln\theta_1(t|x) \\
&= \ln\lambda_1(t,x) + \ln\psi(t^*,x) + \ln\mathbb{E}[r(x,V)|T_2 \geq t, T_1 = t^*, x] \\
&- \ln\lambda_2(t,x) - \ln\mathbb{E}[r(x,V)|T_1 \geq t, x] \\
&= \ln\psi(t^*,x) + \ln\lambda_1(t,x) - \ln\lambda_2(t,x),
\end{aligned}
$$

where in the last equality we used assumption A8. This proves (i). Thus, $\psi$ is identified up to a scale that depends on $x$. Under the normalization assumption in this proposition, the scale $w(x)$ is identified. Thus, $\psi$ is identified.

Next, consider equality (54) in step 1 of the proof of proposition 5. Since we identified $\psi$, $\partial_t\psi$ is also identified. For each $x$, choose $t_1^*$ such that $\lambda_1(t_1^*,x)$ is known. This leads to expression 55 in the proof of proposition 5, where now the constants $a, b$ and $c$ are identified and hence known. Identifying $\lambda_2$ now follows the same steps as in the proof of proposition 5. The identification of $\lambda_1$ follows from equality (54). This proves (ii). Finally, $r$ follows from (60) analogously as in the proof of proposition 5. ■

## C.5 Proofs of propositions in section 4.2

**Proof of proposition 9.** Define the crude survival functions

$$
\begin{aligned}
&Q_{T,S}(t,s) := P\{T > t, S > s, T > S\} \\
&Q_S(s) := Q_{T,S}(0,s) = P\{S > s, T > S\} \\
&Q_T(t) := P\{T > t, T < S\}.
\end{aligned}
$$

All three quantities are nonparametrically identified from the data: the first two because $T > S$ implies that both $T$ and $S$ are observed and the third one because $T$ is observed. For $t > s$, it holds

$$
\begin{aligned}
&\mathbb{1}\{T > t\}\mathbb{1}\{S > s\} = \mathbb{1}\{T > t\}\mathbb{1}\{T > s\}\mathbb{1}\{S > s\} \\
&= \mathbb{1}\{T > t\}\mathbb{1}\{T > s\}\mathbb{1}\{S > s\}(\mathbb{1}\{S > T\} + \mathbb{1}\{T > S\}) \\
&= \mathbb{1}\{T > t\}\mathbb{1}\{T > s, S > T\} + \mathbb{1}\{T > t\}\mathbb{1}\{S > s, T > S\} \\
&= \mathbb{1}\{T > t, S > T\} + \mathbb{1}\{T > t, S > s, T > S\},
\end{aligned}
$$

so that

$$
P\{T > t, S > s\} = Q_T(t) + Q_{T,S}(t,s) \tag{81}
$$

and hence

$$
\frac{\partial P\{T > t, S > s\}}{\partial t} = Q_T'(t) + \frac{\partial Q_{T,S}(t,s)}{\partial t}. \tag{82}
$$

Equality (82) implies that $\frac{\partial P\{T>t,S>s|x\}}{\partial t}$ is identified. All relationships above hold also conditionally on $x$, which implies that $\frac{\partial P\{T>t,S>s|x\}}{\partial t}$ is also identified. In the following, we omit the dependence on $x$ whenever clear from the context.

Further, define $\Lambda_i(t,x) = \int_0^t \lambda_i(r,x)dr$ for $i = T, S$, $\mathcal{Y}(t|s,x) := \int_s^t \lambda_T(r,x)\delta(r,s,x)dr$, $V_T^x := r_T(x, V_T)$, $V_S^x := r_S(x, V_S)$ and let $g_{V_T^x, V_S^x}$ be the joint density of $V_T^x, V_S^x$. For the observed conditional joint survival function $P\{T > t, S > s|x\}$, we have for all $t > s$

$$
\begin{aligned}
&P\{T > t, S > s|x\} = P\{T > t|S > s|x\}P\{S > s|x\} &(83)\\
&= \int_0^\infty \int_0^\infty \exp\{-(\Lambda_T(s,x) + \mathcal{Y}(t|s,x))v_T^x - \Lambda_S(s,x)v_S^x\} g_{V_T^x,V_S^x}(v_T^x, v_S^x)dv_T^x dv_S^x, &(84)
\end{aligned}
$$

and therefore

$$\frac{\partial P\{T > t, S > s|x\}}{\partial t} = -\frac{\partial \mathcal{Y}(t|s,x)}{\partial t} D^T \mathcal{L}_{V_T^x,V_S^x}(\Lambda_T(s,x) + \mathcal{Y}(t|s,x), \Lambda_S(s,x)), \tag{85}$$

which is equivalent to

$$-\frac{\frac{\partial P\{T>t,S>s|x\}}{\partial t}}{D^T \mathcal{L}_{V_T^x,V_S^x}(\Lambda_T(s,x) + \mathcal{Y}(t|s,x), \Lambda_S(s,x))} = \frac{\partial \mathcal{Y}(t|s,x)}{\partial t} \tag{86}$$

whenever the denominator is different than 0.

In the rest part of the proof, we show that $\mathcal{Y}(t|s,x)$ is a unique solution to a differential equation. Consider $x$ and $s$ as fix and define the function

$$g(t,y) = -\frac{\frac{\partial P\{T>t,S>s\}}{\partial t}}{D^T \mathcal{L}_{V_T^x,V_S^x}(c_1 + y, c_2)} \tag{87}$$

where we omit $x$ and $s$ for simplicity of notation. If $c_1$ and $c_2$ are two known constants, then the function $g$ is fully known. $\mathcal{L}_{V_T^x,V_S^x}$ has derivatives of all orders which implies that for any $t$ $g(t,y)$ is Lipschitz continuous w.r.t to $y$ on any compact set where it is defined. Using standard theory of differential equations (see e.g. chapter 12.3 in Fitzpatrick (2009)), it follows that the differential equation

$$g(t, \mathcal{Y}(t)) = \frac{\partial \mathcal{Y}(t)}{\partial t} \tag{88}$$

has a unique solution $\mathcal{Y}(t)$ (note that $\mathcal{Y}(0) = 0$). Setting $c_1 = \Lambda_T(s,x)$ and $c_2 = \Lambda_S(s,x)$ completes the proof. ■

# 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(3):701–710.

Abbring, J. H. and Van Den Berg, G. J. (2003b). The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517.

Abrevaya, J. and Hausman, J. A. (1999). Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells. *Annals of Economics and Statistics*, (55-56):243–275.

Aït-Sahalia, Y. (2007). *Estimating Continuous-Time Models with Discretely Sampled Data*, volume 3 of *Econometric Society Monographs*, page 261327. Cambridge University Press.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1996). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer New York.

Bergemann, A., Pohlan, L., and Uhlendorff, A. (2017). The impact of participation in job creation schemes in turbulent times. *Labour Economics*, 47:182 – 201. EALE conference issue 2016.

Bergström, R. and Edin, P.-A. (1992). Time aggregation and the distributional shape of unemployment duration. *Journal of Applied Econometrics*, 7(1):5–30.

Boizot, C., Robin, J., and Visser, M. (2001). The demand for food products: An analysis of interpurchase times and purchased quantities. *The Economic Journal*, 111(470):391–419.

Bound, J., Brown, C., Duncan, G. J., and Rodgers, W. L. (1989). Measurement error in cross-sectional and longitudinal labor market surveys: Results from two validation studies. Working Paper 2884, National Bureau of Economic Research.

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

Brinch, C. N. (2011). Non-parametric identification of the mixed proportional hazards model with interval-censored durations. *The Econometrics Journal*, 14(2):343–350.

Caliendo, M., Cobb-Clark, D. A., and Uhlendorff, A. (2015). Locus of control and job search strategies. *The Review of Economics and Statistics*, 97(1):88–103.

Chesher, A. (2002). Semiparametric identification in duration models. CeMMAP working paper CWP20/02, Centre for Microdata Methods and Practice, London, UK.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5):1405–1441.

Chesher, A. (2007). *Identification of Nonadditive Structural Functions*, volume 3 of *Econometric Society Monographs*, page 116. Cambridge University Press.

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

Cockx, B. and Picchio, M. (2013). Scarring effects of remaining unemployed for long-term unemployed school-leavers. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 176(4):951–980.

Doiron, D. and Gørgens, T. (2008). State dependence in youth labor market experiences, and the evaluation of policy interventions. *Journal of Econometrics*, 145(1):81 – 97. The use of econometrics in informing public policy makers.

Dorsett, R. and Lucchino, P. (2018). Young people's labour market transitions: The role of early experiences. *Labour Economics*, 54:29 – 46.

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

Eriksson, S. and Rooth, D.-O. (2014). Do employers use unemployment as a sorting criterion when hiring? evidence from a field experiment. *The American Economic Review*, 104(3):1014–1039.

Evdokimov, K. (2010). Nonparametric identication of a nonlinear panel model with application to duration analysis with multiple spells. Discussion Paper.

Farber, H. S., Silverman, D., and von Wachter, T. (2016). Determinants of Callbacks to Job Applications: An Audit Study. *American Economic Review*, 106(5):314–318.

Fitzpatrick, P. (2009). *Advanced calculus*, volume 5. American Mathematical Soc.

Frijters, P. (2002). The non-parametric identification of lagged duration dependence. *Economics Letters*, 75(3):289 – 292.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, 51(2):186–192.

Heckman, J. and Borjas, G. (1980). Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica*, 47(187):247–83.

Heckman, J. J. (1991). Identifying the hand of the past: distinguishing state dependence from heterogeneity. *The American Economic Review*, 81:75–79.

Heckman, J. J. (2008). Econometric causality. *International Statistical Review*, 76(1):1–27.

Heckman, J. J. and Honoré, B. E. (1989). The identifiability of the competing risks model. *Biometrika*, 76(2):325–330.

Heckman, J. J. and Singer, B. (1984). The identifiability of the proportional hazard model. *Review of economic studies*, 51:231–241.

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

Horny, G. and Picchio, M. (2010). Identification of lagged duration dependence in multiple-spell competing risks models. *Economics Letters*, 106(3):241 – 243.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64(1):103–137.

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

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer, Heidelberg.

Kalbfleisch, J. and Prentice, R. (1980). *The statistical analysis of failure time data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.

Knaus, M., Lechner, M., and Strittmatter, A. (2017). Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach. *arXiv e-prints*, page arXiv:1709.10279.

Kroft, K., Lange, F., and Notowidigdo, M. J. (2013). Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment. *The Quarterly Journal of Economics*, 128(3):1123–1167.

Kyyrä, T. and Ollikainen, V. (2008). To search or not to search? The effects of UI benefit extension for the older unemployed. *Journal of Public Economics*, 92(10-11):2048–2070.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47:939–956.

Lancaster, T. (1985). Simultaneous equations models in applied search theory. *Journal of Econometrics*, 28(1):113 – 126.

Lancaster, T. (1990). *The econometric analysis of transition data.* Econometric Society Monographs. Cambridge University Press, Cambridge.

Lewbel, A. (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica*, 66(1):105–121.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903.

Lockwood, B. (1991). Information externalities in the labour market and the duration of unemployment. *The Review of Economic Studies*, 58(4):733–753.

Lundborg, P., Plug, E., and Rasmussen, A. W. (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, 107(6):1611–37.

Mathiowetz, N. A. and Duncan, G. J. (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *Journal of Business & Economic Statistics*, 6(2):221–229.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5):1339–1375.

Matzkin, R. L. (2007). Chapter 73 nonparametric identification. volume 6 of *Handbook of Econometrics*, pages 5307 – 5368. Elsevier.

McCall, B. P. (1994). Identifying state dependence in duration models. american statistical association 1994. *Proceedings of the Business and Economics Section*, 14.

Melino, A. and Sueyoshi, G. T. (1990). A simple approach to the identifiability of the proportional hazards model. *Economics Letters*, 33(1):63 – 68.

Picchio, M. (2012). Lagged duration dependence in mixed proportional hazard models. *Economics Letters*, 115(1):108 – 110.

Portugal, P. and Addison, J. T. (2008). Six Ways To Leave Unemployment. *Scottish Journal of Political Economy*, 55(4):393–419.

Ridder, G. (1986). The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence. Mimeo, University of Amsterdam.

Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *The Review of Economic Studies*, 57(2):167–181.

Ruf, J. and Wolter, J. L. (2019). Nonparametric identification of the mixed hazard model using martingale-based moments. *Econometric Theory*, page 116.

Tsiatis, G. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. 72*, 72(1):20–22.

Van den Berg, G. J. (2001). Duration models: specification, identification, and multiple durations. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5, chapter 55, pages 3381–3460. Elsevier.

Van den Berg, G. J. and Van der Klaauw, B. (2006). Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment. *International Economic Review*, 47(3):895–936.

Weber, S. (2014). Human capital depreciation and education level. *International Journal of Manpower*, 35(5):613–642.