

Petyo Bonev*

Nonparametric Identification in Nonseparable Duration Models with Unobserved Heterogeneity

<https://doi.org/10.1515/jbnst-2024-0001>

Received January 4, 2024; accepted July 10, 2025

Abstract: This paper studies nonparametric identification of nonseparable duration models with unobserved heterogeneity. The models considered here 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. Identification is shown for a comprehensive account of settings. In particular, identification is shown in single-spell models with and without time-varying covariates, in multiple-spells 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;

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, Hausman and Woutersen (2008). 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. Furthermore, in models with competing risks, separability has been used to break the nonidentification result of Tsiatis (1975).

***Corresponding author: Petyo Bonev**, Agroscope and University of St. Gallen, St. Gallen, Switzerland, E-mail: petyo.bonev@unisg.ch. <https://orcid.org/0000-0002-6744-8374>

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, Jun et al. (2016). While unobserved effect heterogeneity has been addressed by numerous studies in the context of regression models (see e.g. Matzkin 2007), it has not been considered in duration context.

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

$$\theta(t | X, V) = \lambda(t, X)r(X, V), \quad (1)$$

where θ 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 λ) and (ii) arbitrary interaction between observed and unobserved covariates (through the function r). The functions λ and r are unknown to the econometrician and are not assumed to belong to a parametric family.

This paper shows identification strategies for a comprehensive account of empirically relevant settings. In particular, identification is shown in 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. The first one is based on the 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 paper is the first 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) and Chiappori et al. (2015), or at least on the

multiplicative separability of the unobserved heterogeneity as in the Mixed Hazard (MH) model, Brinch (2007) and McCall (1994).

The second strategy is based on multiple spells and a mild fixed-effects assumption that restricts unobserved heterogeneity to enter the hazard of at least two subsequent spells in the same way. The model is very general and does not require a finite mean of the unobserved heterogeneity. Moreover, previous spells are allowed to impact the hazard of subsequent spells (a so-called lagged duration dependence). Thus, these multiple-spell models can be interpreted as a hazard version of a dynamic panel model with fixed-effects. Importantly, the lagged duration effect is allowed to be arbitrary heterogeneous with respect to observed characteristics. This assumption is motivated by recent empirical findings in the unemployment duration literature, see e.g. Cockx and Picchio (2013). 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). The models in this paper also nest a shared frailty model, which is widely used in demographics and appropriate when multiple individuals from the same group share unobserved characteristics, Hougaard (2000). Existing multiple-spells 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 the fixed-effects assumption is sufficient to avoid these restrictive assumptions. In particular, the fixed-effects assumption is used to derive a hazard analogue of a difference-in-differences approach.

Next, identification of nonseparable competing risks models is studied. 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. Since only the first exit is observed, the marginal distributions are in general not identified. This paper shows that the nonseparable hazard structure (1) leads to identification when either there are multiple spells and the fixed-effects assumption is satisfied or when the researcher has access to a special regressor of the type discussed above. Previous studies have largely relied on the MPH structure, see Heckman and Honoré (1989), Abbring and Van Den Berg (2003a), Colby and Rilstone (2004) and Horny and Picchio (2010). One exception is the paper by Lee and Lewbel (2013) which nonparametrically identifies a bivariate accelerated failure time competing risk model.

This paper also contributes to the literature on treatment effects in the context of duration models. Much of this literature is discussed by Beyhum et al. (2024). In many

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.

studies, treatment is modeled as a binary variable, and is assumed that the treatment status realizes upon individuals entering into the state of interest (Blanco et al. 2020; Kastoryano and Beyhum 2020; Sant’Anna 2021). This assumption precludes many problems that are endemic to duration models. Most importantly, this assumption states that the treatment status is observed. This assumption may be too restrictive for many cases such as unemployment search assistance, in which the treatment may realize at any time of the elapsed duration and thus will be censored if an individual exists unemployment prior to treatment, and, which makes identification more complicated, censoring is endogenous since the exit is co-determined by individual unobserved characteristics. Recent papers that deal with endogenous censoring of a dynamic treatment are for example Van den Berg et al. (2020a, 2020b) and Beyhum et al. (2024). A common assumption of these papers, as well as of the literature that assumes the treatment to be binary and realized upon inflow, is that the effect of the treatment does not depend on the duration of exposure to the treatment. In that sense, the effect is assumed to realize instantaneously. This assumption is technically important. Relaxing it requires identifying both the hazard function of the time to treatment from endogenously censored data and the duration dependence on the time spent in treatment. A binary treatment that realizes right at the beginning of a spell avoids these challenges. However, this assumption may also be restrictive in many important cases. The success of an Active Labor Market Policy such as training will depend on how long an individual is trained (Vooren et al. 2019). The success of many psychological therapies may depend on the length of exposure to the therapy (American Psychological Association 2017). The seminal paper by Abbring and Van Den Berg (2003b) allows the time to treatment to be endogenously censored and also allows that the treatment effect depends on the exposure to the treatment. The cost of doing so is that it is necessary to identify the full model, i.e. the underlying hazard functions of two duration variable (main duration and time to treatment). To that end, Abbring and Van Den Berg (2003b) have relied on the separability of the MPH model. This paper generalizes their results to a nonseparable context. Specifically, drawing on the results for nonseparable competing risks models shown here, the paper demonstrates that the treatment is identified if the time to treatment and the time to exit follow an augmented bivariate nonseparable duration model based on (1). The major cost for this generality is that the paper falls short of suggesting estimation procedures.

In all models discussed in this paper, identification of the function r draws on insights from the literature on nonadditive 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 generalization 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.

Nonparametric estimation of nonseparable duration models combines two challenges. First, nonparametric estimation of duration models, in particular the mixed proportional hazard model, is nontrivial. Second, nonseparable econometric models in general are hard to estimate (reference are given below). This paper restricts its focus to identification, which is also the major limitation of the paper.

The paper is structured as follows. In Section 2, the nonseparable model (1) is formally introduced and motivated. Identification results in single- and multiple spell models are presented in Section 3. Section 4 presents identification of competing risks models and Section 5 presents identification of duration treatment effect models. Section 6 concludes. All proofs and further supplementary results are in the Online Appendix.

2 Model and Motivation

For illustrative purposes, we build our exposition on a labor market example.³ 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 identically distributed. Let $\theta(t)$ be the unconditional hazard of T_i at elapsed length of unemployment $t \geq 0$, $\theta(t) = \lim_{h \rightarrow 0} (P\{T \in [t, t+h] | T \geq t\} / h)$ (the index i is omitted 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. Time-varying covariates are considered 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 start of the individual spell.

Henceforth, $\theta(t|X, V)$ denotes the individual hazard with notation in analogy to conditional probabilities. Conditionally on X and V , it is assumed fully specified. The

³ Throughout the paper, while empirical examples that aim to introduce a setup are incorporated into the text flow, examples that aim to justify or clarify a model choice or assumption are introduced as separate enumerated text entities.

empirical (or observed) hazard is denoted by $\theta(t|X)$. 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. This paper follows both Lancaster (1979) and Heckman (1991) and assumes that the residual randomness is due to idiosyncratic factors that 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 is referred here to as the **generalized mixed hazard (GMH) model**:

$$\theta(t|X, V) = \lambda(t, X)r(X, V) \quad (2)$$

The functions λ and r are unknown to the econometrician. Two important special cases are the mixed proportional hazard (MPH) model,

$$\theta(t|X, V) = \lambda_{\text{MPH}}(t)\theta_0(X)V \quad (3)$$

and the mixed hazard (MH) model,

$$\theta(t|X, V) = \lambda_{\text{MH}}(t, X)V. \quad (4)$$

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 λ to depend on observed covariates.

Section A in the Appendix briefly discusses the relation of the GMH model to other existing models. In a nutshell, the accelerated failure time model and its generalizations the generalized AFT by Ridder (1990) and the extended generalized AFT by Brinch (2011) are not nested in the GMH model. Conversely, the GMH model is also not nested in those models.

The following examples motivate the GMH model.

Example 1: unobserved treatment heterogeneity. Van den Berg and Van der Klaauw (2006) evaluate the effect of counseling and monitoring 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 counseling and monitoring ($Z = 1$) 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 implies that the effect of the training does not depend on unobserved noncognitive abilities such as locus of control or on intrinsic motivation. However, recent empirical and theoretical evidence suggests that individuals with higher levels of locus of control tend to make higher effort when searching for a job Caliendo et al. (2015).⁴ This finding suggests that individuals with higher locus of control may benefit more from the counseling and monitoring process through more active participation. The simplest model that can capture this relationship is

$$\theta(t|X, Z, V) = \lambda(t) \exp\{X\beta + \delta Z + \gamma Z \ln V + \ln V\}. \quad (5)$$

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

Example 2: heterogeneous duration dependence. The baseline hazard function λ_{MPH} in model (3) 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 (Nüß 2018). For example, screening models predict that employers use the length of the current unemployment spell as a 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

⁴ To establish this result, the paper of Caliendo et al. (2015) links survey measurements on non-cognitive abilities to job arrival rate for newly unemployed individuals in Germany.

across levels of tightness of the local labor market. These empirical findings cannot be incorporated in a separable model. In particular, separability of λ and X implies that individual characteristics only shift the level of duration dependence.

The nonseparable duration dependence $\lambda(t, X)$ in model (2), on the contrary, allows for a flexible interaction between time and observed characteristics. As an example, consider the Weibull specification

$$\lambda(t, x) = \lambda_0 \lambda_1(x) t^{\lambda_1(x)-1} \quad (6)$$

with a scale parameter λ_0 and a shape parameter λ_1 whose value is now allowed to depend on the value of the observed characteristics.⁵ The findings of Farber et al. (2016) can be modeled as $\lambda_1 < 1$ for shorter employment histories and $\lambda_1 = 1$ for long employment histories. Alternatively, flexibility can be achieved in a piecewise-constant baseline hazard, in which the coefficients depend on x .

Example 3: heterogeneous measurement 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 resulting from a 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, which is now demonstrated. Denote by T^* the latent duration variable which might be measured imprecisely. Assume that the observed duration T satisfies

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

where η is a measurement error, k is some unknown function and $V \perp (X, T^*)$. Let the conditional distribution of the latent variable follow a Weibull specification, $P\{T^* \leq t | X = x\} = 1 - \exp\{-t^\alpha \psi(x)\}$ with α being a (unknown) scalar and ψ an unknown function. Then the individual hazard of T satisfies $\theta(t|x, v) = \alpha t^{\alpha-1} \psi(x) k(x, v)^\alpha$, which is a special case of model (2) with $\lambda(t, x) = \alpha t^{\alpha-1} \psi(x)$ and $r(x, v) = k(x, v)^\alpha$.

⁵ The Weibull specification in which λ_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.

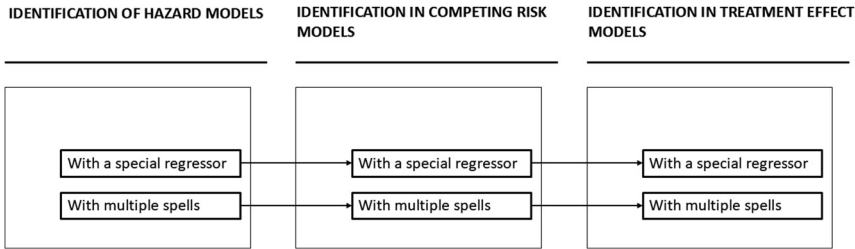


Figure 1: Overview of identification models and strategies.

Roadmap. The intuition behind the choice of models/settings in this paper is the following. The ultimate goal of the paper is to show identification of nonseparable treatment effect models, in which a treatment S is administered at random points and maybe censored by the outcome T . It will be thus unobserved for those individuals who exit the state of interest before being treated. A major insight here is that before the treatment kicks in (i.e., for realizations t, s of T, S with $t < s$), the bivariate hazard model of T, S is a competing risk model. Thus, identification in competing risks models need to be studied before identification in treatment effects models. In turn, the building blocks of a competing risk model are hazard models for each of the competing risks. The paper thus presents first two different strategies for identification of hazard models of the type (2). The first strategy consists of a special regressor that is modelled separably in certain sense. The second strategy relies on multiple spells. Then the paper applies these strategies to competing risks models, and ultimately, applies the result for competing risks models to identify treatment effect models. This procedure is depicted in Figure 1.

3 Identification Under Random Right Censoring

3.1 Two Basic Assumptions

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 . In this section, T is assumed to be observed in the sense that its survival function $S(t|x) = P(T > t | X = x)$ can be consistently estimated. This definition of observability allows for random right-censoring. Furthermore, following a convention in the analysis of identification, the paper ignores sample error: the distribution of the observables $F_{T,X}$ is assumed to be known to the

econometrician, see e.g. Lewbel (2019). We call a tripple $\mathcal{S} = (\lambda, r, F_{X,V})$ that satisfies (2) 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 λ , 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 (2) is identified under a given set of assumptions, if no two structures $\mathcal{S}, \mathcal{S}', \mathcal{S} \neq \mathcal{S}'$ that satisfy these assumptions are observationally equivalent.

Consider the following two assumptions.

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

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 state of interest (e.g., unemployment a certain disease, etc.), 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 with respect to time fulfills

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

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. A1 is trivially fulfilled for the MPH and MH models.

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 $\tilde{r} : \mathbb{X} \times \mathbb{R}^+ \rightarrow \mathbb{R}$, $\tilde{r}(x, y) = r(x, G^{-1}(y))$. Furthermore, let $\tilde{V} = G(V)$. Then \tilde{V} is uniformly distributed on $[0, 1]$, \tilde{r} is nonnegative and strictly increasing in its second argument, and it holds

$$\tilde{r}(x, \tilde{V}) = r(x, V) \tag{7}$$

for each $x \in \mathbb{X}$. Thus, a normalization assumption is necessary.

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

Assumption A2 (i) is similar to normalization assumptions in nonseparable regression context, see Matzkin (2003). Assumption A2 (ii) is a standard assumption in the literature on hazard mixture models. Existing single-spell identification results crucially depends on it. Honoré (1993) shows that multiple-spell MH models are identified without imposing A2 (ii). This point is discussed in Section 3.3 below.

Example 1 (unobserved treatment heterogeneity) continued. The scale-location model (5) satisfies assumption A1 under the restriction $\gamma Z \geq 0$.

Under A1 and A2, identification boils down to retrieving λ 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 λ and r , it is possible to shift separable components depending on X between λ and r without changing the DGP. As an example, the structures (λ_1, r_1) , (λ_2, r_2) with

$$\lambda_1(t, x) = \psi(t) \exp\{\beta_1 x_1 + \beta_2 x_2\}, \quad r_1(x, v) = \exp \beta_3 x_3 + v \quad (8)$$

$$\lambda_2(t, x) = \psi(t) \exp\{\beta_1 x_1\}, \quad r_2(x, v) = \exp \beta_2 x_2 + \beta_3 x_3 + v \quad (9)$$

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. The following lemma states the (well understood) fact that the MH model is not identified without further assumptions.

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 $\tilde{V} = r(V)$. Denote by \tilde{G} the distribution of \tilde{V} . Then, there exists a generalized hazard $\bar{\lambda}$ and a distribution \tilde{G} of a nonnegative random variable \tilde{V} , such that the structures (λ, \tilde{G}) and $(\bar{\lambda}, \tilde{G})$ are observationally equivalent.

Thus, identification is hampered by two distinct problems. The first one arises from the interplay of λ and r and the second one from the MH model. We present solutions to these problems in the next sections.

3.2 Nonparametric Identification in Single-Spell Models

3.2.1 Identification Under a Special Regressor Assumption

At the core of the single-spell approach presented in this paper is the assumption that at least one regressor, say X_1 , can 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

$$r(x_1, x_2, v) = r(x_1^*, x_2, v) \quad (10)$$

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 (unobserved treatment heterogeneity) 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 assigned to the individual. Evidence for interaction of labor market treatments and characteristics of the case worker 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 with respect to most of the case worker characteristics. For example, the estimated effects are homogeneous with respect to tenure, age and gender of the case worker, as well as with respect to whether case worker and the unemployed have the same gender. These results could be used to make an informed model choice when case worker characteristics are unobserved.

For the further discussion, the following additional notation is necessary. For a fixed x_2 , define $\Lambda_{x_2} \cdot = \Lambda(\cdot, \cdot, x_2)$ and $V_{x_2} := r(x_2, V)$. Further, let G_{x_2} and \mathcal{L}_{x_2} be the distribution of V_{x_2} and the corresponding Laplace transform.

To characterize identification under assumption A3, consider the x_2 -strata model (Λ_{x_2}, G_{x_2}) , where x_2 is a given value of X_2 . This sub-model can be identified using known results for the identification of the MH model. Such a procedure is common in empirical studies and not specific to duration models: models for subpopulations, indexed by the values of some regressor, are identified and estimated separately. As an example, researchers almost always estimate unemployment duration models separately for men and women.

However, it is not clear that this procedure says anything about the causal effect of the variable used to index the subpopulations. As an example, studies on stigma effects within groups with different education levels are silent on the effect of education on stigma, see example 2 above for references. The following proposition states that under assumptions A1, A2 and A3 (and under additional regularity

assumptions), identifying the submodels (Λ_{x_2}, G_{x_2}) can be used to identify the function r – and thus the causal effect of the indexing variable X_2 .

Proposition 1: (Identification with a special regressor). *Suppose that model (2) holds.*

(i) *If λ is known and the moment generating function of $r(x, V)$ is finite for all $x \in \mathbb{X}$, then under assumptions A1 and A2, the function r is identified.*

(ii) *Suppose that assumptions A1–A3 are satisfied and that the moment generating function of $r(x_2, V)$ is finite for all $x_2 \in \mathbb{X}_2$. Then model (2) is nonparametrically 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 (λ_{x_2}, G_{x_2}) is nonparametrically identified.*

To interpret part (i) of Proposition 1, transform model (2) in the following way

$$\ln \int_0^T \theta(t|X, V) dt = \ln \int_0^T \lambda(t, X) dt + \ln r(X, V). \quad (11)$$

By rearranging, we obtain

$$Y = k(X, V) + \varepsilon, \quad (12)$$

where $Y := \ln \int_0^T \lambda(t, X) dt$, $k(X, V) := -\ln r(X, V)$ and $\varepsilon := \ln \varepsilon = \ln \int_0^T \theta(t|X, V) dt$. Y is observed by assumption. The error term ε is a unit exponential random variable and is independent of X and V .⁶ Note that the identification result remains valid under any known distribution of ε as long as its moment generating function is finite. Thus, Proposition 1 implies that the identification result of Matzkin (2003) continues to hold if we add a second source of unobserved stochastic variation with a known distribution. Hence, the doubly-stochastic regression model (12) with a known (or identified) transformation behaves as a standard nonseparable regression model. The assumptions needed for identification are identical in those two cases. The distribution of $k(X, V)$ can be deduced from the distributions of Y and ε . Under the normalization assumption A2 (i), knowledge of the distribution of k identifies k .

⁶ This statement follows from $F(T|x, v) \sim U[0, 1]$ for any x, v , so that the random variable $F(T|X, V)$ is uniformly distributed and independent from X, V . Then, because the distribution function is linked to the hazard function via the relationship $\int_0^T \theta(t|X, V) dt = -\log(1 - F(T|X, V))$, it holds

$$\begin{aligned} P\left(\int_0^T \theta(t|X, V) dt < z\right) &= P(-\ln(1 - F(T|X, V)) < z) = P(1 - F(T|X, V) > \exp(-z)) \\ &= 1 - \exp(-z), \end{aligned}$$

where the last equality follows from the uniform distribution of $F(T|X; V)$.

The important statement behind part (ii) of Proposition 1 is that only strategies that identify all submodels indexed by x_2 on their own leads (under assumption A3) to identification of λ . One implication of this result is the following. Since each strategy to identify the nonseparable model must be able to identify each strata MPH model on its own, and since each strata-MPH model requires a normalization assumption, the nonseparable model requires a set of normalization assumptions, for each x_2 -strata one. I discuss this in detail in the next section.

3.2.2 Two Identification Results Using the Special Regressor Assumption

Results (i) and (ii) of Proposition 1 require the moment generating functions of $(V_{x_2})_{x_2 \in \mathbb{X}_2}$ to be finite. This considerably restricts the class of models for which these results hold. As an example, it rules out the log-normal distribution and the Gamma distributions for particular parameter values.⁷

We present now two strategies that complement assumption A3 and drop the requirement that the moment generating functions are finite. The choice of these strategies is motivated by two reasons. First, they involve well known sources of identification. The first one utilizes separable variation of at least one covariate. The second one uses time-varying covariates. Thus, practically any strategy to identify the MH model will be used to identify the nonseparable GMH model. Second, these two strategies differ on their success in identifying a competing risks model, which is discussed in Section 4 below.

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

$$\lambda(t, X) = \mu(t, X_2)\phi(X_1, X_2) \quad (13)$$

(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 (2) can be written as

$$\theta(t|X, V) = \theta(t|X_1, X_2, V) = \mu(t, X_2)\phi(X_1, X_2)r(X_2, V). \quad (14)$$

For a fixed $X_2 = x_2$, (14) reduces to an MPH model. Equation (14) can be therefore interpreted as a generalized MPH model. A4 (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

⁷ To be specific, let V be a random variable with a Gamma distribution $\Gamma(\alpha, \beta)$, where α is the shape parameter and β is the rate parameter. Then for $t > \beta$, the moment generating function $M_V(t)$ of V takes infinite values.

assumption 6b in Van den Berg (2001).⁸ The variation in ϕ 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 .

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 $S\{t_0|z_1, x_2\} \neq S\{t_0|z_2, x_2\}$, where $S(t|x_1, x_2)$ 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 the current value of X_1 “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).⁹ 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

$$\theta(t | X, V) = \theta(t | X_1(t), X_2, V) = \lambda(t, X_1(t), X_2)r(X_2, V). \quad (15)$$

⁸ The paper of Elbers and Ridder (1982) presents the first nonparametric identification result for the MPH model, relying on the properties of the Laplace transform and a finite mean of the frailty.

⁹ The paper of Brinch (2007) presents an identification result in a mixed hazard model, in which the duration dependence and the covariates do not enter the hazard function in a separable way. Identification is achieved via time-varying covariates.

There are two important implications of (15). 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, note that the hazard function at t does not depend on the total path of X_1 but only on its value $X(t)$. Thus, information about future values $X_1(\bar{t})$, $\bar{t} > t$ are not allowed to impact the hazard at t , unless they are accounted for by $X_1(t)$. This can be viewed as a no-anticipation assumption (Abbring and Van Den Berg 2003b).

Example 4: Job search. To give an example of a case with predictability and no anticipation, consider again the case, in which an individual is searching for a job, and let T be the duration until a new job is found while searching on the job and $X_1(t)$ the number of children. In certain cases, $X_1(t)$ might be random and depending on some extrinsic uncertainty that is beyond the domain of impact of the individual. As an example, this might be random variation in the number of children caused by the randomness of an in vitro fertilization (IVF) process.¹⁰ It holds $X_1(t) \perp V$ since the IVF process is idiosyncratic and hence 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 λ 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.

In the context of assumption A4, assumption A5 requires μ to be integrable for each $x_2 \in \mathbb{X}_2$. In the context of assumption A4', it requires the integral $\int_0^t \lambda(s, x_1(s), x_2) ds$ to exist for every finite positive t .

¹⁰ Lundborg et al. (2017) use exogenous variation in the number of children as an instrument to identify the effect of number of children on labor market outcomes. However, they do not consider duration outcomes.

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, μ and ϕ would be identified only up to a scale for each value x_2 . The following example from Abbring (2002) demonstrates this.

Example 2 (heterogeneous duration dependence) continued. Consider the Weibull specification in equation (6) and let X_2 be equal to some specific value x_2 . Under assumption 4, this specification is now equal to

$$\lambda(t, x_1, x_2) = \lambda_0 \lambda_1(x_2) t^{\lambda_1(x_2)-1}. \quad (16)$$

Assume that the function $\phi(x_1, x_2)$ is equal to $\exp\{x_1\beta_1 + x_2\beta_2\}$ with β_1, β_2 being deterministic vectors. Also, assume that for each x_2 , the random variable $V_{x_2} := r(x_2, V)$ follows a Gamma distribution with parameters that depend on the value x_2 . Fixing X_2 to a given value reduces the model to an MPH model. It can be easily shown (see derivations for the MPH model in Abbring 2002) that in this case, the observed survival function $S(t|X_1, X_2 = x_2)$ depends on the quantity $\frac{\lambda_0 \exp\{x_2\beta_2\}}{\alpha_{x_2}}$, where α_{x_2} is the scale parameter of the Gamma distribution corresponding to the value $X_2 = x_2$. In this case, the parameters $(\lambda_0, \beta_2, \alpha_{x_2})$ cannot be identified, only the total quantity $\frac{\lambda_0 \exp\{x_2\beta_2\}}{\alpha_{x_2}}$. Alternatively, to identify any of the three parameters, it is necessary to make a normalization assumption on the other two quantities. This is precisely what assumption 6 does. As analogy, this is a normalization similar to the normalization in a separable regression model, which assumes that the error term has an expectation equal to 0. The value 0 is set for convenience and can be replaced by any arbitrary value. In the nonseparable hazard model, the researcher might be willing to choose t^*, x_1^* independently of x_2 , for example $t^* = x^* = 0$.

The merit of assumption A6' is similar. If A6' is dropped, the class of observationally equivalent structures (Λ, r) can be described in the following way. For any $c > 0$, it holds

$$S(t | x_1, x_2) = L_{x_2}(\Lambda_{x_2}(t, x_1)) = \mathcal{L}_{x_2} \left(\frac{1}{c} (c\Lambda_{x_2}(t, x_1)) \right).$$

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

Finally, A7 is also a normalization assumption. Without A7, it can no longer be distinguished between duration dependence (baseline hazard) and unobserved heterogeneity in the strata models Ridder (1990). Yet, in many cases, this assumption

can be justified with economic theory. As an example, in job search models, individuals with very low hazard rates (i.e., those who are exceptionally bad at finding jobs) are eventually forced to exit the labor market, go on long-term welfare, or adjust their search strategies. This prevents an infinitely persistent pool of unemployed individuals, justifying a finite mean for heterogeneity.

With these assumptions, we can state the following results.

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

Proposition 3. *Under assumptions A1–A3, A4', A5 and A6', the GMH model (2) 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 λ_{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 (λ_{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).¹¹ The function r is identified over quantiles of the identified distributions $(G_{x_2})_{x_2 \in \mathbb{X}_2}$.

Remark 1. Note that under A4', the finite mean assumption A7 is not necessary for identification. This is not surprising in view of the identification results by Honoré (1990) for the MPH model and by Brinch (2007) for the MH model.¹²

3.3 Nonparametric Identification in Multiple-Spells Models

The identification strategy in this section relies on availability of multiple spells and on a fixed-effects assumption. Let T_1 and T_2 be two duration variables. The hazard functions of T_1 and T_2 are modeled as

$$\theta_1(t|X, V) = \lambda_1(t, X)r(X, V) \quad (17)$$

¹¹ An alternative approach to use time-varying covariates in order to identify the MH model is described in Ruf and Wolter (2019). Their approach relies on martingale properties of a correctly specified MH model.

¹² The paper by Honoré (1990) shows how the standard models Elbers and Ridder (1982) and Heckman and Singer (1984) can be generalized to multi-spell models with lagged duration dependence. The paper shows that if the model does not include lagged duration dependence, then no assumption on the tail or expectation of the frailty distribution is necessary for identification.

$$\theta_2(t|X, T_1 = t_1, v) = \lambda_2(t, X)\psi(t_1, X)r(X, V) \quad (18)$$

The hazard model for the first duration is the GMH model (2). 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. Notably, λ_1, λ_2 are allowed to be different. V and the function r , on the contrary, are restricted to be the same in both spells. In the following, this restriction is referred to as “fixed effects”.

We discuss two distinct setups of interest which differ w.r.t to the a priori assumption on ψ . 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 beginning 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. Thus, the spells are not required to be of the same type. The notation implies that X and V realize prior to the beginning of the first spell. The function ψ captures the so called lagged duration effect. Hence, model (17) and (18) 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). When T_1, T_2 represent unemployment spells, ψ is referred to as “scarring effects”. 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 θ_2 right from the beginning of the second spell.¹³ Note that random censoring of T_2 does not change the identification results below.

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 and Frederiksen et al. (2007) for an application in job duration context. 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 5, this assumption is relaxed. Note that

¹³ This model will be contrasted to the model in Section 5 where T_1 arrives as a surprise *during* the spell of T_2 .

random censoring of either T_1 and T_2 or both does not change the identification results below.

Following the conventions in the literature, the two setups are referred to 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, ψ can be restricted to be equal to 1. A similar empirical finding holds for the case when T_1 represents past employment spells and T_2 subsequent unemployment spells. Specifically, Doiron and Gørgens (2008) find that 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.

In the following, $\Lambda_j(t, x)$ denotes the integrated hazard of T_j , $\Lambda_j(t, x) := \int_0^t \lambda_j(s, x) ds$, $j = 1, 2$. Consider the following assumptions.

Assumption MS4. (i) For each $j = 1, 2$, the function λ_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 ψ takes only positive values and is differentiable in t for each $x \in \mathbb{X}$.

Assumption MS5. For each $x \in \mathbb{X}$ and $j = 1, 2$ there exists known $t_j^* = t_j^*(x) \in \mathbb{R}^+$ such that (i.) $\lambda_1(t_1^*, x)$ and $\Lambda_2(t_2^*, x)$ are known and (ii) $\psi(t_1^*, x), \partial_t \psi(t_1^*, x)$ are also known, where ∂_t denotes the derivative with respect to t .

Assumption MS4 (i) is an innocuous regularity assumption. It generalizes a standard assumption in mixture hazard models, see e.g. assumption 2 in Elbers and Ridder (1982). MS4 (i) can be relaxed by requiring that t can be varied such that $\Lambda_2(t, x)$ obtains all values in an open interval. MS4 (ii) requires smoothness of ψ . This is a mild assumption. In particular, consider the case in which T_1 is modeled as any other covariate in the hazard function of T_2 (i.e. T_1 is an element of the vector X). Then, under the standard MPH specification $\lambda(t, x) = h(t) \exp x\beta$, assumption MS4 (ii) is fulfilled.

MS5 (i) is a generalization of a standard scale normalization assumption. Specifically, under an MPH specification, MS5 (i) is equivalent to assumption 7 in Van den Berg (2001). It corresponds to assumption A6 in the context of single spell models. Assumption MS5 (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 Λ_2 is needed.

We can now state the main result of this section.

Proposition 4. (i) Under assumptions A1, MS4 and MS5, the functions λ_1 , λ_2 , ψ are identified. (ii) If in addition A2 (i) is satisfied, then also the function r is identified.

The intuition of this result can be stated as follows. Identification of ψ can be interpreted as a hazard version of a difference-in-differences approach, which proceeds in two steps. In a first step, the fixed-effects assumption is used to show that the difference of the marginal exit rates $\ln \rho(t_1, t_2|x) = \ln \partial_{t_1} P(T_1 > t_1, T_2 > t_2|x) - \ln \partial_{t_2} P(T_1 > t_1, T_2 > t_2|x)$ does not depend on V and r . This step corresponds to first-differencing in panel data models. Thus, the first difference $\ln \rho$ is taken with respect to different periods and eliminates the impact of the unobserved heterogeneity. The second difference $\rho(t, \cdot) - \rho(0, \cdot)$ is taken with respect to different points in time of the first spell and eliminates the impact of the duration dependence λ_1 of the first hazard.

Proposition 4 is now compared to the existing literature on identification of duration models with multiple spells. I compare the following features of my result: nonseparability, heterogeneous duration dependence, whether fixed effects are assumed, whether independence between X and V is required, whether a finite mean of V is assumed, and how many spells are necessary for identification.

First and most importantly, the only other study that allows observed and unobserved factors to enter the hazard in a nonseparable way is Evdokimov (2010). His model, however, rules out lagged duration dependence.

Second, as in Picchio (2012), the lagged duration dependence ψ in model (18) is allowed to depend in an arbitrary way on observed characteristics X .

Third, a fixed-effects assumption is adopted also by Picchio (2012) and in the shared frailty model (model 1) in Honoré (1993), while Evdokimov (2010) and model 3 in Honoré (1993) allow for spell-specific unobservables.

Fourth, Proposition 4 allows X and V to be dependent. This result is of particular importance since independence is often hard to defend. Furthermore, no finite mean assumption on the unobservables V_x is required. These two features – allowing for dependence and for infinite mean of V_x – are not surprising. They follow from the fixed-effects assumptions and are shared by the other two papers that assume fixed effects. However, model 1 in Honoré (1993) does not allow for lagged duration dependence and Picchio (2012) assumes that there are recurrent data for each outcome (i.e. at least two observations for each T_i , $i = 1, 2$). Independence and finite mean of unobservables are also imposed by the lagged duration model (model 3) in Honoré (1993). Note that the paper of Evdokimov (2010) allows for dependence and an infinite mean of V without imposing fixed effects. The price for the generality of his model is that his identification result relies on access to at least three spells for each

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/aspect of identification. (a) Model is nonseparable in X and V . (b) Independence of X and V not needed. (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 adopted. (g) Identification relies on more than two spells per individual.

individual and that (as noted above) lagged duration dependence is precluded. Finally, both the model of Evdokimov (2010) and model 3 in Honoré (1993) allow the unobserved heterogeneity to differ between spells. Thus, their models are richer than a standard shared-frailty model.

Table 1 provides a summary of this discussion. Each row corresponds to a given paper and each column corresponds to a distinct feature/assumption. As an example, column 1, row 1 states that the model in Evdokimov (2010) is allowed to be non-separable in X and V .

4 Identification Under Endogenous Censoring: The Case of 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. 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. Put differently, the duration in each destination is potentially censored by the exit in a different destination. This censoring case is more involved than the case of random right censoring considered in the previous section. Specifically, since the distribution functions of each of the competing risks might depend on the same unobserved features of the individual, censoring is potentially endogenous. Without further structure, the “observed” joint (bivariate) survival function $S(t_1, t_2)$ is nonparametrically nonidentified, Tsiatis

(1975), which greatly complicates identification of the underlying structural model. This section studies identification of competing risks models when the hazard of each of the latent variables follows the GMH model (2).

Single-spell setting. Consider a setting with 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 $1\{T_A < T_B\}$ 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 $P\{T_i > t, T_j > t_j\}$, $i, j = A, B$, $i \neq j$. 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 $X \perp (V_A, V_B)$. This is the standard setup in competing risks models with covariates, see Heckman and Honoré (1989). The hazards θ_A and θ_B of T_A and T_B , respectively, are modeled as

$$\theta_A(t|X, V_A, V_B) = \theta_A(t|X, V_A) = \lambda_A(t, X)r_A(X, V_A) \quad (19)$$

$$\theta_B(t|X, V_A, V_B) = \theta_B(t|X, V_B) = \lambda_B(t, X)r_B(X, V_B). \quad (20)$$

Notably, r_i and λ_i are allowed to be risk-specific, i.e. it is not required that $\lambda_A = \lambda_B$ or $r_A = r_B$.

The following result states that the GMH structure identifies the single-spell competing risk model when there is a special regressor (assumption A3) which is common for both risks A, B and when in addition this special regressor is separable from the duration dependence function λ (assumption A4).

Proposition 5. Suppose that V_A, V_B fulfill assumptions A2 and A7, and let each of the hazards θ_A, θ_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, μ_i, θ_i, r_i , $i = A, B$ are identified from the data.

Assumption A4 (ii)′ modifies assumption A4 (ii) to the competing risks setting. It can be relaxed along the lines of the discussion of assumption A4 (ii) above. Note that identification of ϕ_A, ϕ_B does not require the variation condition A4 (ii)′ and independence of X and V_A, V_B .

Discussion on nonidentification. Unlike assumption A4, assumption A4′ cannot be used to obtain identification of the competing risks model. The following discussion should provide the intuition. Due to the censoring problem, identification

relies on the identified crude survival functions. These functions vary along a single time argument t . Identification with two risk-specific unobservables, however, requires either variation in two time arguments on an open subset of $\mathbb{R}^+ \times \mathbb{R}^+$ or separable variation in x . 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'. This implies that when there is only one spell per individual, assumptions A3 and A4 are necessary for identification.

Multiple-spells setting. Consider now a setting with competing risks and multiple spells. For $i = A, B$, $T_{i,1}, T_{i,2}$ denote two consecutive spells of the latent variable under risk i . The hazards of the latent variables are modeled as

$$\theta_{i1}(t|X, V_i) = \lambda_{i1}(t, X)r_i(X, V_i) \quad (21)$$

$$\theta_{i2}(t|X, T_{i1} = t_{i1}^*, V_i) = \lambda_{i2}(t, X)\psi_i(t_{i1}^*, X)r_i(X, V_i). \quad (22)$$

The researcher observes the tripple $(X, \tilde{T}_k = \min\{T_{Ak}, T_{Bk}\}, 1\{T_{Ak} < T_{Bk}\})$, $k = 1, 2$. (T_{A1}, T_{A2}) are assumed 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} . Specification (21) and (22) contains a risk-specific fixed-effects assumption. In particular, for each risk i , the unobservables V_i and the generalized error function r_i are the same for both spells. However, both the unobservables and the function r are allowed to differ between risks. Furthermore, as in the single-spell competing-risks setup, V_A and V_B are allowed to be dependent.

The following proposition states that the fixed effects assumption used to identify the multiple-spells model in the previous section leads to identification in a competing risk setting.

Proposition 6. Assume that both (T_{A1}, T_{A2}, X, V_A) and (T_{B1}, T_{B2}, X, V_B) satisfy the conditions of Proposition 4. (i) Then for $i = A, B, k = 1, 2$, λ_{ik} and ψ_i are identified from the data. (ii) If in addition the following condition holds:

MS4 (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.

Proposition 6 is based on similar assumptions and a similar difference-in-differences approach as Proposition 4 and so the intuition is common for both results. In particular, the use of the fixed-effects assumption is that the ratio of derivatives of

crude survival functions is free of r and V . Furthermore, distinguishing between λ_{i1} and ψ_i is possible, because the ratio of the subsurvival functions of the two spells depends on ψ_i in a way that varies with the elapsed duration in the second spell, which is not the case for λ_{i1} . Identification of r_i is achieved after the identification of λ_{i1} and ψ_i .

The closest setup to the one considered in Proposition 6 is the one studied in Horny and Picchio (2010). They use multiple spells to prove identification in a competing risk model with a lagged duration dependence. Their result generalizes the single-risk results by Honoré (1993). In the models of Honoré (1993) and Horny and Picchio (2010), separable variation in X is crucial to identify the joint distribution of the unobservables. Identification of this distribution is necessary to distinguish between the lagged duration dependence ψ_i and the genuine duration dependence in the first spell λ_{i1} . The separable variation of X is provided by the structure of the MPH model. Model (21) and (22) does not require such variation.

A model that uses a fixed-effects assumption in a multiple-spell MPH context is developed in Abbring and Van Den Berg (2003a). The precise assumption there is $V_{i1} = V_{i2}$. However, model (21) and (22) is substantially richer. In particular, X and V can interact nonseparably and a lagged duration dependence is allowed. Both these aspects are not allowed in Abbring and Van Den Berg (2003a).

5 Nonparametric Identification of Treatment Effects when the Treatment is Assigned During the Spell

Consider a setup, in which a treatment is assigned during the ongoing spell in the state of interest. As an example, ALMPs such as subsidized temporary work in the public sector are administered during unemployment. Related setups arise in epidemiology when a given medical treatment, such as kidney transplantation, is administered during an ongoing health condition.¹⁴ In such cases, the effect might vary with time to and duration of exposure.

To formalize this setup, let the random variable S denote time since entry into the state of interest (e.g. unemployment) until the start of treatment exposure. Contrary to X , V , which realize prior to or at begin of the spell in the state of interest, S realizes during the spell. As before, T denotes the duration of the underlying condition. In the example above, T is the duration of unemployment and S is the random

¹⁴ See De Luna and Johansson (2010) for a dynamic evaluation of the effect of kidney transplantation on life duration in discrete times.

time from the begin of unemployment until the individual starts temporary work as part of an ALMP. Define the vector $V := (V_T, V_S)$, where V_T, V_S are two unobserved scalar random variables. Consider the following bivariate hazard model:

$$\theta_T(t|S=s, X=x, V) = \begin{cases} \lambda_T(t, x)r_T(x, V_T) & \text{if } t < s \\ \lambda_T(t, x)\delta(t, s, x)r_T(x, V_T) & \text{if } t \geq s. \end{cases} \quad (23)$$

$$\theta_S(t|X=x, V) = \lambda_S(t, x)r_S(x, V_S). \quad (24)$$

In equation (24), the hazard $\theta_S(t|X=x, V)$ of the treatment variable S is specified as the nonseparable model (2). θ_S depends on V only through the S -specific unobservables V_S . Equation (23) specifies the hazard of T . It is an augmented GMH model. Before the treatment is administered ($t < s$), this is simply the GMH model (2). After the begin of the exposure, the hazard function is modified by the component δ , which can be interpreted as a treatment effect. In particular, conditionally on X and V , S can impact T (only) through δ . δ is allowed to depend on time to exposure s on its duration. The latter is inferred from the difference between t and s . δ is also allowed to depend on observed heterogeneity x . V_S and V_T might be dependent, so that the model allows for endogenous selection into the treatment. Furthermore, the functions r_T, r_S can be different (and analogously for λ_T, λ_S).

Model (23) and (24) is a generalization of the widely-used treatment effect model by Abbring and Van Den Berg (2003b). Identification in Abbring and Van Den Berg (2003b) relies on separable MPH structure of each of the two hazards θ_T, θ_S – an assumption that is relaxed here.

For the further discussion, it is informative to review the differences between the treatment effect model (23), (24) and the lagged duration model (17) and (18). The latter imposes that the two durations T_1, T_2 realize sequentially. This implies that both durations are observed. In the treatment effect model, S is potentially censored by T . Moreover, in model (17) and (18), the hazard of T_2 depends on T_1 through ψ for potentially all values t_1, t_2 (i.e. ψ might be $\neq 1$ for all t_1, t_2). In the treatment effect model, on the other hand, δ in (23) is equal to 1 when $s > t$. This condition is plausible when the treatment is not anticipated by the individuals. As an example, ALMPs often have random unanticipated variation that arises from the random order in which case workers treat the unemployed, see e.g. Sianesi (2004). Alternatively, individuals might be aware of the future time of the treatment but they might not be able to act upon this information. In the kidney example, patients might have little other choice but to wait for their turn to come.

Identification of the treatment effect δ proceeds in two steps. First, consider model (23) and (24) for values $t < s$. For such values, the model represents a GMH competing risks model. Under the assumptions from Section 4, $\lambda_j, r_j, j = T, S$ and the

joint distribution of $r_T(X, V_T)$, $r_S(X, V_S)$ can be identified. This is stated in the following proposition.

Proposition 7. *Suppose that for $t < s$, the bivariate GMH competing-risks model (23) and (24) is characterized by the assumptions of either Propositions 5 or 6. Then, λ_j , r_j , $j = S, T$ and the joint distribution of $r_T(X, V_T)$ and $r_S(X, V_S)$ are identified from the data.*

In a second step, knowledge of the components of the competing risks model is used to identify the treatment effect δ . The following result takes a general standpoint and assumes directly knowledge of these components (i.e. the source of this knowledge is ignored).

Proposition 8. *Assume that the functions λ_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 δ is identified from the data.*

Corollary 1. *Suppose that the assumptions of Proposition 7 are satisfied. Then, the treatment effect δ is identified.*

The proof of Proposition 8 is novel. Intuitively, when the competing risk model is identified, the derivative of the joint survival function of S and T with respect to t (which can be identified through its relationship to the crude survival functions) can be linked through a differential equation to the treatment effect δ . This differential equation is shown to have a unique solution, which leads to identification. Thus, it is the change in the marginal hazard $\partial P(T > t, S > s)/\partial t$ over elapsed durations t that identifies the treatment effect. On the contrary, Abbring and Van Den Berg (2003b) use variation of the marginal exit rate with respect to s , $\partial P(T > t, S > s)/\partial s$. Their strategy requires smoothness of the treatment effect. This assumption is not necessary here.

6 Discussion

This paper has provided identification results for nonseparable duration models. These 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, identification has been shown 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. The latter contribution is also the paper's most important one. Proposition 8 and its Corollary 1 relax common assumptions on the values and timing on the treatment. Specifically, while most papers require the treatment to be binary and to realize at the beginning of the duration spell, this paper allows both the treatment to be continuous and to realize at later points in time. The former aspect is important when the duration of exposure to treatment matters. The latter aspect is relevant for most practical applications: hardly any treatment is applied right at the onset of a disease; no job searching is administered on day 1 of unemployment. Building on Abbring and Van den Berg (2003b), who allow for endogenous treatment timing and exposure-dependent effects under separability, this paper extends identification results to a nonseparable setting. Specifically, it shows that treatment effects can still be identified in a bivariate nonseparable duration model.

A natural follow-up of these results would be to develop estimation techniques and to additionally allow unobserved heterogeneity to interact with duration dependence. These topics remain for future research where the last equality follows from the uniform distribution of $F(T|X, V)$.

Acknowledgments: 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 Efthraimidis for his helpful ideas and suggestions.

Conflict of interest: The author reports there are no competing interests to declare.

References

- Abbring, J. H. 2002. *Econometric Duration and Event-History Analysis*. Lecture Notes, University of Chicago.
- Abbring, J. H., and G. J. Van Den Berg. 2003a. "The Identifiability of the Mixed Proportional Hazards Competing Risks Model." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (3): 701–10.
- Abbring, J. H., and G. J. Van Den Berg. 2003b. "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71 (5): 1491–517.
- Abrevaya, J., and J. A. Hausman. 1999. "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells." *Annals of Economics and Statistics* (55–56): 243–75.
- Ait-Sahalia, Y. 2007. *Estimating Continuous-Time Models with Discretely Sampled Data*, volume 3 of *Econometric Society Monographs*, 261–327. Cambridge University Press.

- American Psychological Association. 2017. "How Long Will it Take for Treatment to Work." *American Psychological Association*. <http://www.apa.org/ptsd-guideline/patients-and-families/length-treatment> (accessed April 22, 2023).
- Andersen, P., O. Borgan, R. Gill, and N. Keiding. 1996. "Statistical Models Based on Counting Processes." In *Springer Series in Statistics*. Springer New York.
- Bergström, R., and P.-A. Edin. 1992. "Time Aggregation and the Distributional Shape of Unemployment Duration." *Journal of Applied Econometrics* 7 (1): 5–30.
- Beyhum, J., S. Centorrino, J.-P. Florens, and I. Van Keilegom. 2024. "Instrumental Variable Estimation of Dynamic Treatment Effects on a Duration Outcome." *Journal of Business & Economic Statistics* 42 (2): 732–42.
- Blanco, G., X. Chen, C. A. Flores, and A. Flores-Lagunes. 2020. "Bounds on Average and Quantile Treatment Effects on Duration Outcomes Under Censoring, Selection, and Noncompliance." *Journal of Business & Economic Statistics* 38 (4): 901–20.
- Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers. 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: 349–54.
- Brinch, C. N. 2011. "Non-Parametric Identification of the Mixed Proportional Hazards Model with Interval-Censored Durations." *The Econometrics Journal* 14 (2): 343–50.
- Caliendo, M., D. A. Cobb-Clark, and A. Uhlendorff. 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. London: Centre for Microdata Methods and Practice.
- Chesher, A. 2003. "Identification in Nonseparable Models." *Econometrica* 71 (5): 1405–41.
- Chesher, A. 2007. *Identification of Nonadditive Structural Functions*, volume 3 of *Econometric Society Monographs*, 1–16. Cambridge University Press.
- Chiappori, P.-A., I. Komunjer, and D. Kristensen. 2015. "Nonparametric Identification and Estimation of Transformation Models." *Journal of Econometrics* 188 (1): 22–39.
- Cockx, B., and M. Picchio. 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): 951–80.
- Colby, G., and P. Rilstone. 2004. "Nonparametric Identification of Latent Competing Risks Models." *Econometric Theory* 20 (5): 883–90.
- De Luna, X., and P. Johansson. 2010. "Non-Parametric Inference for the Effect of a Treatment on Survival Times with Application in the Health and Social Sciences." *Journal of Statistical Planning and Inference* 140 (7): 2122–37.
- Doiron, D., and T. Gørgens. 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 P. Lucchino. 2018. "Young People's Labour Market Transitions: The Role of Early Experiences." *Labour Economics* 54: 29–46.
- Elbers, C., and G. Ridder. 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *The Review of Economic Studies* 49: 403–9.
- Eriksson, S., and D.-O. Rooth. 2014. "Do Employers Use Unemployment as a Sorting Criterion when Hiring? Evidence from a Field Experiment." *The American Economic Review* 104 (3): 1014–39.
- Evdokimov, K. 2010. "Nonparametric Identification of a Nonlinear Panel Model with Application to Duration Analysis with Multiple Spells." Discussion Paper.
- Farber, H. S., D. Silverman, and T. von Wachter. 2016. "Determinants of Callbacks to Job Applications: An Audit Study." *The American Economic Review* 106 (5): 314–8.

- Frederiksen, A., B. E. Honoré, and L. Hu. 2007. "Discrete Time Duration Models with Group-Level Heterogeneity." *Journal of Econometrics* 141 (2): 1014–43.
- Frijters, P. 2002. "The Non-Parametric Identification of Lagged Duration Dependence." *Economics Letters* 75 (3): 289–92.
- Hausman, J. A., and T. M. Woutersen. 2008. "The Proportional Hazard Model." In *The New Palgrave Dictionary of Economics*, edited by S. N. Durlauf, and L. E. Blume. Basingstoke: Palgrave Macmillan.
- He, X. 1997. "Quantile Curves Without Crossing." *The American Statistician* 51 (2): 186–92.
- Heckman, J. J. 1991. "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity." *The American Economic Review* 81: 75–9.
- Heckman, J., and G. Borjas. 1980. "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* 47 (187): 247–83.
- Heckman, J. J., and B. E. Honoré. 1989. "The Identifiability of the Competing Risks Model." *Biometrika* 76 (2): 325–30.
- Heckman, J. J., and B. Singer. 1984. "The Identifiability of the Proportional Hazard Model." *The Review of Economic Studies* 51: 231–41.
- Honoré, B. E. 1990. "Identification Results for Duration Models with Multiple Spells or Time-Varying Covariates." Part of this paper was later published in the 1993 Review of Economics Studies paper.
- Honoré, B. E. 1993. "Identification Results for Duration Models with Multiple Spells." *The Review of Economic Studies* 60 (1): 241–6.
- Horný, G., and M. Picchio. 2010. "Identification of Lagged Duration Dependence in Multiple-Spell Competing Risks Models." *Economics Letters* 106 (3): 241–3.
- Horowitz, J. L. 1999. "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity." *Econometrica* 67: 1001–28.
- Hougaard, P. 2000. *Analysis of Multivariate Survival Data*. Heidelberg: Springer.
- Jun, S. J., Y. Lee, and Y. Shin. 2016. "Treatment Effects with Unobserved Heterogeneity: A Set Identification Approach." *Journal of Business & Economic Statistics* 34 (2): 302–11.
- Kalbfleisch, J., and R. Prentice. 1980. "The Statistical Analysis of Failure Time Data." In Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.
- Kastoryano, S., and J. Beyhum. 2020. "Decomposing Causal Mechanisms in Duration Models with Unobserved Heterogeneity".
- Knaus, M., M. Lechner, and A. Strittmatter. 2017. "Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach." *arXiv e-prints*: arXiv:1709.10279.
- Kroft, K., F. Lange, and M. J. Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment." *The Quarterly Journal of Economics* 128 (3): 1123–67.
- Kyyrä, T., and V. Ollikainen. 2008. "To Search or Not to Search? The Effects of UI Benefit Extension for the Older Unemployed." *Journal of Public Economics* 92 (10-11): 2048–70.
- Lancaster, T. 1979. "Econometric Methods for the Duration of Unemployment." *Econometrica* 47: 939–56.
- Lancaster, T. 1985. "Simultaneous Equations Models in Applied Search Theory." *Journal of Econometrics* 28 (1): 113–26.
- Lancaster, T. 1990. *The Econometric Analysis of Transition Data*. Econometric Society Monographs. Cambridge: Cambridge University Press.
- Lee, S., and A. Lewbel. 2013. "Nonparametric Identification of Accelerated Failure Time Competing Risks Models." *Econometric Theory* 29 (5): 905–19.
- Lewbel, A. 1998. "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors." *Econometrica* 66 (1): 105–21.

- 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): 733–53.
- Lundborg, P., E. Plug, and A. W. Rasmussen. 2017. "Can Women Have Children and a Career? IV Evidence from IVF Treatments." *The American Economic Review* 107 (6): 1611–37.
- Mathiowetz, N. A., and G. J. Duncan. 1988. "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment." *Journal of Business & Economic Statistics* 6 (2): 221–9.
- Matzkin, R. L. 2003. "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica* 71 (5): 1339–75.
- Matzkin, R. L. 2007. "Chapter 73 Nonparametric Identification." *Handbook of Econometrics* 6: 5307–68.
- McCall, B. P. 1994. "Identifying State Dependence in Duration Models. American Statistical Association 1994." *Proceedings of the Business and Economics Section* 14.
- Nüß, P. 2018. "Duration Dependence as an Unemployment Stigma: Evidence from a Field Experiment in Germany." Technical report, Economics Working Paper.
- Picchio, M. 2012. "Lagged Duration Dependence in Mixed Proportional Hazard Models." *Economics Letters* 115 (1): 108–10.
- 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): 167–81.
- Ruf, J., and J. L. Wolter. 2019. "Nonparametric Identification of the Mixed Hazard Model Using Martingale-Based Moments." *Econometric Theory* 1–16.
- Sant'Anna, P. H. 2021. "Nonparametric Tests for Treatment Effect Heterogeneity with Duration Outcomes." *Journal of Business & Economic Statistics* 39 (3): 816–32.
- Sianesi, B. 2004. "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s." *The Review of Economics and Statistics* 86 (1): 133–55.
- Tsiatis, G. 1975. "A Nonidentifiability Aspect of the Problem of Competing Risks." *Proceedings of the National Academy of Sciences* 72 (1): 20–2.
- Van den Berg, G. J. 2001. "Duration Models: Specification, Identification, and Multiple Durations." In *Handbook of Econometrics*, Vol. 5, chapter 55, edited by J. Heckman, and E. Leamer, 3381–460. Elsevier.
- Van den Berg, G. J., and B. Van der Klaauw. 2006. "Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment." *International Economic Review* 47 (3): 895–936.
- Van den Berg, G. J., P. Bonev, and E. Mammen. 2020a. "Nonparametric Instrumental Variable Methods for Dynamic Treatment Evaluation." *The Review of Economics and Statistics* 102 (2): 355–67.
- Van den Berg, G. J., A. Bozio, and M. Costa Dias. 2020b. "Policy Discontinuity and Duration Outcomes." *Quantitative Economics* 11 (3): 871–916.
- Vooren, M., C. Haelermans, W. Groot, and H. Maassen van den Brink. 2019. "The Effectiveness of Active Labor Market Policies: A Meta-Analysis." *Journal of Economic Surveys* 33 (1): 125–49.
- Weber, S. 2014. "Human Capital Depreciation and Education Level." *International Journal of Manpower* 35 (5): 613–42.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/jbnst-2024-0001>).