A Duration Model with Dynamic Unobserved Heterogeneity*

Irene Botosaru†

November 1, 2013

Abstract

This paper considers a new class of single-spell duration models in which, first, unobserved heterogeneity changes during the duration of the spell and, second, changes in unobserved heterogeneity may have different effects on the probability of exit depending on their timing during the spell. In contrast, unobserved heterogeneity in standard duration analysis is time invariant and timing effects cannot be analyzed. The aims of the paper are: to provide a modeling strategy for duration analysis when shocks accumulate during the duration of a spell, to show the identification of the primitives entering the hazard function, to discuss the trade-offs between the nonparametric and the semiparametric identification of the model, and to provide a feasible estimation procedure.

JEL classification: C14, C41

Keywords: duration analysis, Levy process, stochastic heterogeneity, stochastic hazard function.

*I am deeply thankful to Donald Andrews, Donald Brown, Xiaohong Chen, Yuichi Kitamura, and Peter Phillips for their suggestions, advice, and support. I have benefited from discussions with Jaap Aabring, Jean-Pierre Florens, James Heckman, Bo Honore, Lena Janys, Simon Lee, Chris Muris, Krishna Pendakur, Franco Peracchi, Christoph Rothe, and Francis Vella. I thank seminar participants at Yale, Mannheim, Montreal, Toulouse, Georgia, Brandeis, ESEM Oslo 2011, and MOOD Rome 2010 for questions and comments. Financial support of Cowles Foundation via the Carl Arvid Anderson Prize is gratefully acknowledged.

†Tel: +33 5 61 12 88 16. Email: irene.botosaru@tse-fr.eu. Address: Toulouse School of Economics, 21 allee de Brienne, 31015 Toulouse Cedex 6, France.

1 Introduction

This paper considers a class of duration models that differs from standard duration models in two significant ways. First, unobserved heterogeneity changes during the duration of the spell. Second, changes in unobserved heterogeneity that happen earlier on in the spell are allowed to affect the probability of exit differently than changes that happen later on in the spell. Both aspects appear to be new to the economics literature dealing with duration analysis, although there is some precedence for this in biostatistics, see Gjessing, Aalen, and Hjort (2003). I will refer to this class of models as dynamic heterogeneity (DH).

I introduce now the terminology specific to duration analysis. Let T be a continuous random variable denoting the duration of individual i in a specific state and let $P(T \leq t)$ be the distribution function of T. Individual i's hazard function, or instantaneous probability of exit, is defined as:

$$h(t_i) = \lim_{dt \to 0} \frac{P(t_i \leq T \leq t_i + dt | T \geq t_i)}{dt} \tag{1}$$

The usual specification for (1) in duration analysis is similar to a regression model in which the dependent variable is the rate at which an event occurs, and the independent variables are time, observed covariates, and unobserved heterogeneity. For example, in the popular mixed proportional hazard (MPH) model the individual hazard function has the following form:

$$h_M(t_i | x_i, z_i) = \phi_M(x_i) \lambda(t_i) z_i \tag{2}$$

where $t_i \in \mathbb{R}_+$ is the time individual i spends in a state, $x_i \in \mathbb{R}^d$ are observed covariates, $\phi_M \in \mathbb{R}_+$ is a function of observed covariates, $\lambda \in \mathbb{R}_+$ is a function of time, also known as the baseline hazard, and z_i are realizations of a positive random variable $Z \in \mathbb{R}_+$ modeling unobserved heterogeneity.

The MPH model has been extensively applied in economics and the theory developing from applied analyses using it has yielded important insights, for an overview see Van den Berg (2001). The MPH has also had important applications in fields where survival and event history analyses are of interest, such as in biostatistics, reliability theory, and sociology, see Aalen, Borgan, and Gjessing (2008).

In this paper, I modify (2) in two ways. First, I introduce time varying unobserved heterogeneity by modeling the unobserved heterogeneity as a stochastic process. Second, I allow for changes in unobserved heterogeneity to have different effects on the hazard function depending on their timing during the spell. I specify the hazard function as:

$$h(t_i | x_i, z_i(t_i)) = \phi(x_i) \int_0^{t_i} f(u) \, dZ_i(u) \tag{3}$$

where $x_i \in \mathbb{R}^d$ are observed covariates,[1] $\phi \in \mathbb{R}_+$ is a function of observed covariates, $Z(t)$ is a stationary stochastic process of bounded variation and $z_i(t_i)$ are its realizations, and $f \in \mathbb{R}_+$ is a square integrable function with respect to the sample paths of $Z(t)$. The process $Z(t)$ stands for unobserved heterogeneity and the function f models the effect of the timing of changes in unobserved heterogeneity. I assume that the distribution of $Z(t)$ is the same among all individuals, but that each individual i has his/her own realization $z_i(t_i)$. The primitives of the model are the two functions ϕ and f, and the distribution of $Z(t)$.

I motivate below the DH framework via an example that cannot be handled by the MPH. For two more examples, see Section A.1. The main reason for which the MPH framework cannot be used is that in the example below the unobserved heterogeneity changes stochastically through time. The MPH framework is appropriate when the unobserved heterogeneity stays constant throughout the duration of the spell. For a detailed description of the differences between the DH and the MPH, see Section 2.1.

Example 1 (Human capital accumulation and the probability of promotion) *During the spell of employment, workers accumulate task-specific human capital by learning-by-doing. The accumulated capital may be transferable to similar jobs and it may be valued by multiple firms, see Kambourov and Manovskii (2009). The value of the capital may then be reflected in a promotion, which can mean either receiving a higher wage for performing the same job or moving to a higher-paid job using similar skills, see Gibbons and Waldman (2004). Moreover, given similar levels of human capital, the timing of additional capital may have different effects on the probability of promotion. For example, in fast track jobs, individuals who accumulate skills earlier on in the spell may have a higher probability of promotion than those who accumulate skills later on in the spell. Ariga, Ohkusa, and Brunello (1999) label the first possibility as "star picking" and the latter as "late selection." Tenure promotion in academia is a concrete example of a fast track job where the timing of human capital directly affects the probability of promotion.*

In the context of duration analysis, the MPH may be inadequate to study how the probability of promotion varies with human capital accumulation since (2) regards the skill level, z_i, as having been realized at the beginning of the employment spell and then as being held constant over the duration of the job. If the job does not terminate quickly, it seems more realistic to specify the hazard of promotion as (3), where $Z(t)$ models accumulating job-specific human capital and f models the possible effects of the timing of skills. As I explain in Section 2, if f is decreasing, then skills that are accumulated earlier on in the spell are rewarded faster, and the type of job can be regarded as being "fast track."

Hazard models with stochastic unobserved heterogeneity have been suggested before. Kebir (1991) was one of the first to mention the relevance of stochastically evolving unobserved heterogeneity. Singpurwalla (1995) gives an overview of models with stochastic randomness, while Singpurwalla (2006) presents an overview

[1] X can be time varying. Appendix A.4 considers this case.

of models in which the hazard rate is viewed as a stochastic process. Models with stochastic frailties are introduced in Aalen and Hjort (2002) and Gjessing, Aalen, and Hjort (2003) introduce a model similar to (3). In their paper, Gjessing, Aalen, and Hjort (2003) describe the statistical implications of the survival function resulting from their model, but they do not study the identification or the estimation of the model.

The main contribution of the paper is to show the identification of the primitives entering (3). The identification strategy is new to duration analysis. It is based on first formulating the survival function in terms of the Laplace transform of the distribution of $Z(t)$ and then on solving the resulting nonlinear Volterra integral equation of the first kind with unknown kernel. The identification results vary from the nonparametric to the semiparametric. For example, both ϕ and the mean of the distribution of $Z(t)$ are identified nonparametrically, when the distribution of $Z(t)$ is unknown. Since the remaining primitives, f and the entire distribution of $Z(t)$, cannot be jointly nonparametrically identified, I show two different semiparametric identification results. First, I show the identification of f when the distribution of $Z(t)$ is known up to its mean, which had been previously nonprametrically identified. Second, I show the identification of the distribution of $Z(t)$ when f is either known or simply not included in the analysis.

I then propose estimation procedures for both identification strategies. When the distribution of $Z(t)$ is parametrized up to its mean, semiparametric maximum likelihood (ML) is a natural estimation procedure for ϕ, f, and the mean of the distribution of $Z(t)$. Because of the nonlinearity of the problem, both functions ϕ and f are approximated by sieves. Since this is the numerically feasible estimator, I show its consistency and its small sample performance is illustrated in Monte Carlo simulations. I also discuss an estimation procedure for the distribution of $Z(t)$ when f is known.

The paper is organized as follows. Section 2 discusses the DH model and compares it with other popular duration models. In this section, I show the interpretation of the primitives in (3) and the implications of modeling the hazard function as in (3). Subsection 2.1 compares the DH model to the MPH and Subsection 2.2 compares it to the mixed hitting time (MHT) model. Section 3 introduces the class of stochastic processes considered in this paper. I present some properties of these processes that are necessary for the analysis presented in the paper. Section 4 presents the identification results and Section 5 contains a collection of remarks regarding the DH model. Section 6 introduces the estimators: Subsection 6.1 describes the sieve ML estimators and their consistency, Subsection 6.2 describes an estimation procedure for the distribution of $Z(t)$, and Subsection 6.3 presents Monte Carlo results for the sieve ML estimators. Section 7 concludes. All proofs are found in the Appendix.

2 Description of the DH model and comparison with other duration models

In this section, I describe each element entering the hazard function (3).

Unobserved heterogeneity Unobserved heterogeneity is modeled as a positive Lévy process, or subordinator, in order to preserve the positivity of the hazard function.[2] The class of subordinators considered in this paper is described in details in Section 3. I discuss here the implications of assuming that $Z(t)$ is a subordinator, since it is these implications that provide the fundamental differences between the DH and other duration models.

The itial value of subordinators is $Z(0) = 0$. Then the first implication of modeling unobserved heterogeneity as a subordinator is that there is no unobserved heterogeneity at the beginning of the spell. The process generating the unobserved heterogeneity starts once individuals enter the spell and it continues until the individuals exit the spell. The second implication is that unobserved heterogeneity accumulates in jumps during the duration of the spell. These jumps can be regarded as permanent shocks. The size of the shocks can be small or large and the rate at which the shocks happen can be fast or slow. Both the size and the rate are controlled by the distribution of the subordinator. For example, if the process is the gamma process, shocks accumulate in tiny and frequent increments, while if the process is the compound Poisson process, shocks are large and rare. These two implications are in sharp contrast with those following from the MPH model, as it is explained in Section 2.1. As such, these two classes of models cannot realistically describe the same set-up.

The function modeling timing effects Function f in (3) is a weight function, henceforth known as a timing function, and the resulting process $\int_0^{t_i} f(u)\, dZ_i(u)$ is known as a weighted stochastic process. First, the shape of f facilitates inference about possible timing effects of shocks to unobserved heterogeneity. Let $\{t_0 = 0, t_1, t_2, ..., t_T = t\}$ be a partition of $(0, t]$. At each $\{t_j\}_{j=0}^{T}$, the unobserved heterogeneity changes the hazard function by $dZ(t_j)$. The jump or shock, $dZ(t_j)$, is weighted by $f(t_j)$. Then for a given sequence of realizations of unobserved heterogeneity, $\{dZ(t_j)\}_{j=0}^{T}$, the process of exit is accelerated if f is decreasing. This happens because shocks to unobserved heterogeneity that take place earlier on in the spell receive a greater weight relative to shocks that take place later on in the spell. Alternatively, if f is increasing, the process of exit is decelerated, while if f is flat, shocks to unobserved heterogeneity have no timing effects,

[2] For a function of locally bounded variation, $Z(t)$, and for a non-random continuous function, f, for $0 = t_0 < t_1 < ... < t_n = t$, with $\max_i |t_i - t_{i-1}| \to 0$, we have the following definition:

$$\int_0^t f(u)\, dZ(u) \equiv \lim_{n\to\infty} \sum_{i=1}^{n} f(t_i)\left[Z(t_i) - Z(t_{i-1})\right]$$

Since the hazard function is positive for each t_i, both f and the difference $Z(t_i) - Z(t_{i-1})$ should be positive.

since each shock receives the same weight.

Second, f allows for richer dynamics of the resulting process modeling unobserved heterogeneity. Even if $Z(t)$ is a time-homogeneous process, which means that the number of expected jumps in a time interval is constant, the weighted stochastic process is time-inhomogeneous, which means that the number of expected jumps varies over time as a function of f. For example, if f is decreasing, the probability of exit is accelerated since on average there are more jumps in the weighted stochastic process that take place towards the beginning of the spell than towards the end, and a higher number of jumps, increases the hazard function, ceteris paribus.

2.1 Comparison with the MPH Model

The MPH has been studied extensively, with powerful identification results resting on the assumption of multiplicative time invariant unobserved heterogeneity. Identification of (2) is shown by Elbers and Ridder (1982) and Heckman and Singer (1984) under assumptions on either the mean or the tail of the distribution of unobserved heterogeneity. Although standard duration models are flexible statistical models, from an interpretability point of view such models are limited to applications in which the unobserved heterogeneity is time invariant, as discussed in the introduction.

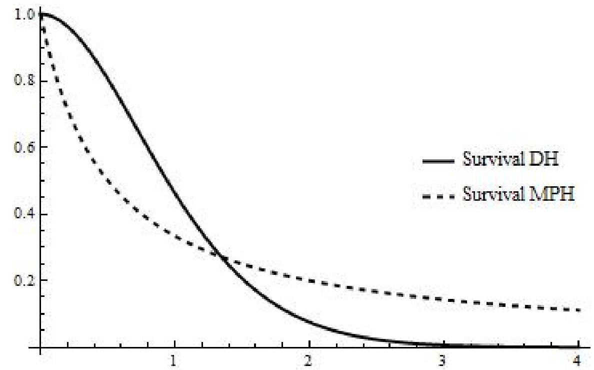
In the MPH set-up, individuals enter the spell with a given level of unobserved heterogeneity. Once the spell begins, unobserved heterogeneity does not change anymore. That is, the process that generated the initial differences among individuals stops once the individuals enter the spell. In this framework, it is individuals with high *initial* levels of heterogeneity who leave the sample faster, ceteris paribus, and as time elapses, it is individuals with low initial levels of heterogeneity who remain in the sample. On the other hand, in the DH set-up, individuals enter with zero levels of unobserved heterogeneity, and the process generating unobserved heterogeneity starts once the spell begins. As time elapses, it is those with the most *accumulated* heterogeneity who leave the sample faster, ceteris paribus. It is safe to say that in the MPH setting, individuals remaining in the sample become more and more homogeneous, while in the DH setting, individuals remaining in the sample become more and more heterogeneous.

To illustrate the fact that sorting takes place faster in the MPH than in the DH setting, consider example 2. I define first the average survival function since it is this function that is observed in the data rather than the individual hazard function. The average survival function is connected to (1) via the exponentiation formula:

$$P(T \geq t_i) = E\left[\exp\left(-\int_0^{t_i} h(s)\,ds\right)\right] \tag{4}$$

Example 2 *Let $Z \sim Ga(\rho, \nu)$ and $Z(t) \sim Ga(\rho t, \nu)$, and suppose there are no observed covariates, and that $f(t) = 1$ and $\lambda(t) = 1$ for all t. Then (2) is given by $h_M(t, z) = z$ and (3) is given by $h_{DH}(t, z) = Z(t)$. Note that both Z and $Z(t)$ have the same distribution with the same scale and shape parameters. Let $\mathcal{L}$ be the*

Figure 1: Survival Functions: Sorting takes place earlier in the MPH setting, and it accelerates as time elapses in the DH setting.



Laplace transform of the density of Z *and let* Ψ *be the Laplace exponent of* $Z(t)$*. The two survival functions are given by, respectively:*

$$\begin{aligned} S_M(t;\rho,\nu) &= \mathcal{L}_Z(t) = (1+\rho t)^{-\nu} \\ S_{DH}(t;\rho,\nu) &= \exp\left[-\int_0^t \Psi(t-u)\,du\right] = \exp\left[-\int_0^t \rho\log\left(1+\frac{t-u}{\nu}\right)du\right] \end{aligned}$$

Please refer to Van den Berg (2001) for the derivation of S_M *and to A.2 for the derivation of* S_{DH}*. I plot in Figure 1 the two survival functions with* $(\rho,\nu)=(2,1)$*.*

Another difference between the two classes of models is that there is no proportionality in the DH model, even with time-invariant observed covariates. Proportionality refers to the fact that the relative risk for two individuals is time-invariant. The proportionality of hazards in the MPH setting is an implication of the time-invariance of unobserved heterogeneity and deterministic time variation of the baseline hazard function. This type of behavior is believed to not be realistic. For example, it is usually found that when firms enter into a market, smaller firms exit faster, which would imply diverging hazards. Alternatively, in the DH setting, transition rates can be converging, diverging, or crossing during the duration of the spell.

Further, the baseline hazard λ entering (2) is fundamentally different from f entering (3). λ depends on the actual duration of the spell and it is a weight function applied to the hazard function. f is a weight applied to the unobserved heterogeneity and it weighs each increment in unobserved heterogeneity, $dZ(t)$, by a potentially different value, as explained in the introduction. Additionally, λ can exist no matter if there or if there is no unobserved heterogeneity. f depends on the time-scale of the stochastic process, so that it is defined as long as the process generating unobserved heterogeneity exists.

In the MPH, the hazard function evolves deterministically with time. There is a one-time change in unobserved heterogeneity and once λ is known, the entire evolution of the hazard function is known. As a result, individual risks at beginning and at the end of the spell are perfectly correlated. In contrast, in DH, the hazard function evolves stochastically, and there is a decreasing correlation between individual risk at the end of the spell and that at the beginning of the spell. As long as the entire history of realizations of unobserved heterogeneity during the spell is unknown, knowing f does not pin down the hazard function. As such, there is more flexibility in hazard functions at the individual level.

As the comparison above suggests, the MPH and the DH models are non-nested. The two frameworks describe different environments. When there is an unobserved shock that happens at the beginning of the spell and that has long lasting effects on the probability of exit, then the hazard function is more realistically modeled by the MPH. When there are unobserved shocks that accumulate during the duration of the spell and that affect the hazard function directly, it is the DH framework that provides the more realistic description of individual risk.

2.2 Comparison with the MHT Model

The MHT model is a first passage model in which individuals leave the initial state as soon as a risk process, $Y(t)$, hits a barrier, B. The process $Y(t)$ is modeled as a spectrally negative Lévy process, which is a time-continuous process, with independently and identically distributed increments, and no positive jumps. Duration is defined as:

$$T = \inf\{t \geq 0 : Y(t) > B\} \tag{5}$$

When $B \equiv \phi_{HT}(x) z_{HT}$, where $\phi_{HT}(x)$ is a function of observed covariates and Z_{HT} denotes unobserved heterogeneity, Abbring (2012) showed the nonparametric identification of $\phi_{HT}(x)$, and of the distributions of $Y(t)$ and Z_{HT}. The identification strategy took advantage of the multiplicative structure of the time invariant threshold, which is similar to that used for the identification of the MPH.

The MHT can be regarded as a way to allow unobserved heterogeneity to vary with time when the underlying process $Y(t)$ in (5) is interpreted as unobserved heterogeneity. In this case, unobserved heterogeneity would be modeled as a Brownian motion or as a subordinator perturbed by a diffusion since $Y(t)$ is a spectrally negative Lévy process. Then (5) and (3) describe very different environments. For example, accumulating skills or health damages cannot be described by spectrally negative Lévy processes. On the other hand, stock prices or the value of an investment cannot be realistically modeled as subordinators, unless prices and the value were increasing over time.

Another difference is that in the MHT set-up it is the distance between the unobserved risk process and the barrier that determines when the exit happens. In the DH the exit can happen at any time, with the exit being driven by the hazard function rather than a barrier. The latter specification may be more intuitive when changes in unobserved heterogeneity result directly in changes in the individual hazard function. For example, when $Z(t)$ is the skill level, an increase in skills may increase directly the hazard of promotion, rather than result in a decrease in the amount of skills that that still need to be acquired for a promotion.

Note that in the MHT setting neither the individual hazard nor the survival function can be backed out, while in the DH setting the individual hazard is modeled directly with the survival function being expressed as a function of the primitives of the model, as in the MPH.

Nonetheless there is a connection between the DH and hitting time models. As Singpurwalla (2006) shows, (3) can be thought of as a hitting time model with an exponentially distributed barrier. Let W be a random variable whose distribution is exponential(1) and define:

$$H(t, x, Z(t)) \equiv \int_0^t \left(\phi(x) \int_0^s f(u)\, dZ(s) \right) du$$

Suppose $H(t, x, Z(t))$ and W are independent, then

$$P(T > t) = P(W > H(t)) = E_Z(\exp -H(t))$$

which is the average survival function associated with the hazard function (3), see Lemma (1).

3 Lévy processes

Following Bertoin (1996), the formal definition of a Lévy process is:

Definition 1 (Lévy Process) *Let P be a probability measure on $(\Omega, \mathcal{F})$. $\{Z(u)\}_0^t$ is a Lévy process for $(\Omega, \mathcal{F}, P)$ if for every $s, t \geq 0$, the increment $Z(s+t) - Z(t)$ is independent of the process $\{Z(v)\}_0^t$ and has the same law as $\{Z(s)\}_0^t$.*

Lévy processes are Markov processes and examples include the Brownian motion, the gamma process, the Poisson process, the compound Poisson process.

The class of processes considered in this paper is that of positive Lévy processes, also known as subordinators. Subordinators take values in $\mathbb{R}_+$ which implies that their sample paths are increasing. The gamma process and the compound Poisson process are standard examples of subordinators.

Example 3 (Gamma Process) *Let Z be a gamma random variable with shape parameter $\rho > 0$ and scale parameter $\nu > 0$, and let $Z(t)$ be a gamma subordinator. Then for $t \geq 0$:*

$$\begin{aligned} Z &\sim Ga(\rho, \nu) \\ Z(t) &\sim Ga(\rho t, \nu) \\ \int_0^t f(u)\, dZ(u) &\sim Ga(\rho t, \nu f(t)) \end{aligned}$$

For the derivation of the distribution of the weighted stochastic process, see Dykstra and Laud (1981).

Definition 2 (Laplace Exponent) *Let $Z(t)$ be a subordinator. The Laplace exponent of $Z(t)$, $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ is given by:*

$$E\left(\exp\left[-\chi Z(t)\right]\right) = \exp\left[-t\Psi(\chi)\right], \ \chi > 0.$$

Remark 1 *Subordinators are characterized uniquely by their Laplace exponent. As such, saying that the primitive of the model is the distribution of $Z(t)$ is equivalent to saying that the primitive of the model is the Laplace exponent of $Z(t)$*

Laplace exponents have the following property, see Gnedin and Pitman (2008):

Property 1 A function $\Psi(\chi)$, $\chi \geq 0$, is the Laplace exponent of a subordinator if: (i) $\Psi(\chi)$ is infinitely differentiable with respect to $\chi \geq 0$; (ii) $\Psi(0) = 0$; and (iii) $(-1)^n \frac{\partial^n}{\partial \chi^n} \Psi(\chi) \leq 0$ for every n.

Both the Laplace exponent and the concept of cumulant will play an important role in this paper.

Definition 3 (Cumulants) *The j^{th} cumulant of the subordinator $Z(t)$ is defined as*

$$\left.\frac{d}{d\chi^j}\Psi_j(\chi)\right|_{\chi=0} = k_j \tag{6}$$

Cumulants are the coefficients in the Taylor expansion about the origin of the log of the moment generating function. As such, there is a one-to-one relationship between cumulants and moments. For example, the first cumulant of the process is the mean of the distribution, while the second cumulant is the variance. Let m_j be the j^{th} moment of the distribution of $Z(t)$. The relationship between the first four moments and cumulants is:

$$\begin{aligned}
m_1 &= k_1 \\
m_2 &= k_1^2 + k_2 \\
m_3 &= k_1^3 + 3k_1k_2 + k_3 \\
m_4 &= k_1^4 + 6k_1^2k_2 + 3k_2^2 + 4k_1k_3 + k_4
\end{aligned}$$

This section is finished off with a few examples of subordinators, their Laplace exponents, and their first cumulants.

Example 4 (Gamma Process) *Let $Z(t) \sim Ga(\rho t, \nu)$. The Gamma process has an infinity of very small jumps in any time interval. As such, it is commonly used to model processes that take place gradually in* *time, such as erosion and wear-and-tear. The Laplace exponent of $Z(t)$ is*

$$\Psi(\chi) = \log\left(1 + \frac{\chi}{v}\right)^{\rho}$$

and the mean of the distribution of $Z(t)$ is given by

$$\left.\frac{d}{d\chi}\Psi(\chi)\right|_{\chi=0} \equiv k_1 = \frac{\rho}{\nu}$$

Example 5 (Poisson Process) *The Poisson process plays an important role in risk analysis and it is used to model shocks to a market, accidents, and natural disasters. The Poisson process of intensity κ is a counting process with independent and stationary increments, where the increments are exponentially distributed with rate κ. If events can happen at any time, and event arrivals are independent of one another and past arrivals do not influence future arrivals, then κt events are expected in an interval of length t. The Laplace exponent of such a process is*

$$\Psi(\chi) = \kappa(1 - \exp(-\chi))$$

with the mean of the distribution of $Z(t)$ equal to

$$\left.\frac{d}{d\chi}\Psi(\chi)\right|_{\chi=0} = \kappa$$

Example 6 (Compound Poisson Process) *The Poisson process has increments that have unit size. The compound Poisson process allows the size of the jumps to be a random variable, with a given distribution. This type of process usually models situations in which, say, the number of claims in a time interval is a Poisson process, but the monetary size of the claims is a random variable. Then the total amount of money spent on the claims up to some time follows a compound Poisson. If the jumps are distributed as $Ga(\rho, \nu)$, the Laplace exponent is*

$$\Psi(\chi) = \kappa\left(1 - \left(\frac{\nu}{v + \chi}\right)^{\rho}\right)$$

and the mean of the distribution of $Z(t)$ is

$$\left.\frac{d}{d\chi}\Psi(\chi)\right|_{\chi=0} = \frac{\kappa\rho}{\nu}$$

4 Identification

This section explains the identification of the primitives of the model. The section also discusses trade-offs in identification. Theorem 1 shows the nonparametric identification of ϕ and of the mean of the distribution of

$Z(t)$. Since f and the distribution of $Z(t)$ enter multiplicatively under an integral operator, which smooths out their individual variations, f and $Z(t)$ cannot be jointly identified. Essentially, we need to solve a nonlinear integral equation where the kernel is the Laplace exponent of $Z(t)$ which is unknown. This is where the trade-off arises: If the distribution of $Z(t)$ is parametrized up to its mean, we can identify f. If f is assumed known or if f is not of interest, we can identify the distribution of $Z(t)$.

As mentioned in the introduction, what is observed in the data is the conditional survival function associated to (3), call it $S(t|x)$. As such, the identification strategy begins by calculating $S(t|x)$.

Assume the following:

Assumption A $T|(x, Z(t))$ is a random variable with an absolutely continuous distribution function (wrt the Lebesgue measure).

Assumption B f is a continuous, square integrable function with respect to the sample paths of $Z(t)$

Assumption A excludes jumps in the conditional survival function induced by changes in X and in the filtration of $Z(t)$. It allows us to work with density functions. Assumption B is a regularity condition.

Lemma 1 *Given assumptions A and B, the conditional survival function associated to* (3) *is given by*

$$S(t|x) = \exp\left[-\int_0^t \Psi(\phi(x) f(u)(t-u))\, du\right] \tag{7}$$

Proof. By assumption A, the usual exponential formula (4) is applicable. Then substituting the definition of h given by (3), interchanging the order of integration, and applying the properties of independent and stationary increments of $Z(t)$, obtains:

$$\begin{aligned}
S(t|x) &= E_Z \exp\left[-\int_0^t h(s|x, Z(s))\, ds\right] \\
&= E_Z \exp\left[-\int_0^t \left(\phi(x)\int_0^s f(u)\, dZ(s)\right) du\right] \\
&= E_Z \exp\left[-\int_0^t \left(\int_u^t \phi(x) f(u)\, ds\right) dZ(u)\right] \\
&= \exp\left[-\int_0^t \Psi(\phi(x) f(u)(t-u))\, du\right]
\end{aligned}$$

The detailed derivation is shown in Appendix A.2. ■

Let the true survival function be $S_0(t|x)$. Then $S_0(t|x) = S(t|x)$. Taking log of both sides, letting $H(t,x) \equiv -\log S_0(t|x)$, and rearranging obtains:

$$H(t,x) = \int_0^t \Psi(\phi(x) f(u)(t-u))\, du \tag{8}$$

The identification of the parameters of interest is based on solving (8) for all $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$.

Assumption ID1 (i) $\Psi(\chi)$, $\chi \geq 0$, is differentiable at zero; (ii) $\Psi_1(0) \equiv \frac{d}{d\chi}\Psi(\chi)\Big|_{\chi=0} \neq 0$.

Assumption ID2 (i) $\lim_{t\to 0} f(t) = 1$; (ii) $\phi(0) = 1$.

Assumption ID1 implies that the mean of the distribution of $Z(t)$ exists and is nonzero. This assumption is key in solving (8): It allows us to obtain an integral equation of the second kind, which under additional assumptions, can be solved uniquely for the parameters of interest. The assumption excludes subordinators with fat tails such as stable processes and it can be thought of as the analogue of the finite mean assumption in Elbers and Ridder (1982). Note that $\Psi_1(0)$ is the mean of the distribution of $Z(t)$, see (3). Assumption ID2(i) is a normalization assumption on the weight function which is needed to identify ϕ up to the mean of the distribution of $Z(t)$, while ID2(ii) is a normalization assumption used to pin down the mean of the distribution.

Theorem 1 *Let assumptions A, B, ID1, and ID2 hold and let the distribution of the subordinator $Z(t)$ be unknown, and define $H_{tt}(t,x) \equiv \frac{\partial^2}{\partial t^2}H(t,x)$. The function ϕ and the mean of the distribution of $Z(t)$, call it k_1, are identified and given by:*

$$\phi(x) = \frac{\lim_{t\downarrow 0} H_{tt}(t,x)}{\lim_{t\downarrow 0} H_{tt}(t,0)} \tag{9}$$

$$k_1 \equiv \Psi_1(0) = \lim_{t\downarrow 0} H_{tt}(t,0) \tag{10}$$

Proof. Let $\Psi_{11}(\chi) \equiv \frac{\partial^2}{\partial \chi^2}\Psi(\chi)$. Differentiating (8) twice wrt t obtains:

$$H_{tt}(t,x) = \phi(x) f(t) \Psi_1(0) + \phi^2(x) \int_0^t f^2(u) \Psi_{11}(\phi(x) f(u)(t-u))\, du \tag{11}$$

Letting $t \downarrow 0$ in (11), by ID1(ii) and ID2(ii), obtains (10) and by ID2(i) obtains (9). ■

4.1 Failure of Joint Identification

In this subsection I show it is not possible to jointly identify the timing function, f, and the distribution of the stochastic process, $Z(t)$.

Consider the following assumptions:

Assumption IDz (i) $f(t)$ is $s-$times differentiable for all $t \in R_+$; (ii) $\Psi(\chi)$, $\chi \geq 0$ is $s+1$ many times differentiable at zero.

Assumption IDz restricts both the timing function to be $s-$times smooth and the class of subordinators to that for which $s+1$ moments exist.

Lemma 2 *Let assumptions A, B, ID1, ID2, and IDz hold and let the distribution of the subordinator $Z(t)$ be unknown. The timing function f and the distribution of $Z(t)$ cannot be jointly identified.*

Proof. I will use the concept of cumulant introduced in (6) and the following notation: For $i \in \{1, 2, ..., s\}$, define:

$$
\begin{aligned}
H_{(i)} &\equiv \lim_{t \to 0} \frac{\partial^i}{\partial t^i} H(t, 0) && (12) \\
f_{(i)} &\equiv \lim_{t \to 0} \frac{\partial^i}{\partial t^i} f(t) && (13)
\end{aligned}
$$

Consider (11) where $x = 0$, and differentiate the resulting expression $s-$times with respect to t. Evaluate the answer in the limit as $t \downarrow 0$ to obtain the following system:

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
f_{(1)} & 1 & 0 & 0 & \ldots & 0 \\
f_{(2)} & 2f_{(1)} & 1 & 0 & \ldots & 0 \\
f_{(3)} & 3f_{(2)} & 2f_{(1)} & 1 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
f_{(s-1)} & (s-1) f_{(s-2)} & (s-2) f_{(s-3)} & (s-3) f_{(s-4)} & \ldots & 1
\end{bmatrix}
\begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ \ldots \\ k_s \end{bmatrix}
=
\begin{bmatrix} H_{(2)} \\ H_{(3)} \\ H_{(4)} \\ H_{(5)} \\ \ldots \\ H_{(s+1)} \end{bmatrix}
\qquad (14)
$$

When the distribution of $Z(t)$ is unknown, the vector of cumulants $\{k_j\}_{j=1}^{s}$ is unknown. If both $f(t)$ and $\{k_j\}_{j=1}^{s}$ are unknown, the system has an infinity of solutions. ■

4.2 Identification of the Timing Function

In order to identify f and $Z(t)$ stronger restrictions will have to be imposed on any one of them.

Showing the identification of f involves solving equation (8) for f. If the distribution of $Z(t)$ were known, then Ψ would be known. In this case, (8) would be a nonlinear Volterra integral equation of the first kind for f.

In general, Volterra integral equations of the first kind do not have unique solutions. However, it is possible to show they have a unique solution if they can be transformed into Volterra integral equations of the second kind. Under certain regularity assumptions, a Volterra integral equation of the first kind is transformed to one of the second kind by differentiation with respect to the upper limit of integration until an additive term is obtained. This term usually depends on the functions of interest, and it allows one to solve for the unknown functions by the method of contraction mappings. Once the existence and uniqueness of the solution to the Volterra integral equation of the second kind has been shown, the solution also solves uniquely the Volterra integral equation of the first kind by the Fundamental Theorem of Calculus. This is the solution strategy adopted in this section.

For the analysis that follows, I assume that the distribution of $Z(t)$ is known up to its mean, k_1, so that $\Psi(\chi)$ becomes $\Psi(\chi, k_1), \chi > 0$, and the true survival function becomes $S_0(t|x, k_1)$. I show the identification of f. Solving (11) for f and using ID2(ii) obtains

$$f(t) = \frac{1}{k_1}\left[H_{tt}(t, 0, k_1) - \int_0^t f^2(u)\Psi_{11}(f(u)(t-u), k_1)\,du\right] \tag{15}$$

where now $\Psi_{11}(\chi, k_1) \equiv \frac{\partial^2}{\partial\chi^2}\Psi(\chi, k_1)$.

Let C_w^s be the Banach space of s-times continuously differentiable functions endowed with the appropriate norm, weighted by a continuous, positive function, $w(t)$, which will be defined later. Define the following operator:

$$(Tf)(t) = \frac{1}{k_1}\left[H_{tt}(t, 0, k_1) - \int_0^t f^2(u)\Psi_{11}(f(u)(t-u), k_1)\,du\right] \tag{16}$$

Let the following assumptions hold:

Assumption ID3 $f(t) \in C_w^s(\mathbb{R}_+)$ and $0 < f(t) \leq M < \infty$, for all $t \in \mathbb{R}_+$.

Assumption ID4 (i) $\Psi(\chi, k_1)$ is $s+1-$times continuously differentiable in $\chi \in R_+$ for all k_1; (ii) There exists a constant $\delta > 0$ such that $k_1 \geq \delta$; and (iii) $\left|\frac{\partial^3}{\partial\chi^3}\Psi(\chi, k_1)\right| \leq B$ for all $\chi > 0$ and k_1.

Assumption ID3 restricts timing effects to be smooth and bounded. This implies that the effect of previous jumps on the hazard cannot be infinite, and so individuals cannot exit the sample due to a too high weight on the increment. Assumptions ID4 implies the second partial derivative of $\Psi(\chi, k_1)$ wrt χ is Lipschitz continuous with Lipschitz constant B, i.e. for $\chi_1 \neq \chi_2$

$$|\Psi_{11}(\chi_1, k_1) - \Psi_{11}(\chi_2, k_1)| \leq B|\chi_1 - \chi_2|$$

Lipschitz continuity is needed in order guarantee the kernel of (15) is Lipschitz continuous, which will further reflect into the operator T being Lipschitz continuous with a bounded Lipschitz constant. Assumption ID4(ii) is needed in order to guarantee that the operator T is a contraction. For subordinators, k_1 is bounded from below by a positive number δ. Examples 1 and 2 in A.2 show assumption ID4 is satisfied for both the gamma and the compound Poisson processes.

Theorem 2 *Assume the distribution of the stochastic process $Z(t)$ is known up to its mean. Under assumptions A, B, ID1 to ID4, the function $f(t) \in C_w^s$ is identified, where $w(t)$ is given by*

$$w(t) = \exp\left(\frac{3BM^2}{\delta}t\right) \tag{17}$$

The solution is found by the successive approximation method.

Proof. First, it is shown via the Banach Fixed Point Theorem that there is a unique solution $f(t) \in C_w^0$. Then it is shown that $f(t) \in C_w^s$ by applying an induction argument on the smoothness parameter, s. For details, see A.3. ■

4.3 Identification of the Marginal Distribution of Unobserved Heterogeneity

Suppose that either $f(t)$ is known or that it is not relevant to the model, i.e. $f(t) = 1$ for all t. Then all cumulants of the distribution of $Z(t)$ are identified as $t \downarrow 0$. Since there is a one-to-one relationship between cumulants and moments, see Section 3, we can then identify the moments of the distribution of $Z(t)$. If the moments satisfy a condition stated below, then the sequence of moments uniquely determines the distribution function with those moments.[3] The problem is then called determinate.

Let m_j be the j^{th} moment of the distribution of $Z(t)$ and define the Hankel matrix $M = (M_{i,j})_{1 \leq i,j \leq s}$ where $M_{i,j} = m_{i+j-1}$.Berg, Chen, and Ismail (2002) show that the moment problem is determinate if and only if the smallest eigenvalue of the Hankel matrix tends to zero as the number of moments, s, tends to infinity.

Assumption IDz' $\Psi(\chi)$, $\chi \geq 0$ is infinitely many times differentiable at zero.

Assumption IDz' is sufficient for the Hankel matrix of moments to exist and it implies that *all* moments of the distribution of $Z(t)$ exist. This assumption excludes processes with fat tails, such as the stable processes.

Theorem 3 *Let assumptions A, B, ID1, ID2, and IDz' hold such that the smallest eigenvalue of the Hankel matrix tends to zero as the number of moments tends to infinity. Consider* (7) *where* $f(t)$ *is known and* $\phi(0) = 1$. *The distribution of the stochastic process* $Z(t)$ *is uniquely determined.*

Proof. Consider system (14) . The system has a unique solution for the sequence of cumulants $\{k_j\}_1^s$ since the determinant of the matrix of coefficients equals unity. Letting $s \to \infty$ and given the one-to-one relationship between cumulants and moments, all moments of the process are identified. ■

5 Remarks

Remark 2 (Censoring) *When data is right censored and the censoring is non-informative, the identification strategy presented in this paper remains unchanged.*

[3] This is precisely the moment problem: Given a sequence of real numbers that are the moments of some distribution, is there a positive measure uniquely determined by those moments?

Remark 3 (Non-Positive Duration Dependence) *Suppose that a priori it is believed that there is non-positive duration dependence at the individual level. Then one could specify the hazard function as*

$$h\left(t|x,W\left(t\right)\right)=\phi\left(x\right)b\left(t\right)W\left(t\right) \tag{18}$$

where $W\left(t\right)$ *is a subordinator. If there is negative duration dependence, then* $b\left(t\right)$ *is a function such that* $b'\left(t\right)<0$ *for all* t*. If* $b\left(t\right)$ *is known, the identification results of this paper remain unchanged. If* $b\left(t\right)$ *is unknown and if the distribution of* $W\left(t\right)$ *is unknown, then* $b\left(t\right)$ *cannot be identified since time effects due to* $b\left(t\right)$ *and* $W\left(t\right)$ *cannot be disentangled. If the distribution of* $W\left(t\right)$ *were known, then the equation that would need to be solved for* $b\left(t\right)$ *has the form*

$$-\log S\left(t|x\right)=\int_{0}^{t}\Psi\left(\phi\left(x\right)B\left(t,u\right)\right)du$$

where $B\left(t,u\right)=\int_{u}^{t}b\left(s\right)ds$ *and* Ψ *takes on a known functional form. The resulting equation is a complicated nonlinear Volterra integral equation of the first kind, which may or may not have a solution. Basically, if one believes there is negative duration dependence at the individual level, one could use a parametric specification for* $b\left(t\right)$ *and then fit the model with* (18)*. Nonetheless, the interpretation of* $b\left(t\right)$ *would be very different from that of* $f\left(t\right)$ *in* (3)*.*

Remark 4 (Time deformed unobserved heterogeneity) *Modeling dynamics of unobserved heterogeneity in this set-up is more flexible than in the MPH. For example, expressing* $Z\left(t\right)$ *as a function of time varying observed covariates is a by product of modeling unobserved heterogeneity as a stochastic process. Let* $x\left(t\right)$ *be time-varying covariates, and let* g *be an unknown function of* $x\left(t\right)$ *such that unobserved heterogeneity evolves as a function of* $g\left(x\left(t\right)\right)$*. That is, unobserved heterogeneity evolves as* $Z\left(g\left(x\left(t\right)\right)\right)$*. Stochastic processes of this form as known as time-deformations since the time scale of the process is not calendar time,* t*, anymore but some data-driven time scale,* $g\left(x\left(t\right)\right)$*. Stock (1988) mentions how there are contexts when it may be more realistic to model certain phenomena as evolving in operational or economic time. He gives the example of how the output of a factory may be thought to take place on a time scale based on days the factory was actually open and operating rather than on a time scale based on weeks. When* $Z\left(t\right)$ *is unobserved and modeling the wear-and-tear effects of occupational risk, it may make more sense to model the aging process as taking place on a time scale based on the number of actual hours worked than on a calendaristic time scale. In a different paper, I discuss the interpretability and the identification of* g *when the hazard function is specified as*

$$\phi\left(x\right)\int_{0}^{t}dZ\left(g\left(x\left(u\right)\right)\right)du$$

6 Estimation

This section proposes estimation procedures for the two identification strategies included above. I first describe the estimation procedure for the case that f is unknown but the distribution of $Z(t)$ is known up to its mean, and I show that the estimators proposed are consistent. Second, I describe an estimation procedure for the case that function f is known but the distribution of $Z(t)$ is unknown.

6.1 Estimation when the Distribution of Heterogeneity is Known

Suppose the distribution of $Z(t)$ is known up to its mean, k_1. The parameters of interest are (ϕ, f, k_1). When the distribution of $Z(t)$ is known, maximum likelihood seems to be the natural estimation procedure. Since interest lies in estimating two infinite dimensional parameters, ϕ and f, under shape restrictions and which enter the criterion function nonlinearly, I will approximate ϕ and f by positive transformations of a linear span of known basis functions. Both k_1 and the coefficients in the linear expansions are simultaneously estimated by maximizing the log-likehood over a sequence of approximating spaces. This estimation procedure is known as sieve estimation.

Let $\{(X_i, T_i)\}_{i=1}^n$ be iid draws from the distribution of (X, T) with bounded support $\mathcal{X} \times \mathcal{T} = [0,1] \times (0,1]$. Let the distribution of $Z(t)$ be known up to k_1. The survival function is given by

$$S(t|x; \phi, f, k_1) = \exp - \int_0^t \Psi(\phi(x) f(u)(t-u), k_1)\, du$$

and the conditional distribution of $T|X$ is given by $p(t|x; \phi, f, k_1) = -\frac{\partial}{\partial t} S(t|x, k_1)$.

The true value $\alpha_0 = (\phi_0, f_0, \rho_0) \in \mathcal{A} = \Phi \times \mathcal{F} \times \Theta$ solves

$$\begin{aligned} \alpha_0 &= \arg \max_{(\phi, f, k_1) \in \mathcal{A}} Q(\phi, f, k_1) \\ &= \arg \max_{(\phi, f, k_1) \in \mathcal{A}} E_{x,t} \log p(t|x; \phi, f, k_1) \end{aligned} \tag{19}$$

where $k_1 \in \Theta$, a compact subset of $\mathbb{R}_+ - \{0\}$, while the two functions are assumed to belong to the following spaces:

$$\Phi = \{\phi(x) \in C^{s_1}(\mathcal{X}, \mathbb{R}_+) : \phi(0) = 1\} \tag{20a}$$

$$\mathcal{F} = \left\{ f(t) \in C^{s_2}(\mathcal{T}, \mathbb{R}_+) : \lim_{t \to 0} f(t) = 1 \right\} \tag{20b}$$

A sieve ML estimator is proposed for $\alpha_0 \in \mathcal{A}$ by replacing $\mathcal{A}$ by a sieve space $\mathcal{A}_n$ that is compact, linear, finite dimensional space and that becomes dense in $\mathcal{A}$ as $n \to \infty$. Let $B_j(.)$ be a sequence of known univariate basis functions. Then $\mathcal{A}_n$ is a linear span of finitely many $B_j(.)$. For sieve approximation, I consider the functions ϕ and f in finite dimensional spaces Φ_n and $\mathcal{F}_n$, respectively, defined as:

$$\Phi_n = \left\{ \phi_n(x) \in \Phi : \phi_n(x) = \exp \sum_{j=1}^{m_n} a_j B_j(x) \right\} \tag{21a}$$

$$\mathcal{F}_n = \left\{ f_n(t) \in \mathcal{F} : f_n(t) = \exp \sum_{j=1}^{m_n} b_j B_j(t) \right\} \tag{21b}$$

where m_n is the dimension of the sieve spaces, such that $m_n \to \infty$ with $\frac{m_n}{n} \to 0$. The exponential transformation serves to impose the positivity of the functions. The sieve spaces are open and convex, with approximation rate of order $O(n^{-s_1})$ and $O(n^{-s_2})$, respectively.

The sieve ML estimator $\widehat{\alpha}_n = \left(\widehat{\phi}_n, \widehat{f}_n, \widehat{k}_n\right) \in \mathcal{A}_n = \Phi_n \times \mathcal{F}_n \times \Theta$ maximizes the sample analog of (19) with α restricted to the sieve space $\mathcal{A}_n$. Then, the sieve ML estimator satisfies

$$\widehat{Q}_n(\widehat{\alpha}_n) \geq \sup_{\alpha \in \mathcal{A}_n} Q_n(\alpha) - O_p(\eta_n), \ \eta_n = o(1)$$

The following assumption is made on the parameter space, $\mathcal{A}$.

Assumption C0. (i) $\mathcal{A}$ is connected in the sense that for any $\alpha_1, \alpha_2 \in \mathcal{A}$, there exists a continuous path $\{\alpha(\tau) : \tau \in [0,1]\}$ in $\mathcal{A}$ such that $\alpha(0) = \alpha_1$ and $\alpha(1) = \alpha_2$. (ii) The parameter space is convex at α_0, such that for any $\alpha \in \mathcal{A}$, $(1-\tau)\alpha_0 + \tau\alpha \in \mathcal{A}$ for small $\tau > 0$. (iii) For almost all (X, T), $p(t|x, (1-\tau)\alpha_0 + \tau\alpha)$ is continuously differentiable at $\tau = 0$.

The consistency of the estimators is established under metric $||.||_\infty$ defined below. For any $\alpha \in \mathcal{A}$:

$$||\alpha - \alpha_0||_\infty = \sup_x |(\phi - \phi_0)(x)| + \sup_t \left| \int_0^t (f - f_0)(u)\, du \right| + ||k||_E \tag{22}$$

where $||.||_E$ is the Euclidean norm. To establish the consistency of the estimators, it is assumed that:

Assumption C1. (i) The functions ϕ and f are such that $(20a)$ and $(20b)$ hold. (ii) f is bounded from above. (iii) ϕ and f are bounded away from zero for all x and all t, respectively.

Assumption C2. Let $\Psi(\chi, k_1)$ be such that for all $\chi > 0$ and k, the following partial derivatives are bounded below and above:

$$\begin{aligned}
0 &< m_1 \leq \frac{\partial}{\partial \chi}\Psi(\chi, k_1) \equiv \Psi_1(\chi, k_1) \leq M_1 < \infty \\
-\infty &< m_{11} \leq \frac{\partial^2}{\partial \chi^2}\Psi(\chi, k_1) \equiv \Psi_{11}(\chi, k_1) \leq 0 \\
-\infty &< m_{12} \leq \frac{\partial^2}{\partial \chi \partial k_1}\Psi(\chi, k_1) \equiv \Psi_{12}(\chi, k_1) \leq M_{12} < \infty \\
0 &< m_2 \leq \frac{\partial^2}{\partial k_1^2}\Psi(\chi, k_1) \equiv \Psi_2(\chi, k_1) \leq M_2 < \infty
\end{aligned}$$

where the partial derivatives are evaluated at $\chi = \phi(x) f(u)(t-u)$.

Let $\mathcal{A}_o$ be an open and convex space such that

$$\mathcal{A}_o = \Phi_o \times \mathcal{F}_o \times \Theta_o = \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_\infty = o(1)\}$$

Assumption C3. $\Psi(\chi, k_1)$ is pathwise differentiable with respect to $\lambda \in \mathcal{A}_o$ for all $t \in \mathcal{T}$ and for all $k_1 \in \Theta$ and continuously differentiable in $k_1 \in \Theta_o$ for all $\chi \in \mathcal{A}$ in the norm $||.||_w$ defined in 22.

Assumption C4. $\Psi(\lambda, k_1)$ is monotonic in k_1.

Remark 5 *Assumption C1(ii) implies that the hazard function is bounded away from zero. As noted by Dabrowska (2006), this assumption holds if the covariates are bounded and the regression coefficients vary over a bounded neighborhood of the true parameter, conditions which hold by construction in this paper. The uniform boundedness assumption on the functions of interest is used to verify the continuity of the sample criterion function in the consistency norm. The assumption controls the behavior of a term that explodes as the product of the functions ϕ and f approaches zero.*

Assumptions C2 and C3 imply the Laplace exponent and its first partial derivatives with respect to λ and k are Lipschitz continuous in λ and k. Define the following infima and suprema:

$$(m_\phi,\ M_\phi,\ m_f,\ M_f) \equiv (\inf \phi,\ \sup \phi,\ \inf f,\ \sup f)$$

In the problem, $\chi = \phi(x) f(u)(t-u)$, where $\phi : \mathcal{X} \to [m_\phi, M_\phi] \subset \mathbb{R}_+$, $f : \mathcal{T} \to [m_f, M_f] \subset \mathbb{R}_+$, and $t \in \mathcal{T}$. Although the partial derivatives of Ψ are continuous on $[m_\phi, M_\phi] \times [m_f, M_f] \times \mathcal{T}$, the range of χ is not closed, so that the partial derivatives are not bounded unless C2 holds. Note that C2 holds for both gamma and compound Poisson processes.

Assumption C4 is needed in order to derive the bracketing number of the class of functions indexing the criterion function. For the gamma and compound Poisson processes, assumption C4 holds automatically, see 3.

Theorem 4 *Under Assumptions C0-C4 above*

$$||\widehat{\alpha}_n - \alpha_0||_\infty = o_p(1) \quad as\ n \to \infty$$

Proof. To show the consistency of the estimators I verify the conditions of Lemma B.1 of Chen and Pouzo (2012). First, I present Lemma B.1 adapted to our model. Then I verify its conditions. The proof can be found in A.5. ■

6.2 Estimation when the Distribution of Heterogeneity is Unknown

Suppose f is known and let the distribution of $Z(t)$ be unknown. Function ϕ and the cumulants of $Z(t)$ could be estimated by the following procedure: First, estimate the survival function by a smooth nonparametric kernel estimator and call the estimator $\widehat{S}(t|x)$. Let $\widehat{H}(t,x) = -\log \widehat{S}(t|x)$. Differentiate $\widehat{H}(t,x)$ twice with respect to t, call it $\widehat{H}_{tt}(t,x)$. Then

$$\widehat{\phi}(x) = \frac{\lim_{t\downarrow 0} \widehat{H}_{tt}(t,x)}{\lim_{t\downarrow 0} \widehat{H}_{tt}(t,0)}$$

Now consider (14). Since f is known, the matrix of derivatives of f is known. $\widehat{H}(t,x)$ and its partial derivatives can be calculated from the data. Then the cumulants of the distribution can be obtained by inverting the matrix of derivatives of f:

$$\begin{bmatrix} \widehat{k}_1 \\ \ldots \\ \widehat{k}_s \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & 0 \\ f_{(s-1)} & (s-1)\, f_{(s-2)} & \ldots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{H}_{(2)} \\ \ldots \\ \widehat{H}_{(s+1)} \end{bmatrix}$$

where I used notation (12) and (13). Once the cumulants are calculated, I can calculate the moments of the distribution.

Although possible in theory, in practice this estimation procedure may be problematic for a few reasons. First, one would have to differentiate repeatedly a nonparametric estimator for the survival function, which will lead to numerical error. Second, the number of times the survival function should be differentiated should approach infinity, which in practice, would involve trimming s.

6.3 Monte Carlo Simulations for the Sieve ML Estimators

This section presents simulation results. I run three main cases: Case 1 corresponds to the true data generating process (DGP) being given by (3) with $Z(t) \sim Ga(\rho t, \nu)$ and the estimation being the one described in Section 6.1. Cases 2 and 3 are misspecification studies: For Case 2, the DGP is the one associated to (3) but the estimation procedure is that of the classical MPH, while for Case 3, the DGP is that of the MPH but the estimation procedure is that described in 6.1. In each case, Monte Carlo simulations with 500 repetitions are conducted.

6.3.1 Case 1

For the first case, $\{X_i\}_{i=1}^{n=1000} \sim U[0,1]$ and duration t_i associated to each x_i is calculated by solving (7) for t_i. The process $Z(t) \sim Gamma(t,1)$. Two different sub-cases are presented, Study 1 and Study 2, which are summarized Table 1 below:

Table 1: Simulation Studies

True Parameters	Study 1	Study 2
$\phi(x)$	$\exp(2x-3x^2)$	$1+\sqrt{x}-x^3$
$f(t)$	$1-t+\frac{2}{3}t^3$	$1-t+\frac{2}{3}t^3$
$k1$	1	1

The functions ϕ and f are approximated by polynomial splines of the second degree:

$$\begin{aligned}\phi_n(x) &= \exp\left(\sum_{j=0}^{2} a_j x^j + b_1 \max\{x-q_1^x,0\}^2\right),\ a_0=0 \\ f_n(t) &= \exp\left(\sum_{j=0}^{2} c_j x^j + \sum_{j=1}^{3} d_j \max\left\{t-q_j^t,0\right\}^2\right),\ c_0=0\end{aligned}$$

where q_1^x is the 0.5 quantile of x and q_1^t, q_2^t, q_3^t are the 0.2, 0.5, and 0.8 quantiles of t.

Figure 1 shows simulation results for $\widehat{k}_1$ and for $\left(\widehat{\phi}_n, \widehat{f}_n\right)$. The average of $\widehat{k}_1$ over 500 repetitions and its standard error (in parenthesis) appears in the caption of each figure. The averages of $\widehat{\phi}_n$ and $\widehat{f}_n$ over 500 repetitions are represented as continuous lines, while the bands represent 90% confidence intervals. The results suggest that the sieve estimators capture quite well the shape of the functions. The bias of the mean estimator is negligible compared to the standard error.

6.3.2 Case 2

Case 2 is the first misspecification study. The data is generated according to the DH model of Study 1 but the estimation procedure is that of the MPH. That is, I fit the DH data with the MPH where I parametrize the distribution of $Z \sim Gamma(\rho,1)$, ϕ_M and λ are estimated by a second degree polynomial splines in Study 3, and by the following functions in Study 4:

$$\begin{aligned}\phi_M(x) &= \exp\left(ax+bx^2\right) \\ \lambda(t) &= \alpha_1\alpha_2(\alpha_3+t)^{\alpha_2-1},\ \alpha_1,\alpha_2,\alpha_3>0\end{aligned}$$

Study 4 are the functional forms most commonly used in practice, with λ being the generalized Weibull function.

Since there is no f function in the MPH framework, I show simulation results for ϕ_M and for the survival functions: The true survival function calculated according to (7) with the gamma distribution and the estimated survival function calculated according to the standard formula

$$S_M(t|x) = \mathcal{L}_Z(\phi_M(x)\Lambda(t))$$

where $\mathcal{L}_Z$ is the Laplace transform of the gamma distribution and $\Lambda(t) = \int_0^t \lambda(s)\,ds$ is the integrated baseline hazard.

The results for Study 3 are in Figure 2 and those of Study 4 in Figure 3. As the results show the MPH can estimate quite well the true ϕ function when the baseline hazard is flexible (the band represents again the 90% confidence interval). However, it does not estimate well the survival function as it cannot capture the sorting over time as explained in 2.1. When the baseline hazard is specified as the generalized Weibull function, neither ϕ nor the survival function are approximated well.

6.3.3 Case 3

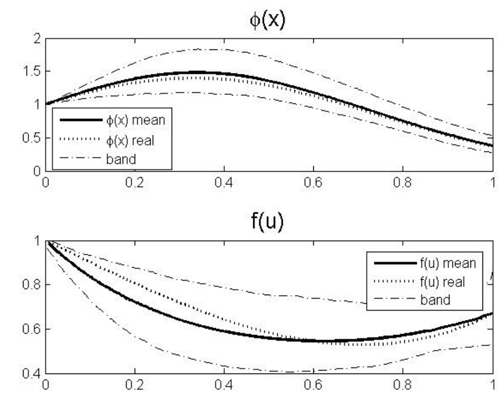
For the second misspecification study, the DGP is the one for the MPH with $Z \sim Gamma(1,1)$ and

$$\begin{aligned} \phi_M(x) &= \exp\left(2x - 3x^2\right) \\ \lambda(t) &= 2t \end{aligned}$$

while the estimation procedure is as described in 6.1 with $Z(t) \sim Gamma(\rho t, 1)$ and ϕ and f as described in Study 1. The results can be found in Figure 4. The DH appears to estimate well both ϕ and the survival function when the data was generated according to the MPH model.

Figure 2: Simulation Results for Case 1

(a) Results for Study 1: $\hat{k}_1 = 1.09$ (0.136)



(b) Results for Study 2: $\hat{k}_1 = 1.13$ (0.189)

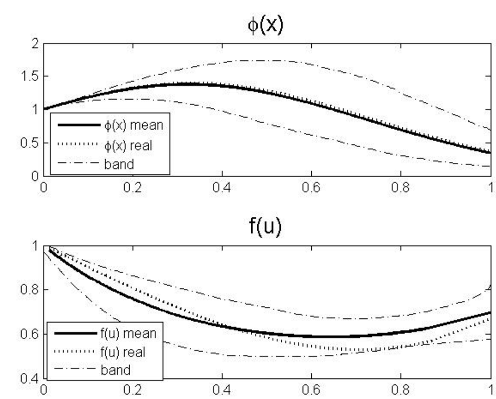
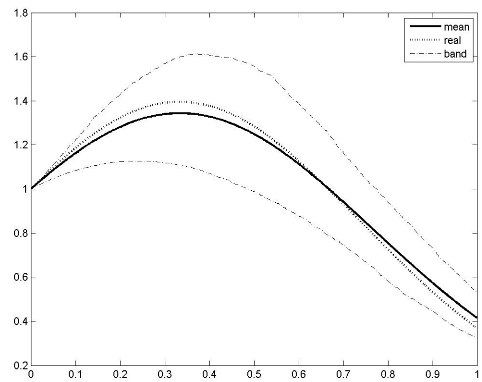


Figure 3: Simulation Results for Case 2

(a) Results for Study 3: $\phi(x)_M$



(b) Results for Study 3: $S(t|x)_M$

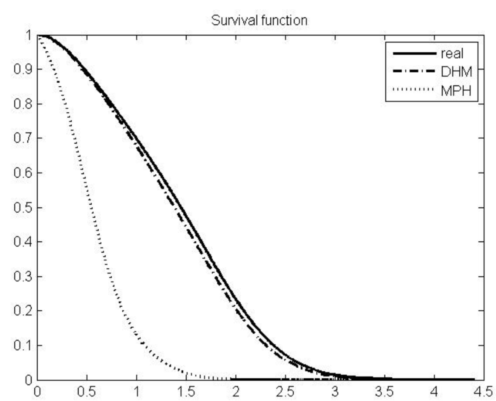
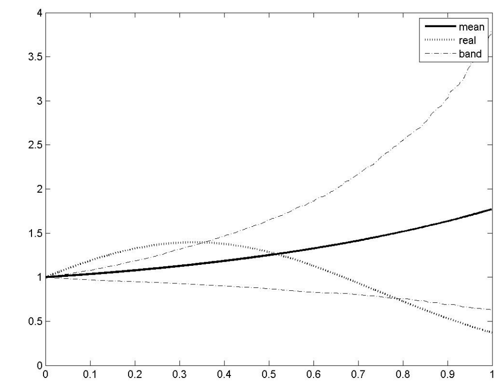


Figure 4: Simulation Results for Case 2

(a) Results for Study 4: $\phi(x)_M$



(b) Results for Study 4: $S(t|x)_M$

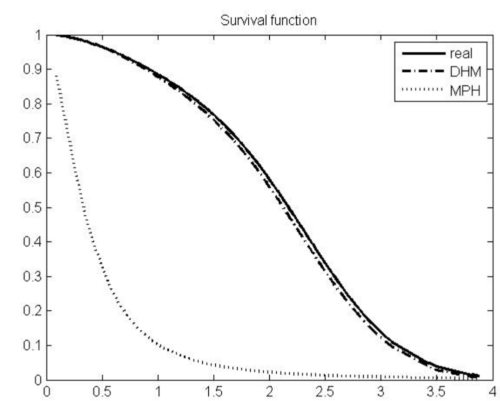
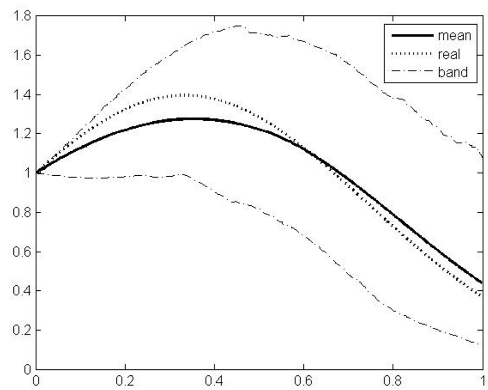
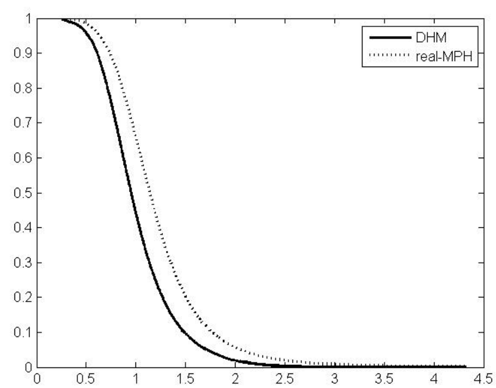


Figure 5: Simulation Results for Case 3

(a) Results for $\phi(x)$



(b) Results for $S(t|x)$



7 Conclusion

In this paper, I considered a new class of duration models in which unobserved heterogeneity changes stochastically over the duration of the spell. The changes in heterogeneity have both permanent effects on the hazard function and timing effects on the probability of exit. Standard duration models, with time invariant unobserved heterogeneity and deterministically time-varying hazard functions, cannot accommodate the type of set-ups described by the dynamically changing unobserved heterogeneity framework. I outlined the differences between major duration models and the DH class, and I showed that these models are competing in the sense that they cannot model realistically the same environments.

The paper showed the identification of the new class of models. The identification method is new to the duration literature and it is based on solving a nonlinear Volterra integral equation of the first kind. The results vary from the nonparametric to the parametric, depending on how the effects of time can be differentiated. Trade-offs in identification strategies were discussed. Estimators for the identification strategies were developed, with the more easily applicable of the two analyzed in more details. This estimator is a sieve ML estimator, which is shown to be consistent. Monte Carlo simulations show its performance in small samples.

For future research, I intend to explore a generalization of the model to allow unobserved heterogeneity to be a semi-martingale. That is, unobserved heterogeneity would be modeled as a positive transformation of a general Lévy process. This would allow for negative duration dependence at the individual level, which would enhance the applicability of the DH class. Since the process would not have independent increments, the form of the survival function would be changed: The new functional form of the survival function would need to include a term for the quadratic variation of the process. It is conjectured that the identification strategy presented in this paper would still apply.

A Appendix section

A.1 Motivating Examples

Underlying latent health and the probability of early retirement There is a large literature of econometric studies that stresses the importance of health in the decision to retire early, see Bound, Stinebrickner, and Waidmann (2010), Christensen and Kallestrup-Lamb (2012). Specifically, deteriorating health has often been cited as one of the leading causes of early retirement. Take for example the early retirement decision of registered nurses. The culprits for the retirement decision are believed to be occupational stress and the high risk of occupational injuries. In terms of occupational stress, nurses face a large risk of infectious diseases and physical violence from patients and their family members, see Gerbrich, Church, McGovern, Hansen, Nachreiner, Geisser, Ryan, Mongin, and Watt (2004). In the American Nurse Association survey from May 2004, nurses cite overwork and stress as their top reasons for early retirement. In terms of occupational injuries, BLS (2008) ranks nursing as having the highest risk of musculoskeletal disorders and as being sixth in terms of the greatest risk of strains and sprains. For more anecdotal evidence, consider what the NY State Nurses Association states: "The age at which nurses retire [...] is determined by the physical and stress-related demands of the profession."[4] In face of this anecdotal evidence, it seems important to account for on-the-job wear-and-tear effects that accumulate and that directly affect the probability of early retirement of registered nurses.

In a duration analysis setting, this example would fit (3) as follows. Occupational injuries and stress that accumulate over the duration of the job would be modeled by $Z(t)$, while f would denote if the probability of early retirement is driven by injuries and stress that happen earlier or later in the spell. It is not possible to allow for these considerations via the MPH framework.

Occupational choice and mortality In epidemiologic studies, impairment of pulmonary functions (chronic bronchitis, lung cancer) have been linked to exposure to insoluble respirable particles, such as diesel exhaust and particulate air pollution, see USEPA (2003), Pope, Burnett, Thun, Calle, Krewski, Ito, and Thurston (2002). Medical studies have shown that prolonged exposure to respirable particles leads to pulmonary deposition of the particles, see McConnochie (1990). By combining epidemiology, radiology, and physiology, it is possible to "construct a dose-response relationship where one can predict the amount of disease likely to be caused by exposure to a given amount of dust" (McConnochie (1990), page 386). However, such studies are expensive and difficult to run in all settings. One such setting is coal mining, where cumulative exposure to coal-mine dust has been linked to loss in pulmonary function, see Attfield and Hodous (1992), but where radiographic detection may not always be performed on a regular basis. Epidemiological studies that have analyzed the loss of pulmonary function and morbidity of coal miners have used standard duration models, with or without controlling for unobserved heterogeneity.[5] The model introduced in this paper is particularly well suited to the setting just described: During the spell of employment in a coal mine, dust inhalation leads to pulmonary accumulation of toxic substances. The individual rate of accumulation is unobserved and the timing of exposure may have different effects on an individual's probability of developing lung cancer: For example, early exposure may increase susceptibility to the effects of particle exposures leading to an earlier onset of pulmonary problems, or its long-term effects may be delayed, leading to a later onset.

To illustrate the interpretation of the elements entering (3), consider the examples introduced earlier. In the second example, $Z(t)$ models In the final example, $Z(t)$ stands for the accumulating damage due to

[4] http://www.nysna.org/images/pdfs/advocacy/tierVI/physTaxing_reTierVI.pdf

[5] For research papers on this topic, please refer to the National Institute for Occupational Safety and Health: http://www.cdc.gov/niosh/

inhalation of coal dust, while f models the possible increase or decrease in susceptibility to early exposure to coal dust.

A.2 Survival Function: The Laplace Exponent

Below, I present the derivation of the survival function (7). Let $F(u,t,x) = \int_u^t \phi(x) f(u)\, ds$ be square integrable with respect to the distribution of $\{Z(u)\}_0^t$. Using that $\{Z(u)\}_0^t$ has independent increments and letting $0 = u_{n,0} < u_{n,1} < ... < u_{n,n} = t$, $n = 1, 2, ...$ and a fixed $u^*_{n,j} \in [u_{n,j-1}, u_{n,j}]$, $j = 1, 2, ..., n$, obtains in mean square limit:

$$
\begin{aligned}
S(t|x) = E_Z \exp\left[-\int_0^t F(u,t,x)\, dZ(u)\right] & \\
= \quad & E_Z \exp\left[-\lim_{n\to\infty}\sum_{j=1}^n F\left(u^*_{n,j},t,x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right] && (23)\\
= \quad & E_Z \lim_{n\to\infty} \exp\left[-\sum_{j=1}^n F\left(u^*_{n,j},t,x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right] && (24)\\
= \quad & \lim_{n\to\infty} E_Z \exp\left[-\sum_{j=1}^n F\left(u^*_{n,j},t,x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right] && (25)\\
= \quad & \lim_{n\to\infty} E_Z \prod_{j=1}^n \exp\left(-F\left(u^*_{n,j},t,x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right) && (26)\\
= \quad & \lim_{n\to\infty} \prod_{j=1}^n E_Z\left[\exp\left(-F\left(u^*_{n,j},t,x\right)\left(Z(u_{n,j}) - Z(u_{n,j-1})\right)\right)\right] && (27)\\
= \quad & \lim_{n\to\infty} \prod_{j=1}^n \exp\left(-\left(u_{n,j} - u_{n,j-1}\right)\Psi\left(F\left(u^*_{n,j},t,x\right)\right)\right) && (28)\\
= \quad & \lim_{n\to\infty} \exp\left[-\sum_{j=1}^n \left(u_{n,j} - u_{n,j-1}\right)\Psi\left(F\left(u^*_{n,j},t,x\right)\right)\right] && (29)\\
= \quad & \exp\left(-\int_0^t \Psi\left(\int_u^t f(u,x)\, ds\right)\right) du && (30)\\
= \quad & \exp\left(-\int_0^t \Psi\left(f(u,x)(t-u)\right)\right) du && (31)
\end{aligned}
$$

(24) holds since $\exp(.)$ is a continuous function, so that:

$$
\exp\left(-\lim_{n\to\infty}\sum_{j=1}^n G_j\right) = \lim_{n\to\infty}\exp\left(-\sum_{j=1}^n G_j\right)
$$

(25) follows by the Bounded Convergence Theorem since:

$$\left|\exp\left(-\sum_{j=1}^{n} F\left(u_{n,j}^{*},t,x\right)\left(Z\left(u_{n,j}\right)-Z\left(u_{n,j-1}\right)\right)\right)\right| \leq 1$$

(27) follows by the independence of the increments, while (28) follows by the definition of the Laplace exponent. Since the process has independent increments it holds that:

$$\begin{aligned} E_Z \exp\left[-u\left(Z\left(t\right)-Z\left(s\right)\right)\right] &= E_Z \exp\left[-uZ\left(t-s\right)\right] \\ &= \exp\left[-\left(t-s\right)\Psi\left(u\right)\right] \end{aligned}$$

which in our problem obtains (28) .The calculation is finished off by switching back to integral notation in (31) .

Example 7 *Let $Z(t)$ be the gamma process with rate ρt and scale ν with $0 < \underline{\nu} \leq \nu \leq \overline{\nu} < \infty$ and $0 < \underline{\rho} \leq \rho \leq \overline{\rho} < \infty$. The first moment is $k_1 = \frac{\rho}{\nu}$. Assumption ID2 is satisfied with $k_1 \in \left[\frac{\underline{\rho}}{\overline{\nu}}, \frac{\overline{\rho}}{\underline{\nu}}\right]$.Let the weight function $f(x,t)$ be such that $0 < f(x,t) \leq M < \infty$. The Laplace exponent is $\Psi(\chi, k_1) = \rho \log\left(1+\frac{\chi}{\nu}\right)$, so that assumption ID4 is verified with*

$$\left|\frac{\partial^3}{\partial \chi^3}\Psi\left(\chi,k_1\right)\right| \leq \frac{2\overline{\rho}}{\underline{\nu}^3} = B$$

Example 8 *Let $Z(t)$ be the compound Poisson process with scale parameter ν, rate parameter ρt, and expected number of jumps κ with $0 < \underline{\kappa} \leq \kappa \leq \overline{\kappa} < \infty$, $0 < \underline{\nu} \leq \nu \leq \overline{\nu} < \infty$ and $0 < \underline{\rho} \leq \rho \leq \overline{\rho} < \infty$. The first moment is $k_1 = \frac{\kappa\rho}{\nu}$ so that Assumption ID2 is satisfied with $k_1 \in \left[\frac{\underline{\kappa\rho}}{\overline{\nu}}, \frac{\overline{\kappa\rho}}{\underline{\nu}}\right]$. Let the weight function $f(x,t)$ be such that $0 < f(x,t) \leq M < \infty$. The Laplace exponent is $\Psi(\chi, k_1) = \kappa\left(1-\frac{\nu}{\nu+\chi}\right)^{\rho}$, so that assumption ID4 is satisfied with*

$$\left|\frac{\partial^3}{\partial \chi^3}\Psi\left(\chi,k_1\right)\right| \leq \frac{\overline{\kappa\rho}\left(\overline{\rho}+1\right)\left(\overline{\rho}+2\right)}{\underline{\nu}^3} \equiv B$$

A.3 Proof of Theorem 2

First, the operator satisfies the inclusion $TC_w^0 \subset C_w^0$ under ID1, ID3 to ID4, and by the definition of $w(t)$. I now show the operator is a contraction. Letting $f, g \in C_w^0$ such that $f \neq g$, we have:

$$\begin{aligned}
& ||(Tf)(t) - (Tg)(t)||_{\infty,w} \\
= \; & \left| w(t)^{-1} \left[(Tf)(t) - (Tg)(t) \right] \right| \\
\leq \; & \frac{1}{k_1} \sup_{t,x} \left[\frac{3BM^2}{w(t)} \int_0^t w(u) \left| \frac{1}{w(u)} (f(u) - g(u)) \right| du \right] && (32) \\
\leq \; & \frac{1}{k_1} \sup_{t,x} \left[\frac{3BM^2}{w(t)} \int_0^t w(u)\, du \right] ||f - g||_{\infty,w} && (33) \\
= \; & \frac{\delta}{k_1} ||f - g||_{\infty,w} && (34)
\end{aligned}$$

where (32) follows by the calculation below, (33) follows by assumption ID2, and (34) follows by the definition of $w(t)$.

First, consider the calculation for (32):

$$\begin{aligned}
& \left| f^2(u) \Psi_{11}(f(u)v, k_1) - g^2(u) \Psi_{11}(g(u)v, k_1) \right| \\
\leq \; & \left| f^2(u) \Psi_{11}(f(u)v, k_1) - g^2(u) \Psi_{11}(f(u)v, k_1) \right| \\
& + \left| g^2(u) \Psi_{11}(f(u)v, k_1) - g^2(u) \Psi_{11}(g(u)v, k_1) \right| \\
\leq \; & |f(u) - g(u)| \, |f(u) + g(u)| \, |\Psi_{11}(f(u)v, k_1)| \\
& + \left| g^2(u) \right| |\Psi_{11}(f(u)v, k_1) - \Psi_{11}(g(u)v, k_1)| \\
\leq \; & 3M^2 B \, |f(u) - g(u)|
\end{aligned}$$

where the last inequality follows by ID3 and ID4.

Note that (32) holds whenever the stronger inequality (33) holds. Then an appropriate $w(t)$ needs to be defined such that

$$3BM^2 \frac{1}{w(t)} \int_0^t w(u)\, du = \delta \tag{35}$$

Once $w(t)$ is formulated according to (35), (34) holds. The solution to (35) is given by the solution to the following differential equation:

$$\frac{w(t)}{w'(t)} = \frac{3BM^2}{\delta}$$

which is

$$w(t) = \exp\left(\frac{3BM^2}{\delta} t \right)$$

Therefore, the solution exists and is an element of C_w^0. Since C_w^0 is a complete Banach space, the solution is unique.

To show $f(t) \in C_w^s$, I apply an induction argument on the smoothness parameter, s. The argument is as follows: For $s = 0$, it was shown that $f(t) \in C_w^0$. For $s \geq 1$, let the inductive hypothesis be that $f(t) \in C_w^{s-1}$ where $f(t)$ is defined by (15). By ID4 and by the inductive hypothesis:

$$f^2(u)\,\Psi_{11}(f(u)\,v, k_1) \in C_w^{s-1}$$

so that by the Fundamental Theorem of Calculus:

$$\int_0^t f^2(u)\,\Psi_{11}(f(u)\,v, k_1)\,du \in C_w^s$$

Additionally, by $ID4$, $H_{tt}\left(t, 0, k_1^0\right) \in C_w^s$, so that $f(t) \in C_w^s$.

A.4 Identification with Time Varying Covariates

When observed covariates are time-varying, it is assumed as in the standard MPH literature (see Honoré (1991), Heckman and Taber (1994)) that $x(t)$ are jump variables. That is, they are realizations of stochastic processes with continuous sample paths. When time varying observed covariates enter the hazard function multiplicatively, the effects of time coming in through the observed covariates cannot be separated from those entering through the unobservables without imposing stronger conditions on the unobserved stochastic process. Thus, in order to identify the function of covariates, it is assumed the stochastic process is known entirely. Under the assumptions below, the covariate function $\phi(x(t))$ is parametrically identified. The proof follows that of Theorem 5 in Heckman and Taber (1994).

Assumption P (i) The processes $\{x(u)\}_0^t$ and $\{Z(u)\}_0^t$ are independent; (ii) There is no contemporaneous feedback between $x(t)$ and $Z(t)$; and (iii) Future values of the two processes do not affect duration.

Assumption V (i) There are two different values for $x(t)$ at time t, $x_1(t) \neq x_2(t)$, such that these two different realizations at time t have the same sample paths up to t^-:

$$\{x_1(u)\}_0^{t^-} = \{x_2(u)\}_0^{t^-}$$

(ii) $\phi(x(t^-)) = 1$

Theorem 5 *Assume the distribution of the stochastic process* $\{Z(u)\}_0^t$ *is entirely known. Under Property 3(i), assumptions P and V, the covariate function* $\phi(x(t))$ *is identified.*

Proof. Define

$$\int_u^t \phi(x(s))\, ds = \Phi(t,u)$$

As before, the survival function is written as

$$\begin{aligned} S\left(t|\{x(u)\}_0^t\right) &= E_Z \exp\left[-\int_0^t \left[\phi(x(s))\int_0^s dZ(u)\right] ds\right] \\ &= E_Z \exp\left[-\int_0^t \Phi(t,u)\, dZ(u)\right] \\ &= \exp\left[-\int_0^t \Psi(\Phi(t,u))\, du\right] \end{aligned}$$

Then

$$\begin{aligned} &\frac{\partial}{\partial t} S\left(t|\{x(u)\}_0^t\right) \\ = &-\left[\phi(x(t))\int_0^t \Psi_1(\Phi(t,u))\, du\right] \exp\left[-\int_0^t \Psi(\Phi(t,u))\, du\right] \end{aligned} \tag{36}$$

Evaluating (36) at the same t for two different values $x_1(t)$ and $x_2(t)$ that have the same sample path up to t^-, obtains:

$$\frac{\frac{\partial}{\partial t} S\left(t|\{x_1(u)\}_0^t\right)}{\frac{\partial}{\partial t} S\left(t|\{x_2(u)\}_0^t\right)} = \frac{\phi(x_1(t))}{\phi(x_2(t))}$$

Using assumption V(ii), $\phi(x(t))$ is identified on the support of $x(t)$ for all $t \in \mathbb{R}_+$. ■

A.5 Consistency of Sieve Semiparametric MLE

Lemma 3 (B.1) *Let* $\widehat{\alpha}_n = \left(\widehat{\phi}_n, \widehat{f}_n, \widehat{\rho}_n\right)$ *be such that* $\widehat{Q}_n(\widehat{\alpha}_n) \geq \sup_{\alpha\in\mathcal{A}_n} \widehat{Q}_n(\alpha) - O_p(\eta_n)$ *with* $\eta_n = o_p(1)$. *Suppose the following conditions hold:*

B.1.1 (i) $Q(\alpha_0) < \infty$;

(ii) $\liminf_{n\to\infty} \sup_{\alpha\in\mathcal{A}_n: ||\alpha-\alpha_0||\geq\varepsilon} Q(\alpha) < Q(\alpha_0)$ uniformly in $\varepsilon > 0$.

B.1.2 (i) $\mathcal{A} \subseteq \mathbf{A}$ and $(\mathbf{A},\ ||.||)$ is a metric space;

(ii) $\mathcal{A}_n \subseteq \mathcal{A}_{n+1} \subseteq ... \subseteq \mathcal{A}$ for all $n \geq 1$, and there exists a sequence $\Pi_n\alpha_0 \in \mathcal{A}_n$ such that $||\Pi_n\alpha_0 - \alpha_0|| \to 0$ as $n \to \infty$.

B.1.3 (i) $\widehat{Q}_n(\alpha)$ is a measurable function of the data $\{x_i, t_i\}_{i=1}^n$ for all $\alpha \in \mathcal{A}_n$;

(ii) $\widehat{\alpha}_n$ is well defined and measurable.

B.1.4 (i) Let $c(m(n)) = \sup_{\alpha\in\mathcal{A}_n} \left|\widehat{Q}_n(\alpha) - Q(\alpha)\right| = o_p(1)$;

(ii) Uniformly over $\varepsilon > 0$

$$\max\left\{c\left(m\left(n\right)\right),\ \eta_n,\ \left|Q\left(\Pi_n\alpha_0\right) - Q\left(\alpha_0\right)\right|\right\} = o\left(1\right)$$

Then $d\left(\widehat{\alpha}_n, \alpha_0\right) = o_p\left(1\right)$.

Note that since there is no penalty term, $\overline{Q}_n\left(.\right) = \overline{Q}\left(.\right) = Q\left(.\right)$ in the original Lemma B1.

Let us now check the conditions of Lemma $B.1$ above.

Condition $B.1.1$(i) is satisfied by assumptions $C1$ and $C2$. In order for the criterion function $Q\left(\alpha_0\right) = E_{t,x}\log p\left(t|x,\alpha_0\right) < \infty$ and in anticipation of the information inequality, I show that

$$E_{t,x}p\left(t|x,\alpha_0\right) < \infty$$

where $p\left(t|x,\alpha_0\right) > 0$. The joint probability distribution of T and X is denoted as $P\left(t,x\right)$, while the marginal densities of $X|T$ and of T are denoted as $\mu_t\left(x\right)$ and $\pi\left(t\right)$, respectively. Then

$$\begin{aligned}
E_{t,x}p\left(t|x,\alpha_0\right) &= \int_{\mathcal{X}\times\mathcal{T}} p\left(s|w,\alpha_0\right)dP\left(s,w\right) && (37)\\
&= \int_{\mathcal{T}}\left[\int_{\mathcal{X}} p\left(s|w,\alpha_0\right)\mu_t\left(w\right)dw\right]\pi\left(s\right)ds && (38)\\
&\leq M_\phi M_f M_1\int_{\mathcal{T}}\left[\int_{\mathcal{X}}\mu_t\left(w\right)dw\right]\pi\left(s\right)ds && (39)\\
&= M_\phi M_f M_1\int_{\mathcal{T}}\mu\left(s\right)ds < \infty && (40)
\end{aligned}$$

where (39) follows since

$$\begin{aligned}
p\left(t|x,\alpha_0\right) &= S\left(t|x;\phi_0,f_0,k_0\right)\int_0^t \phi_0\left(x\right)f_0\left(u\right)\Psi_1\left(\phi_0\left(x\right)f_0\left(u\right)\left(t-u\right),k_0\right)du\\
&\leq \int_0^t \phi_0\left(x\right)f_0\left(u\right)\Psi_1\left(\phi_0\left(x\right)f_0\left(u\right)\left(t-u\right),k_0\right)du\\
&\leq M_\phi M_f M_1
\end{aligned}$$

where I used that $\sup_x \phi_0\left(x\right) \equiv M_\phi$, $\sup_t f\left(t\right) \equiv M_f$, and $\sup_{\lambda,k}\Psi_1\left(\lambda,k\right) \equiv M_1$. Equality (40) follows since $\int_{\mathcal{X}}\mu_t\left(w\right)dw = 1$ a.s. in w. Since $E_{t,x}p\left(t|x,\alpha_0\right) < \infty$, the stronger condition that $E_{t,x}\log p\left(t|x,\alpha_0\right) < \infty$ is satisfied. Thus $Q\left(\alpha_0\right) < \infty$ so that Condition $B.1.1\left(i\right)$ is satisfied.

Condition $B.1.1$(ii) is implied by assumptions $C1$ and $C2$. Since α_0 is identified and $E_{t,x}p\left(t|x,\alpha_0\right) < \infty$, by the information inequality, $Q\left(\alpha\right) - Q\left(\alpha_0\right) < 0$ for $\alpha \in \mathcal{A}_n$ with $\alpha \neq \alpha_0$.

Define:

$$\delta\left(m\left(n\right),\varepsilon\right) \equiv \sup_{\alpha\in\mathcal{A}_n:||\alpha-\alpha_0||_\infty\geq\varepsilon} Q\left(\alpha\right) - Q\left(\alpha_0\right)$$

Since $\mathcal{A}_n$ is compact, there exists a $\alpha_n^* \in \mathcal{A}_n$ with $||\alpha_n^* - \alpha_0||_\infty \geq \varepsilon > 0$ such that

$$\alpha_n^* = \arg \max_{\alpha \in \mathcal{A}_n : ||\alpha_n^* - \alpha_0||_\infty \geq \varepsilon} Q(\alpha)$$

Then, for some constant $C > 0$,

$$\begin{aligned} \delta(m(n), \varepsilon) &= Q(\alpha_n^*) - Q(\alpha_0) \\ &= Q(\alpha_n^*) - Q(\Pi_n \alpha_0) + Q(\Pi_n \alpha_0) - Q(\alpha_0) \\ &\geq C\, ||\alpha_n^* - \Pi_n \alpha_0||_\infty^2 + o(1) \end{aligned}$$

Suppose $Q(\alpha_n^*) - Q(\alpha_0) \to 0$, then $||\alpha_n^* - \Pi_n \alpha_0||_\infty^2 \to 0$. However, since

$$||\alpha_n^* - \alpha_0||_\infty^2 \leq ||\alpha_n^* - \Pi_n \alpha_0||_\infty^2 + ||\Pi_n \alpha_0 - \alpha_0||_\infty^2$$

then $||\alpha_n^* - \alpha_0||_\infty^2 \to 0$, which is a contradiction to $||\alpha_n^* - \alpha_0||_\infty^2 \geq \varepsilon > 0$. Therefore

$$\lim \inf_{n \to \infty} \delta(m(n), \varepsilon) > 0$$

Condition $B.1.2$ is implied by the way the parameter and the sieve spaces are defined in $(20a) - (20b)$ and $(21a) - (21b)$.

Condition $B.1.3$ is implied by assumptions C1, C2, and C3. In order to check B.1.3 I apply Remark B.1(1)(a) in Chen and Pouzo (2012). First note that by construction, $\mathcal{A}_n$ is a compact subset of $\mathcal{A}$ for each n under the norm defined in (22). Before showing the continuity of the criterion function in the consistency norm, let $\alpha = (\gamma, k)$ and define the following terms:

$$\begin{aligned} \gamma(u, x) &= \phi(x) f(u) \\ \Upsilon(\gamma, t, k) &= \int_0^t \gamma(x, u)\, \Psi_1(\gamma(t-u), k)\, du \\ \Gamma_1(\gamma(t-u), k) &= f(u)(1 - (t-u)\Upsilon(\gamma, t, k))\Psi_1(\gamma(t-u), k) \\ &\quad + \phi(x) f^2(u)(t-u)\Psi_{11}(\gamma(t-u), k) \\ \Gamma_2(\gamma(t-u), k) &= \phi(x)(1 - (t-u)\Upsilon(\gamma, t, \rho))\Psi_1(\gamma(t-u), k) \\ &\quad + \phi^2(x) f(u)(t-u)\Psi_{11}(\gamma(t-u), k) \\ \Gamma_3(\gamma(t-u), k) &= \phi(x) f(u)\Psi_{12}(\gamma(t-u), k) \\ &\quad - \Upsilon(\gamma, t, k)\Psi_2(\gamma(t-u), k) \end{aligned}$$

By assumption C2 and by letting $M_f \equiv \sup_t f(t)$ and $M_\phi \equiv \sup_x \phi(x)$, we have that

$$\begin{aligned} \sup \Gamma_1 (\gamma(t-u), k) &\leq M_f M_1 \\ \sup \Gamma_2 (\gamma(t-u), k) &\leq M_\phi M_1 \\ \sup \Gamma_3 (\gamma(t-u), k) &\leq M_\phi M_f M_{12} \end{aligned}$$

and by assumption C1(i), we have that for all α and x and almost all t :

$$\Upsilon(\gamma, t, k) \geq \xi \neq 0$$

By a mean value expansion of $\widehat{Q}_n(\alpha_1)$ about $\alpha = (\phi, f, k)$, with $\widetilde{\alpha}$ the mean value between α_1,$\alpha \in \mathcal{A}$ obtains:

$$\left|\widehat{Q}_n(\alpha_1) - \widehat{Q}_n(\alpha)\right| \tag{41}$$

$$\leq \frac{1}{n}\sum_i \left|\frac{1}{\Upsilon\left(\gamma_i, t_i, \widetilde{k}\right)}\right| \left[\begin{array}{l} |(\phi_1-\phi)(x_i)| \int_0^{t_i} \left|\Gamma_1\left(\widetilde{\gamma}_i(t_i-u), \widetilde{k}\right)\right| du \\ + \left(\sup_{0<u\leq t} \left|\Gamma_2\left(\widetilde{\gamma}_i(t_i-u), \widetilde{k}\right)\right|\right) \\ \times \left(\int_0^{t_i} |(f_1-f)(u)| du\right) \\ + |\rho_1-\rho| \int_0^{t_i} \left|\Gamma_3\left(\widetilde{\gamma}_i(t_i-u), \widetilde{k}\right)\right| du \end{array}\right] \tag{42}$$

$$\leq \left[\frac{1}{\xi}\left(|(M_f + c_1 M_\phi) M_1| + |M_\phi M_f M_{12}|\right)\right] ||\alpha_1 - \alpha||_\infty$$

Condition $B.1.4$(i) is implied by assumptions C1 through C3. To show the uniform convergence of the criterion function over the sieve space, we have to show that:

$$\sup_{\alpha \in \mathcal{A}_{m(n)}} \left|\widehat{Q}_n(\alpha) - Q(\alpha)\right| = o_p(1)$$

which holds if the class of functions indexing the criterion is Glivenko-Cantelli. That is, we need to show that the class of functions (43) is Glivenko-Cantelli

$$\mathcal{L} = \{l(t|x, \alpha) = \log p(t|x, \alpha) : \alpha \in \mathcal{A}_n\} \tag{43}$$

By Theorem 2.4.1 of van der Vaart and Wellner (1986) if the bracketing number $N_{[]}(\varepsilon, \mathcal{L}, L_1)$ is finite for all $\varepsilon > 0$, then $\mathcal{L}$ is Glivenko-Cantelli. I proceed now to calculate the bracketing number of the class $\mathcal{L}$.

Define

$$\begin{aligned} v &= t-u \\ \gamma_j^L &= \phi_j^L(x) f_j^L(u) \\ \gamma_j^U &= \phi_j^U(x) f_j^U(u) \end{aligned}$$

where $\gamma_j^L < \gamma_j^U$ for some $j = 1, ..., m(n)$ and $i = 1, ..., k$, where it is known that the minimum value of k is of order $O(1/\varepsilon)$, $\varepsilon > 0$. For $k^L \leq k \leq k^U$ such that $\left|k_i^U - k_i^L\right| \leq \varepsilon,\ i = 1, ..., \frac{c}{\varepsilon}$, define:

$$\begin{aligned} l_{ij}^U(x,t,\gamma,k) &= \log \int_0^t \gamma_j^U \Psi_1\left(\gamma_j^U v, k_i^U\right) du - \int_0^t \Psi\left(\gamma_j^L v, k_i^L\right) du \\ l_{ij}^L(x,t,\gamma,k) &= \log \int_0^t \gamma_j^L \Psi_1\left(\gamma_j^L v, k_i^L\right) du - \int_0^t \Psi\left(\gamma_j^U v, k_i^U\right) du \end{aligned}$$

By assumption C3, $\Psi(\lambda, k)$ is increasing in both λ and k, so for each $\alpha \in \mathcal{A}_n$ and for some $j = 1, ..., m(n)$ and $i = 1, ..., k$:

$$l_{ij}^L(x,t,\gamma,k) \leq l(x,t,\gamma,k) \leq l_{ij}^U(x,t,\gamma,k)$$

Furthermore, letting $\overline{\gamma}_j$ and $\overline{\rho}_i$ be mean values between $\left(\gamma_j^L, \gamma_j^U\right)$ and $\left(k_i^L, k_i^U\right)$ respectively, a mean value expansion obtains:

$$\begin{aligned} & \left|\int_0^t \Psi\left(\gamma_j^U v, k_i^U\right) du - \int_0^t \Psi\left(\gamma_j^L v, k_i^L\right) du\right| \\ \leq\ & t \sup_{0<u\leq t} \left|v\Psi_1\left(\overline{\gamma}_j v, \overline{k}_i\right)\right| \left[\int_0^t \left|\gamma_j^U - \gamma_j^L\right| du\right] \\ & + \left|k_i^U - k_i^L\right| \int_0^t \left|\Psi_2\left(\overline{\gamma}_j v, \overline{k}_i\right)\right| du \end{aligned} \tag{44}$$

and

$$\begin{aligned} & \left|\log \int_0^t \gamma_j^U \Psi_1\left(\gamma_j^U v, k_i^U\right) du - \log \int_0^t \gamma_j^L \Psi_1\left(\gamma_j^L v, k_i^L\right) du\right| \\ \leq\ & \frac{t \sup_{0<u\leq t}\left|\Psi_1\left(\overline{\gamma}_j v, \overline{k}_i\right) + \overline{\gamma}_j v \Psi_{11}\left(\overline{\gamma}_j v, \overline{k}_i\right)\right|}{\int_0^t \left|\overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{k}_i\right)\right| du} \left[\int_0^t \left|\gamma_j^U - \gamma_j^L\right| du\right] \\ & + \left|\frac{\int_0^t \overline{\gamma}_j \Psi_{12}\left(\overline{\gamma}_j v, \overline{k}_i\right) du}{\int_0^t \overline{\gamma}_j \Psi_1\left(\overline{\gamma}_j v, \overline{k}_i\right) du}\right| \left|k_i^U - k_i^L\right| \end{aligned} \tag{45}$$

Combining (44) and (45) obtains

$$\begin{aligned}&\left|l_{ij}^{U}\left(t,\gamma,k\right)-l_{ij}^{L}\left(t,\gamma,k\right)\right|\\ &\leq t\left[\frac{\sup_{0<u\leq t}\left|\Psi_1\left(\overline{\gamma}_j v,\overline{k}_i\right)+\overline{\gamma}_j v\Psi_{11}\left(\overline{\gamma}_j v,\overline{k}_i\right)\right|}{\int_0^t\left|\overline{\gamma}_j\Psi_1\left(\overline{\gamma}_j v,\overline{k}_i\right)\right|du}+\sup_{0<u\leq t}\left|v\Psi_1\left(\overline{\gamma}_j v,\overline{k}_i\right)\right|\right]\\ &\times\int_0^t\left|\gamma_j^U-\gamma_j^L\right|du\\ &+\left|\int_0^t\Psi_2\left(\overline{\gamma}_j v,\overline{k}_i\right)du+\frac{\int_0^t\overline{\gamma}_j\Psi_{12}\left(\overline{\gamma}_j v,\overline{k}_i\right)du}{\int_0^t\overline{\gamma}_j\Psi_1\left(\overline{\gamma}_j v,\overline{k}_i\right)du}\right|\left|k_i^U-k_i^L\right|\end{aligned}$$

Let

$$C=\sqrt{c_2}\left(\frac{M_1}{\xi}+M_1\right)+\frac{M_\phi M_f\left|M_{12}\right|}{\xi}+M_2$$

for all x and almost all t.

By using a result of Shen and Wong (1994) (page 597) and by using a bracketing entropy preservation result of Kosorok (2008) (2008, Lemma 9.25) I show below that

$$\left|\left|l_{ij}^{U}\left(t,\gamma,k\right)-l_{ij}^{L}\left(t,\gamma,k\right)\right|\right|_\infty\leq C\varepsilon$$

First, notice that $\left|k_i^U-k_i^L\right|\leq\varepsilon/2$ holds as k is a finite dimensional parameter and the covering number of Θ is of order $O\left(\frac{1}{\varepsilon}\right)$. Then I show that $\int_0^t\left|\gamma_j^U-\gamma_j^L\right|du\leq\varepsilon/2$ holds. According to a result on page 597 of Shen and Wong (1994), the bracketing entropy of Φ_n is bounded by

$$\log N_{[]}\left(\tfrac{\varepsilon}{2M_\phi},\Phi_n,||.||_\infty\right)\leq C^{'}m_n\log\left(\tfrac{2M_\phi}{\varepsilon}\right)$$

where the envelope of the class of functions indexing Φ_n is M_ϕ and where I used that if $\mathcal{F}$ is a class of functions with envelope equal to 1, then $M\mathcal{F}$, where M is a constant, has $N_{[]}\left(\varepsilon M,\mathcal{F},||.||\right)=N\left(\frac{\varepsilon}{M},\mathcal{F},||.||\right)$. Also, $\mathcal{F}_n^{int}$, the space of functions indexed by $\int_0^t f_n\left(u\right)du$ is a finite dimensional linear space with envelope $\int_0^t f_n\left(u\right)du\leq M_f$. Applying the same result in Shen and Wong (1994), we have that the bracketing entropy of $\mathcal{F}_n^{int}$ is bounded by

$$\log N_{[]}\left(\tfrac{\varepsilon}{2M_f},\mathcal{F}_n^{int},||.||_\infty\right)\leq C^{''}m\left(n\right)\log\left(\frac{2M_f}{\varepsilon}\right)$$

By bracketing entropy preservation results,[6] since both $\frac{\phi_n(x)}{M_\phi}$ and $\frac{1}{M_f}\int_0^t f_n\left(u\right)du$ are uniformly bounded by 1, letting $K=\max\left(C^{'},C^{''}\right)$ and defining the class of functions indexed by $\phi\left(x\right)\int_0^t f_n\left(u\right)du$ as Δ, we

[6] Let $\mathcal{F}$ and $\mathcal{G}$ be classes of measurable functions. Then for any probability measure P and any $1\leq r\leq\infty$, provided $f\in\mathcal{F}:|f|\leq L$ and $g\in\mathcal{G}:|g|\leq K$

$$N_{[]}\left(\varepsilon,\mathcal{F}\cdot\mathcal{G},L_r\left(P\right)\right)\leq N_{[]}\left(\frac{\varepsilon}{2L},\mathcal{F},L_r\left(P\right)\right)N_{[]}\left(\frac{\varepsilon}{2K},\mathcal{G},L_r\left(P\right)\right)\tag{46}$$

have that the class Δ is bounded by

$$\log N_{[]}\left(\varepsilon, \Delta, ||.||_{\infty}\right) \leq K m_n \log\left(\frac{4M_\phi M_f}{\varepsilon}\right)$$

which means there exists a set of functions $\left\{\phi_j^L f_j^L,\ \phi_j^U f_j^U\right\}_{j=1}^{(4M_\phi M_f/\varepsilon)^{Km_n}}$ such that the following two expressions hold for some $j = 1, ..., \left(\frac{4M_\phi M_f}{\varepsilon}\right)^{Km_n}$

$$\phi_j^L f_j^L \leq \Lambda \leq \phi_j^U f_j^U$$
$$\left|\left|\phi_j^L f_j^L - \phi_j^U f_j^U\right|\right|_{\infty} \leq \varepsilon/2$$

Then, the class of functions $\mathcal{L}$ is bounded by

$$\begin{aligned}\log N_{[]}\left(\varepsilon, \mathcal{L}, ||.||_{\infty}\right) &\leq \log N_{[]}\left(\varepsilon, \Delta, ||.||_{\infty}\right) + \log N_{[]}\left(\varepsilon, \Theta, ||.||_{E}\right) \\ &= K m_n \log\left(\frac{4M_\phi M_f}{\varepsilon}\right) + \log\left(\frac{2}{\varepsilon}\right)\end{aligned}$$

so that the class $\mathcal{L}$ is Glivenko-Cantelli. Moreover, $\mathcal{L}$ is Donsker. Then we can find $c^{\widehat{Q}}(m_n)$ explicitly by calculating the integral below

$$\int_0^1 \sqrt{K m_n \log\left(\frac{4M_\phi M_f}{\varepsilon}\right) + \log\left(\frac{2}{\varepsilon}\right)} d\varepsilon$$

which obtains a result of order $O\left(\sqrt{1+m_n}\right)$. Therefore:

$$c^{\widehat{Q}}(m_n) = \left(\frac{1+m_n}{n}\right)^{1/2}$$

(ii) The second part of condition B.1.4 states that

$$\begin{aligned} c^{\widehat{Q}}(m_n) &= o(1) && (47a) \\ \left|Q\left(\Pi_n \alpha_0\right) - Q\left(\alpha_0\right)\right| &= o(1) && (47b) \\ \eta_n &= o(1) && (47c) \end{aligned}$$

$(47a)$ holds since d_θ is fixed and by construction $m_n \to \infty$ at a rate slower than n. $(47b)$ is satisfied by the continuity of $Q(\alpha)$ and by $\Pi_n \alpha_0 \to \alpha_0$ from Condition B.1.2(ii). $(47c)$ holds with η_n small enough by the uniform convergence of the criterion function.

References

Aalen, O., O. Borgan, and H. Gjessing (2008): *Survival and Event History Analysis: A Process Point of View*. Springer.

Aalen, O., and N. Hjort (2002): "Frailty Models that Yield Proportional Hazards," *Statistics and Probability Letters*, 58, 335–342.

Abbring, J. (2012): "Mixed Hitting-Time Models," *Econometrica*, 80, 783–819.

Ariga, K., Y. Ohkusa, and G. Brunello (1999): "Fast Track: Is It in the Genes? The Promotion Policy of a Large Japanese Firm," *Journal of Economic Behavior and Organization*, 38(4), 385–402.

Attfield, M., and T. Hodous (1992): "Pulmonary Function of US Coal Miners Related to Dust Exposure Estimates," *American Review of Respiratory Diseases*, 145(3), 605 – 609.

Berg, C., Y. Chen, and M. Ismail (2002): "Small Eigenvalues of Large Henkel Matrices: The Indeterminate Case," *Mathematica Scandinavica*, 91(1), 67–81.

Bertoin, J. (1996): *Lévy Processes*. Cambridge University Press.

BLS (2008): *Musculoskeletal Disorders and Days Away from Work in 2007*Washington D.C. Bureau of Labor Statistics.

Bound, J., T. Stinebrickner, and T. Waidmann (2010): "Health, Economic Resources, and the Work Decision of Older Men," *Journal of Econometrics*, 156(1), 106–129.

Chen, X., and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 20(1), 277–321.

Christensen, B., and M. Kallestrup-Lamb (2012): "The Impact of Health Changes on Labor Supply: Evidence from Merged Data on Individual Objective Medical Diagnosis Codes and Early Retirement Behavior," *Health Economics*, 21, 56 – 100.

Dabrowska, D. (2006): "Estimation in a Class of Semiparametric Transformation Models," *Lecture Notes-Monograph Series*, 49, 131–169.

Dykstra, R., and P. Laud (1981): "A Bayesian Nonparametric Approach to Reliability," *Annals of Statistics*, 9(2), 356–367.

Elbers, C., and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49(3), 403–409.

Gerbrich, S., T. Church, P. McGovern, H. Hansen, N. Nachreiner, M. Geisser, A. Ryan, S. Mongin, and G. Watt (2004): "An Epidemiological Study of the Magnitude and Consequences of Work Related Violence: The Minnesota Nurses Study," *Occupational and Environmental Medicine*, 61(6), 495 – 503.

Gibbons, R., and M. Waldman (2004): "Task-Specific Human Capital," *The American Economic Review*, 94(2), 203–207.

Gjessing, H., O. Aalen, and N. Hjort (2003): "Frailty Models Based on Lévy Processes," *Advances in Applied Probability*, 35, 532–550.

Gnedin, A., and J. Pitman (2008): "Moments of Convex Distribution Functions and Completely Alternating Sequences," in *Probability and Statistics: Essays in Honor of David A. Freedman*, ed. by D. Nolan, and T. Speed, pp. 30–41. Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Heckman, J., and B. Singer (1984): "Econometric Duration Analysis," *Journal of Econometrics*, 24(1-2), 63–132.

Heckman, J., and C. Taber (1994): "Econometric Mixture Models and More General Models for Unobservables in Duration Analysis," *Statistical Methods in Medical Research*, 3(3), 279–302.

Honoré, B. (1991): "Identification Results for Duration Models with Multiple Spells or Time-Varying Covariates," working paper, Northwestern University.

Kambourov, G., and I. Manovskii (2009): "Occupational Specificity of Human Capital," *International Economic Review*, 50(1), 63–115.

Kebir, Y. (1991): "On Hazard Rate Processes," *Naval Research Logistics*, 6(38), 865–876.

Kosorok, M. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer: New York.

McConnochie, K. (1990): "Environmental Lung Disorders: Mineral Pneumoconiosis," in *Radiologic Diagnosis of Chest Disease*, ed. by M. Sperber, M., pp. 386–401. Springer US.

Pope, C., R. Burnett, M. Thun, E. Calle, D. Krewski, K. Ito, and G. Thurston (2002): "Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution," *Journal of the Americal Medical Association*, 287, 1132–1141.

Shen, X., and W. Wong (1994): "Convergence Rate of Sieve Estimates," *The Annals of Statistics*, 22(2), 580–615.

Singpurwalla, N. (1995): "Survival in Dynamic Environments," *Statistical Science*, 10(1), 86–103.

——— (2006): "The Hazard Potential: Introduction and Overview," *Journal of the American Statistical Association*, 101(476), 1705–1717.

Stock, J. (1988): "Estimating Continuous-Time Processes Subject to Time Deformation: An Application to Postwar U.S. GNP," *Journal of the American Statistical Association*, 83(401), 77–85.

USEPA (2003): *Fourth External Review Draft of Air Quality Criteria for Particulate Matter*US Environmental Protection Agency, National Center for Environmental Assessment, Office of Research and Development, Research Triangle Park NC.

Van den Berg, G. (2001): "Duration Models: Specification, Identification and Multiple Durations," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5 of *Handbook of Econometrics*, chap. 55, pp. 3381–3460. Elsevier.

van der Vaart, A., and J. Wellner (1986): *Weak Convergence and Empirical Processes with Application to Statistics*. New York: Springer-Verlag.