Johannes Ruf and James Lewis Wolter

Nonparametric identification of the mixed hazard model using martingale-based moments

Article (Accepted version)
(Refereed)

NONPARAMETRIC IDENTIFICATION OF THE MIXED HAZARD MODEL USING MARTINGALE-BASED MOMENTS

JOHANNES RUF AND JAMES LEWIS WOLTER

Abstract. Nonparametric identification of the Mixed Hazard model is shown. The setup allows for covariates that are random, time-varying, satisfy a rich path structure and are censored by events. For each set of model parameters, an observed process is constructed. The process corresponding to the true model parameters is a martingale, the ones corresponding to incorrect model parameters are not. The unique martingale structure yields a family of moment conditions that only the true parameters can satisfy. These moments identify the model and suggest a GMM estimation approach. The moments do not require use of the hazard function.

JEL Classifications: C01, C14, C34, C41.

1. Introduction

An important aspect of economic durations is the influence of unobserved heterogeneity. When relevant variables are unobserved, the result is survivorship bias. Failure to account for this leads to biased estimates and faulty inference. This issue is well known and a large econometric literature addresses it.

The archetype economic example is unemployment durations; see, for example, Cornelißen and Hübler (2011), Caliendo et al. (2013) Hausman and Woutersen (2014), and

Johannes Ruf, Department of Mathematics, London School of Economics and Political Science, Columbia House, Houghton St, London WC2A 2AE, UK.

James Wolter, Lord, Abbett & Co. LLC, 90 Hudson Street, Jersey City, NJ 07302, US.

E-mail address: j.ruf@lse.ac.uk, jwolter@lordabbett.com.

Date: December 25, 2018.

Key words and phrases. Identification, Heterogeneity, Hazard Models, Nonparametrics, Martingales.

Farber and Valletta (2015). All of these papers incorporate either unobserved heterogeneity, time-varying covariates or both. This matches the model features considered in this paper. Kroft et al. (2013) analyses unemployment durations with an experimental approach explicitly designed to avoid the problem of unobserved heterogeneity.

A large portion of the literature studying heterogeneity in economic durations considers the Mixed Proportional Hazard (MPH) model. Initially proposed by Lancaster (1979), it has been the topic of substantial econometric analysis. This includes both identification results and estimation approaches. Lancaster (1979) proposed a parametric model. The literature that followed questions this assumption. Many papers extend the MPH to semiparametric or nonparametric cases with corresponding identification and estimation results. See Hausman and Woutersen (2014) for a guide to the literature.

In this paper, we consider identification of the Mixed Hazard (MH) model and its variants. The MH model removes the MPH's proportionality assumption between the baseline hazard and the covariates. The heterogeneity still acts multiplicatively on the other components.

Heckman (1991) identifies part of the MH model, the hazard rate at $t = 0$. This is done assuming time-constant covariates. McCall (1994) identifies the full model with time-varying covariates by assuming the hazard rate is zero (or arbitrarily close to zero) over an interval $[0, t)$. This has to hold for some, but not all, realizations of the covariates. Both of these approaches identify parameters by considering limits approaching portions of the random events which are not influenced by the sorting phenomena. At time $t = 0$ there is no sorting because no one has left the sample. When the hazard rate is zero on $[0, t)$ there is also no sorting for the same reason. In an important theoretical advance, Brinch (2007) identifies the full MH model with time-varying covariates avoiding arguments such as in Heckman (1991) and McCall (1994).

None of the above papers presents an estimation approach. Moreover, while McCall (1994) and Brinch (2007) identify the full model, their results raise a number of issues. First, the arguments are based on the survival probabilities past a given time t conditional on a path of the covariates up to time t. When the covariates $Z = (Z_s)$ are random and time-varying, this object is only directly observed from the data if the random events τ do not censor our observation of Z. If censoring is present, we cannot observe these objects directly. Hence, previous results need to be modified for this case.

Second, the argument depends on the survival probabilities conditioning on a fixed path z of the covariates. In realistic data scenarios, observations will have random covariate paths which do not necessarily match across observations. Thus, applying previous results requires that survival probabilities conditional on z can be constructed from the data. Certain model specifications can be ruled out by particular paths z. An estimation approach requires a way of systematically dealing with observed random paths to rule out incorrect models. This would connect observed random paths with identification based on fixed paths, the result being a consistency proof.

In this paper, we show nonparametric identification of the MH model when covariates are time-varying and censored by events. This is done avoiding the use of hazard functions. Moments which identify the model are constructed based on the conditional survival probability form. These are adjusted for censoring so that only observables are required. The moments also show how to identify the model allowing for random covariate paths. This connects identification with fixed paths to a data based approach that can be used in realistic finite sample estimation.

It is first shown that rich enough (random) time-variation in the covariates leads to identification of the conditional survival processes. These objects are only partially observed because events censor observation of the covariates. This initial result is then used to define a class of fully observed processes indexed by the potential parameters of the model. We then argue that only the parameters generating the data make their corresponding process a martingale. This unique martingale property is used to find certain moment conditions that only the true parameters can satisfy. This set of moment conditions suggests a GMM estimation approach.

The proposed identification approach is general and applicable to a large class of MH models. This includes models where events can happen with positive probability at specific times. For example, at known fixed times the event can happen with positive probability. Alternatively, the event can happen with positive probability when other random times occur. In addition to identifying the standard MH model, we also show identification of these more complicated models using this paper's approach. As a result, the same moment conditions give identification.

The remainder of the paper is organized as follows. Section 2 presents the basic Mixed Hazard setup and assumptions and provides preliminary results on conditional survival probabilities. Section 3 shows the general identification approach and identification for a standard Mixed Hazard model. The appendices[1] present several proofs and technical results. They also contain counterexamples showing the importance of our assumptions.

The notation in this article follows standard notation for continuous-time stochastic processes. For an excellent treatment, see Jacod and Shiryaev (2003).

2. Heterogeneous Hazard Setup and Preliminary Results on Conditional Survival Probabilities

We focus on the single-spell case throughout.

All random variables are defined on a fixed completed probability space $(\Omega, \mathscr{H}, \mathsf{P})$. This probability space supports a stochastic d–dimensional covariate process[2] Z, where $d \in \mathbb{N}$. It also supports a family $\mathcal{V}$ of scalar nonnegative random variables, all independent of Z. To avoid trivialities, we assume that the deterministic random variable $V = 0$ is not in $\mathcal{V}$. The probability space also supports a random time τ with support $[0, \infty]$.

For a random variable $V \in \mathcal{V}$, we denote its Laplace transform by

$$\mathcal{L}_V(t) = \int \mathrm{e}^{-vt} \mathrm{d}\mathsf{P}_V(v), \qquad t \geq 0.$$

2.1. Mixed Hazard Setup.

The following statement is the usual definition of a (mixed) hazard model with time-varying covariates Z. There exists an appropriate function α such

[1]Appendices B, C, and D can be found in an online appendix.

[2]The notation Z is used to represent the stochastic process $(Z_t)_{t\in[0,\infty)}$. We use Z regardless of whether or not Z varies through time. This is in keeping with standard notation in the stochastic process literature. This notational convention is used for all stochastic processes throughout the sequel.

that

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathsf{P}[t < \tau \leq t + \Delta t \mid \{\tau > t\}, V, Z_s \colon s \in [0,t]] = V\alpha(t, Z_t). \tag{2.1}$$

The setup of (2.1) is a special case of the more general model described in Subsection 2.2.

More precisely, we shall formulate conditions on Z and α (but no further conditions on $\mathcal{V}$) which allow for identification. We start by stating the following condition concerning the covariate process Z.

Assumption (P)**.** The covariate process Z takes values in some open set $\mathcal{Z} \subset \mathbb{R}^d$. Moreover, one of the following three conditions holds.

(1) The covariate process Z is piecewise constant with update times described by positive Poisson arrival rates, uniformly bounded away from zero. At each event a new value is drawn from a distribution with support $\mathcal{Z}$.

(2) The process Z is a diffusion; that is,

$$\mathrm{d}Z_t = \mu(t, Z_t)\mathrm{d}t + \sigma(t, Z_t)\mathrm{d}W_t, \qquad t \geq 0,$$

where W is d–dimensional Brownian motion, the drift $\mu : [0,\infty) \times \mathcal{Z} \to \mathbb{R}^d$ and the dispersion matrix $\sigma : [0,\infty) \times \mathcal{Z} \to \mathbb{R}^{d\times d}$ are locally Lipschitz-continuous, and $\sigma(t,z)$ is invertible for all $(t,z) \in [0,\infty) \times \mathcal{Z}$.

(3) We have $d = 1$ and the process Z is a continuous, regular strong Markov process.

□

The covariate structure in Assumption (P) is assumed mostly for convenience. Any covariates with a similarly rich path structure will also lead to identification. Next, we state the condition on the process α.

Assumption (A)**.** With $\mathcal{Z} \subset \mathbb{R}^d$ denoting the range of Z as in Assumption (P), assume that the function $\alpha : [0,\infty) \times \mathcal{Z} \to [0,\infty)$ is continuous and satisfies

$$\sup_{z \in \mathcal{Z}} \int_0^t \alpha(s,z)\mathrm{d}s < \infty, \qquad t \geq 0.$$

Moreover, assume that there exist $t_1, t_2 > 0$ with $t_1 < t_2$ such that $\alpha(t,z) > 0$ for all $(t,z) \in [t_1,t_2] \times \mathcal{Z}$, and that $\alpha(t_1,\cdot)$ is not constant.

Let $\mathcal{A}$ denote all these processes that satisfy the condition of Assumption (A).

We now make the representation in (2.1) more precise. To this end, we define a filtration $(\mathscr{F}_t)$ by

$$\mathscr{F}_t = \sigma\{Z_s \colon s \in [0,t]\}.$$

In the Mixed Hazard setup, we shall then work under the assumption that there exists a pair $(V^\circ, \alpha^\circ) \in \mathcal{V} \times \mathcal{A}$ such that the random time τ satisfies

$$\mathsf{P}[\tau > t \mid \mathscr{F}_t \vee \sigma(V^\circ)] = \mathrm{e}^{-V^\circ \int_0^t \alpha^\circ(s,Z_s)\mathrm{d}s}, \qquad t \geq 0, \tag{2.2}$$

where $\mathscr{F}_t \vee \sigma(V^\circ)$ is the smallest sigma algebra that contains $\mathscr{F}_t$ and makes V° measurable.[3] The representation in (2.2) forces the hazard rates to have multiplicative unobservables V°, an almost universal assumption in the literature.

As can be seen from (2.2), there are two components which determine the random times τ. These are the unobserved factor $V \in \mathcal{V}$ and the function $\alpha \in \mathcal{A}$. A general statement of the identification problem is: if $(V^\circ, \alpha^\circ) \in \mathcal{V} \times \mathcal{A}$, can we tell which members generated the observed data? Identification is of functions $\alpha \in \mathcal{A}$ and distributions for $V \in \mathcal{V}$.

From (2.2), it is clear that it usually is impossible to completely identify V and α from the data. Indeed, multiplying V by a constant and α by its reciprocal yields exactly the same distribution of the data. Hence, we can only expect identification up to a constant. To emphasize this point we shall write that (V,F) equals to (W,G) modolo

[3]Assume, for example, that there exists a pair $(V^\circ, \alpha^\circ) \in \mathcal{V} \times \mathcal{A}$ and an exponentially distributed random variable ζ, independent of $\mathscr{F}_\infty \vee \sigma(V^\circ)$ such that the representation

$$\tau = \inf\left\{t \geq 0 : V^\circ \int_0^t \alpha^\circ(s,Z_s)\mathrm{d}s \geq \zeta\right\}$$

holds. We then have

$$\mathsf{P}[\tau > t | \mathscr{F}_t \vee \sigma(V^\circ)] = \mathsf{P}\left[V^\circ \int_0^t \alpha^\circ(s,Z_s)\mathrm{d}s < \zeta \,\middle|\, \mathscr{F}_t \vee \sigma(V^\circ)\right] = \mathrm{e}^{-V^\circ \int_0^t \alpha^\circ(s,Z_s)\mathrm{d}s}, \qquad t \geq 0;$$

hence, (2.2) follows.

"trivial normalization" if there exists a constant $\kappa > 0$ such that

$$(\mathcal{L}_V, \alpha) = \left(\mathcal{L}_{\kappa W}, \frac{1}{\kappa}\beta\right).$$

Corollary 2.1. *Suppose Assumptions (P) and (A) hold. Let $(V, \alpha), (W, \beta) \in \mathcal{V} \times \mathcal{A}$. Then we have the equivalence*

$$(\mathcal{L}_V, \alpha) = \left(\mathcal{L}_{\kappa W}, \frac{1}{\kappa}\beta\right) \text{ for some } \kappa > 0$$

if and only if

$$\mathsf{P}\left[\mathcal{L}_V\left(\int_0^t \alpha(s, Z_s)\mathrm{d}s\right) = \mathcal{L}_W\left(\int_0^t \beta(s, Z_s)\mathrm{d}s\right)\right] = 1.$$

The corollary follows from Theorem 2.2 in Subsection 2.2 below; for details see Proposition 2.3 and the discussion that follows.

This result says that, if the pairs $(\mathcal{L}_V, \alpha)$ and $(\mathcal{L}_W, \beta)$ are not equal modolo trivial normalization, then the conditional survival processes $\mathcal{L}_V(F)$ and $\mathcal{L}_W(G)$ are not equal (in a pathwise sense with positive probability).

2.2. **A More General Setup.** Note that each $\alpha \in \mathcal{A}$ defines an $(\mathscr{F}_t)$–predictable process

$$F_t(\alpha) = \int_0^t \alpha(s, Z_s)\mathrm{d}s, \qquad t \geq 0.$$

and that (2.2) can be written as

$$\mathsf{P}[\tau > t \mid \mathscr{F}_t \vee \sigma(V^\circ)] = \mathrm{e}^{-V^\circ F_t(\alpha^\circ)}, \qquad t \geq 0. \tag{2.3}$$

It turns out that identification can occur not only in the family $\{F(\alpha) \colon \alpha \in \mathcal{A}\}$ but among a much larger family of processes. Those processes serve as the integrated observed part of the hazard rate. They do not need to be absolutely continuous and can even be discontinuous.

To describe the general model, we let $\mathcal{F}$ be an arbitrary family of stochastic processes F satisfying the following conditions:

(1) F is $(\mathscr{F}_t)$–predictable;
(2) F is nondecreasing;
(3) F is right-continuous;
(4) $F_0 = 0$.

For example, we could use

$$\mathcal{F} = \{F(\alpha) \colon \alpha \in \mathcal{A}\}. \tag{2.4}$$

Then the general setup corresponds exactly to the Mixed Hazard setup of Subsection 2.1.

In general, the class $\mathcal{F}$ can contain many other processes. In particular, $\mathcal{F}$ can allow for discontinuities in F. This creates times where τ will happen with positive probability. For example, at a known fixed time there can be a positive probability of an event. The more general setup also allows for a positive probability of τ happening when another random time happens. See Subsection 2.3 for an example.

Similarly to (2.2) and (2.3), in this generalized setup, we assume the existence of a pair $(V^\circ, F^\circ) \in \mathcal{V} \times \mathcal{F}$ such that the random time τ satisfies

$$\mathsf{P}[\tau > t \mid \mathscr{F}_t \vee \sigma(V^\circ)] = \mathrm{e}^{-V^\circ F^\circ_t}, \qquad t \geq 0. \tag{2.5}$$

There is no hope to get a general identification result without further assumptions on $\mathcal{F}$. Indeed, Appendix A contains a specific counterexample. In the following, we formulate an assumption that will yield identification. It is stated in generality to allow for several different applications.

To give intuition, we provide some interpretation of the following necessary notation. Fix two right-continuous finite-valued paths $f, g : [0, \infty) \to \mathbb{R}$. These paths can be interpreted as realizations of elements in $\mathcal{F}$. Next, fix $t, \varepsilon \in [0, \infty]$. The functions f and g will be paired with realizations of elements in $\mathcal{F}$ over the interval $[t, t + \varepsilon)$. To this end, fix $c, c' \in \mathbb{R}$. These two constants describe how different the realizations of F and G need to be compared with f and g, respectively. Finally, fix $n \in \mathbb{N}$. We are now ready to define the set

$$O_{f,g,t,c,c',\varepsilon,n} = \Big\{\widetilde{\omega} \in \Omega : \sup_{s\in[t,t+\varepsilon)} \big(|c + f_s - F_s| + |c' + g_s - G_s|\big) < \frac{1}{n}\Big\}.$$

This notation allows us to introduce the core assumption of this paper.

Assumption (R)**.** For each pair $(F, G) \in \mathcal{F} \times \mathcal{F}$ there exist a $[0,\infty]$–valued random variable ρ, a $(0,\infty)$–valued random variable $\mathcal{E}$, and an $\mathbb{R}^2 \setminus \{(0,0)\}$–valued random vector (C, C') such that the intersection of the following two events has positive probability:

$$\begin{aligned} A_1 &= \{F_\rho < F_s, \quad \text{for each } s \in (\rho, \rho+\mathcal{E})\}; \\ A_2 &= \bigcap_{n\in\mathbb{N}} \left\{\mathsf{P}\left[O_{f,g,t,c,c',\varepsilon,n}\right]\big|_{f=F,g=G,t=\rho,c=C,c'=C',\varepsilon=\mathcal{E}} > 0\right\}. \end{aligned} \tag{2.6}$$

That is, we have $\mathsf{P}[A_1 \cap A_2] > 0$.[4] □

Let us provide an interpretation of Assumption (R). To this end, fix a pair $(F, G) \in \mathcal{F} \times \mathcal{F}$. For each $n \in \mathbb{N}$, any $\omega \in \Omega$ can be matched to the event

$$O_{F(\omega),G(\omega),\rho(\omega),C(\omega),C'(\omega),\mathcal{E}(\omega),n}.$$

This is the set of all $\widetilde{\omega} \in \Omega$ such that $F_s(\widetilde{\omega})$ and $G_s(\widetilde{\omega})$ do not differ too much (measured in terms of n) on the interval $s \in [\rho(\omega), \rho(\omega) + \mathcal{E}(\omega))$ from $(C(\omega) + F_s(\omega))$ and $(C'(\omega) + G_s(\omega))$, respectively. If the probability of the matched event $O_{f,g,t,c,c',\varepsilon,n}$ is greater than zero for all choices of n, we include ω in A_2. Observe that (C, C') is assumed to be $\mathbb{R}^2 \setminus \{(0,0)\}$–valued. Thus, for sufficiently large $n \in \mathbb{N}$, a state $\omega \in \Omega$ cannot be matched to an event that contains ω itself.

For two pairs

$$(V, F), (W, G) \in \mathcal{V} \times \mathcal{F},$$

write $(\mathcal{L}_V, F) = (\mathcal{L}_W, G)$ if $\mathcal{L}_V = \mathcal{L}_W$ and $\mathsf{P}[F = G] > 0$. We are now ready to state the general identification result.

[4]Several measurability results are required for this assumption to be well stated. The details are provided in Appendix C.

Theorem 2.2. *Suppose Assumption (R) holds.*[5] *Let* $(V, F), (W, G) \in \mathcal{V} \times \mathcal{F}$. *Then we have the equivalence*

$$(\mathcal{L}_V, F) = \left(\mathcal{L}_{\kappa W}, \frac{1}{\kappa} G\right) \text{ for some } \kappa > 0$$

if and only if

$$\mathsf{P}[\mathcal{L}_V(F) = \mathcal{L}_W(G)] = 1.$$

The proof of Theorem 2.2 is contained in Subsection 2.2.

This result says that the pairs $(\mathcal{L}_V, F)$ and $(\mathcal{L}_W, G)$ are equal modulo trivial normalization if the conditional survival processes $\mathcal{L}_V(F)$ and $\mathcal{L}_W(G)$ are equal (in a pathwise sense with positive probability). In Appendix A, a counterexample is presented where the conclusion of Theorem 2.2 fails if Assumption (R) fails. It may be possible to simplify or relax assumption (R), but it is also possible to lose the desired result when doing so.

The following proposition yields Corollary 2.1, provided we show that Theorem 2.2 holds.

Proposition 2.3. *If Assumptions (P) and (A) hold along with* (2.5)*; i.e., if we are in the Mixed Hazard setup, then Assumption (R) is satisfied.*

The proof of this proposition can be found in Appendix B. This proposition, in conjunction with Thereom 2.2, then yields Corollary 2.1. To see this, note that under Assumption (P), any two functions $\alpha_1, \alpha_2 \in \mathcal{A}$ with $\alpha_1 \neq \alpha_2$ yield corresponding processes $F(\alpha_1)$ and $F(\alpha_2)$ that have different paths with positive probability. This can be argued in the same way as Proposition 2.3 itself. Hence identification of the process F yields also identification of the function α.

Before finishing this subsection, let us give some intuition how Theorem 2.2 is proven, in the spirit of Brinch (2007). Suppose that we are given four right-continuous functions

[5]Assumption (R) could be weakened and the results below would still be correct. For example, we could allow that the matching happens at different times. That is, given a pair $(F, G) \in \mathcal{F} \times \mathcal{F}$ and $\omega \in \Omega$, another state $\widetilde{\omega} \in \Omega$ could be paired to it if the shifted paths $(F_{h+\cdot}(\widetilde{\omega}), G_{h+\cdot}(\widetilde{\omega}))$ are "similar" to $(F_\cdot(\omega), G_\cdot(\omega))$, for some (possibly random) $h \geq 0$. However, the current version of Assumption (R) is already challenging and covers all applications we have in mind.

$f, \widetilde{f}, g, \widetilde{g}$, defined on $[0, \infty)$ and taking values in $[0, \infty)$. Assume furthermore that there exist some $t \geq 0$ and $\varepsilon > 0$ such that

$$
\begin{aligned}
f_t &\neq \widetilde{f}_t; \\
f_t &< f_s && \text{for all } s \in (t, t+\varepsilon); \\
f_s - f_t &= \widetilde{f}_s - \widetilde{f}_t && \text{for all } s \in (t, t+\varepsilon); \\
g_s - g_t &= \widetilde{g}_s - \widetilde{g}_t && \text{for all } s \in (t, t+\varepsilon).
\end{aligned}
$$

We are given, moreover, two scalar nonnegative random variables $V, W \in \mathcal{V}$. Provided that $(V, f, \widetilde{f})$ does not equal $(W, g, \widetilde{g})$ modulo trivial normalization we then first argue that $\mathcal{L}_V(f) \neq \mathcal{L}_W(g)$ and/or $\mathcal{L}_V(\widetilde{f}) \neq \mathcal{L}_W(\widetilde{g})$.

So far, the argument is purely analytic as it does not include any probabilistic statement. We now consider two processes $F, G \in \mathcal{F}$ and make use of Assumption (R). This allows us to pair sufficiently many $\omega \in \Omega$ with some $\widetilde{\omega} \in \Omega$ such that the functions

$$
f_s = F_s(\omega); \qquad \widetilde{f}_s = F_s(\widetilde{\omega}); \qquad g_s = G_s(\omega); \qquad \widetilde{g}_s = G_s(\widetilde{\omega}), \qquad s \geq 0
$$

satisfy the conditions above, at least approximately. Hence, for two realizations, the corresponding paths are different at time t but increase at the same rate for some period after t.

2.3. **Example: Positive Default Probability at Certain Times.**

We provide now a generalization of the Mixed Hazard setup of Subsection 2.1, for which Theorem 2.2 can also be applied.

For simplicity, assume in this subsection that the covariate process Z has left-hand limits. We now let $\tau^\dagger$ be an $(\mathscr{F}_t)$–predictable time. By this we mean that $\tau^\dagger$ is an $(\mathscr{F}_t)$–stopping time that is announced.[6] For example, if $\tau^\dagger$ is a nonnegative deterministic constant, it is automatically an $(\mathscr{F}_t)$–predictable time. Many other cases are possible. Importantly, $\tau^\dagger$ is adapted to $(\mathscr{F}_t)$ and therefore is contained in this information.

[6]This means $\tau^\dagger$ can be written as the limit of a strictly increasing sequence of $(\mathscr{F}_t)$–stopping times.

We again work under Assumptions (P) and (A), but now consider, for any continuous function $\gamma : \mathcal{Z} \mapsto [0,\infty)$, the process

$$F_t(\alpha,\gamma) = \begin{cases} \int_0^t \alpha(s, Z_s)\mathrm{d}s, & \text{if} \quad t < \tau^\dagger; \\ \int_0^t \alpha(s, Z_s)\mathrm{d}s + \gamma(Z_{\tau^\dagger -}), & \text{if} \quad t \geq \tau^\dagger, \end{cases} \qquad t \geq 0.$$

All possible choices of α and γ then yield the family

$$\mathcal{F}^\dagger = \{F(\alpha,\gamma) \colon \alpha \in \mathcal{A},\ \gamma : \mathcal{Z} \mapsto [0,\infty) \text{ (continuous)}\}.$$

Whereas standard hazard models have probability zero that an event happens at any given time, this setup allows for a positive probability of events at observable times. Indeed, when a separate random time $\tau^\dagger$ happens, there is positive probability that τ happens immediately. The unobservable factor V and the observed covariates Z partially determine the probability of τ happening at $\tau^\dagger$. There is no need to restrict the number of times $\tau^\dagger$ to be one. We assume one stopping time $\tau^\dagger$ here for simplicity.

Exactly as in Proposition 2.3 we may now argue that Assumption (R) holds. If we additionally assume that the support of $Z_{\tau^\dagger -}$ equals $\mathcal{Z}$, then we obtain identification modulo trivial normalization. That is, for all $(V, F(\alpha,\gamma)), (W, F(\beta,\upsilon)) \in \mathcal{V} \times \mathcal{F}^\dagger$ we then have the equivalence

$$(\mathcal{L}_V, \alpha, \gamma) = \left(\mathcal{L}_{\kappa W}, \frac{1}{\kappa}\alpha, \frac{1}{\kappa}\upsilon\right) \text{ for some } \kappa > 0$$

if and only if

$$\mathsf{P}[\mathcal{L}_V(F(\alpha,\gamma) = \mathcal{L}_W(F(\beta,\upsilon)] = 1.$$

3. Martingale Properties and Identifying Moments

3.1. **Censoring and Observation Filtration.** It is important to emphasize that the processes $\mathcal{L}_V(F)$ are usually not observed in data. For any given observation, τ can happen before time t. As we do not observe Z_t after τ, we do not know $\mathcal{L}_V(F_t)$. This is true for any possible pair $(V,F) \in \mathcal{V} \times \mathcal{F}$ regardless of what the true parameters (V°, F°) are.

The same issue can be seen in a different way by considering the problem of estimating $\mathcal{L}_{V^\circ}(F_t^\circ) = \mathsf{P}[\tau > t \mid Z_s = z_s : s \in [0,t]]$ from observed data. In the case where there is no censoring of Z, there exists a simple estimator: count the number of observations with realized path $Z = z$ for $s \in [0,t]$ with $\tau > t$ and divide by the total number of observations with this path. In the case where there is censoring, this estimator is impossible. For some observations, Z will follow the path z up until τ, but τ will happen before t. When τ occurs, it censors Z. For these cases, it is not known if $Z_s = z_s$ for all $s \in [0,t]$, only that $Z_s = z_s$ for all $s \in [0,\tau]$. Therefore, we cannot construct an estimator of $\mathsf{P}[\tau > t \mid Z_s = z_s : s \in [0,t]]$ directly in the censoring case. We can infer these objects from the conditional hazard rate at time t. An estimation approach based on this would require recovering the conditional survival probabilities from estimated hazard rates. Our approach will avoid this and use a censored version of $\mathcal{L}_V(F_t)$ directly.

To make headway, let $(\mathscr{G}_t)$ denote the filtration given by

$$\mathscr{G}_t = \sigma\{Z_{s\wedge\tau}, \mathbf{1}_{\{\tau\le s\}} \colon s \in [0,t]\}, \qquad t \ge 0.$$

This filtration contains the information whether or not the event τ has happened. It also contains the observed covariates censored by the event.[7] The filtration $(\mathscr{G}_t)$ corresponds to the observed information in the censoring case. Identification results should only depend on this.

3.2. **Martingale Moments.** Mostly for convenience, in the following we will assume that

$$F_\infty = \lim_{t\uparrow\infty} F_t < \infty. \tag{3.1}$$

[7] It is possible to work with the filtration which observes Z up until just before time τ and derive the same identification results in the sequel. This is needed if the event τ prevents observation of Z. To this end, let us write

$$Z_t^{\tau-} = Z_t\mathbf{1}_{\{t<\tau\}} + \liminf_{s\uparrow\tau} Z_s\mathbf{1}_{\{s\ge\tau\}}, \qquad t \ge 0.$$

That is, $Z^{\tau-}$ corresponds to the process Z up to the time just before τ occurs, and afterwards just stays constant. Let us define the corresponding filtration $(\mathscr{G}_t^*)$ by

$$\mathscr{G}_t^* = \sigma\{Z_s^{\tau-}, \mathbf{1}_{\{\tau\le s\}} \colon s \in [0,t]\}, \qquad t \ge 0..$$

Therefore, $(\mathscr{G}_t^*)$ contains all information that $(\mathscr{F}_t)$ contains before time τ. We have $\mathscr{G}_t^* \subset \mathscr{G}_t$. The identification results below would all remain true if we replaced $(\mathscr{G}_t)$ by $(\mathscr{G}_t^*)$.

In the context of the Mixed Hazard setup, this means that

$$\int_0^T \alpha(s, z)\mathrm{d}s < \infty$$

for some $T > 0$ and $\alpha(s, z)$ is zero for larger times. This corresponds to an upper bound on the duration lengths considered.

In the following, we consider processes of the form $M = 1/\mathcal{L}_W(G)\mathbf{1}_{[\![0,\tau[\![}$ where $(W, G) \in \mathcal{V} \times \mathcal{F}$. Notice that the process $M = 1/\mathcal{L}_W(G)\mathbf{1}_{[\![0,\tau[\![}$ is observable, that is, $(\mathscr{G}_t)$–adapted. When τ happens, M jumps to zero. Hence, the fact that τ censors Z does not prevent us from constructing M from $(\mathscr{G}_t)$. The process also does not depend directly on the hazard rate. The following corollary shows that M is only a $(\mathscr{G}_t)$–martingale if $(\mathcal{L}_W, G)$ equals $(\mathcal{L}_{V^\circ}, F^\circ)$ modulo trivial normalization.

Corollary 3.1. *Suppose that Assumption (R) and* (3.1) *hold, fix a pair* $(W, G) \in \mathcal{V} \times \mathcal{F}$, *and define the process* $M = 1/\mathcal{L}_W(G)\mathbf{1}_{[\![0,\tau[\![}$. *Then the following conditions are equivalent:*

(i) $(\mathcal{L}_{\kappa W}, G/\kappa) = (\mathcal{L}_{V^\circ}, F^\circ)$ *for some* $\kappa > 0$.
(ii) M *is a uniformly integrable* $(\mathscr{G}_t)$*–martingale.*

Proof. This is a direct consequence of Theorems 2.2 and D.1 below. □

In the sequel, the moment conditions that identify the model depend on the fact that only if $(\mathcal{L}_W, G)$ is a trivial normalization of $(\mathcal{L}_{V^\circ}, F^\circ)$, then M is a uniformly integrable $(\mathscr{G}_t)$–martingale. An essential role for this result is played by the filtration $(\mathscr{G}_t)$. Indeed, the result fails if the filtration generated by M is used instead of $(\mathscr{G}_t)$, as Example D.2 in the online appendix illustrates.

The martingale structure in Corollary 3.1 leads to a set of moment conditions which only the true hazard parameters (V°, F°) can satisfy. This identifies the model. In order to derive these moment conditions, we introduce the filtration $(\mathscr{O}_t)$ given by

$$\mathscr{O}_t = \sigma\{Z_{s\wedge\tau} \colon s \in [0, t]\}, \qquad t \geq 0.$$

Note that $\mathscr{O}_t \subset \mathscr{G}_t$ and therefore we observe the information in $(\mathscr{O}_t)$.[8] The identifying moment conditions are based on a set of $\mathscr{O}_t$–stopping times. These stopping times take the following form: for $t \geq 0$ and $A \in \mathscr{O}_t$, we consider $\eta_{t,A} = t\mathbf{1}_A + \infty\mathbf{1}_{\Omega\setminus A}$. For nonnegative processes M we set $M_\infty = \liminf_{t\uparrow\infty} M_t$. The family of nonnegative dyadic numbers $\mathcal{D} = (k2^{-j})_{k,j\in\mathbb{N}}$ will be needed in the sequel.

Theorem 3.2. *Suppose Assumption (R) and* (3.1) *hold, let* $(W, G) \in \mathcal{V} \times \mathcal{F}$, *and define the process* $M = 1/\mathcal{L}_W(G)\mathbf{1}_{[\![0,\tau[\![}$. *Then the following statements are equivalent:*

(i) $(\mathcal{L}_{\kappa W}, G/\kappa) \neq (\mathcal{L}_{V^\circ}, F^\circ)$ *for all* $\kappa > 0$.

(ii) *There exist* $t \in \mathcal{D}$ *and* $A \in \mathscr{O}_t$ *such that*

$$\mathsf{E}[M_{\eta_{t,A}}] \neq 1.$$

Proof. If (ii) holds, M cannot be a uniformly integrable $(\mathscr{G}_t)$–martingale. An application of Corollary 3.1 yields (i). Now, assume (i) holds. Then, by Corollary 3.1, M is not a uniformly integrable $(\mathscr{G}_t)$–martingale. Indeed, by stopping the processes G and F° we get that M is not even a $(\mathscr{G}_t)$–martingale. This in conjunction with the right-continuity of M then guarantees the existence of $s, u \in \mathcal{D}$ with $s < u$ and $B \in \mathscr{G}_s$ such that $\mathsf{E}[M_s\mathbf{1}_B] \neq \mathsf{E}[M_u\mathbf{1}_B]$. Thanks to the special structure of the sigma algebra $\mathscr{G}_s$, this also yields the existence of $A \in \mathscr{O}_s \subset \mathscr{O}_u$ such that $\mathsf{E}[M_s\mathbf{1}_A] \neq \mathsf{E}[M_u\mathbf{1}_A]$. In particular, one of the two quantities cannot equal $\mathsf{E}[M_\infty\mathbf{1}_{\Omega\setminus A}]$, yielding (ii). □

The previous theorem shows that for any pair $(\mathcal{L}_W, G)$, different from $(\mathcal{L}_{V^\circ}, F^\circ)$ modulo trivial normalization, one of the moment conditions must be violated. However, the number of moments is large. We have to consider $\eta_{t,A}$ for all $t \in \mathcal{D}$ and $A \in \mathscr{O}_t$. It is not obvious how to systematically choose $A \in \mathscr{O}_t$ in an estimation strategy. In the following corollary, we reduce the number of moments needed to identify the model. This leads to a tractable set of moment conditions for estimation.

---

[8] As in Footnote 7, we could also consider the slightly smaller filtration $(\mathscr{O}_t^*)$, given by

$$\mathscr{O}_t^* = \sigma\{Z_s^{\tau-}, s \in [0,t]), \qquad t \geq 0.$$

All results below would remain true if we replaced $(\mathscr{O}_t)$ by $(\mathscr{O}_t^*)$.

Corollary 3.3. *With the assumptions and notation of Theorem 3.2, suppose that for each $t \in \mathcal{D}$ there is a π-system $\mathbb{S}_t$ generating $\mathscr{O}_t$. Then* (i) *and* (ii) *in Theorem 3.2 are equivalent to the following condition:*

(iii) *There exist $t \in \mathcal{D}$ and $A \in \mathbb{S}_t \cup \{\emptyset, \Omega\}$ such that*

$$\mathsf{E}[M_{\eta_{t,A}}] \neq 1.$$

Proof. It suffices to argue that a failure of (iii) implies a failure of Theorem 3.2(ii). Thus, let us assume that

$$\mathsf{E}[M_\infty] = 1 \qquad \text{and} \qquad \mathsf{E}[M_t] = 1 \qquad \text{and} \qquad \mathsf{E}[M_{\eta_{t,A}}] = 1$$

for all $t \in \mathcal{D}$ and $A \in \mathbb{S}_t$. Then a standard application of Dynkin's π-λ-argument concludes. □

This corollary states that identification follows from the smaller set of moment conditions defined by $t \in \mathcal{D}$ and $A \in \mathbb{S}_t \cup \{\emptyset, \Omega\}$. If the π-systems $\mathbb{S}_t$ are defined in an appropriate way, it is possible to find a convenient set of moment conditions for estimation. We now give an example of such a π-system $\mathbb{S}_t$.

Consider the event

$$B(s, x, j) = \left\{\|x - Z_{s\wedge\tau}\|_1 \leq d2^{-j}\right\},$$

where $s \geq 0$, $x \in \mathbb{R}^d$, and $j \in \mathbb{N}$. Here, $\|\cdot\|_1$ denotes the L^1–metric. Hence, $B(s, x, j)$ is the event that, at time s, the stopped covariate process Z^τ is within $d2^{-j}$ of the point x in L^1–distance. We call $B(s, x, j)$ a "checkpoint" and say the realization of the process $Z^\tau(\omega)$ passes the checkpoint if $\omega \in B(s, x, j)$.

Next, define the event

$$A(k, j) = \bigcap_{\ell=0}^{2^j} B\left(2^{-j}\ell t, 2^{-j}k^{(\ell)}, j\right),$$

where $k = (k^{(0)}, \cdots, k^{(2^j)}) \in (\mathbb{Z}^d)^{2^j+1}$, and $j \in \mathbb{N}$. In words, $A(k, j)$ is the event where the stopped covariate process $Z^\tau(\omega)$ passes a sequence of checkpoints. If the process fails to pass any $B\left(2^{-j}\ell t, 2^{-j}k^{(\ell)}, j\right)$ then $\omega \notin A(k,j)$. These checkpoints are at times $0, 2^{-j}t, 2\times 2^{-j}t, \cdots, t$ and are centered at $\mathbb{R}^d$–valued vectors; the ℓ-th center being $2^{-j}k^{(\ell-1)}$. Therefore, at time $2^{-j}\ell t$ the process $Z^\tau(\omega)$ must be within $d2^{-j}$ of the point $2^{-j}k^{(\ell)}$.

We can now define

$$\mathbb{S}_t^j = \bigcup_{k\in\{-j2^j,\cdots,j2^j\}^{d\times(2^j+1)}} A(k,j),$$

and

$$\mathbb{S}_t = \bigcup_{j\in\mathbb{N}} \mathbb{S}_t^j.$$

It is relatively simple to show that $\mathbb{S}_t$ is indeed a π-system. Moreover, if we assume that the covariate process is either left- or right-continuous, then its paths are fully determined by their values at dyadic numbers. Hence, under such a weak additional continuity assumption, $\mathbb{S}_t$ also generates the sigma algebra $\mathscr{O}_t$.

Corollary 3.3 shows that the correct hazard parameters $(\mathcal{L}_{V^\circ}, F^\circ)$ can be identified with moment conditions derived from the π-systems $\mathbb{S}_t$ described above. Any incorrect specification will violate $\mathsf{E}[M_{\eta_{t,A}}] = 1$ for some $t\in\mathcal{D}$ and $A\in\mathbb{S}_t$.[9] If a moment condition from $\mathbb{S}_t^j$ is violated, and $j'>j$, then a moment condition from $\mathbb{S}_t^{j'}$ must also be violated. This is because for any $A\in\mathbb{S}_t^j$, there exist disjoint sets in $\mathbb{S}_t^{j'}$ whose union is A. It follows that if $\mathsf{E}[M_{\eta_{t,A}}]\neq 1$ then for one of the sets $A'\in\mathbb{S}_t^{j'}$ we have $\mathsf{E}[M_{\eta_{t,A'}}]\neq 1$.

For a given set of incorrect parameters $(W,G)\in\mathcal{V}\times\mathcal{F}$ there exist $t\in\mathcal{D}$ and $j\in\mathbb{N}$ such that $\mathsf{E}[M_{\eta_{t,A(k,j)}}]\neq 1$ for some $A(k,j)\in\mathbb{S}_t^j$. Therefore, for each incorrect specification, only a finite set of moment conditions $\mathbb{S}_t^j$ is needed to rule it out. However, for a given $(W,G)\in\mathcal{V}\times\mathcal{F}$ the number j is unknown.

In the case where $\mathcal{Z}=\mathbb{R}^d$, the covariates can satisfy all possible sets of checkpoints. When $\mathcal{Z}\subset\mathbb{R}^d$, some of the sets $A(k,j)$ defined above will be empty. This simply makes

[9]Here and in the sequel, when we write $\mathbb{S}_t$ or $\mathbb{S}_t^j$ referring to moment conditions it is implicitly assumed that we include $\{\emptyset,\Omega\}$ in these sets. This is done to ease notation.

the corresponding stopping times $\eta_{A(k,j)}$ equal to ∞ with probability one. All of the stated results are still true.

Corollary 3.3 and the observations afterwards suggest a GMM estimation strategy based on a set of moment conditions which grows with the amount of observed data. A fixed set of moments corresponding to $\mathbb{S}_t^j$, $t \geq 0$ may not rule out an individual incorrect model $(W, G) \in \mathcal{V} \times \mathcal{F}$ for a given j, but as j tends to ∞ they will eventually rule out all incorrect models. An estimation approach based on this idea would require considering each of the moment conditions (or a subset of them) corresponding to $\mathbb{S}_t^j$, $t \geq 0$ for a given amount of data. The set of moments is increased by increasing j as the number of observations gets larger. In the limit, it may be possible to achieve a consistent estimator. This could use a sieve estimation approach where the relevant functions are estimated with an increasingly large number of basis functions. Specifics of this estimation strategy are beyond the scope of the paper and left to future research.

3.3. **Additional Random Censoring.** In practice, there might be additional random censoring. That is, if C denotes an exogenous censoring time, then the stopped process Z^C would be observed and replace Z in the arguments above. All the results in this section will still go through with mild assumptions on C. In particular, this would require that the stopped process F^C satisfies Assumption (R). In the special setup of the Mixed Hazard model, this would imply that $C > t_2$ with positive probability (in the notation of Assumption (A)).

Acknowledgements

We thank an anonymous referee, Sokbae (Simon) Lee as the co-editor, and Peter Phillips as the editor for very helpful remarks that improved this paper. We are also grateful to the Oxford-Man Institute of Quantitative Finance for their hospitality.

Appendix A. An Example for Non-Identification

First, let us illustrate that the statement of Theorem 2.2 would be wrong, in general, if one did not make Assumption (R).

Example A.1. In this example we construct three random variables $V^a, V^b, V^c \in \mathcal{V}$, satisfying $\mathsf{E}[V^a] = \mathsf{E}[V^b] = \mathsf{E}[V^c] = 1$, and three non-deterministic processes $F^a, F^b, F^c \in \mathcal{F}$ such that $\mathcal{L}_{V^a}(F^a) = \mathcal{L}_{V^b}(F^b) = \mathcal{L}_{V^c}(F^c)$, but $(L_{V^i}, F^i) \neq (L_{V^j}, F^j)$ for all $i, j \in \{a, b, c\}$ with $i \neq j$.

To start, set V^a to a constant; to wit, $\mathsf{P}[V^a = 1] = 1$. Next, V^b is assumed to be exponentially distributed; to wit, $\mathsf{P}[V^b \in \mathrm{d}v] = \mathbf{1}_{v \geq 0} \mathrm{e}^{-v} \mathrm{d}v$. Moreover, assume that V^c be distributed as a mixture with $\mathsf{P}[V^c \in \mathrm{d}v] = 1/2 \mathrm{d}\delta_0 + 1/4 \mathbf{1}_{v \geq 0} \mathrm{e}^{-v/2} \mathrm{d}v$, where δ_0 denotes the Dirac measure at zero. That is, V^c can be generated by a fair coin toss $X \in \{0, 1\}$ and an exponentially distributed random variable Y with expectation 2; to wit, $\mathsf{P}[V^c = XY] = 1$. In particular, we then have $\mathsf{E}[V^i] = 1$, thus $V^i \in \mathcal{V}$, for each $i \in \{a, b, c\}$, and also the following equalities:

$$\begin{aligned}
\mathcal{L}_{V^a}(s) &= \mathsf{E}[\mathrm{e}^{-sV^a}] = \mathrm{e}^{-s}, && s \geq 0; \\
\mathcal{L}_{V^b}(s) &= \mathsf{E}[\mathrm{e}^{-sV^b}] = \frac{1}{1+s}, && s \geq 0; \\
\mathcal{L}_{V^c}(s) &= \mathsf{E}[\mathrm{e}^{-sV^c}] = \frac{1+s}{1+2s}, && s \geq 0.
\end{aligned}$$

Next, introduce the (deterministic) functions

$$\begin{aligned}
f^a(t) &= \log\left(2 - \frac{1}{1+t}\right) \wedge \log\left(\frac{3}{2}\right), && t \geq 0; \\
f^b(t) &= \frac{t}{1+t} \wedge \frac{1}{2}, && t \geq 0; \\
f^c(t) &= t \wedge 1, && t \geq 0.
\end{aligned}$$

It is easy to check that now $\mathcal{L}_{V^a} \circ f^a = \mathcal{L}_{V^b} \circ f^b = \mathcal{L}_{V^c} \circ f^c$. Introduce also the functions

$$\begin{aligned}
f^{a\prime}(t) &= \log(1+t) \wedge \log\left(\frac{3}{2}\right), && t \geq 0; \\
f^{b\prime}(t) &= t \wedge \frac{1}{2}, && t \geq 0; \\
f^{c\prime}(t) &= \frac{t}{1-t} \wedge 1, && t \geq 0.
\end{aligned}$$

Then again $\mathcal{L}_{V^a} \circ f^{a\prime} = \mathcal{L}_{V^b} \circ f^{b\prime} = \mathcal{L}_{V^c} \circ f^{c\prime}$.

We now assume that there exists a (measurable) partition (D, D') of Ω such that D is independent of V^a, V^b, V^c and such that $\mathsf{P}[D] \in (0,1)$. We then introduce the process $F^i = f^i \mathbf{1}_D + f^{i'} \mathbf{1}_{D'}$ for each $i \in \{a,b,c\}$. Then we have $\mathcal{L}_{V^a} \circ F^a = \mathcal{L}_{V^b} \circ F^b = \mathcal{L}_{V^c} \circ F^c$. Thus, the conclusion of Theorem 2.2 does not hold.

Let us now argue that Assumption (R) fails so that this example does not give a contradiction to Theorem 2.2. To this end, set $F = F^a$, $G = F^b$, and fix any $[0,\infty]$–valued random variable ρ such that $\mathsf{P}[\rho < \infty] > 0$. Fix also a $(0,\infty)$–valued random variable $\mathcal{E}$ and an $\mathbb{R}^2 \setminus \{(0,0)\}$–valued random vector (C, C'). It is now sufficient to argue that for each $\omega \in \{\rho < \infty\}$ there exists some $n \in \mathbb{N}$ such that

$$\mathsf{P}\left[O_{f,g,t,c,c',\varepsilon,n}\right]\big|_{f=F,g=G,t=\rho,c=C,c'=C',\varepsilon=\mathcal{E}} = 0. \tag{A.1}$$

To make headway, fix $\omega \in \{\rho < \infty\}$. Let us assume that $\omega \in D$ and $C(\omega) \neq 0$. Indeed, if $\omega \in D'$ or if $C(\omega) = 0$ (but $C'(\omega) \neq 0$), the argument follows in exactly the same manner. Next, let us observe that $F(\omega) = f^a$ and $G(\omega) = f^b$ and let us set $t = \rho(\omega)$, $c = C(\omega)$, $c' = C'(\omega)$, and $\varepsilon = \mathcal{E}(\omega)$. We then have, for each $n \in \mathbb{N}$,

$$\begin{aligned} O_{f^a,f^b,t,c,c',\varepsilon,n} &\subset \left\{\omega \in \Omega : \sup_{s \in [t,t+\varepsilon)} |c + f^b(s) - G_s| < \frac{1}{n}\right\}, \\ &= \left\{\omega \in D : \ |c| < \frac{1}{n}\right\} \cup \left\{\omega \in D' : \sup_{s \in [t,t+\varepsilon)} |c + f^b(s) - f^{b'}(s)| < \frac{1}{n}\right\}. \end{aligned} \tag{A.2}$$

Clearly, for $n > 1/|c|$, the first event in (A.2) is the empty set. Since $\varepsilon > 0$ is fixed and $f^b - f^{b'}$ non-constant on $[0,1]$ and equal to zero on $[1,\infty)$, also the second event in (A.2) is the empty set for sufficiently large $n \in \mathbb{N}$. Hence, we have $O_{f^a,f^b,t,c,c',\varepsilon,n} = \emptyset$, which then yields (A.1) for sufficiently large $n \in \mathbb{N}$. □

References

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

Bruggeman, C. and Ruf, J. (2016). A one-dimensional diffusion hits points fast. *Electronic Communications in Probability*, 21(22):1–7.

Caliendo, M., Tatsiramos, K., and Uhlendorff, A. (2013). Benefit duration, unemployment duration and job match quality: a regression-discontinuity approach. *Journal of Applied Econometrics*, 28:604–627.

Cornelißen, T. and Hübler, O. (2011). Unobserved individual and firm heterogeneity in wage and job-duration functions: Evidence from German linked employer-employee data. *German Economic Review*, 12(4):469–489.

Farber, H. and Valletta, R. (2015). Do extended unemployment benefits lengthen unemployment spells? *Journal of Human Resources*, 50(4):873–909.

Feller, W. (1968). On Müntz' theorem and completely monotone functions. *American Mathematics Monthly*, 75:342–359.

Hausman, J. and Woutersen, T. (2014). Estimating a semi-parametric duration model without specifying heterogeneity. *Journal of Econometrics*, 178:114–131.

Heckman, J. (1991). Identifying the hand of past: Distinguishing state dependence from heterogeneity. *American Economic Review*, 81(2):75–79.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2nd edition.

Kroft, K., Lange, F., and Notowidigdo, M. (2013). Duration dependence and labor market conditions: Evidence from a field experiment. *Quarterly Journal of Economics*, 128(3):1123–1167.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47(4):939–956.

McCall, B. (1994). Identifying state dependence in duration models. In *American Statistical Association 1994, Proceedings of the Business and Economics Section*, pages 14–17. American Statistical Association.

Perkowski, N. and Ruf, J. (2015). Supermartingales as Radon-Nikodym densities and related measure extensions. *The Annals of Probability*, 43(6):3133–3176.

Stroock, D. W. and Varadhan, S. R. S. (1972). On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability Theory*, pages 333–359. University of California Press, Berkeley, Calif.

Widder, D. (1946). *The Laplace Transform*. Princeton University Press, London.

Yoeurp, C. (1976). Décompositions des martingales locales et formules exponentielles. In *Séminaire de Probabilités, X*, pages 432–480. Springer, Berlin.

ONLINE APPENDIX TO: NONPARAMETRIC IDENTIFICATION OF THE MIXED HAZARD MODEL USING MARTINGALE-BASED MOMENTS

JOHANNES RUF AND JAMES LEWIS WOLTER

Appendix B. The Proofs of Theorem 2.2 and Proposition 2.3

The proof of Theorem 2.2 requires two analytic results, which we provide in the following two lemmas.

Lemma B.1. *Let* $(c, c') \in \mathbb{R}^2 \setminus \{(0,0)\}$, $V, W \in \mathcal{V}$, *and let* $f, g : [0,\infty) \mapsto [0,\infty)$ *denote two right-continuous functions. Assume that there exist some* $t \geq 0$ *and* $\varepsilon > 0$ *such that*

$$f_t < f_s, \qquad \text{for all } s \in (t, t+\varepsilon) \tag{B.1}$$

and define the two right-continuous functions

$$\widetilde{f}_s = f_s + c\mathbf{1}_{s\in[t,t+\varepsilon)}, \qquad s \geq 0; \tag{B.2}$$

$$\widetilde{g}_s = g_s + c'\mathbf{1}_{s\in[t,t+\varepsilon)}, \qquad s \geq 0. \tag{B.3}$$

Then we have the equivalence

$$(\mathcal{L}_V, f, \widetilde{f}) = \left(\mathcal{L}_{\frac{c'}{c}W}, \frac{c}{c'}g, \frac{c}{c'}\widetilde{g}\right) \qquad \text{if and only if} \qquad (\mathcal{L}_V \circ f, \mathcal{L}_V \circ \widetilde{f}) = (\mathcal{L}_W \circ g, \mathcal{L}_W \circ \widetilde{g}).$$

Proof. The "only if" direction is trivial; hence let us assume that $(\mathcal{L}_V \circ f, \mathcal{L}_V \circ \widetilde{f}) = (\mathcal{L}_W \circ g, \mathcal{L}_W \circ \widetilde{g})$. Since this implies that

$$\mathcal{L}_V(f_s) - \mathcal{L}_V\left(\widetilde{f}_s\right) = \mathcal{L}_W(g_s) - \mathcal{L}_W\left(\widetilde{g}_s\right)$$

we then have $c \neq 0$ and $c' \neq 0$. By swapping f with $\widetilde{f}$ (and g with $\widetilde{g}$) we may assume that $c > 0$. The monotonicity of $\mathcal{L}_V$ and $\mathcal{L}_W$ then also yields that $c' > 0$. Next, define the functions

$$\varphi : (\mathcal{L}_V(\infty), 1] \to \mathbb{R}; \quad s \mapsto \mathcal{L}_V\left(\mathcal{L}_V^{-1}(s) + c\right);$$
$$\psi : (\mathcal{L}_W(\infty), 1] \to \mathbb{R}; \quad s \mapsto \mathcal{L}_W\left(\mathcal{L}_W^{-1}(s) + c'\right)$$

and observe

$$\varphi\left(\mathcal{L}_V(f_s)\right) = \mathcal{L}_V\left(\widetilde{f}_s\right) = \mathcal{L}_W\left(\widetilde{g}_s\right) = \psi\left(\mathcal{L}_W(g_s)\right) = \psi\left(\mathcal{L}_V(f_s)\right), \quad s \in [t, t+\varepsilon).$$

Now, (B.1) guarantees that the range of the function $[t, t+\varepsilon) \ni s \mapsto \mathcal{L}_V(f_s)$ contains an open non-empty interval. This then yields that $\varphi = \psi$ since the two functions φ and ψ are analytic; see Chapter II.5 in Widder (1946).

Johannes Ruf, Department of Mathematics, London School of Economics and Political Science, Columbia House, Houghton St, London WC2A 2AE, United Kingdom.

James Wolter, Lord, Abbett & Co. LLC, 90 Hudson Street, Jersey City, NJ 07302, US.

E-mail address: j.ruf@lse.ac.uk, jwolter@lordabbett.com.

Date: December 25, 2018.

Next, note that $\mathcal{L}_V(0) = \mathcal{L}_W(0)$ and proceed by induction as follows. Assume that we have argued $\mathcal{L}_V((n-1)c) = \mathcal{L}_W((n-1)c')$ for some $n \in \mathbb{N}$. Then we get

$$\mathcal{L}_V(nc) = \varphi(\mathcal{L}_V((n-1)c)) = \psi(\mathcal{L}_W((n-1)c')) = \mathcal{L}_W(nc'), \qquad n \in \mathbb{N}.$$

Now, consider the random variable $W' = (c'/c)W \geq 0$ with Laplace transform $\mathcal{L}_{W'}(x) = \mathcal{L}_W(xc'/c)$ for all $x \geq 0$. We thus have

$$\mathcal{L}_V(nc) = \mathcal{L}_{W'}(nc), \qquad n \in \mathbb{N}.$$

Since, moreover, $\sum_{n\in\mathbb{N}} 1/(cn) = \infty$, the Müntz theorem yields $\mathcal{L}_V = \mathcal{L}_{W'} = \mathcal{L}_{(c'/c)W}$; see Feller (1968), in particular, Theorem 2 and the representation of (1.7) in that paper. We then also get that

$$f = \mathcal{L}_V^{-1}(\mathcal{L}_V(f)) = \mathcal{L}_V^{-1}(\mathcal{L}_W(g)) = \mathcal{L}_V^{-1}(\mathcal{L}_{(c/c')V}(g)) = \mathcal{L}_V^{-1}\left(\mathcal{L}_V\left(\frac{c}{c'}g\right)\right) = \frac{c}{c'}g,$$

and similarly, that $\widetilde{f} = (c/c')\widetilde{g}$. Hence, the statement follows. □

The proof of Lemma B.1 is inspired by Brinch (2007).

Lemma B.2. *Recall the setup and notation of Lemma B.1. Then there exists $n \in \mathbb{N}$ such that, for all right-continuous functions $\overline{f}, \overline{g} : [0,\infty) \mapsto [0,\infty)$ satisfying*

$$\sup_{s\in[t,t+\varepsilon)} \left(|c + f_s - \overline{f}_s| + |c' + g - \overline{g}_s|\right) < \frac{1}{n} \tag{B.4}$$

and

$$(\mathcal{L}_V, f, \overline{f}) \neq \left(\mathcal{L}_{\kappa W}, \frac{1}{\kappa}g, \frac{1}{\kappa}\overline{g}\right) \quad \textit{for each constant } \kappa > 0,$$

we have

$$(\mathcal{L}_V \circ f, \mathcal{L}_V \circ \overline{f}) \neq (\mathcal{L}_W \circ g, \mathcal{L}_W \circ \overline{g}),$$

Proof. We may assume in the proof that $\mathcal{L}_V \circ f = \mathcal{L}_W \circ g$; otherwise nothing is to be argued. We recall the right-continuous functions $\widetilde{f}$ and $\widetilde{g}$ from (B.2) and (B.3). Let us first assume that $\mathcal{L}_V \circ \widetilde{f} = \mathcal{L}_W \circ \widetilde{g}$ on $[t, t+\varepsilon)$. Then we also have $\mathcal{L}_V \circ \widetilde{f} = \mathcal{L}_W \circ \widetilde{g}$. Lemma B.1 now yields that $\mathcal{L}_V = \mathcal{L}_{(c'/c)W}$. This then yields the statement.

Hence, let us now assume that $\mathcal{L}_V \circ \widetilde{f} \neq \mathcal{L}_W \circ \widetilde{g}$ on $[t, t+\varepsilon)$. Thanks to the continuity properties of the two functions $\mathcal{L}_V$ and $\mathcal{L}_W$, there exists $n \in \mathbb{N}$ such that $\mathcal{L}_V \circ \overline{f} \neq \mathcal{L}_W \circ \overline{g}$ whenever (B.4) holds. This concludes the proof. □

Proof of Theorem 2.2. We only need to argue one direction. Hence, let us assume that $(\mathcal{L}_V, F) \neq (\mathcal{L}_{\kappa W}, G/\kappa)$ for each constant $\kappa > 0$. In the notation of Assumption (R), fix $\omega \in A_1 \cap A_2$, set $f = F(\omega), g = G(\omega), t = \rho(\omega), c = C(\omega), c' = C'(\omega)$, and $\varepsilon = \mathcal{E}(\omega)$. Then Lemma B.2 yields the existence of $N(\omega)$ such that

$$(\mathcal{L}_V \circ f, \mathcal{L}_V \circ F(\widetilde{\omega})) \neq (\mathcal{L}_W \circ g, \mathcal{L}_W \circ G(\widetilde{\omega})), \qquad \widetilde{\omega} \in O_{f,g,t,c,c',\varepsilon,N(\omega)}. \tag{B.5}$$

Setting $N(\omega) = 1$ for all $\omega \notin A_1 \cap A_2$ then yields a mapping $N : \Omega \to \mathbb{N}$. It can be checked that N is measurable, hence N is a random variable.

To proceed with the argument, let us consider the product space $(\Omega\times\Omega, \mathscr{H}\times\mathscr{H}, \mathsf{P}\times\mathsf{P})$ and identify all random variables $X : \Omega \to \mathbb{R}$ with random vectors $(X^1, X^2) : \Omega\times\Omega \to \mathbb{R}^2$ by $X^1(\omega_1,\omega_2) = X(\omega_1)$ and $X^2(\omega_1,\omega_2) = X(\omega_2)$ for all $(\omega_1,\omega_2) \in \Omega\times\Omega$. Next, define the set

$$B = \widetilde{A} \cap \{(\omega_1,\omega_2) : \omega_2 \in O_{F^1,G^1,\rho^1,C^1,C'^1,\mathcal{E}^1,N^1}\} \subset \Omega\times\Omega, \tag{B.6}$$

where

$$\widetilde{A} = \{(\omega_1, \omega_2) : \omega_1 \in A_1 \cap A_2\} \in \mathscr{H} \times \mathscr{H}.$$

Similar as in the proof of Lemma C.1 below, we can see that $B \in \mathscr{H} \times \mathscr{H}$.

Next, note that Assumption (R) guarantees that $(\mathsf{P} \times \mathsf{P})[B] > 0$. Moreover, on B we have

$$|\mathcal{L}_V(F^1) - \mathcal{L}_W(G^1)| + |\mathcal{L}_V(F^2) - \mathcal{L}_W(G^2)| > 0,$$

thanks to (B.5). Since the distribution of (F, G) under P is the same as the one of (F^1, G^1) under the product measure $\mathsf{P} \times \mathsf{P}$, as well as the one of (F^2, G^2), we now obtain

$$\mathsf{E}[|\mathcal{L}_V(F) - \mathcal{L}_W(G)|] = \frac{1}{2}\mathsf{E}_{\mathsf{P}\times\mathsf{P}}\left[|\mathcal{L}_V(F^1) - \mathcal{L}_W(G^1)| + |\mathcal{L}_V(F^2) - \mathcal{L}_W(G^2)|\right] > 0,$$

which proves that $\mathcal{L}_V(F) \neq \mathcal{L}_W(G)$, and hence the statement follows. □

Remark B.3. In Assumption (R), each $\omega \in A_1 \cap A_2$ is matched with an event that has positive probability. After reading the proof of Theorem 2.2, the diligent reader might possibly wonder why it does not suffice to consider those paths that can be paired with a single path $\widetilde{\omega}$ (instead of an event with positive probability). To require now that the family of those ω's has positive probability is less restrictive than requiring that the event $A_1 \cap A_2$ in Assumption (R) has positive probability. Instead of considering the product measure in the proof of Theorem 2.2 one could conjecture that it suffices to use

$$|\mathcal{L}_V(F(\omega)) - \mathcal{L}_W(G(\omega))| + |\mathcal{L}_V(F(\widetilde{\omega})) - \mathcal{L}_W(G(\widetilde{\omega}))| > 0$$

for all such pairs $(\omega, \widetilde{\omega})$. However, we were not able to construct a measurable selection to pick the "right partner." Indeed, if $\omega \in \Omega$ can be paired with some candidate $\widetilde{\omega}$, it can also usually be paired with uncountably many other candidates $\widetilde{\omega}$ and it is not clear how to pick one in a measurable way. □

Proof of Proposition 2.3. Fix some continuous functions $\alpha, \beta \in \mathcal{A}$ and set

$$F_t = \int_0^t \alpha(s, Z_s)\mathrm{d}s \qquad \text{and} \qquad G_t = \int_0^t \beta(s, Z_s)\mathrm{d}s$$

By assumption there exist $x, y \in \mathcal{Z}$ and $t_1 > 0$ such that $\alpha(t_1, x) \neq \alpha(t_2, y)$. Set now $\rho = t_1$ and $\mathcal{E} = t_2 - t_1$. We directly get that $\mathsf{P}[A_1] = 1$ since the function α was assumed to be strictly positive.

Next, observe that the continuity of α implies that there exists $\delta \in (0, t_1)$ such that

$$\int_{t_1-\delta}^{t_1} \alpha(s, x)\mathrm{d}s \neq \int_{t_1-\delta}^{t_1} \alpha(s, y)\mathrm{d}s.$$

Hence, we can construct, in a measurable way, a random variable $\overline{Z}$, taking values in $\{x, y\}$, such that

$$C = \int_{t_1-\delta}^{t_1} \left(\alpha(s, \overline{Z}) - \alpha(s, Z_s)\right) \mathrm{d}s \neq 0, \tag{B.7}$$

where the inequality is with probability 1. Hence, C is $\mathbb{R} \setminus \{0\}$–valued. Similarly, we define the random variable

$$C' = \int_{t_1-\delta}^{t_1} \left(\beta(s, \overline{Z}) - \beta(s, Z_s)\right) \mathrm{d}s.$$

We now want to argue that $\mathsf{P}[A_2] > 0$ with this choice of random variables $\rho, \mathcal{E}$, and (C, C'). Indeed, we will argue that $\mathsf{P}[A_2] = 1$. To this end, fix $n \in \mathbb{N}$ and $\omega \in \Omega$ such that $z = Z(\omega)$ is in the support of the process Z, which happens with probability one. Moreover, set

$$f = F(\omega) = \int_0^\cdot \alpha(s, z_s)\mathrm{d}s, \qquad g = G(\omega) = \int_0^\cdot \beta(s, z_s)\mathrm{d}s,$$

$t = \rho(\omega)$, $c = C(\omega)$, $c' = C'(\omega)$, and $\varepsilon = \mathcal{E}(\omega)$. Next, let $\widetilde{z}$ denote a $\mathcal{Z}$–valued path with

$$\widetilde{z}_s = z_s \mathbf{1}_{s<t_1-\delta \text{ or } s>t_1} + \overline{Z}(\omega)\mathbf{1}_{t_1-\delta\leq s\leq t_1}, \qquad s \geq 0.$$

Hence $\widetilde{z}$ equals the given path z outside of the interval $[t_1 - \delta, t_1]$. On that interval, $\widetilde{z}$ is constant and takes value x or y. We now want to prove that the event

$$\begin{aligned} O_{f,g,t,c,c',\varepsilon,n} &= \left\{\widetilde{\omega} \in \Omega : \sup_{s\in[t,t+\varepsilon)} \big(|c + f_s - F_s| + |c' + g_s - G_s|\big) < \frac{1}{n}\right\} \\ &= \left\{\widetilde{\omega} \in \Omega : \sup_{s\in[t,t+\varepsilon)} \left(\left|\int_0^s (\alpha(u, \widetilde{z}_u) - \alpha(u, Z_u(\widetilde{\omega})))\,\mathrm{d}u\right|\right.\right. \\ &\qquad\left.\left. + \left|\int_0^s (\beta(u, \widetilde{z}_u) - \beta(u, Z_u(\widetilde{\omega})))\,\mathrm{d}u\right|\right) < \frac{1}{n}\right\} \end{aligned}$$

has positive probability. Indeed, with this representation, Lemma B.4 below yields that $\mathsf{P}[O_{f,g,t,c,c',\varepsilon,n}] > 0$, which concludes the proof. □

Lemma B.4. *Suppose Assumptions (P) and (A) hold along with* (2.5)*; i.e., we are in the mixed hazard setup. Assume, moreover, that $\alpha \in \mathcal{A}$ and that z be a $\mathcal{Z}$–valued function in the support of the observation process Z. Fix $n \in \mathbb{N}$, $\overline{z} \in \mathcal{Z}$, $\delta \in (0, t_1)$, and $T > 0$, and define the $\mathcal{Z}$–valued function*

$$\widetilde{z}_s = z_s \mathbf{1}_{s<t_1-\delta \text{ or } s>t_1} + \overline{z}\mathbf{1}_{t_1-\delta\leq s\leq t_1}, \qquad s \geq 0.$$

Then we have

$$\mathsf{P}\left[\sup_{s\in[0,T]} \left(\int_0^s |\alpha(u, \widetilde{z}_u) - \alpha(u, Z_u)|\,\mathrm{d}u\right) < \frac{1}{n}\right] > 0.$$

Proof. Thanks to the continuity of the function α, it suffices to show for some appropriately chosen sufficiently small $\varepsilon > 0$ that the event

$$\left\{\sup_{s\in[0,t_1-\varepsilon)} |Z_s - z_s| < \varepsilon\right\} \cap \left\{\sup_{s\in(t_1+\varepsilon,t_2-\varepsilon)} |Z_s - \overline{z}| < \varepsilon\right\} \cap \left\{\sup_{s\in(t_2+\varepsilon,T]} |Z_s - z_s| < \varepsilon\right\}$$

has positive probability. Any of the three conditions in Assumption (P) yields this. Indeed, if the covariate process Z is piecewise constant with Poisson update times this probability can be written as the product of positive probabilities of the following three events: (1) the event that up to time $t_1 - \varepsilon$ the observations stay close to z; (2) the event that a jump occurs around time t_1 to a neighbourhood of $\overline{z}$, and that another jump occurs around time t_2 to a neighbourhood of z; (3) the event that after time $t_2 + \varepsilon$ the observations stay close to z.

If the covariate process is a diffusion, the result follows from the support theorem; see, for example, Stroock and Varadhan (1972). In the case of one-dimensonsal Markov processes, a similar argument yields the claim; see for example Bruggeman and Ruf (2016). □

Appendix C. Some Results on Measurability

Assumption (R) implicitly uses two subtle but essential facts:

(1) The set $O_{f,g,t,c,c',\varepsilon,n}$ is measurable such that the probability in (2.6) is well-defined.
(2) The set $A_1 \cap A_2$ is measurable.

In this subsection, we provide the necessary arguments to justify these facts.

Lemma C.1. *With the notation of Assumption (R), the set $O_{f,g,t,c,c',\varepsilon,n}$ is an event; more precisely,*

$$O_{f,g,t,c,c',\varepsilon,n} \in \mathscr{F}_\infty.$$

Proof. Consider the right-continuous $\mathscr{F}_\infty$–measurable process

$$H_s = \mathbf{1}_{s\in[t,t+\varepsilon)}\big(|c+f_s-F_s| + |c'+g_s-G_s|\big), \qquad s \geq 0.$$

Then we have

$$O_{f,g,t,c,c',\varepsilon,n} = \bigcap_{q\in\mathbb{Q}\cap[0,\infty)} \{H_q < 1/n\} \in \mathscr{F}_\infty,$$

which concludes the argument. □

Lemma C.2. *With the notation of Assumption (R), we have $A_1 \cap A_2 \in \mathscr{H}$.*

Proof. It suffices to fix $n \in \mathbb{N}$ and check the measurability of the following set:

$$A^{(n)} = A_1 \cap \left\{\mathsf{P}\left[O_{f,g,t,c,c',\varepsilon,n}\right]\big|_{f=F,g=G,t=\rho,c=C,c'=C',\varepsilon=\mathcal{E}} > 0\right\},$$

where $A_1 \in \mathscr{H}$. We now use the same argument as in the proof of Theorem 2.2. Consider the product space $(\Omega\times\Omega, \mathscr{H}\times\mathscr{H}, \mathsf{P}\times\mathsf{P})$ and define, as in (B.6),

$$B^{(n)} = \overline{A} \cap \left\{(\omega_1,\omega_2) : \omega_2 \in O_{F^1,G^1,\rho^1,C^1,C'^1,\mathcal{E}^1,n}\right\} \in \mathscr{H}\times\mathscr{H},$$

where

$$\overline{A} = \left\{(\omega_1,\omega_2) : F^1_\rho < F^1_s, \quad \text{for each } s \in (\rho^1, \rho^1+\mathcal{E}^1)\right\} \in \mathscr{H}\times\mathscr{H}.$$

Next, let Y describe the conditional expectation of $\mathbf{1}_{B^{(n)}}$ given the sigma algebra $\mathscr{H} \times \{\emptyset,\Omega\}$; to wit,

$$\begin{aligned} Y &= \mathsf{E}\left[\mathbf{1}_{B^{(n)}} | \mathscr{H}\times\{\emptyset,\Omega\}\right] \\ &= \mathbf{1}_{\overline{A}} \times \mathsf{P}\left[O_{f,g,t,c,c',\varepsilon,n}\right]\big|_{f=F^1,g=G^1,t=\rho^1,c=C^1,c'=C'^1,\varepsilon=\mathcal{E}^1}. \end{aligned}$$

Then we see that

$$A^{(n)} = \{\omega_1 : Y(\omega_1,\omega_2) > 0\}.$$

Since we assumed that the sigma algebra $\mathscr{H}$ is complete, we get $A^{(n)} \in \mathscr{H}$, concluding the proof. □

Appendix D. Some Martingale Properties Of The Hazard Model

Theorem D.1. *Fix a pair $(W,G) \in \mathcal{V}\times\mathcal{F}$, assume that $\int_0^\infty F^\circ_t \mathrm{d}t < \infty$, and define the process $M = 1/\mathcal{L}_W(G)\mathbf{1}_{[\![0,\tau[\![}$. Then the following conditions are equivalent:*

(i) *$\mathcal{L}_W(G) = \mathcal{L}_{V^\circ}(F^\circ)$.*
(ii) *M is a uniformly integrable $(\mathscr{G}_t)$–martingale.*

Proof. Assume that (i) holds. Now fix $s, t \in [0, \infty]$ with $s < t$ and $A \in \mathscr{G}_s$. Then we can write $A \cap \{\tau > s\} = B \cap \{\tau > s\}$ for some $B \in \mathscr{F}_s$. Hence, we get

$$\begin{aligned}\mathsf{E}[M_t \mathbf{1}_A] &= \mathsf{E}[M_t \mathbf{1}_B] = \mathsf{E}\left[\frac{1}{\mathcal{L}_{V^\circ}(F_t^\circ)} \lim_{u \uparrow t} \mathbf{1}_{\{\tau > u\} \cap B}\right] = \mathsf{E}\left[\frac{1}{\mathcal{L}_{V^\circ}(F_t^\circ)} \mathbf{1}_B \mathrm{e}^{-V^\circ F_t^\circ}\right] = \mathsf{P}[B] \\ &= \mathsf{E}\left[\frac{1}{\mathcal{L}_{V^\circ}(F_s^\circ)} \mathbf{1}_B \mathrm{e}^{-V^\circ F_s^\circ}\right] = \mathsf{E}\left[\frac{1}{\mathcal{L}_{V^\circ}(F_s^\circ)} \mathbf{1}_{\{\tau > s\} \cap B}\right] = \mathsf{E}[M_s \mathbf{1}_B] = \mathsf{E}[M_s \mathbf{1}_A].\end{aligned}$$

Here, we used (2.5). Thus, M is indeed a uniformly integrable $(\mathscr{G}_t)$–martingale.

Let us now assume that (ii) holds, but $\mathcal{L}_W(G) \neq \mathcal{L}_{V^\circ}(F^\circ)$. Then (2.5) yields the existence of $t > 0$ such that

$$\mathsf{P}\left[\{\tau > t\} \cap \{\mathcal{L}_W(G) \neq \mathcal{L}_{V^\circ}(F^\circ)\}\right] > 0.$$

Thus, we have

$$M \neq \frac{1}{\mathcal{L}_{V^\circ}(F^\circ)} \mathbf{1}_{[\![0,\tau[\![}.$$

The left-hand side is a $(\mathscr{G}_t)$–martingale by assumption, the right-hand side by the implication from (i) to (ii). This, however, contradicts the uniqueness of the multiplicative decomposition of the nonnegative $(\mathscr{G}_t)$–supermartingale $\mathbf{1}_{[\![0,\tau[\![}$ as a product of a local martingale and a predictable nonincreasing process:

$$\mathbf{1}_{[\![0,\tau[\![} = M \times \mathcal{L}_W(G); \qquad \mathbf{1}_{[\![0,\tau[\![} = \frac{1}{\mathcal{L}_{V^\circ}(F^\circ)} \mathbf{1}_{[\![0,\tau[\![} \times \mathcal{L}_{V^\circ}(F^\circ);$$

see also Yoeurp (1976) and Appendix B in Perkowski and Ruf (2015). Hence, we have the implication from (ii) to (i). □

Example D.2. The choice of filtration is essential in the statement of Theorem D.1, even if there is no unobserved factor. To see this, we provide now a setup that satisfies Assumption (R). However, in this specific setup there exists $G \in \mathcal{F}$ such that the process $M = 1/\mathcal{L}_{V^\circ}(G) \mathbf{1}_{[\![0,\tau[\![}$ is a uniformly integrable $(\mathscr{E}_t)$–martingale, but $\mathcal{L}_{V^\circ}(G) \neq \mathcal{L}_{V^\circ}(F^\circ)$. Here, $(\mathscr{E}_t) \subset (\mathscr{G}_t)$ denotes the filtration generated by M itself.

Suppose that $\Omega = \{w_1, w_2\} \times [0, \infty)$, that $\zeta(w, r) = r$ for all $(w, r) \in \Omega$ and that

$$\mathsf{P}\left[\{w_1\} \times (t, \infty)\right] = \frac{1}{2} \mathrm{e}^{-t} = \mathsf{P}\left[\{w_2\} \times (t, \infty)\right], \qquad t \geq 0.$$

In particular, ζ is exponentially distributed. Moreover, let $V^\circ = 1$ and $\mathcal{F} = \{F^\circ, G\}$, where

$$F_t^\circ(w_1, r) = t \wedge 3; \qquad G_t(w_1, r) = \log\left(\frac{2}{1 + \mathrm{e}^{-(1 \wedge t)}}\right) + (t-1)^+ \wedge 2, \qquad t, r \geq 0;$$

$$F_t^\circ(w_2, r) = (t-2)^+ \wedge 1; \qquad G_t(w_2, r) = \log\left(\frac{2}{1 + \mathrm{e}^{-(1 \wedge t)}}\right) + (t-2)^+ \wedge 1, \qquad t, r \geq 0.$$

Next, define τ in the same way as in Footnote 3. Then the basic setup is satisfied. It is also easy to check that Assumption (R) is satisfied with $\rho = C = 2$ and $\mathcal{E} = C' = 1$. However, clearly we have $\mathcal{L}_{V^\circ}(G) \neq \mathcal{L}_{V^\circ}(F^\circ)$.

Let us observe that

$$M_t = \frac{2}{1 + \mathrm{e}^{-t}} \mathbf{1}_{[\![0,\tau[\![}, \qquad t \in [0, 1];$$

$$M_t = \frac{2}{(1 + \mathrm{e}^{-1}) \mathrm{e}^{F_1^\circ}} \times \mathrm{e}^{F_t^\circ} \mathbf{1}_{[\![0,\tau[\![}, \qquad t \geq 1.$$

Next, let us check that M is a uniformly integrable $(\mathscr{E}_t)$–martingale. To this end, for $s, t \in [0, 1]$ with $s < t$ we have, on the event $\{\tau > s\} \in \mathscr{E}_s$,

$$\mathsf{E}[M_t|\mathscr{E}_s] = \frac{2}{1+\mathrm{e}^{-t}}\mathsf{P}[\tau > t|\tau > s] = \frac{2}{1+\mathrm{e}^{-t}} \times \frac{1+\mathrm{e}^{-t}}{1+\mathrm{e}^{-s}} = M_s.$$

For $s, t \in [1, \infty]$ with $s < t$ it is also easy to check that $\mathsf{E}[M_t|\mathscr{E}_s] = M_s$. Hence, M is indeed a uniformly integrable $(\mathscr{E}_t)$–martingale.

Let us now double-check that M is indeed not a $(\mathscr{G}_t)$–martingale. The event $A = \{w_2\} \times [0, \infty)$ is $\mathscr{G}_{1/2}$–measurable, but we have

$$\begin{aligned}\mathsf{E}[M_1\mathbf{1}_A] &= \frac{2}{1+\mathrm{e}^{-1}}\mathsf{P}[\{\tau > 1\} \cap A] = \frac{2}{1+\mathrm{e}^{-1}}\mathsf{P}[A] = \frac{1}{1+\mathrm{e}^{-1}} \\ &> \frac{1}{1+\mathrm{e}^{-1/2}} = \frac{2}{1+\mathrm{e}^{-1/2}}\mathsf{P}[A] = \mathsf{E}[M_{1/2}\mathbf{1}_A].\end{aligned}$$

Here we used the fact that $\tau(\omega_2, r) = r + 2 \geq 2$, for all $r \geq 0$, on A. Thus, M is not a $(\mathscr{G}_t)$–martingale, which is consistent with the assertion of Theorem D.1. □

References

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

Bruggeman, C. and Ruf, J. (2016). A one-dimensional diffusion hits points fast. *Electronic Communications in Probability*, 21(22):1–7.

Feller, W. (1968). On Müntz' theorem and completely monotone functions. *American Mathematics Monthly*, 75:342–359.

Perkowski, N. and Ruf, J. (2015). Supermartingales as Radon-Nikodym densities and related measure extensions. *The Annals of Probability*, 43(6):3133–3176.

Stroock, D. W. and Varadhan, S. R. S. (1972). On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability Theory*, pages 333–359. University of California Press, Berkeley, Calif.

Widder, D. (1946). *The Laplace Transform.* Princeton University Press, London.

Yoeurp, C. (1976). Décompositions des martingales locales et formules exponentielles. In *Séminaire de Probabilités, X*, pages 432–480. Springer, Berlin.