



TI 2007-057/3

Tinbergen Institute Discussion Paper

Mixed Hitting-Time Models

Jaap H. Abbring

VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at
<http://www.tinbergen.nl>.

Mixed Hitting-Time Models*

Jaap H. Abbring[†]

August 2009

Abstract

We study mixed hitting-time models, which specify durations as the first time a Lévy process—a continuous-time process with stationary and independent increments—crosses a heterogeneous threshold. Such models are of substantial interest because they can be reduced from optimal-stopping models with heterogeneous agents that do not naturally produce a mixed proportional hazards structure. We show how strategies for analyzing the identifiability of the mixed proportional hazards model can be adapted to prove identifiability of a hitting-time model with observed covariates and unobserved heterogeneity. We discuss inference from censored data and give examples of structural applications. We conclude by discussing the relative merits of both models as complementary frameworks for econometric duration analysis.

*First draft: July 2007. I thank Jeffrey Campbell, Andrew Chesher, Frank de Jong, Geert Ridder; seminar participants at Tinbergen Institute Amsterdam, Toulouse, Paris School of Economics, Chicago, Northwestern, Tilburg, IZA Bonn, and RWI Essen; and participants in the Cowles 75th Anniversary Conference at Yale, the Econometric Society European Meeting in Milan, and the Heterogeneity in Labor Markets conference at Chicago's Milton Friedman Institute for helpful suggestions and discussion.

[†]CentER, Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: J.H.Abbing@uvt.nl. [Http://center.uvt.nl/staff/abbring](http://center.uvt.nl/staff/abbring).

Keywords: duration analysis, hitting time, identifiability, Lévy process, mixture.

JEL codes: C14, C41.

1 Introduction

Mixed hitting-time (MHT) models are mixture duration models that specify durations as the first time a latent stochastic process crosses a heterogeneous threshold. In this paper, we explore the empirical content of an MHT model in which the latent process is a spectrally-negative Lévy process, a continuous-time process with stationary and independent increments and no positive jumps, and the threshold is proportional in the effects of observed covariates and unobserved heterogeneity. We show that existing strategies for analyzing the identifiability of Lancaster's (1979) mixed proportional hazards (MPH) model can be adapted to prove this model's identifiability. In particular, we show that the latent Lévy process, the covariate effect on the threshold, and the distribution of the unobserved heterogeneity in the threshold are uniquely determined by data on durations and covariates. Some assumption on the tails of the heterogeneity distribution or the latent process is required for full identification. Some conditions for identification that may or may not be satisfied in the analogous MPH problem here follow from the Lévy structure and do not require additional assumptions. Finally, multiple-spell data facilitate identification of much more general models, with arbitrary interactions of the latent process and unobserved heterogeneity with covariates.

Mixed hitting-time models are of substantial interest because they are closely related to economic models in which agents optimally time discrete actions, with payoffs driven by Brownian motion (Dixit and Pindyck, 1994; Stokey, 2009) or a more general Lévy process (Kyprianou, 2006; Boyarchenko and Levendorskiĭ, 2007). Such models' optimal decision rules routinely involve thresholds, and heterogeneity in their primitives generates threshold heterogeneity. In this paper, we develop a range of examples. In the simplest of these, agents are endowed with an option to invest in a project, at a time of their choice. Investment incurs a given cost; in return, the agent receives the project's value at the time of the investment. The log of this value follows a Brownian motion. At each point in time, the agent weighs the direct payoffs of investing in the project, net of the amount

to be invested, against the value of retaining the option of investing later, given the primitive parameters and the history of project values. The agent maximizes his expected discounted payoffs by investing when the project's value hits a time-invariant threshold. Primitive heterogeneity; such as variation in initial project values, investment costs, and discount rates across agents; induces heterogeneity in the threshold. Consequently, data on investment times and covariates can be analyzed with an MHT model, and our identification results show that this yields estimates of the latent process for project values and the agents' investment decision rules. These estimates may be of interest by themselves, or can be used as inputs in a further analysis of the model's remaining primitives. Similar results are found for model variants in which the latent process induces a flow of payoffs, such as wages or profits, and extensions in which the duration of interest is embedded in a multistate transition model, such as match durations in a search-matching model.

Hitting-time models based on Brownian motion or more general Lévy processes do not generally predict hazard rates that are proportional in the effects of elapsed duration and those of observed and unobserved heterogeneity. Because proportionality is key to the identifiability of the MPH model (Van den Berg, 2001), estimates of an MPH model on data from an MHT model are not likely to be informative on true state dependence and heterogeneity. Thus, there are *structural* reasons to use an MHT model in applications in which agents are assumed to solve an optimal stopping problem driven by Brownian motion or a more general Lévy process. In addition, there may be *statistical* reasons: We will give examples of MHT specifications to which no observationally equivalent MPH specifications exist.

The MHT approach to continuous-time duration analysis is inspired by the literature on discrete-time discrete choice models pioneered by Heckman (1981a,c). As in this literature, we explicitly build a statistical model for dynamic discrete outcomes on a latent process that can serve as the state in a dynamic discrete choice problem. In particular, Heckman and Navarro (2007) discuss a general discrete-time mixture duration model

based on a latent process crossing thresholds (see [Abbring and Heckman, 2007, 2008](#), for reviews). They emphasize the distinction between this model and a discrete-time MPH model and its extensions, and study its identifiability and its relation to dynamic discrete choice. This paper complements theirs with an analysis in continuous time. This paper's continuous-time setting facilitates a different approach to the identification analysis and connects our work to the popular continuous-time MPH model and to continuous-time economic models.

Applications in labor economics include the analysis of job tenure, strikes, and unemployment. In his classic text book on econometric duration analysis, [Lancaster \(1990, Sections 3.4.2, 5.7 and 6.5\)](#) reviews a canonical special case of our model, a reduced-form marginal duration model that specifies durations as the first-passage times of a Brownian motion with drift, and relates it to [Jovanovic's \(1979;1984\)](#) job tenure model. In [Lancaster \(1972\)](#), he applies this model to strike durations, interpreting the gap between the Brownian motion and the threshold as the level of disagreement. [Shimer \(2008\)](#) more recently analyzed unemployment durations using [Alvarez and Shimer's \(2008\)](#) model of search and rest unemployment, which involves a threshold rule for transitions between rest unemployment and work. Possible applications in other fields of economics include marriage and divorce, firm entry and exit, and credit default.

Statisticians have increasingly been studying continuous-time duration models based on latent processes, including MHT models that are special cases of this paper's model (e.g. [Singpurwalla, 1995](#); [Aalen and Gjessing, 2001](#); [Lee and Whitmore, 2004, 2006](#)). This literature is very informative on the descriptive implications of such models, but is silent about their identifiability. Our contribution to both the econometrics and the statistics literatures is a rigorous analysis of the empirical content of a nonparametric class of MHT models with covariates.

The paper is organized as follows. Section [2](#) introduces the MHT model. Section [3](#) develops the paper's main ideas for the well-understood, and therefore instructive, special

case in which the latent Lévy process is a Brownian motion with drift. In particular, the MHT model structure is explored, and the key connection between the analysis of its empirical content and the MPH identification literature is highlighted. Section 4 presents the general MHT model’s implications for the data and the main identification results. Section 5 discusses estimation from complete and censored data. Section 6 presents examples of economic models that can be analyzed using the MHT model. Section 7 discusses extensions with time-varying covariates, and to latent processes with nonstationary and dependent increments. Finally, Section 8 concludes with some discussion of the relative merits of the MHT and MPH models as complementary frameworks for econometric duration analysis.

2 The Model

We model the distribution of a random duration T conditional on observed covariates X by specifying T as the first time a real-valued Lévy process $\{Y\} \equiv \{Y(t); t \geq 0\}$ crosses a threshold that depends on X and some unobservables V .

A Lévy process is the continuous-time equivalent of a random walk: It has stationary and independent increments. [Bertoin \(1996\)](#) provides a comprehensive exposition of Lévy processes and their analysis. Formally, we have

Definition 1. *A Lévy process is a stochastic process $\{Y\}$ such that the increment $Y(t + \Delta) - Y(t)$ is independent of $\{Y(t'); 0 \leq t' \leq t\}$ and has the same distribution as $Y(\Delta)$, for every $t, \Delta \geq 0$.*

We take $\{Y\}$ to have right-continuous sample paths with left limits. Note that Definition 1 implies that $Y(0) = 0$ almost surely.

An important example of a Lévy process is the scalar Brownian motion with drift, in which case $Y(\Delta)$ is normally distributed with mean $\mu\Delta$ and variance $\sigma^2\Delta$, for some scalar parameters $\mu \in \mathbb{R}$ and $\sigma \in [0, \infty)$. Brownian motion is the single Lévy process with

continuous sample paths. In general, Lévy processes may have jumps. The jump process $\{\Delta Y\}$ of a Lévy process $\{Y\}$ is a Poisson point process with characteristic measure Υ such that $\int \min\{1, x^2\}\Upsilon(dx) < \infty$, and any Lévy process $\{Y\}$ can be written as the sum of a Brownian motion with drift and an independent pure-jump process with jumps governed by such a point process (Bertoin, 1996, Chapter I, Theorem 1). The characteristic measure of $\{Y\}$'s jump process is called its *Lévy measure* and, together with the drift and dispersion parameters of its Brownian motion component, fully characterizes $\{Y\}$'s distributional properties. Key examples of pure-jump Lévy processes are compound Poisson processes, which have independently and identically distributed jumps at Poisson times. In fact, in distribution, each Lévy process can be approximated arbitrary closely by a sequence of compound Poisson processes (Feller, 1971, Section IX.5, Theorem 2).

Let $T(y)$ denote the first time that the Lévy process $\{Y\}$ exceeds a threshold $y \in [0, \infty)$: $T(y) \equiv \inf\{t \geq 0 : Y(t) > y\}$. Here, we use the convention that $\inf \emptyset \equiv \infty$; that is, we set $T(y) = \infty$ if $\{Y\}$ never exceeds y . For completeness, we set $T(\infty) = \infty$. The (proportional) mixed hitting-time (MHT) model specifies that T is the first time that $Y(t)$ crosses $\phi(X)V$, or

$$T = T[\phi(X)V]; \tag{1}$$

for some observed covariates X with support $\mathcal{X} \in \mathbb{R}^k$, measurable function $\phi : \mathcal{X} \mapsto (0, \infty)$, and positive random variable V , with (X, V) independent of $\{Y\}$.

The hitting times $T(y)$ characterize durations for given thresholds $y \in [0, \infty)$, and thus for given individual characteristics (X, V) . Their analysis is particularly straightforward in the case that $\{Y\}$ is *spectrally negative*. In this case, $\{Y\}$ has no positive jumps; that is, its Lévy measure Υ has negative support. Because $\{Y\}$ is continuous from the right, this implies that $\{Y\}$ equals the threshold at each finite hitting time: $Y[T(y)] = y$ if $T(y) < \infty$. In turn, this ensures that $T(y)$ is easy to characterize in terms of the parameters of $\{Y\}$ (see Section 4.1). Throughout the paper's remainder, we assume that

$\{Y\}$ is spectrally negative. Note that this includes Brownian motion with drift as a special case.

Variation in $\phi(X)V$ corresponds to heterogeneity in individual thresholds. The factor V is an unobserved individual effect and is assumed to be distributed independently of X with distribution G on $(0, \infty]$. This explicitly allows for an unobserved subpopulation $\{V = \infty\}$ of *stayers*, on which $T = T(\infty) = \infty$. In addition, there may be *defecting movers*: For some specifications of $\{Y\}$, $T = \infty$ with positive probability on $\{V < \infty\}$. The distinction between stayers and defecting movers can be of substantial interest (see [Abbring, 2002](#), for discussion). We exclude the two trivial cases in which $T = \infty$ almost surely, the case in which the population consists of only stayers ($\Pr(V < \infty) = 0$) and the case in which all movers defect ($\{Y\}$ is nonincreasing). Because $\{Y\}$ has only negative shocks, this requires that either $\mu > 0$ or $\sigma > 0$.

For expositional convenience, we have assumed that the threshold $\phi(X)V$ is almost surely positive. This avoids a mass of agents who employ a zero threshold and have zero durations. Appendix [A](#) shows that this assumption, and the assumption that $\phi(X)$ is finite, can be relaxed.

We will pay some specific attention to a version of this model without covariates, that is $\phi = 1$. Such a model can be applied to strata defined by the covariates, without restrictions across the strata, and can thus be interpreted as a more general, nonproportional MHT model.

Because the increments of the Lévy process are independent of its history, in particular its initial condition, an equivalent model arises if we take the initial condition $Y(0)$ to be heterogeneous, say equal to $-\phi(X)V$, and fix the threshold at a common value of zero. Similarly, we can redistribute a linear drift μt from $\{Y\}$ to the threshold without changing the implications for T . In the Lévy-based MHT model, all that matters to the specification of T is the first time that $\phi(X)V - Y(t)$ falls below zero. In different applications, different interpretations in terms of heterogeneous initial conditions and heterogeneous and time-

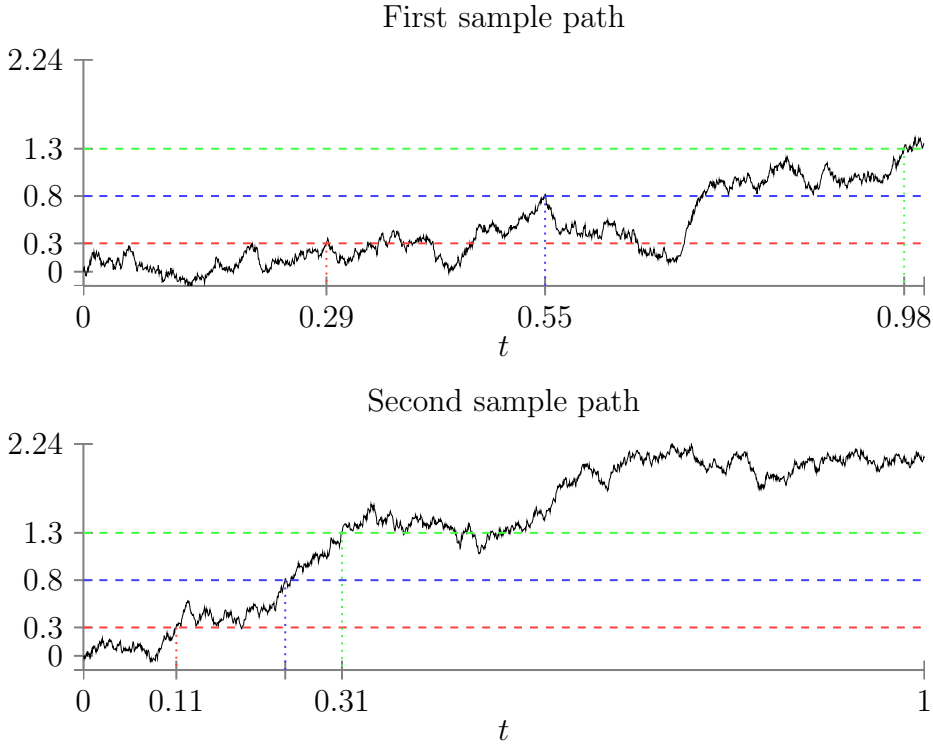


Figure 1: Two sample paths of $Y(t) = t + W(t)$, three possible thresholds, and the corresponding first hitting times.

varying thresholds may be appropriate. Section 6’s structural examples illustrate this.

3 Gaussian Example

We illustrate some of this paper’s key ideas with the canonical example in which $\{Y\}$ is a Brownian motion with upward drift. In this case, we can write $Y(t) = \mu t + \sigma W(t)$, for some $\mu \in (0, \infty)$ and $\sigma \in [0, \infty)$; with $W(t)$ a standard Brownian motion, or Wiener process, and $W(0) = 0$. Note that the Lévy measure $\Upsilon = 0$ in this example. For expositional convenience, in this section only, we assume that $V < \infty$ almost surely. With $\mu > 0$, so that $T(y) < \infty$ for $y \in [0, \infty)$, this ensures that T is nondefective.

3.1 Characterization

Figure 1 plots two sample paths of $\{Y\}$ for the case in which $\mu = \sigma = 1$, with three possible exit thresholds; 0.3, 0.8, and 1.3. For a given threshold y , the time that each path first crosses that threshold is a realization of $T(y)$.

If $\sigma > 0$, the distribution of $T(y)$, $y \in [0, \infty)$, is inverse Gaussian with location parameter y/μ and scale parameter $(y/\sigma)^2$ (Cox and Miller, 1965). Its survival function is

$$\bar{F}(t|y) \equiv \Pr [T(y) > t] = \Phi \left(\frac{y - \mu t}{\sigma \sqrt{t}} \right) - \exp \left(\frac{2\mu y}{\sigma^2} \right) \Phi \left(-\frac{y + \mu t}{\sigma \sqrt{t}} \right), \quad (2)$$

and its Lebesgue density

$$f(t|y) = \frac{y}{\sigma \sqrt{2\pi t^3}} \exp \left(-\frac{(y - \mu t)^2}{2\sigma^2 t} \right). \quad (3)$$

Here, Φ denotes the standard normal cumulative distribution function. In this case, conditional on the observed covariates X only, the MHT model specifies $T = T[\phi(X)V]$ as a mixture of inverse Gaussian distributions. This is the duration model reviewed by Lancaster (1990, Sections 4.2 and 5.7), extended with observed and unobserved heterogeneity in its parameters.

In the polar case with $\sigma = 0$, we have that $Y(t) = \mu t$, and $T(y) = \mu^{-1}y$ is a deterministic linear function of the threshold y . Then, $T = \mu^{-1}\phi(X)V$, and the MHT model reduces to the accelerated failure time (AFT) model for $T|X$: V takes the role of a “baseline” duration variable, which is “accelerated” or “decelerated” by the covariate-dependent factor $\mu^{-1}\phi(X)$ (see Equation (45) and its discussion in Cox, 1972, pp. 200–201). An interpretation of the AFT model based on the MHT model is that it attributes all variation in durations for given X to *ex ante* unobserved heterogeneity. The fact that the MHT model can capture situations in which little or no uncertainty is resolved during the spell is appealing. Meyer (1990), for example, entertains this possibility (using a model due

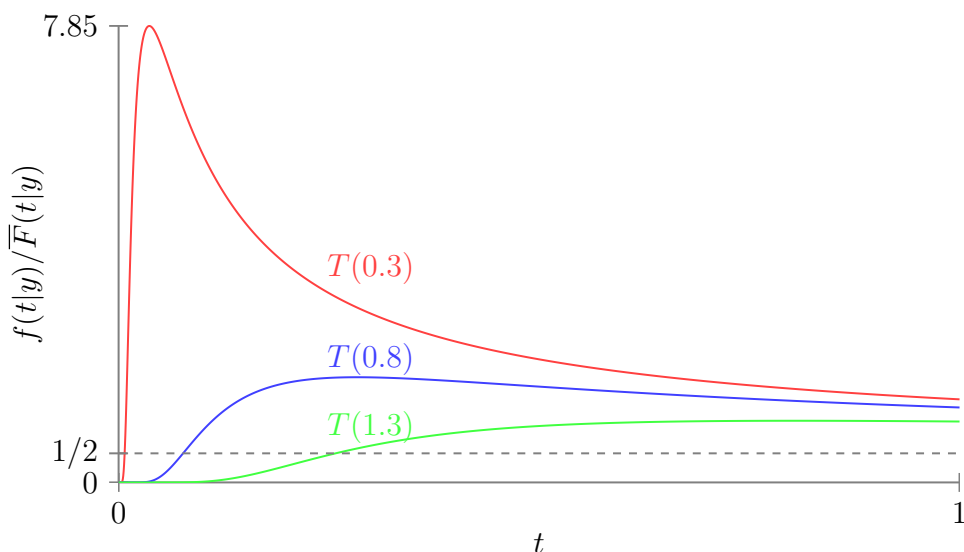


Figure 2: Hazard rates of $T(y)$ for Figure 1's three thresholds y and $Y(t) = t + W(t)$.

to [Moffitt and Nicholson, 1982](#)) as an alternative for a job search model in his study of unemployment insurance and durations.

Although the hazard rate of $T(y)$, $f(t|y)/\bar{F}(t|y)$, is not a primitive of the MHT model, it is useful to display it for comparison with hazard-based models like the MPH model. Figure 2 plots the hazard-rate paths for Figure 1's three threshold levels y ; 0.3, 0.8, and 1.3; again for the case in which $\mu = 1$ and $\sigma = 1$. The hazard paths have a hump-shaped pattern: They start at 0, rise to a maximum that is attained between $y^2/(3\sigma^2) = y^2/3$ and $2y^2/(3\sigma^2) = 2y^2/3$, and then fall towards a limit $\mu^2/(2\sigma^2) = 1/2$. The hazard rate corresponding to the lowest threshold ($y = 0.3$) is falling at most times, whereas that corresponding to the highest threshold ($y = 1.3$) is increasing for nearly all plotted times. Clearly, the hazard rates are not proportional; in this sense, the MHT model is structurally different from the MPH model.

By mixing over thresholds, a wide variety of duration distributions can be generated. Take, for example, the polar case with $\sigma = 0$, in which $T = \mu^{-1}\phi(X)V$. If $\phi = 1$, then T is independent of X , and we can match any distribution of $T = \mu^{-1}V$, by setting G equal to the corresponding distribution of μT . If ϕ is not trivial, and T depends on X ; then we can still match any distribution of $T|(X = x_0)$, by setting G in a similar way, for given $x_0 \in \mathcal{X}$.

However, the required specification of G depends on x_0 , through $\phi(x_0)$. Consequently, this construction cannot be repeated to match an arbitrary distribution of $T|X$ over the entire support \mathcal{X} of X without violating the assumption that V is independent of X . In this polar case, the distribution of $T|X$ is necessarily a rescaled version of that of $T|(X = x_0)$, with scale factor $\phi(X)/\phi(x_0)$. In general, the MHT model does not restrict the distribution of $T|(X = x_0)$ for given $x_0 \in \mathcal{X}$, but does restrict the way T depends on X .

3.2 Identifiability

This takes us to the question whether the model's structural determinants; μ , σ , ϕ , and G ; can be uniquely determined (“identified”) from large-sample data, the distribution of $T|X$. The latter is uniquely characterized by its Laplace transform, $\mathcal{L}_T(s|X) \equiv \mathbb{E}[\exp(-sT)|X]$, $s \in [0, \infty)$ (Feller, 1971, Section XIII.1, Theorem 1). It turns out to be particularly convenient, both in this Gaussian example and in the general case, to study identification of the model's determinants in terms of $\mathcal{L}_T(\cdot|X)$.

This requires that we express $\mathcal{L}_T(\cdot|X)$ in the model determinants μ , σ , ϕ , and G ; and check whether the latter are uniquely determined by $\mathcal{L}_T(\cdot|X)$. To this end, note that the (unconditional) Laplace transform of $T(y)$ is given by

$$\mathcal{L}_{T(y)}(s) = \exp[-y\Lambda(s)], \quad \text{with } \Lambda(s) \equiv \begin{cases} \frac{\sqrt{\mu^2 + 2\sigma^2 s} - \mu}{\sigma^2} & \text{if } \sigma > 0; \\ s/\mu & \text{if } \sigma = 0. \end{cases} \quad (4)$$

so that $\mathcal{L}_T(s|X, V) = \exp[-\phi(X)V\Lambda(s)]$. Here, $\mathcal{L}_{T(y)}$ and $\mathcal{L}_T(\cdot|X, V)$ are defined analogously to $\mathcal{L}_T(\cdot|X)$. The Laplace transform of the data; in terms of the model determinants μ , σ , ϕ , and G ; follows by taking the expectation of $\mathcal{L}_T(\cdot|X, V)$ conditional on X :

$$\mathcal{L}_T(s|X) = \mathcal{L}[\phi(X)\Lambda(s)]. \quad (5)$$

Here, \mathcal{L} the Laplace transform of the distribution G of V .

One trivial identification problem requires our attention upfront. Take the time T implied by (1) if $\{Y\}$ is a Brownian motion with parameters μ and σ , with threshold $\phi(X)V$. Clearly, the process $\{\kappa\nu Y\}$ and threshold $\kappa\phi(X)\nu V$; with $\kappa, \nu \in (0, \infty)$; produce the same time T . Thus, they imply the same distribution of $T|X$ and are *observationally equivalent*. Like the latent error and index in static discrete-choice models, the latent process and threshold in the MHT model are only identified up to scale. At best, we can determine the distribution of $\{Y\}$, ϕ , and G up to two innocuous scale normalizations.

Key to this paper’s identifiability analysis is an analogy with the analysis of the MPH model. To appreciate this, note that the right-hand side of (5) equals the survival function— rather than the Laplace transform— of $T|X$ in an MPH model with integrated baseline hazard Λ , covariate effect $\phi(X)$, and unobserved-heterogeneity distribution G . It is easily checked that, in the MHT model, Λ is an increasing function such that $\lim_{s \rightarrow \infty} \Lambda(s) = \infty$ and that, in this example, $\Lambda(0) = 0$. We can therefore borrow insights from the MPH identification literature pioneered by [Elbers and Ridder \(1982\)](#), [Heckman and Singer \(1984a\)](#), and [Ridder \(1990\)](#); exploiting the structure imposed by the MHT model on, in particular, Λ .

Consider the case that $\phi(X) = \exp(X'\beta)$ for some parameter vector $\beta \in \mathbb{R}^k$. Note that Λ is differentiable on $(0, \infty)$ and that $0 < \lim_{s \downarrow 0} \Lambda'(s) = \mu^{-1} < \infty$. Thus, [Ridder and Woutersen’s \(2003\) Proposition 1](#) implies that μ, σ, β , and G are uniquely determined from $\mathcal{L}_T(\cdot|X)$ under support conditions on the covariates X , up to the two scale normalizations discussed earlier. In the next section, we extend this result to general spectrally-negative Lévy processes and general distributions G . Doing so, we rely on the key insight that the representation (5) of the data in terms of the model primitives continues to hold, but with a more general, semiparametric specification of Λ . We show that the special structure of Λ facilitates sharper identification results than those available for the MPH model.

Note that, even in this Gaussian special case, covariate variation is crucial to identifiability. For example, take again the polar case with $\sigma = 0$. Suppose that $\phi = 1$ and $\mu = 1$,

so that $T = V$. Clearly, if V has an inverse Gaussian distribution with location parameter $\tilde{\mu}^{-1}$ and scale parameter $\tilde{\sigma}^{-2}$; with $\tilde{\mu}, \tilde{\sigma} \in (0, \infty)$; then an alternative specification with a latent process $\{\tilde{Y}\}$ such that $\tilde{Y}(t) = \tilde{\mu}t + \tilde{\sigma}W(t)$ and a homogeneous unit threshold is observationally equivalent.

4 Empirical Content

We now return to the general framework of Section 2. So, suppose that $\{Y\}$ is a spectrally-negative Lévy process, but not necessarily a Brownian motion, and that G is general, with possibly $\Pr(V < \infty) < 1$.

4.1 Characterization

We first characterize the hitting-time process $\{T\} \equiv \{T(y); y \geq 0\}$ implied by $\{Y\}$. Its distribution can be characterized in terms of its Laplace transform, which we now define as

$$\mathcal{L}_{T(y)}(s) \equiv \mathbb{E} \left[\exp[-sT(y)] \cdot I[T(y) < \infty] \right], \quad s \in [0, \infty); \quad (6)$$

with $I(\cdot) = 1$ if \cdot is true, and 0 otherwise. The factor $I[T(y) < \infty]$ makes explicit that the distribution of $T(y)$ may be defective. Note that $\Pr[T(y) = \infty] = 1 - \mathcal{L}_{T(y)}(0)$.

Before we can derive $\mathcal{L}_{T(y)}$, we first have to introduce a common probabilistic characterization of the latent Lévy process. Recall from Section 2 that $\{Y\}$ can be decomposed in a Brownian motion with drift and an independent pure-jump process with jumps $\{\Delta Y\}$ following a Poisson point process. Therefore, $\{Y\}$ is fully characterized by the drift and dispersion coefficients μ and σ of its Brownian motion component and the characteristic (Lévy) measure Υ of $\{\Delta Y\}$. The latter satisfies $\int \min\{1, x^2\} \Upsilon(dx) < \infty$ and, because we exclude positive jumps, has negative support. It follows (Bertoin, 1996, Section VII.1) that $\mathbb{E}[\exp(sY(t))] = \exp[\psi(s)t]$, for $s \in \mathbb{C}$ with nonnegative real part, with the *Laplace*

exponent ψ given by the Lévy-Khintchine formula,

$$\psi(s) = \mu s + \frac{\sigma^2}{2} s^2 + \int_{(-\infty, 0)} [e^{sx} - 1 - sxI(x > -1)] \Upsilon(dx). \quad (7)$$

The Laplace exponent, as a function on $[0, \infty)$, is continuous and convex, and satisfies $\psi(0) = 0$ and $\lim_{s \rightarrow \infty} \psi(s) = \infty$. Therefore, there exists a largest solution $\Lambda(0) \geq 0$ to $\psi[\Lambda(0)] = 0$, and an inverse $\Lambda : [0, \infty) \rightarrow [\Lambda(0), \infty)$ of the restriction of ψ to $[\Lambda(0), \infty)$. Theorem 1 of Bertoin (1996, Chapter VII) implies that $\{T\}$ is a *killed subordinator* with Laplace exponent Λ :

$$\mathcal{L}_{T(y)}(s) = \exp[-\Lambda(s)y]. \quad (8)$$

That is, $\{T\}$ is a nondecreasing Lévy process with Laplace exponent $\Lambda - \Lambda(0)$, forced to equal ∞ (the *graveyard state*) from some random threshold level $E_{\Lambda(0)}$ up if $\Lambda(0) > 0$. Here, $E_{\Lambda(0)}$ has an exponential distribution with parameter $\Lambda(0)$, and is independent from $(\{Y\}, X, V)$. Note that the probability $\Pr(E_{\Lambda(0)} \leq y) = 1 - \exp[-\Lambda(0)y]$ that $\{T\}$ has been *killed* at or below threshold level y equals the share $1 - \mathcal{L}_{T(y)}(0)$ of defecting movers at threshold level y .

If, for example, $\{Y\}$ is a Brownian motion with general drift coefficient $\mu \in \mathbb{R}$ and dispersion coefficient $\sigma \in (0, \infty)$, we have that $\psi(s) = \mu s + \sigma^2 s^2/2$, so that $\Lambda(0) = \max\{0, -2\mu/\sigma^2\}$ and $\Lambda(s) = \left[\sqrt{\mu^2 + 2\sigma^2 s} - \mu \right] / \sigma^2$. If $\mu \geq 0$, then $\Lambda(0) = 0$, $T(y)$ is nondefective, and substituting in (8) gives the Laplace transform (4) of Section 3's Gaussian example. If $\mu < 0$, on the other hand, $\Lambda(0) = -2\mu/\sigma^2 > 0$ and the distribution of $T(y)$ has a defect of size $1 - \exp(2y\mu/\sigma^2)$. Note that in this case, $\sigma = 0$ is excluded to avoid the trivial outcome that $T(y) = \infty$ almost surely. Either way, $\{T\}$ is an inverse Gaussian subordinator, killed at an independent exponential rate $\Lambda(0)$ if $\Lambda(0) > 0$.

Not every subordinator is the hitting-time process of a spectrally-negative Lévy process. For example, consider the *stable subordinator* of index $\rho \in (0, 1]$; that is, the Lévy

process with Laplace exponent $\Lambda_\rho(s) \equiv s^\rho$ (Bertoin, 1996, Section III.1). Proposition 2(i) in Bertoin (1996, Chapter I) implies that $\lim_{s \rightarrow \infty} s^{-2}\psi(s) = \sigma^2/2 \in [0, \infty)$ if ψ is the Laplace exponent of a spectrally-negative Lévy process. Consequently, if $\rho \in (0, 1/2)$; $\Lambda_\rho^{-1}(s) = s^{1/\rho}$ cannot be the Laplace exponent of a spectrally-negative Lévy process. This suggests that, when estimating the MHT model, it is more convenient to parameterize the model in terms of ψ , than to specify Λ directly through the Lévy-Khintchine formula for subordinators. We will come back to this in Section 5.

Now define $\mathcal{L}_T(\cdot|X, V)$, $\mathcal{L}_T(\cdot|X)$, and \mathcal{L} analogously to $\mathcal{L}_{T(y)}$ in (6), explicitly allowing for defects. From (1) and (8), it follows that

$$\mathcal{L}_T(s|X, V) = \exp[-\Lambda(s)\phi(X)V], \quad (9)$$

so that

$$\mathcal{L}_T(s|X) = \mathcal{L}[\Lambda(s)\phi(X)]. \quad (10)$$

Note that this expression for the Laplace transform of $T|X$ is the same as that for Section 3's Gaussian example in (5). However, in the general case here, we do not require that Λ has equation (4)'s inverse Gaussian two-parameter specification. Instead, we have semiparametrically specified Λ as the inverse of the latent process's Laplace exponent ψ in (7). This way, we now also allow for defecting movers, $\Lambda(0) > 0$. Moreover, there can be a mass of stayers, if the distribution G of V has a mass point at ∞ .

4.2 Identifiability

The distribution of $T|X$ implied by the MHT model only depends on its primitives $(\mu, \sigma^2, \Upsilon)$ and (ϕ, G) through the triplet $(\Lambda, \phi, \mathcal{L})$. In this section, we study the fundamental question under what conditions the model triplet $(\Lambda, \phi, \mathcal{L})$ can be uniquely determined from a “large” data set that gives the distribution of $T|X$.

Because there is a one-to-one relation between $(\Lambda, \phi, \mathcal{L})$ and the MHT model’s primitives, the identification analysis applies without change to these primitives. In particular, G can be uniquely determined from \mathcal{L} by the uniqueness of the Laplace transform (Feller, 1971, Section XIII.1, Theorem 1). The Laplace exponent ψ of $\{Y\}$ is uniquely determined from Λ by inversion and, if $\Lambda(0) > 0$, analytic extension from $[\Lambda(0), \infty)$ to $[0, \infty)$. Subsequently, the parameters $(\mu, \sigma^2, \Upsilon)$ of the latent Lévy process can be uniquely determined from ψ by the uniqueness of the Lévy-Khintchine representation (Bertoin, 1996, Chapter I, Theorem 1).

We focus on the “two-sample” case that $\mathcal{X} = \{0, 1\}$ and $\phi(x) = \beta^x$, for some $\beta \in (0, \infty)$, and we have data on the distributions F_0 of $T|(X = 0)$ and F_1 of $T|(X = 1)$. This assumes minimal covariate variation and thus poses the hardest identification problem (Elbers and Ridder, 1982, use a similar approach in their analysis of the MPH model). We assume that $\beta \neq 1$, so that there is actual variation with the covariates. This assumption can be tested, because $F_0 \neq F_1$ if and only if $\beta \neq 1$. Note that we have implicitly fixed $\phi(0) = 1$, which is an innocuous normalization because the scale of V is unrestricted at this point.

As in Section 3’s Gaussian example, our analysis exploits an analogy with the analysis of the MPH model. Note that the right-hand side of (10) equals the survival function—rather than the Laplace transform—of $T|X$ in a two-sample MPH model with integrated baseline Λ , covariate effect $\phi(X) = \beta^X$, and unobserved-heterogeneity distribution G . Consequently, we can borrow insights from Elbers and Ridder’s (1982) and Ridder’s (1990) analysis of this MPH model. Because of possible defects in the MHT model, their analysis does not apply directly. In particular, the possibility that movers defect, $\Lambda(0) > 0$, creates an identification problem similar to a left-censoring problem in the MPH model. Fortunately, the MHT model’s mover-stayer structure can be identified without further assumptions, and problems caused by defecting movers can be solved by analytic extension.

First, consider identifiability of the mover-stayer structure from (F_0, F_1) .

Proposition 1 (Identifiability of the Share of Stayers). *If two MHT triplets $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ imply the same pair of distributions (F_0, F_1) , then $\tilde{\mathcal{L}}(0) = \mathcal{L}(0)$.*

Proposition 1 directly implies identification of the share of stayers, $\Pr(V = \infty|X) = \Pr(V = \infty) = 1 - \mathcal{L}(0)$. In turn, the proportion of defecting movers in sample x can be uniquely determined from the share of stayers and F_x , using that

$$\Pr(T = \infty, V < \infty|X = x) = \mathcal{L}(0) - \lim_{t \rightarrow \infty} F_x(t); \quad x = 0, 1.$$

Proposition 1's proof, and that of all other results in this section, is given in Appendix B. It exploits that the share of defecting movers, if positive, varies between the two samples and, by the assumed independence of V and X , the share of stayers does not. Intuitively, if the defects of F_0 and F_1 are the same, they equal the share of stayers; and movers never defect, $\Lambda(0) = 0$. Otherwise, $\Lambda(0) > 0$, and it is clear from (10) that the data only provide direct information about \mathcal{L} away from 0; then, the analyticity of the Laplace transform can be used to learn about $\mathcal{L}(0)$. Abbring (2002) proves a related result for the MPH model, but relies on an additional assumption on G .

Our core result on the identifiability of $(\Lambda, \beta, \mathcal{L})$ requires a regularity condition in terms of Karamata's concepts of slow and regular variation (Feller, 1971, Section VIII.8).

Definition 2. *A function $L : (0, \infty) \rightarrow (0, \infty)$ varies slowly at 0 (at ∞) if $L(cs)/L(s) \rightarrow 1$ as $s \downarrow 0$ ($s \rightarrow \infty$) for every fixed $c \in (0, \infty)$. A function $k : (0, \infty) \rightarrow (0, \infty)$ varies regularly if $k(s) = s^\tau L(t)$ for some exponent $\tau \in \mathbb{R}$ and slowly varying function L .*

Note that a slowly varying function is regularly varying with exponent 0. Any function that has a positive (and finite) limit varies slowly; but slowly varying functions may converge to 0 or diverge, such as $L(s) = |\ln(s)|$ and $L(s) = 1/|\ln(s)|$. By Feller (1971, Section VIII.8, Lemma 2), a function k that varies regularly with exponent τ at ∞ (at 0) asymptotically satisfies $s^{\tau-\varepsilon} < k(s) < s^{\tau+\varepsilon}$, for any given $\varepsilon > 0$ ($\varepsilon < 0$).

Definition 3. A MHT triplet $(\Lambda, \beta, \mathcal{L})$ has a regularly varying tail if at least one of the following is true:

- (i). $|\mathcal{L}'|$ varies regularly at 0, with some exponent $\tau \in (-1, 0]$;
- (ii). $|\mathcal{L}'|$ varies regularly at ∞ , with some exponent $\tau \in (-\infty, -1)$;
- (iii). $|\psi'|$ varies regularly at 0, with some exponent $\tau \in (-1, 1]$; or
- (iv). $|\psi'|$ varies regularly at ∞ , with some exponent $\tau \in [0, 1]$.

It is fairly innocuous to require that an MHT triplet has a regularly varying tail, because this is implied by each of the identifying assumptions suggested by the MPH literature. For example, in Section 3's application of [Ridder and Woutersen's \(2003\)](#) result for the MPH model to a Gaussian MHT model with positive drift, $0 < \lim_{s \downarrow 0} \psi'(s) = \mu < \infty$, so that $|\psi'|$ varies slowly at 0. Also, [Elbers and Ridder's \(1982\)](#) finite-mean assumption on G is equivalent to $\lim_{s \downarrow 0} |\mathcal{L}'(s)| < \infty$ and, because $\lim_{s \downarrow 0} |\mathcal{L}'(s)| > 0$, implies that $|\mathcal{L}'|$ varies slowly at 0. Finally, [Heckman and Singer \(1984a\)](#) assume that $|\mathcal{L}'|$ varies regularly at 0 with a prespecified exponent $\tau \in (-1, 0)$.

Following Definition 3, we can say that an MHT triplet has a regularly varying tail without specifying the tail or fixing the exponent. The ranges of the exponents in Definition 3 follow from the properties of the functions involved, and do not constitute additional restrictions, except for the exclusion of the boundary case that $\tau = -1$ in (i)–(iii). In particular, the ranges in (i) and (ii) are determined by the restrictions that $|\mathcal{L}'|$ is decreasing and integrable, and the Lemma in [Feller \(1971, Section VIII.9\)](#). The ranges in (iii) and (iv) follow from the Lévy-Khintchine formula (7) and that same Lemma. We will briefly return to the boundary case that $\tau = -1$ when discussing identification from tail conditions on \mathcal{L} .

We are now ready to state our core result.

Proposition 2 (Identifiability of the MHT Model). *If two MHT triplets $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ both have a regularly varying tail (possibly with different exponents τ and $\tilde{\tau}$), and imply the same pair of distributions (F_0, F_1) ; then*

$$\begin{aligned}\tilde{\beta} &= \beta^\rho, \\ \tilde{\Lambda} &= \kappa\Lambda^\rho, \text{ and} \\ \tilde{\mathcal{L}}(\kappa s^\rho) &= \mathcal{L}(s) \text{ for all } s \in [0, \infty),\end{aligned}$$

for some $\kappa \in (0, \infty)$ and $\rho \in [1/2, 2]$.

Proposition 2 establishes identification up to a power transformation, indexed by ρ , and an innocuous normalization, indexed by κ . It is analogous to Ridder's (1990)'s Theorem 1 for the generalized accelerated failure time model, which encompasses the single-spell MPH identification literature. Our analysis deviates in three important ways from Ridder's. First, our proof makes explicit use of the assumption that both MHT triplets have a regularly varying tail. Ridder's Theorem 1 implicitly requires a similar regularity condition (Abbring and Ridder, 2009). Second, we allow for defective duration distributions, which naturally arise in the context of an MHT model. Third, we use the special structure of the MHT model to show that ρ cannot be any positive number, but lies in $[1/2, 2]$.

The observational equivalence characterized by Proposition 2 can be given an appealing stochastic interpretation. For expositional clarity, we set $\kappa = 1$, and focus on the interpretation of ρ . Without loss of generality, let $\rho \in [1/2, 1)$. Let $\{S_\rho\}$ be an independent stable subordinator of index ρ . Then, if $\{T\}$ is the hitting-time process characterized by Λ , the process $\{T[S_\rho(y)]; y \geq 0\}$, has Laplace exponent $\tilde{\Lambda}$ (Feller, 1971, Section XVII.4(e)). Consequently, for each given threshold level y , $\tilde{\Lambda}(y)$ corresponds to a positive-stable mixture $T[S_\rho(y)]$ over $\{T\}$. Thus, we can interpret $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ as reassigning some of the threshold heterogeneity in $(\Lambda, \beta, \mathcal{L})$ to the individual hitting-time process. Indeed, $|\tilde{\beta} - 1| = |\beta^\rho - 1| < |\beta - 1|$, so that there is less observed variation

in the thresholds between the two samples. Similarly, we can interpret \tilde{G} as specifying less unobserved heterogeneity than G . Suppose, for example, that $\mathcal{L}(0) = 1$ and that $|\mathcal{L}'|$ varies regularly at 0 with exponent $\tau \in (-1, 0)$, as in Heckman and Singer (1984a). Then, it follows from the Lemma in Feller (1971, Section VIII.9), Proposition 2, and Theorem 4 in Feller (1971, Section XIII.5) that $1 - G$ and $1 - \tilde{G}$ vary regularly at ∞ with exponents $-1 < -(\tau + 1) < 0$ and $-(\tau + 1)/\rho < -(\tau + 1)$, respectively. Consequently, \tilde{G} has a thinner right tail than G .

The restriction of ρ to $[1/2, 2]$ in Proposition 2 relies on the special structure of ψ and $\tilde{\psi}$. Recall from Section 4.1 that ψ is convex, and that $\psi(s) \rightarrow \infty$ and $s^{-2}\psi(s) \rightarrow \sigma^2/2 \in [0, \infty)$ as $s \rightarrow \infty$. Now suppose that $\tilde{\Lambda} = \kappa\Lambda^\rho$ characterizes the hitting-time process of a latent process with Laplace exponent $\tilde{\psi}$. From the fact that Λ and $\tilde{\Lambda}$ are the inverses of ψ and $\tilde{\psi}$, respectively, it follows that $\tilde{\psi}(s) = \psi[(s/\kappa)^{1/\rho}]$. Because $\tilde{\psi}$ should at least be of linear order and at most of quadratic order at ∞ , just like ψ , it is necessary that $\rho \in [1/2, 2]$.

Note that Λ^ρ is the Laplace exponent of a (killed) subordinator if Λ is; for all $\rho \in (0, 1]$, and not just for $\rho \in [1/2, 1]$. Proposition 2 provides identification up to $\rho \in (0, \infty)$ for a more general model that requires $\{T\}$ to be a subordinator, but not necessarily the hitting-time process of a spectrally-negative Lévy process. Any strategy for point identification of Λ that exploits the subordinator structure of $\{T\}$, but not its hitting-time structure, will provide overidentifying restrictions that can be used in testing the MHT model.

Point identification of the MHT model requires further assumptions on either \mathcal{L} or Λ . In the MPH literature, this is invariably achieved by not only requiring regular variation of one of their tails, but also fixing the corresponding exponent of regular variation (Ridder, 1990; Abbring and Ridder, 2009).

First, consider tail assumptions on \mathcal{L} . Elbers and Ridder (1982) have proved identifiability of the two-sample MPH model, up to scale, under the assumption that the unobserved factor has a finite mean. Within the context of an MPH model, this is an arbitrary normalization with substantive meaning (Ridder, 1990). In some cases, the cor-

responding assumption on the MHT model, $\lim_{s \downarrow 0} |\mathcal{L}'(s)| = \mathbb{E}[VI(V < \infty)] < \infty$, may follow naturally from optimal stopping models in which threshold heterogeneity is reduced from primitive unobserved heterogeneity (see Section 6). In other cases, it will be a similarly arbitrary normalization. To see how it yields point identification, suppose that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly and belong to observationally equivalent models. Then, Proposition 2 implies that $|\tilde{\mathcal{L}}'(s)| = \kappa^{-1/\rho} s^{\frac{1-\rho}{\rho}} |\mathcal{L}'[(s/\kappa)^{1/\rho}]|$, so that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ can only vary regularly with the same exponent τ , at either 0 or ∞ , if ρ satisfies $(\tau + 1)(\rho - 1) = 0$. Consequently, Elbers and Ridder's finite mean assumption, which implies that $\tau = 0$, fixes $\rho = 1$ and yields identification up to scale. We summarize this result in

Proposition 3 (Identifiability of the MHT Model Under a Finite-Mean Assumption on G). *If two MHT triplets $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ are such that $\lim_{s \downarrow 0} |\mathcal{L}'(s)| < \infty$ and $\lim_{s \downarrow 0} |\tilde{\mathcal{L}}'(s)| < \infty$, and imply the same pair of distributions (F_0, F_1) ; then $\tilde{\beta} = \beta$, $\tilde{\Lambda} = \kappa\Lambda$, and $\tilde{\mathcal{L}}(\kappa s) = \mathcal{L}(s)$ for all $s \in [0, \infty)$, for some $\kappa \in (0, \infty)$.*

Alternatively, following Heckman and Singer (1984a), we could fix $\rho = 1$ by assuming that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at 0 with the same exponent $\tau \in (-1, 0)$. Note that we need to exclude the boundary case that $\tau = -1$, in which $(\tau + 1)(\rho - 1) = 0$ holds for all ρ . Finally, we could make an assumption on the variation of $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ at ∞ . However, we have no examples of economic models that imply either of these alternative assumptions.

In cases in which there is no substantial justification for a finite-mean assumption, the special structure of Λ offers a more attractive approach to point identification in the MHT model. To gain some intuition, first consider the case without defecting movers. Let Λ and $\tilde{\Lambda}$ be two Laplace exponents that belong to observationally equivalent MHT triplets, with $\Lambda(0) = \tilde{\Lambda}(0) = 0$. Assume that $\lim_{s \downarrow 0} \Lambda'(s) < \infty$ and $\lim_{s \downarrow 0} \tilde{\Lambda}'(s) < \infty$. Then, because $\lim_{s \downarrow 0} \Lambda'(s) > 0$ always, $0 < \lim_{s \downarrow 0} \Lambda'(s) < \infty$; and, by the inverse function theorem, $0 < \lim_{s \downarrow 0} \psi'(s) < \infty$, so that $|\psi'|$ varies slowly at 0. A similar analysis applies to $\tilde{\Lambda}$. Consequently, both $|\psi'|$ and $|\tilde{\psi}'|$ vary regularly, Proposition 2 applies, and $\tilde{\Lambda} = \kappa\Lambda^\rho$; for some $\kappa, \rho \in (0, \infty)$. It is easily checked that $0 < \lim_{s \downarrow 0} \Lambda'(s) < \infty$ and $0 < \lim_{s \downarrow 0} \tilde{\Lambda}'(s) < \infty$.

∞ can only both hold if $\rho = 1$. Thus, in this case, the assumption that the Laplace exponent (of $\{T\}$) has a finite derivative at 0 identifies the model up to scale.

This argument does not directly extend to the general case, in which movers may defect and possibly $\Lambda(0) > 0$ and $\tilde{\Lambda}(0) > 0$. However, note that, in the case without defecting movers, we have effectively obtained identification by fixing the exponents of regular variation at 0 of both $|\psi'|$ and $|\tilde{\psi}'|$ to 0, by assuming that they have finite positive limits at 0. This continues to be sufficient for identification in the general case.

Proposition 4 (Identifiability of the MHT Model Based on Conditions on $\{Y\}$).

If two MHT triplets $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ are such that $0 < \lim_{s \downarrow 0} |\psi'(s)| < \infty$ and $0 < \lim_{s \downarrow 0} |\tilde{\psi}'(s)| < \infty$, and imply the same pair of distributions (F_0, F_1) ; then $\tilde{\beta} = \beta$, $\tilde{\Lambda} = \kappa\Lambda$, and $\tilde{\mathcal{L}}(\kappa s) = \mathcal{L}(s)$ for all $s \in [0, \infty)$, for some $\kappa \in (0, \infty)$.

The assumption that $0 < \lim_{s \downarrow 0} |\psi'(s)| < \infty$ requires that $\mathbb{E}[Y(t)] = t \lim_{s \downarrow 0} \psi'(s) \neq 0$ and $\mathbb{E}[Y(t)] > -\infty$ for $t \in (0, \infty)$. In the investment option problem introduced in Section 1, $\mathbb{E}[Y(t)] < 0$ is natural if the project depreciates over time relative to alternative investments, say because technological progress is embodied in new projects. In this case, the agent may end up never investing, and $\Lambda(0) > 0$. In a model of job tenure like Section 6.3's, the accumulation of job-specific skills may lead to a similar pattern; but wear of the job and progress elsewhere may instead imply $\mathbb{E}[Y(t)] > 0$ and $\Lambda(0) = 0$. The condition that $\mathbb{E}[Y(t)] > -\infty$ only has bite if $\Lambda(0) > 0$, and is a restriction on the negative jumps in $\{Y\}$. Because we have excluded positive jumps, $\mathbb{E}[Y(t)] < \infty$ always holds.

Our analysis for the case without defecting movers is similar to [Ridder and Woutersen's \(2003\)](#) analysis of the MPH model (they have no equivalent to our general analysis). In particular, [Ridder and Woutersen](#) use an assumption on the baseline hazard that is analogous to our assumption that $0 < \lim_{s \downarrow 0} \Lambda'(s) < \infty$. Unlike their assumption for the MPH model, however, ours can be related to more primitive conditions on $\{Y\}$.

Next, we revisit Section 3's Gaussian example $Y(t) = \mu t + \sigma W(t)$; but with $\mu \in \mathbb{R}$, $\sigma \in [0, \infty)$, $\sigma > 0$ if $\mu \leq 0$, and general G . In this case, $\psi'(s) = \mu + \sigma^2 s$, so that $|\psi'|$

varies regularly at 0; with exponent 1 if $\mu = 0$, and exponent 0 otherwise. Consequently, Proposition 2 applies. With

$$\Lambda(s) = \begin{cases} \frac{\sqrt{\mu^2 + 2\sigma^2 s} - \mu}{\sigma^2} & \text{if } \sigma > 0 \text{ and} \\ s/\mu & \text{if } \sigma = 0, \end{cases}$$

this gives

Proposition 5 (Identifiability of the Gaussian MHT Model). *If two Gaussian MHT triplets $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ imply the same pair of distributions (F_0, F_1) , then either one of the following is true:*

- (i). $\tilde{\beta} = \beta$, $\tilde{\Lambda} = \kappa\Lambda$, and $\tilde{\mathcal{L}}(s) = \mathcal{L}(\kappa s)$ for all $s \in [0, \infty)$, for some $\kappa \in (0, \infty)$;
- (ii). $\tilde{\beta} = \beta^2$ and, for all $s \in [0, \infty)$. $\tilde{\Lambda}(s) = \kappa\Lambda(s)^2 = \nu s$ and $\tilde{\mathcal{L}}(\kappa s^2) = \mathcal{L}(s)$, for some $\kappa, \nu \in (0, \infty)$; or
- (iii). $\tilde{\beta} = \beta^{1/2}$ and, for all $s \in [0, \infty)$, $\tilde{\Lambda}(s) = \kappa\Lambda(s)^{1/2} = \nu\sqrt{s}$ and $\tilde{\mathcal{L}}(\kappa s^{1/2}) = \mathcal{L}(s)$, for some $\kappa, \nu \in (0, \infty)$.

Thus, if two Gaussian MHT triplets are observationally equivalent, then they are either the same, up to a scale normalization, or one triplet corresponds to a degenerate upward drift and the other to a driftless nondegenerate Brownian motion. In Section 3's example, we ensured identification by requiring upward drift. More generally, identification can be achieved by either requiring $\sigma > 0$ or $\mu \neq 0$. In terms of Proposition 2, the range of possible ρ is restricted to $\{1/2, 1, 2\}$; because the exponents of regular variation of $|\psi'|$ and $|\tilde{\psi}'|$ can only take two values in the Gaussian special case, 0 and 1.

Finally, note that the analogy with the MPH literature stretches beyond the set of basic results exploited so far. For example, consider the case in which we have stratified data, with one shared value of V and observations on two durations, T^1 and T^2 , in each stratum. The two durations may concern a single agent's consecutive spells, or

the single spells of two agents who are known to have the same value of V . Formally, suppose we observe the joint distribution of (T^1, T^2) ; for now, suppress covariates X . Let $T^1 = \inf\{t \geq 0 : Y^1(t) > V\}$ and $T^2 = \inf\{t \geq 0 : Y^2(t) > V\}$, with $\{Y^1\}$ and $\{Y^2\}$ independent spectrally-negative Lévy processes; and V a nonnegative random variable, distributed independently from $(\{Y^1\}, \{Y^2\})$ with distribution G . Denote the Laplace exponent of the hitting-time process corresponding to $\{Y^j\}$ with Λ_j ; $j = 1, 2$. Then, analogously to Section 4.1's analysis for the single-spell case, it can be shown that

$$\mathcal{L}_{T^1, T^2}(s_1, s_2) \equiv \mathbb{E} [I(T^1 < \infty, T^2 < \infty) \exp(-s_1 T^1 - s_2 T^2)] = \mathcal{L} [\Lambda_1(s_1) + \Lambda_2(s_2)].$$

In the case without defecting movers; that is, $\Lambda_1(0) = \Lambda_2(0) = 0$; \mathcal{L}_{T^1, T^2} fully characterizes the distribution of (T^1, T^2) . An expression similar to that for \mathcal{L}_{T^1, T^2} appears in Honoré's (1993) analysis of the MPH model with multiple-spell data, for the joint *survival function* of (T^1, T^2) . In fact, in this special case, Honoré's Theorem 1 applies directly: Its proof applies to the case with stayers, even though it is stated for the nondefective case. However, Honoré does not cover the general case in which possibly $\Lambda_1(0) > 0$ and $\Lambda_2(0) > 0$. In this general case, there may be independent information about the marginal distributions of T^1 and T^2 , and in particular their defects, in the marginal transforms \mathcal{L}_{T^j} of T^j ; $j = 1, 2$; and we have to exploit this information to obtain identification. Moreover, Proposition 1 does not apply here. So, the following result is of independent value.

Proposition 6 (Identifiability of the MHT Model from Stratified Data). *If two two-spell MHT triplets $(\Lambda_1, \Lambda_2, \mathcal{L})$ and $(\tilde{\Lambda}_1, \tilde{\Lambda}_2, \tilde{\mathcal{L}})$ imply the same joint distribution of (T^1, T^2) ; then $\tilde{\Lambda}_1 = \kappa\Lambda_1$, $\tilde{\Lambda}_2 = \kappa\Lambda_2$, and $\tilde{\mathcal{L}}(\kappa s) = \mathcal{L}(s)$ for all $s \in [0, \infty)$, for some $\kappa \in (0, \infty)$.*

Note that this identification result for stratified data, unlike Propositions 3–5 for the single-spell case, does not require additional assumptions on Λ or G . Moreover, it does not rely on external variation with covariates X . Thus, it also applies to a model extended

with covariates X that interact in an unrestricted way with $\{Y^1\}$, $\{Y^2\}$, and V .

4.3 Censoring

The identification analysis so far assumes that the distribution of $T|X$ is known. In practice, duration data are often censored. With independent censoring (Andersen et al., 1993, Section II.1), the distribution of $T|X$ is identified, provided that obvious support conditions are met. This paper’s identification results carry over to such independently censored data without change. A common example is right-censoring at times C that are independent of T given X , and that have unbounded support.

The identification analysis does not immediately carry over to censoring mechanisms that obstruct the identification of the distribution of $T|X$. However, the specific structure implied by the Lévy assumption suggests that identifiability may continue to hold under similar conditions with independent right-censoring, but subject to support qualifications. For example, take the case that $Y(t) = t$ and $\beta = 1$, so that $T = V$. Then, if all durations are censored at some fixed $C \in (0, \infty)$, only the restriction of G to $[0, C]$ is identified.

5 Estimation

So far, we have ignored sampling variation. This section briefly discusses estimation of an MHT model, based on its characterization in Sections 3.1 and 4.1, and standard likelihood and moment methods. Abbring and Salimans (2009) provide a full development of the estimators, and code for their computation.

Let Λ , ϕ and G be specified up to a finite vector of unknown parameters $\alpha \in \mathcal{A}$. We assume that this parameterization is one-to-one, so that α is uniquely determined by $(\Lambda, \phi, \mathcal{L})$. In the two-sample specification $\phi(X) = \beta^X$, it is sufficient that $\mathcal{X} = \{0, 1\}$. More generally, if we have multivariate and continuous covariates, we can specify $\phi(X) = \exp(X'\beta)$. Then, we require the “rank condition” that the support \mathcal{X} of X contains a

nonempty open set in \mathbb{R}^k . Note that this excludes an intercept from $\ln[\phi(X)] = X'\beta$, and thus embodies a scale normalization on ϕ similar to that in the two-sample specification.

The Lévy-Khintchine formula (7) can be used to specify ψ ; Λ then follows by inversion. This ensures that Λ satisfies the model's restrictions. Section 3's Gaussian special case offers an attractive baseline specification, with only the drift parameter μ and dispersion parameter σ , and $\Upsilon = 0$. In applications that require more flexibility, compound Poisson shocks with a finitely discrete shock distribution can be added. Then, the integral in the Lévy-Khintchine formula is a finite sum, so that the resulting specification of ψ is easy to compute. Moreover, because the number of support points of the shock distribution can be freely chosen, it is flexible. In fact, a formal reason to prefer this specification over others is that each Lévy process can be approximated by a sequence of compound Poisson processes (Feller, 1971, Section IX.5, Theorem 2).

The heterogeneity distribution G can be specified as in empirical applications of the MPH model. A finitely discrete specification is particularly popular because of its versatility and computational convenience, and appears in Heckman and Singer's (1984b) influential work on semiparametric estimation of the MPH model. Alternatively, a gamma specification of G combines naturally with the MHT model's mixture-of-exponentials specification of $\mathcal{L}_{T|X}$ (Abbring and Van den Berg, 2007).

Suppose that we have a complete random sample $((T_1, X_1), \dots, (T_N, X_N))$ from the "true" distribution of (T, X) ; which corresponds to the distribution of $T|X$ induced by the parametric MHT model with parameter vector α_0 , and some ancillary marginal distribution of X . Our objective is to estimate the parameters α_0 .

5.1 Maximum Likelihood

First, consider Section 3's Gaussian special case: $\{Y\}$ is a Brownian motion with drift $\mu > 0$, so that, by the analysis in the previous section, $T|X$ has a mixed (nondefective) inverse Gaussian distribution. Assume that $\phi(X)$ is nondegenerate; that is, the threshold

varies with the observed covariates. Then, Proposition 5 ensures that α_0 is uniquely determined from the distribution of $T|X$, provided that we impose scale normalizations on two of the three functions Λ , ϕ , and \mathcal{L} . In this case, we can normalize ϕ as discussed before, and add a normalization on Λ such as $\mu = 1$. Alternatively, we may drop one of these normalizations and fix a scale parameter of \mathcal{L} .

In this special case, it is very easy to estimate α_0 by maximum likelihood. A conditional likelihood $L_N(\alpha)$ of $(T_1, \dots, T_N)|(X_1, \dots, X_N)$ can be constructed using the explicit expression for the density of $T(y)$ in (3): $L_N(\alpha) = \prod_{i=1}^N \int f [T_i|\phi(X_i)v] dG(v)$. Here, the dependence of f , ϕ , and G on the parameter vector α is kept implicit. Under standard regularity conditions, the maximizer of $L_N(\alpha)$ is a \sqrt{N} -consistent and asymptotically normal and efficient estimator of α_0 .

The Gaussian special case can be estimated by maximum likelihood because it comes with explicit expressions for the density and survival function of $T|X$. This feature it shares with many of the models studied in the statistics literature (Lee and Whitmore, 2006). In the general Lévy case, however, such expressions are not available, and maximum likelihood cannot be directly implemented. For this case, we propose a generalized method-of-moments (GMM) estimator.

5.2 Generalized Method of Moments

A GMM estimator can be based on (10), which provides a continuum of conditional moment conditions, one for each point s at which the Laplace transform can be evaluated.

Define $h(t, x; s, \alpha) \equiv \exp(-st)I(t < \infty) - \mathcal{L}[\Lambda(s)\phi(x)]$. Then, it follows from (10) that $\mathbb{E}[h(T, X; s, \alpha_0)|X] = 0$ for all $s \in (0, \infty)$. In our estimation procedure, we will specify an $(M \times 1)$ -vector Z of instruments, and use the unconditional moment conditions

$$\mathbb{E}[h(T, X; s, \alpha_0)Z] = 0, \quad s \in (0, \infty). \quad (11)$$

The canonical example takes $M = K + 1$ and $Z' = [1 \ X']$, which gives $K + 1$ unconditional moment conditions, $\mathbb{E}[h(T, X; s, \alpha_0)] = 0$ and $\mathbb{E}[h(T, X; s, \alpha_0)X] = 0$, for each s . We assume that the set of moment conditions (11) uniquely determines α_0 .

We first construct a consistent GMM estimator with naive weighting of the moments. This estimator is easy to compute; it can serve as the first step in a more efficient two-step estimator, and may be of interest in its own right. Denote the empirical analogue to the moment vector in the left-hand side of (11) with

$$h_N(s, \alpha) \equiv N^{-1} \sum_{i=1}^N h(T_i, X_i; s, \alpha) Z_i. \quad (12)$$

We define a feasible (one-step) GMM estimator $\hat{\alpha}_N$ of α_0 as the value of α that minimizes the quadratic GMM objective function

$$H_N(\alpha; W_N, w) \equiv \int_0^\infty h_N(s, \alpha)' Q_N h_N(s, \alpha) W_N(ds).$$

Here, Q_N is a positive-definite and symmetric $M \times M$ random matrix that converges in probability to a positive-definite fixed matrix Q . For given s , the matrix Q_N weights the various moments corresponding to the M instruments, with weights independent of s . Examples include the $M \times M$ identity matrix and $\left(N^{-1} \sum_{i=1}^N Z_i Z_i'\right)^{-1}$. The function W_N is a random probability measure that converges to a nonrandom measure W . It weights the various moment conditions corresponding to the evaluation points s of the Laplace transform, identically across the instruments in Z_i . It could be finitely discrete, and selecting only a finite number of Laplace evaluation points, or absolutely continuous. Examples of the latter include $W_N(s) = \exp(-\varpi_N s)$ for either a fixed or a random (data-dependent) positive ϖ_N .

The analysis of Carrasco and Florens (2000) can be adapted to prove that, under appropriate regularity conditions, $\hat{\alpha}_N$ is \sqrt{N} -consistent and asymptotically normal. Moreover, Carrasco and Florens's (2002) method for efficient estimation based on empirical charac-

teristic functions can be adapted to produce an GMM estimator of the MHT model that efficiently weights across evaluation points of $\mathcal{L}_{T|X}$, for given finite instrument vector Z . This estimator is a two-step estimator that uses $\hat{\alpha}_N$ as a first-stage estimator.

5.3 Censoring

Section 5.1's maximum likelihood estimator for the Gaussian special case can be straightforwardly applied to independently censored data. For example, an observation i that is independently right-censored at T_i would contribute a factor $\int \bar{F}[T_i|\phi(X_i)v] dG(v)$ to the likelihood, which can be easily computed using the explicit expression (2) for \bar{F} .

The generalization of the GMM estimators to independently censored data is not covered by [Abbring and Salimans \(2009\)](#), but feasible. In the two-sample case; or more generally, in the case that the support \mathcal{X} of X is finite; the GMM estimator can be readily adapted to allow for independent censoring, by nonparametrically correcting the empirical moments in (12) for censoring. To this end, first estimate the distribution of T in each sample using the Nelson-Aalen estimator or, in special cases, the Kaplan-Meier estimator (see e.g. [Andersen et al., 1993](#), Section IV.1). Then, compute the empirical analogue of the moment condition (11) using these nonparametric estimators of the distribution of T ; instead of the empirical distribution function, as in (12). Provided that the censoring mechanism is such that the distribution of T is identified in each sample, its nonparametric estimator is consistent and asymptotically Gaussian, and the properties of the censoring-corrected GMM estimator can be derived in a standard manner.

In the case that $\phi(X) = \exp(X'\beta)$, with \mathcal{X} general, we cannot rely on repeated application of the Nelson-Aalen estimator to each sample. Instead, we need a semiparametric estimator of the distribution of $T|X$ to compute the empirical analogue of the moment condition (11).

6 Structural Examples

The MHT model can be applied to the empirical analysis of heterogeneous agents' optimal stopping decisions. [Dixit and Pindyck \(1994\)](#) and [Stokey \(2009\)](#) analyze and review various models based on Brownian motions and their applications. [Kyprianou \(2006\)](#) and [Boyarchenko and Levendorskiĭ \(2007\)](#) review recent extensions to general Lévy processes.

This section presents some simple examples of such models. With payoffs that are monotonic in a Lévy state variable, threshold rules routinely arise. We primarily focus on the way primitive heterogeneity generates *heterogeneous* threshold rules, and how this squares with the MHT model. We first study the optimal timing of an irreversible investment. This well-studied problem—it is closely related to the analysis of American options in finance—is a good vehicle to introduce the relation between optimal stopping models and the MHT framework. We then study two models of optimal transitions between unemployment and employment. The first is [Dixit's \(1989\)](#) model of entry and exit. The second is a stylized version of the search-matching model that has become the standard in labor economics. Both models extend the first, investment option model by not only specifying the transitions out of the state of interest, but also the transitions into it. This determines the initial conditions for the MHT analysis of the durations in this state, and tightly structures the dependence of the thresholds on primitive heterogeneity.

6.1 Investment Timing

[McDonald and Siegel \(1986\)](#) study the optimal timing of an irreversible investment in a project of which the log value follows a Brownian motion. Their paper is an early and influential example of the large “real options” literature that applies insights from the literature on pricing financial derivatives—in this case, perpetual American call options—to real investments ([Dixit and Pindyck, 1994](#)). Here, we discuss a version of their model, due to [Mordecki \(2002\)](#), in which log project values follow a Lévy process.

Consider an agent with the option of investing an amount $K > 0$ in a project at a

nonnegative time of his choice. If the agent invests at time t , the project returns a gross payoff of $U(t) \equiv U_0 \exp[Y(t)]$ to the agent, where $U_0 > 0$ is the project's initial value. [Mordecki](#) allows $\{Y\}$ to be a general Lévy process; we continue to assume it is spectrally negative. The agent chooses a random investment time T that maximizes expected net payoffs, discounted at a rate R ; $w_M(T) \equiv \mathbb{E}[\exp(-RT) \{U(T) - K\}]$. The agent's choice is restricted to investment times T that are feasible given the information available to the agent, which, at time t , we take to be $\{Y(t'); 0 \leq t' \leq t\}$, K , U_0 , and R . Formally, if $\{\mathcal{F}\}$ is the filtration generated by these variables, then $\{T \leq t\}$ should be adapted to $\{\mathcal{F}\}$.

Suppose that $R > \ln \mathbb{E}[\exp(Y(1))] = \psi(1)$, so that $\Lambda(R) > 1$. For example, in the Brownian motion case, this requires that $R > \mu + \sigma^2/2$. Denote $\bar{Y}(t) \equiv \sup_{t' \in [0,t]} Y(t')$. Let E_R be an independent exponential time with parameter R . Then, because $\{Y\}$ is spectrally negative, $\bar{Y}(E_R)$ has an exponential distribution with parameter $\Lambda(R)$ ([Bertoin, 1996](#), Section VII.1). Using this, Theorem 1 in [Mordecki \(2002\)](#) implies that the agent will invest when $\{Y\}$ first crosses

$$\underline{y}_M \equiv \max \left\{ \ln \left[\frac{K}{U_0} \frac{\Lambda(R)}{\Lambda(R) - 1} \right], 0 \right\}, \quad (13)$$

at time $T(\underline{y}_M)$. Note that \underline{y}_M decreases with the discount rate R . As $R \rightarrow \infty$, the agent will invest as soon as the investment option is *in the money*, $U(t) > K$. If the option is sufficiently deep in the money at time 0; that is, if U_0 is sufficiently larger than K ; then $\underline{y}_M = 0$, and the agent will invest immediately.

A closely-related class of models, due to [Novikov and Shiryaev \(2005\)](#), alternatively specifies the payoffs to T as

$$w_n(T) \equiv \mathbb{E}[\exp(-RT) \max \{U_0 + Y(T) - K, 0\}^n], \quad n \in \mathbb{Z}_+.$$

Here, we can interpret $U_0 + Y(t)$ as a project's value at time t , with K again the investment cost. Theorem 2 in [Kyprianou and Surya \(2005\)](#) gives optimal investment thresholds for

all $n \in \mathbb{Z}_+$. Again applying the simplifications brought by the absence of positive shocks, these thresholds reduce to

$$\underline{y}_1 \equiv \max \left\{ K - U_0 + \frac{1}{\Lambda(R)}, 0 \right\} \quad \text{and} \quad \underline{y}_2 \equiv \max \left\{ K - U_0 + \frac{2}{\Lambda(R)}, 0 \right\},$$

for $n = 1$ and $n = 2$, respectively.

In both specifications, primitive heterogeneity in investment costs K , initial project values U_0 , and discount rates R generates heterogeneous nonnegative investment thresholds. Suppose that we have data on investment times T and covariates X ; that (K, U_0, R) is fully determined by X and an unobserved heterogeneity factor V ; and that $\{Y\}$ is independent of (X, V) . Then, we can apply any of Propositions 2–5 if we assume that the threshold is proportional in the effects of X and that of V (if necessary, using Appendix A’s extension with zero thresholds).

Without further data or assumptions on the model’s primitives, such a direct assumption on the reduced-form dependence of the threshold on X and V needs be made; because thresholds are nonnegative, a proportional specification is a natural first choice. Typically, this implies that the primitive heterogeneity in (K, U_0, R) depends on the parameters of Λ , which is unattractive. For example, in [Novikov and Shiryaev](#)’s specification, with $n = 1$ and $U_0 = K$, we get $\underline{y}_1 = \phi(X)V$ if $R = \Lambda^{-1} [\{\phi(X)V\}^{-1}]$. Note though that we can invoke alternative identification results, yielding identification of more attractive specifications, if we impose additional structure or use more information. For example, if data stratified on V are available, with multiple durations per stratum, Proposition 6 can be applied to establish identification of a model in which X enters in an unrestricted way. This accommodates any specification of the dependence of (K, U_0, R) on X and V .

Either way, under an appropriate set of identifying assumptions, we can separately measure agent-level investment dynamics, coded into Λ , and investment threshold heterogeneity. This provides a theory-based empirical distinction of state dependence and heterogeneity in investment timing. The results can moreover be used to further ex-

plore the model’s primitives. Obviously, without more information on these primitives, or strong assumptions, they are typically not fully identified. Nevertheless, the MHT identification results provide a useful first stage for exploring their second-stage identification, and that of other structural quantities. For example, in [Novikov and Shiryaev](#)’s example with $U_0 = K$ and $n = 1$, the investment option’s value is $w_1[T(\underline{y}_1)] = \exp(-1)\underline{y}_1$. Thus, from the MHT analysis, not only the distribution of R , but also the distribution of option values is identified up to scale if we assume linear utility.

An unattractive feature of this section’s models is that they take the project’s initial value U_0 and the investment size K as primitives. Without further constraints on their distribution in the data; it is clear from [\(13\)](#) that this may lead to masses of agents with zero thresholds, who invest immediately, and nontrivial selection on primitives in the subpopulation with positive thresholds. This complicates the model’s econometric specification, and the interpretation of the empirical results. Such problems are not specific to the MHT framework, but are a special instance of the initial-conditions problem studied by [Heckman \(1981b\)](#). This problem arises if a stochastic process is not sampled from its origin, and is usually solved by somehow modeling the initial conditions of the sample. Within the context of this section’s models, this requires that we model the way agents ended up with their investment option to begin with. To this end, we will explicitly model entry into the state of interest along with exit from this state.

6.2 Unemployment Durations and Heterogeneous Entry and Exit Costs

Consider a labor market in which workers continuously choose between unemployment and employment. A worker earns a flow B when unemployed, and $U(t) \equiv U_0 \exp[\mu t + \sigma W(t)]$ when employed at calendar time t . Note that $U(t)$ is a geometric Brownian motion with drift, and that $\mathbb{E}[U(t)] = U_0 \exp[(\mu + \sigma^2/2)t]$. Workers incur a lump-sum cost $\underline{K} \geq 0$ when they leave their job; and pay $\overline{K} \geq 0$ when they enter a job. They maximize

expected earnings, discounted at a rate $R > \mu + \sigma^2/2$.

This setup is equivalent to [Dixit's \(1989\)](#) model of firm entry and exit, and has many alternative applications, for example to marriage and divorce. From [Dixit's](#) analysis, it follows that an unemployed worker enters employment when $U(t)$ increases above \bar{U} , and resigns when $U(t)$ falls below \underline{U} ; where $\bar{U} = \underline{U}$ if $\bar{K} = \underline{K} = 0$, and $\bar{U} > \underline{U}$ otherwise.

The MHT model applies to an inflow sample of unemployment durations. Normalize the start time of each unemployment spell in the sample to 0. Then, unemployed start the sampled spell with earnings $U(0) = \underline{U}$, and end their spell when earnings hit the exit threshold $\bar{U} \geq \underline{U}$. Define $Y(t) \equiv \ln U(t) - \ln \underline{U}$, and note that $Y(t)$ is a Brownian motion with drift term μt . Then, we can equivalently say that workers initially have normalized log earnings $Y(0) = 0$, and leave for employment when $\{Y\}$ hits $\underline{y}_D \equiv \ln \bar{U} - \ln \underline{U}$. From [Dixit's \(1989\)](#) analysis it follows that \underline{y}_D varies on $[0, \infty)$ with observed and unobserved determinants of \bar{K} and \underline{K} , with $\underline{y}_D = 0$ only in the frictionless limit. Thus, a proportional specification $\underline{y}_D = \phi(X)V$ is natural.

If $\bar{K} < \infty$, then $\underline{y}_D < \infty$ even if $\underline{K} \rightarrow \infty$. This exemplifies that unrestricted primitive heterogeneity may lead to bounded threshold heterogeneity. Then, threshold heterogeneity has a finite mean, $\mathbb{E}[V] < \infty$, and [Proposition 3](#) provides point identification.

6.3 Job Separations and Heterogeneous Search

In [Dixit's \(1989\)](#) model, transaction costs are lump-sum entry and exit costs, earnings are general, and utility is linear. In labor economics, transaction costs are often specified as job search frictions. Moreover, key search models, such as [Mortensen and Pissarides's \(1994\)](#), entertain job-specific shocks. Therefore, we end this section with a basic model of endogenous job separations in the presence of heterogeneous search frictions, job-specific shocks, and nonlinear utility.

Again consider a labor market in which workers are either employed or unemployed. When employed in their j -th job for an amount of time t , workers earn a flow utility

$U^j(t) \equiv U_0 \exp[-\alpha Y^j(t)]$. Here, $\{Y^j\}$ is a Lévy process indexed by *job tenure* t that is distributed identically and independently across jobs j , and $U_0 > 0$ and $\alpha > 0$ are job-invariant parameters. Employed workers cannot search on the job; but, they can leave their jobs for unemployment immediately and at no cost, and will do so when the expected discounted utility of continued employment falls below the expected discounted utility of unemployment. Once they are unemployed, workers can search sequentially for new jobs. We assume that unemployed workers are offered jobs at an exogenous and independent Poisson rate A , and earn a flow utility $B < U_0$. Because all new jobs offer identical earnings prospects, this ensures that unemployed workers accept the first job they are offered. Consequently, search frictions are effectively exogenous, and we can focus on endogenous job separations given search frictions indexed by A .

Denote the expected discounted utility in a job in state Y with $v(Y)$, and the expected discounted utility of unemployment with W . We first provide some explicit results for the special case in which $\{Y^j\}$ is a compound Poisson process with negative jumps and positive drift: $Y^j(t) = \mu t + \Delta Y^j(t)$; with $\mu > 0$, and $\Delta Y^j(t)$ shocks that arrive at a Poisson rate $\lambda > 0$ and have an independent exponential distribution on $(-\infty, 0)$ with parameter $\omega > 0$. To ensure nontrivial job separation strategies; we assume that $\omega > \alpha$, so that $U^j(t)$ has finite expectations for finite t ; and we assume that the discount rate R strictly exceeds the expected utility growth rate in employment, $\alpha \left[\frac{\lambda}{\omega - \alpha} - \mu \right]$, so that the expected discounted utility v^* of staying employed forever exists, and equals $v^*(Y) = \gamma \exp(-\alpha Y)$, with $\gamma \equiv U_0 \left\{ R - \alpha \left[\frac{\lambda}{\omega - \alpha} - \mu \right] \right\}^{-1} > 0$.

From standard contraction arguments, it follows that $v(Y)$ weakly decreases with Y , so that employed workers apply a threshold strategy: They will leave their j -th job for unemployment when $Y^j(t)$ exceeds a threshold \underline{y} . Given W , the expected discounted utility in employment v and the job separation threshold \underline{y} satisfy the Bellman equation

$$(R + \lambda) v(Y) = U_0 \exp(-\alpha Y) + \lambda \int_0^\infty v(Y - e) \omega \exp(-\omega e) de + \mu v'(Y), \quad Y \in (-\infty, \underline{y});$$

with value matching, $\lim_{Y \uparrow \underline{y}} v(Y) = W$; smooth pasting, $\lim_{Y \uparrow \underline{y}} v'(Y) = 0$; and a no-bubble condition, $\lim_{Y \rightarrow -\infty} [v(Y) - v^*(Y)] = 0$. It is straightforward to verify that this implies that $v(Y) = v^*(Y) + \delta(W) \exp(\zeta Y)$ and $\underline{y} = (\zeta + \alpha)^{-1} \ln\left(\frac{\alpha\gamma}{\delta(W)\zeta}\right)$; where $\zeta \equiv \left[R + \lambda - \mu\omega + \sqrt{(R + \lambda - \mu\omega)^2 + 4R\omega\mu} \right] / (2\mu) > 0$, and $\delta(W)$ is implicitly determined by $\delta(W) = \exp(-\zeta\underline{y}) [W - \gamma \exp(-\alpha\underline{y})]$. With $W = \{B + A[\gamma + \delta(W)]\} / (A + R)$, this gives a unique solution (v, W, \underline{y}) . The job separation threshold \underline{y} decreases with A , and $\underline{y} \downarrow 0$ as $A \rightarrow \infty$. That is, smaller job search frictions make the employed less tolerant to decreases in utility from employment; in the frictionless limit, they will not tolerate any utility loss. As $A \rightarrow 0$, \underline{y} may either diverge to ∞ or converge to a finite limit.

If A varies over $[0, \infty)$ in the population, then the job separation threshold \underline{y} has support $(0, \infty]$. As before, under assumptions that ensure that $\underline{y} = \phi(X)V$, the MHT model can be applied to employment duration data to learn about job separations. The fact that \underline{y} is, under some conditions, bounded may be exploited to justify the application of Proposition 3.

As in Section 6.1, deeper parameters can possibly be identified if more data are available. In particular, note that the model specifies that unemployment durations conditional on A are exponential, so that the distribution of A is identified from a random sample of unemployment durations by the uniqueness of the Laplace transform (Feller, 1971, Section XIII.1, Theorem 1). This is a simple example of the MHT and mixed hazard approaches joining forces in structural duration analysis.

A similar analysis can be developed for the case that $\{Y\}$ is a Brownian motion with drift, along the lines of Stokey (2009, Section 6.4). In fact, the results extend to more general Lévy processes (Boyarchenko and Levendorskiĭ, 2007, Chapter 11). Here, we focused on the compound Poisson case to connect to the search-matching literature in labor economics, which often relies on Poisson processes. Mortensen and Pissarides's (1994) model with endogenous job separations, for example, assumes that new match-specific productivity values are drawn independently from a fixed distribution at Poisson times.

This specification is typical of the way much of the search literature models transitions, and ensures a stationary environment in which agents only leave their jobs at the time of a shock. It directly implies a separation hazard, which is the arrival rate of new productivity draws times the time-invariant probability that such a draw is below a separation threshold. This can be contrasted with the specification studied here, which involves persistent idiosyncratic shocks that improve the payoffs in employment, combined with a common continuous drift towards separation. Because shocks can only improve payoffs to employment, separations do not take place at Poisson times, and a hazard specification is not directly implied. Because shocks are persistent, the model implies that individual workers, with given thresholds, have time-varying rates of leaving their jobs.

7 Extensions

This section discusses three important extensions that are beyond the scope of this paper.

7.1 Time-Varying Covariates

Following most of the duration-model identification literature, we have ignored time-varying covariates. Time-varying covariates can be introduced in the MHT model as determinants of a time-varying threshold. However, both the characterization of the corresponding hitting-time process, and its structural interpretation as a reduced form of an optimal stopping model are hard. This suggests that we alternatively treat time-varying covariates as noisy measurements of the latent state process, as in [Abbring and Campbell's \(2005\)](#) discrete-time model of industry dynamics. This complicates the analysis with a filtering problem, but respects much of the current model's structure.

It is well known that time variation in observed covariates can be exploited to relax some of the more controversial identifying assumptions for the MPH model, such as [Elbers and Ridder's \(1982\)](#) finite-mean assumption (see e.g. [Heckman and Taber, 1994](#)). From

this perspective, the case of time-invariant covariates, and in fact a single binary one, can be seen as informing us what can be learned with minimal covariate variation. Additional time-variation in the covariates can only aid identification, as with the MPH model.

7.2 Nonstationary Increments

[Aalen and Gjessing \(2001\)](#) show that hitting-time models based on Brownian motions exhibit *quasi-stationarity*: The distribution of $Y(t)|(T \geq t)$ converges to a gamma distribution and hazard rates corresponding to different thresholds converge to a common limit as time t increases. This both suggests that the MHT model may be too restrictive in some applications and that models with richer time effects may be identifiable. One such model specifies $T \equiv \xi(U)$, for an increasing time transformation $\xi : [0, \infty] \mapsto [0, \infty]$ and the distribution of $U|X$ given by the MHT model. If ξ is linear, this simply gives the MHT model for $T|X$; any nonlinearities correspond to additional duration dependence.

One structural source of nonstationarity that may be captured this way is Bayesian learning, as in [Jovanovic's \(1979; 1984\)](#) model of job tenure. [Lancaster \(1990, Section 6.5\)](#) suggests that we approximate job tenure T predicted by [Jovanovic's](#) theory by $\xi(U)$, with

$$\xi(u) \equiv \begin{cases} \frac{\eta^2 u}{1 - \eta u} & \text{if } u \in [0, \eta^{-1}) \text{ and} \\ \infty & \text{if } u \in [\eta^{-1}, \infty]. \end{cases}$$

Here, U the first time a Brownian motion crosses a threshold that decreases linearly from a positive initial value, which is equivalent to the first time a Brownian motion with upward drift crosses a positive threshold. The probability $\Pr(U \geq \eta^{-1})$ equals the defect $\Pr(T = \infty)$ that arises because some agents will eventually learn that they are in a good match and never leave it. We can extend this framework to include observed and unobserved covariates by replacing the marginal specification of U by a Gaussian MHT model for the distribution of $U|X$. The resulting model is a simple, one-parameter extension of the MHT model that allows for nonstationary increments.

7.3 Generalized Ornstein-Uhlenbeck Processes

Lévy processes are a key component in many process-based duration models in econometrics and statistics. Another frequent choice is the Ornstein-Uhlenbeck process (e.g. [Aalen and Gjessing, 2004](#)). This process allows for mean reversion and may be more appropriate in some applications. A specification for $\{Y\}$ that includes both as special cases is the Ornstein-Uhlenbeck process driven by a Lévy process. Such a process satisfies $dY(t) = -\varrho Y(t)dt + dZ(t)$, with $\varrho \in [0, \infty)$ and $\{Z\}$ a Lévy process. The usual Ornstein-Uhlenbeck process arises if $\{Z\}$ is a Brownian motion and $\varrho > 0$. We explicitly include the boundary case $\varrho = 0$, in which $\{Y\}$ is a Lévy process.

The Laplace transform of the distribution of $T|X$ in a MHT model generalized this way can be derived from [Novikov \(2004\)](#), who provides explicit expressions for the Laplace transform of the hitting-time distribution of an Ornstein-Uhlenbeck process driven by a spectrally-negative Lévy process. However, even though the generalized model adds only one parameter, ϱ , [Novikov's](#) results suggest that an analysis of its identifiability requires more than just a simple variation of the present paper's analysis.

8 Conclusion

This paper's main contribution is to provide fundamental insight in the empirical content of a framework for econometric duration analysis, the MHT model, that is connected to an important class of dynamic economic models with heterogeneous agents. It does so by highlighting and exploiting a close analogy between the identification analysis of the MHT model and that of the MPH model. This way, it extends the applicability of the MPH identification literature to a new, and structurally important, class of duration models.

The analogy between the analysis of the MHT and the MPH models should not be mistaken for a structural similarity between both frameworks. In the MPH model, the (mixed) exponential form arises from the exponential formula for the survival function. In

the MHT model, it arises from the infinite divisibility of the law characterizing the latent Lévy process $\{Y\}$, which, with the assumption that $\{Y\}$ is spectrally negative, ensures that the hitting times $T(y)$ are infinitely divisible.

In fact, as we have noted in the introduction and illustrated with Figure 2, MHT hazard rates are generally not multiplicative in the effects of time and those of heterogeneity. This implies that the empirical analysis of data generated by the MHT model with an MPH framework will generally produce invalid structural conclusions. For example, consider Section 3's Gaussian example of the MHT model, with $\sigma = 0$ and V distributed as a mixture of exponentials: $\Pr(V > v) = \int_0^\infty \exp(-xv)dG^*(x)$, for some distribution G^* . This MHT triplet cannot be statistically distinguished from an MPH model with a constant baseline hazard and an unobserved heterogeneity factor with distribution G^* ; both imply a mixture of exponentials specification of $T|X$. However, the MHT specification assigns all variation between individuals to time-invariant unobserved heterogeneity; the MPH specification instead interprets part of the cross-sectional variation as driven by idiosyncratic, time-homogeneous Poisson shocks. This strongly motivates the use of the MHT model when the MHT structure holds, for example in applications to optimal stopping problems of the type discussed in Section 6.

Of course, the same considerations should lead one to prefer an MPH model when an MPH structure holds. Hazard models are particularly natural in applications to decision processes that are driven by Poisson processes, such as sequential job search or insurance claim behavior. The fact that such processes usually do not generate proportional hazards (Van den Berg, 2001) may cast doubt on the structural applicability of the MPH model, but calls for the use of specific nonproportional hazard models, rather than the MHT model. The fact that Section 6.3's search-matching model combines a hazard model for job search with a hitting-time model for job tenure exemplifies the complementary nature of the hitting-time and hazard approaches to duration analysis.

There may also be statistical reasons to prefer one framework over the other. Both

the MHT and the MPH models are rich descriptive frameworks, which can perfectly fit any duration distribution for a single given value of the observed covariates. They do however impose restrictions on the variation of durations with covariates. To some extent, these restrictions are the same in both models: The mixture of exponentials example shows that they contain nontrivial subclasses of observationally equivalent specifications. However, it is easy to show, by counterexample, that the MHT and MPH models are not observationally equivalent in general. Consider again Section 3's Gaussian example with $\sigma = 0$, but now with V concentrated on a strict subset of $(0, \infty)$, such as $(0, 1)$. Then, the implied support of $T|X$ varies with the covariates X . The MPH model cannot reproduce this statistical implication, because it can only generate gaps in the support of $T|X$ through the baseline hazard, which is common across covariate values X .

An attractive feature of the MHT model is that it includes the AFT model as a special case. In fact, this section's two Gaussian examples with $\sigma = 0$ are both special cases of this standard model from statistics. As discussed in Section 3, the AFT model can be interpreted as a polar specification of the MHT model in which all variation in durations is due to *ex ante* heterogeneity. More generally, the hitting-time structure, with the Lévy assumption on the latent process, tightly specifies agent-level time effects as potentially endogenous outcomes; whereas the MPH model offers direct control over such effects, through the baseline hazard. This tight specification of agent-level dynamics, in terms of a latent process that can be the state in a decision problem, is key to the MHT model's close relation with economic theory. It does however complicate the introduction of time-varying covariates; which, at least from a statistical perspective, can be straightforwardly introduced into a hazard model. Section 7.1 proposes that we respect the basic structure of the MHT model by introducing time-varying covariates as noisy measurements of the latent state. The further development of a theory-based and computationally feasible way to introduce time-varying covariates in the MHT model is a key next step in its analysis.

Appendix

A Extending the Support of the Threshold

If we extend the support of G to $[0, \infty]$, the model allows for an unobserved subpopulation $\{V = 0\}$ of agents using a zero threshold. On this subpopulation, $T = T(0) = 0$ almost surely, that is $\Pr(T = 0, V = 0) = \Pr(V = 0)$, because $\{Y\}$ visits $(0, \infty)$ at arbitrarily small times almost surely (Bertoin, 1996, Chapter VII, Theorem 1).

The case in which V , and therefore T , has a mass point at 0 may be of interest in some applications; but even then, data on immediate transitions may not be available. In applications in which a mass at 0 is indeed relevant, the analysis in the main text applies to the distribution of $V|V > 0$ and all other model components. If data on immediate transitions are available, in addition $\Pr(V = 0)$ can be identified with $\Pr(T = 0)$. Thus, our focus on the case in which $\Pr(V = 0) = 0$ is without loss of generality.

We could also extend the model by allowing for an *observed* subpopulation with a zero threshold, by including 0 in the range of ϕ . Similarly, we could allow for observed stayers by including ∞ in the range of ϕ . Because such subpopulations can be trivially identified from complete data, these extensions are of little interest for the purpose of this paper.

B Proofs

Denote $\mathcal{L}_x(\cdot) \equiv \mathcal{L}_T(\cdot|X = x)$ and note that F_0 and F_1 uniquely determine \mathcal{L}_0 and \mathcal{L}_1 .

Proof of Proposition 1. Without loss of generality, let $\mathcal{L}(0) \leq \tilde{\mathcal{L}}(0)$; and suppose that $\mathcal{L}_0 \leq \mathcal{L}_1$, so that $\beta < 1$ and $\tilde{\beta} < 1$.

Observational equivalence implies that $\mathcal{L} \circ (\beta\mathcal{L}^{-1}) = \mathcal{L}_1 \circ (\mathcal{L}_0^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}\tilde{\mathcal{L}}^{-1})$ on $(0, \mathcal{L}_0(0))$, where \circ denotes function composition. Moreover, by the real analyticity of the Laplace transform (Widder, 1946, Chapter IV, Theorem 3a), the real analytic inverse function theorem (Krantz and Parks, 2002, Theorem 1.5.3), and the real analyticity of

compositions of real analytic functions (Krantz and Parks, 2002, Proposition 1.4.2); $\mathcal{L} \circ (\beta\mathcal{L}^{-1})$ and $\tilde{\mathcal{L}} \circ (\tilde{\beta}\tilde{\mathcal{L}}^{-1})$ are real analytic on $(0, \mathcal{L}(0))$ (see also Kortram et al., 1995, for an alternative, complex analytic approach). Taken together with $\mathcal{L}_0(0) > 0$, using analytic extension (based on e.g. Krantz and Parks, 2002, Corollary 1.2.6), this implies that

$$\mathcal{L} \circ (\beta\mathcal{L}^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}\tilde{\mathcal{L}}^{-1}) \quad \text{on } (0, \mathcal{L}(0)). \quad (14)$$

Note that both sides of (14) map $(0, \mathcal{L}(0))$ into itself. Thus, we can compose each side n times with itself, and find that

$$\mathcal{L} \circ (\beta^n\mathcal{L}^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}^n\tilde{\mathcal{L}}^{-1}) \quad \text{on } (0, \mathcal{L}(0)), \quad n \in \mathbb{N}. \quad (15)$$

Because \mathcal{L} and $\tilde{\mathcal{L}}$ are continuous at 0; evaluating both sides of (15) at a fixed $s \in (0, \mathcal{L}(0))$, and letting $n \rightarrow \infty$, gives $\mathcal{L}(0) = \tilde{\mathcal{L}}(0)$. \square

Proof of Proposition 2. We first show that the claimed result holds for some $\rho \in (0, \infty)$ in each of the regularity condition's four possible cases, and then prove that it holds with $1/2 \leq \rho \leq 2$.

Without loss of generality, suppose that $\mathcal{L}_0 \leq \mathcal{L}_1$, so that $\beta < 1$ and $\tilde{\beta} < 1$.

(i). Suppose that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at 0, with exponents $\tau, \tilde{\tau} \in (-1, 0]$.

By Proposition 1, $\mathcal{L}(0) = \tilde{\mathcal{L}}(0)$. With (15), this implies that $\mathcal{L}(0) - \mathcal{L} \circ (\beta^n\mathcal{L}^{-1}) = \tilde{\mathcal{L}}(0) - \tilde{\mathcal{L}} \circ (\tilde{\beta}^n\tilde{\mathcal{L}}^{-1})$ on $(0, \mathcal{L}(0))$, $n \in \mathbb{N}$. Taking logs, and then derivatives, yields

$$\frac{\mathcal{K}'(s)}{\mathcal{K}(s)} \left(\frac{\beta^n \mathcal{K}(s) \mathcal{L}'[\beta^n \mathcal{K}(s)]}{\mathcal{L}(0) - \mathcal{L}[\beta^n \mathcal{K}(s)]} \right) = \frac{\tilde{\mathcal{K}}'(s)}{\tilde{\mathcal{K}}(s)} \left(\frac{\tilde{\beta}^n \tilde{\mathcal{K}}(s) \tilde{\mathcal{L}}'[\tilde{\beta}^n \tilde{\mathcal{K}}(s)]}{\tilde{\mathcal{L}}(0) - \tilde{\mathcal{L}}[\tilde{\beta}^n \tilde{\mathcal{K}}(s)]} \right)$$

for all $s \in (0, \mathcal{L}(0))$ and $n \in \mathbb{N}$, with $\mathcal{K} \equiv \mathcal{L}^{-1}$ and $\tilde{\mathcal{K}} \equiv \tilde{\mathcal{L}}^{-1}$. Rearranging gives

$$\frac{\tilde{\mathcal{K}}'(s)/\tilde{\mathcal{K}}(s)}{\mathcal{K}'(s)/\mathcal{K}(s)} = \frac{\beta^n \mathcal{K}(s) |\mathcal{L}'[\beta^n \mathcal{K}(s)]| / \{\mathcal{L}(0) - \mathcal{L}[\beta^n \mathcal{K}(s)]\}}{\tilde{\beta}^n \tilde{\mathcal{K}}(s) |\tilde{\mathcal{L}}'[\tilde{\beta}^n \tilde{\mathcal{K}}(s)]| / \{\tilde{\mathcal{L}}(0) - \tilde{\mathcal{L}}[\tilde{\beta}^n \tilde{\mathcal{K}}(s)]\}}, \quad (16)$$

for $s \in (0, \mathcal{L}(0))$ and $n \in \mathbb{N}$. By [Feller \(1971, Section VIII.9, Theorem 1\(a\)\)](#); the numerator in the right-hand side of (16) converges to $\tau + 1 \in (0, 1]$, and the denominator to $\tilde{\tau} + 1 \in (0, 1]$; for each given $s \in (0, \mathcal{L}(0))$, as $n \rightarrow \infty$. Consequently, $\tilde{\mathcal{K}}'/\tilde{\mathcal{K}} = \rho\mathcal{K}'/\mathcal{K}$ on $(0, \mathcal{L}(0))$, where $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in (0, \infty)$. In turn, this implies $\tilde{\mathcal{K}} = \kappa\mathcal{K}^\rho$ on $(0, \mathcal{L}(0))$, for some arbitrary $\kappa \in (0, \infty)$. Using the definition of \mathcal{K} , this gives $\tilde{\mathcal{L}}(\kappa s^\rho) = \mathcal{L}(s)$ for all s . Finally, from $\mathcal{L} \circ \Lambda = \tilde{\mathcal{L}} \circ \tilde{\Lambda}$, we get $\tilde{\Lambda} = \kappa\Lambda^\rho$; and with $\mathcal{L} \circ (\beta\Lambda) = \tilde{\mathcal{L}} \circ (\beta\tilde{\Lambda})$, we find that $\tilde{\beta} = \beta^\rho$.

(ii). Suppose that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at ∞ , with exponents $\tau, \tilde{\tau} \in (-\infty, -1)$.

Observational equivalence implies that $\mathcal{L} \circ (\beta^{-1}\mathcal{L}^{-1}) = \mathcal{L}_0 \circ (\mathcal{L}_1^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}^{-1}\tilde{\mathcal{L}}^{-1})$ on $(0, \mathcal{L}_1(0))$. As in [Proposition 1's](#) proof, this gives $\mathcal{L} \circ (\beta^{-n}\mathcal{L}^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}^{-n}\tilde{\mathcal{L}}^{-1})$ on $(0, \mathcal{L}(0))$, $n \in \mathbb{N}$. Taking logs and derivatives, and rearranging, yields

$$\frac{\tilde{\mathcal{K}}'(s)/\tilde{\mathcal{K}}(s)}{\mathcal{K}'(s)/\mathcal{K}(s)} = \frac{\beta^{-n}\mathcal{K}(s) |\mathcal{L}'[\beta^{-n}\mathcal{K}(s)]| / \{\mathcal{L}[\beta^{-n}\mathcal{K}(s)]\}}{\tilde{\beta}^{-n}\tilde{\mathcal{K}}(s) |\tilde{\mathcal{L}}'[\tilde{\beta}^{-n}\tilde{\mathcal{K}}(s)]| / \{\tilde{\mathcal{L}}[\tilde{\beta}^{-n}\tilde{\mathcal{K}}(s)]\}}, \quad (17)$$

for all $s \in (0, \mathcal{L}(0))$. By [Feller \(1971, Section VIII.9, Theorem 1\(a\)\)](#); the numerator in the right-hand side of (17) converges to $-(\tau + 1) \in (0, \infty)$ and the denominator to $-(\tilde{\tau} + 1) \in (0, \infty)$, so that the right-hand side again converges to $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in (0, \infty)$; for each given $s \in (0, \mathcal{L}(0))$, as $n \rightarrow \infty$. As in [Case \(i\)](#), this gives $\tilde{\mathcal{L}}(\kappa s^\rho) = \mathcal{L}(s)$ for all s , $\tilde{\Lambda} = \kappa\Lambda^\rho$, and $\tilde{\beta} = \beta^\rho$.

(iii). Suppose that $|\psi'|$ and $|\tilde{\psi}'|$ vary regularly at 0, with exponents $\tau, \tilde{\tau} \in (-1, 1]$.

Observational equivalence implies that $\psi \circ (\beta\Lambda) = \mathcal{L}_0^{-1} \circ \mathcal{L}_1 = \tilde{\psi} \circ (\tilde{\beta}\tilde{\Lambda})$ on (\underline{s}, ∞) , with $\underline{s} \equiv \mathcal{L}_1^{-1}[\mathcal{L}_0(0)] = \psi[\beta^{-1}\Lambda(0)] = \tilde{\psi}[\tilde{\beta}^{-1}\tilde{\Lambda}(0)]$. Recall that $\psi(0) = 0$ and $\lim_{s \rightarrow \infty} \psi(s) = \infty$, and note that ψ is either strictly increasing or strictly convex (or both). Consequently, ψ attains a unique minimum. Denote this minimum with $s_m \in (-\infty, 0]$. Similarly, $\tilde{\psi}$ attains a unique minimum $\tilde{s}_m \in (-\infty, 0]$. Without loss of generality, suppose that $s_m \geq \tilde{s}_m$. Note that the inverses of ψ and $\tilde{\psi}$ exist

on (s_m, ∞) , and redefine Λ and $\tilde{\Lambda}$ to be these inverses. Because ψ and $\tilde{\psi}$ are real analytic (Bertoin, 1996, Section VII.1), Λ and $\tilde{\Lambda}$ are real analytic on (s_m, ∞) by the real analytic inverse function theorem, and compositions of real analytic functions are real analytic; $\psi \circ (\beta\Lambda)$ and $\tilde{\psi} \circ (\tilde{\beta}\tilde{\Lambda})$ are real analytic on (s_m, ∞) . With $\underline{s} < \infty$, using analytic extension, this implies that

$$\psi \circ (\beta\Lambda) = \tilde{\psi} \circ (\tilde{\beta}\tilde{\Lambda}) \quad \text{on } (s_m, \infty). \quad (18)$$

Note that both sides of (18) map (s_m, ∞) into itself. Thus, we can compose each side n times with itself, which gives $\psi \circ (\beta^n\Lambda) = \tilde{\psi} \circ (\tilde{\beta}^n\tilde{\Lambda})$ on (s_m, ∞) , $n \in \mathbb{N}$. Applying calculations that parallel those for Cases (i) and (ii), we find that

$$\frac{\tilde{\Lambda}'(s)/\tilde{\Lambda}(s)}{\Lambda'(s)/\Lambda(s)} = \frac{\beta^n\Lambda(s) |\psi'[\beta^n\Lambda(s)]| / \psi[\beta^n\Lambda(s)]}{\tilde{\beta}^n\tilde{\Lambda}(s) |\tilde{\psi}'[\tilde{\beta}^n\tilde{\Lambda}(s)]| / \tilde{\psi}[\tilde{\beta}^n\tilde{\Lambda}(s)]}, \quad (19)$$

for all $s \in (0, \infty)$ and $n \in \mathbb{N}$. Feller (1971, Section VIII.9, Theorem 1(a)) implies that the right-hand side of (19) converges to $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in (0, \infty)$; for each given $s \in (0, \infty)$, as $n \rightarrow \infty$. With continuity of Λ and $\tilde{\Lambda}$ at 0, this gives $\tilde{\Lambda} = \kappa\Lambda^\rho$, for some arbitrary $\kappa \in (0, \infty)$. With observational equivalence, and using analytic extension, it follows that $\tilde{\mathcal{L}}(\kappa s^\rho) = \mathcal{L}(s)$ for all s , and that $\tilde{\beta} = \beta^\rho$.

(iv). Suppose that $|\psi'|$ and $|\tilde{\psi}'|$ vary regularly at ∞ , with exponents $\tau, \tilde{\tau} \in [0, 1]$.

Observational equivalence implies that $\psi \circ (\beta^{-1}\Lambda) = \mathcal{L}_1^{-1} \circ \mathcal{L}_0 = \tilde{\psi} \circ (\tilde{\beta}^{-1}\tilde{\Lambda})$ on $(0, \infty)$. Analogously to the analysis for Case (iii), this can be used to show that (19) extends from $n \in \mathbb{N}$ to all $n \in \mathbb{Z}$. Feller (1971, Section VIII.9, Theorem 1(b)) implies that the right-hand side of (19) converges to $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in [1/2, 2]$; for each given $s \in (0, \infty)$, as $n \rightarrow -\infty$. Consequently, the conclusion of Case (iii) extends to this case, but with $\rho \in [1/2, 2]$.

At least one of these four cases holds by assumption; so their common conclusion that

$\tilde{\beta} = \beta^\rho$, $\tilde{\Lambda} = \kappa\Lambda^\rho$, and $\tilde{\mathcal{L}}(\kappa s^\rho) = \mathcal{L}(s)$ for all s ; for some $\kappa \in (0, \infty)$ and $\rho \in (0, \infty)$; holds.

Remains to show that Case (iv)'s tighter bound on ρ holds generally. To this end, note that both ψ and $\tilde{\psi}$ should satisfy the Lévy-Khintchine formula (7). Because ψ is convex and $\psi(s) \rightarrow \infty$ as $s \rightarrow \infty$ (Bertoin, 1996, Chapter VII, Section 1), $s^{-1}\psi(s)$ either converges to a strictly positive constant or diverges to ∞ as $s \rightarrow \infty$. Moreover, $s^{-2}\psi(s) \rightarrow \sigma^2/2 < \infty$ (Bertoin, 1996, Chapter I, Proposition 2). Obviously, the same asymptotic behavior is displayed by $\tilde{\psi}$. From $\psi[\Lambda(s)] = s = \tilde{\psi}[\tilde{\Lambda}(s)]$, it follows that $\tilde{\psi}(s) = \psi[(s/\kappa)^{1/\rho}]$, $s \in [\tilde{\Lambda}(0), 0)$. Therefore, if $\rho > 2$, then $\lim_{s \rightarrow \infty} s^{-1}\tilde{\psi}(s) = \lim_{s \rightarrow \infty} \kappa^{-1}s^{-\rho}\psi(s) = 0$. Consequently, $\rho \leq 2$ and, by symmetry, $\rho \geq 1/2$. \square

Proof of Proposition 3. Because $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ have finite positive limits at 0, they vary slowly at 0, and the argument for Case (i) in the proof of Proposition 2 holds with $\tau = \tilde{\tau} = 0$. Consequently, Proposition 2's conclusion holds with $\rho = (\tau + 1)/(\tilde{\tau} + 1) = 1$. \square

Proof of Proposition 4. Because $|\psi'|$ and $|\tilde{\psi}'|$ have finite positive limits at 0, they vary slowly at 0, and the argument for Case (iii) in the proof of Proposition 2 holds with $\tau = \tilde{\tau} = 0$. Consequently, Proposition 2's conclusion holds with $\rho = (\tau + 1)/(\tilde{\tau} + 1) = 1$. \square

Proof of Proposition 5. Because the case that both $\mu = 0$ and $\sigma = 0$ is excluded, $|\psi'|$ and $|\tilde{\psi}'|$ vary regularly at 0, with either exponent 0 or exponent 1. Therefore, the argument for Case (iii) in the proof of Proposition 2 holds with $\tau \in \{0, 1\}$ and $\tilde{\tau} \in \{0, 1\}$. Consequently, Proposition 2's conclusion holds with $\rho = (\tau + 1)/(\tilde{\tau} + 1) \in \{1/2, 1, 2\}$. \square

Proof of Proposition 6. Denote $\mathcal{L}^j \equiv \mathcal{L}_{T^j}$ and $\mathcal{L}^{12} \equiv \mathcal{L}_{T^1, T^2}$; and note that \mathcal{L}^1 , \mathcal{L}^2 and \mathcal{L}^{12} are uniquely determined by the distribution of (T^1, T^2) . Denote $\Lambda_{12}(s) \equiv \Lambda_1(s) + \Lambda_2(s)$, $s \in [0, \infty)$.

Observational equivalence implies that

$$\frac{\Lambda'_1(s_1)}{\Lambda'_2(s_2)} = \frac{\partial \mathcal{L}^{12}(s_1, s_2)/\partial s_1}{\partial \mathcal{L}^{12}(s_1, s_2)/\partial s_2} = \frac{\tilde{\Lambda}'_1(s_1)}{\tilde{\Lambda}'_2(s_2)}, \quad (s_1, s_2) \in (0, \infty)^2.$$

Consequently,

$$\tilde{\Lambda}_j - \tilde{\Lambda}_j(0) = \kappa [\Lambda_j - \Lambda_j(0)]; \quad j = 1, 2; \quad (20)$$

for some $\kappa \in (0, \infty)$. Analogously to [Honoré's \(1993\)](#) proof of his Theorem 1, this would provide identification up to scale if we would know that $\Lambda_j(0) = \tilde{\Lambda}_j(0) = 0; j = 1, 2$. However, at this point, $\Lambda_j(0)$ and $\tilde{\Lambda}_j(0); j = 1, 2$; are not yet determined; and (20) only identifies the Laplace exponents up to location and the common scale factor κ .

To resolve this problem, note that observational equivalence also implies that

$$\Lambda_j^{-1} \circ \Lambda_{12} = (\mathcal{L}^j)^{-1} \circ \mathcal{L}^{12} = \tilde{\Lambda}_j^{-1} \circ \tilde{\Lambda}_{12} \quad \text{on } [0, \infty); \quad j = 1, 2. \quad (21)$$

Substituting (21) in (20) gives

$$\tilde{\Lambda}_{12} - \tilde{\Lambda}_j(0) = \kappa [\Lambda_{12} - \Lambda_j(0)]; \quad j = 1, 2. \quad (22)$$

Moreover, (20) implies that

$$\tilde{\Lambda}_{12} - \tilde{\Lambda}_1(0) - \tilde{\Lambda}_2(0) = \kappa [\Lambda_{12} - \Lambda_1(0) - \Lambda_2(0)]. \quad (23)$$

Together, (22) and (23) imply that $\tilde{\Lambda}_j(0) = \kappa \Lambda_j(0); j = 1, 2$. With (20), this gives $\tilde{\Lambda}_j = \kappa \Lambda_j; j = 1, 2$.

Finally, observational equivalence implies that $\tilde{\mathcal{L}}(\kappa s) = \mathcal{L}(s), s \in (\min_j \Lambda_j(0), \infty)$. Because $\min_j \Lambda_j(0) < \infty$, this equality analytically extends to all $s \in (0, \infty)$. Finally, because \mathcal{L} and $\tilde{\mathcal{L}}$ are continuous at 0, we have that $\mathcal{L}(0) = \tilde{\mathcal{L}}(0)$. \square

References

- Aalen, O. O. and H. K. Gjessing (2001). Understanding the shape of the hazard rate: A process point of view. *Statistical Science* 16(1), 1–14. [3](#), [37](#)
- Aalen, O. O. and H. K. Gjessing (2004). Survival models based on the Ornstein-Uhlenbeck process. *Lifetime Data Analysis* 10, 407–423. [38](#)
- Abbring, J. H. (2002). Stayers versus defecting movers: A note on the identification of defective duration models. *Economics Letters* 74, 327–331. [6](#), [16](#)
- Abbring, J. H. and J. R. Campbell (2005). A firm’s first year. Discussion Paper 05-046/3, Tinbergen Institute, Amsterdam. Under revision. [36](#)
- Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects and dynamic discrete choice, and general equilibrium policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 72, pp. 5145–5303. Amsterdam: Elsevier Science. [3](#)
- Abbring, J. H. and J. J. Heckman (2008). Dynamic policy analysis. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data* (third ed.), Chapter 24, pp. 797–865. Dordrecht: Kluwer Academic Publishers. [3](#)
- Abbring, J. H. and G. Ridder (2009). A note on the non-parametric identification of generalized accelerated failure-time models. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. [18](#), [19](#)
- Abbring, J. H. and T. Salimans (2009). Estimating mixed hitting-time models. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. [24](#), [28](#)
- Abbring, J. H. and G. J. Van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94, 87–99. [25](#)
- Alvarez, F. and R. Shimer (2008). Search and rest unemployment. Working Paper 13772,

- National Bureau of Economic Research, Cambridge, MA. [3](#)
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag. [24](#), [28](#)
- Bertoin, J. (1996). *Lévy Processes*. Number 121 in Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press. [4](#), [5](#), [12](#), [13](#), [14](#), [15](#), [30](#), [41](#), [44](#), [45](#)
- Boyarchenko, S. and S. Levendorskiĭ (2007). *Irreversible Decisions under Uncertainty: Optimal Stopping Made Easy*. Berlin: Springer-Verlag. [1](#), [29](#), [35](#)
- Carrasco, M. and J.-P. Florens (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* *16*, 797–834. [27](#)
- Carrasco, M. and J.-P. Florens (2002). Efficient GMM estimation using the empirical characteristic function. Mimeo, Department of Economics, University of Rochester. [27](#)
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B* *34*, 187–202. [8](#)
- Cox, D. R. and H. D. Miller (1965). *The Theory of Stochastic Processes*. London: Methuen. [8](#)
- Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy* *97*(3), 620–638. [29](#), [33](#)
- Dixit, A. K. and R. S. Pindyck (1994). *Investment under Uncertainty*. Princeton University Press. [1](#), [29](#)
- Elbers, C. and G. Ridder (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies* *64*, 403–409. [11](#), [15](#), [17](#), [19](#), [20](#), [36](#)
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (Second ed.), Volume II. New York: Wiley. [5](#), [10](#), [15](#), [16](#), [17](#), [18](#), [19](#), [25](#), [35](#), [43](#), [44](#)
- Heckman, J. J. (1981a). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies*

- in Labor Markets*. University of Chicago Press. [2](#)
- Heckman, J. J. (1981b). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 179–85. Cambridge, MA: MIT Press. [32](#)
- Heckman, J. J. (1981c). Statistical models for discrete panel data. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178. Cambridge, MA: MIT Press. [2](#)
- Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics* *136*(2), 341–396. [2](#)
- Heckman, J. J. and B. Singer (1984a). The identifiability of the proportional hazard model. *Review of Economic Studies* *51*, 231–241. [11](#), [17](#), [19](#), [20](#)
- Heckman, J. J. and B. Singer (1984b). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* *52*, 271–320. [25](#)
- Heckman, J. J. and C. Taber (1994). Econometric mixture models and more general models for unobservables in duration analysis. *Statistical Methods in Medical Research* *3*, 279–302. [36](#)
- Honoré, B. E. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies* *60*, 241–246. [23](#), [46](#)
- Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy* *87*, 972–990. [3](#), [37](#)
- Jovanovic, B. (1984). Matching, turnover, and unemployment. *Journal of Political Economy* *92*, 108–122. [3](#), [37](#)
- Kortram, R., A. Lenstra, G. Ridder, and A. van Rooij (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica* *49*, 269–281. [42](#)

- Krantz, S. G. and H. R. Parks (2002). *A Primer of Real Analytic Functions* (Second ed.). Boston: Birkhäuser Verlag. [41](#), [42](#)
- Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Berlin: Springer-Verlag. [1](#), [29](#)
- Kyprianou, A. E. and B. A. Surya (2005). On the Novikov-Shiryaev optimal stopping problems in continuous time. *Electronic Communications in Probability* *10*, 146–154. [30](#)
- Lancaster, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society Series A* *135*(2), 257–271. [3](#)
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* *47*, 939–956. [1](#)
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press. [3](#), [8](#), [37](#)
- Lee, M.-L. T. and G. A. Whitmore (2004). First hitting time models for lifetime data. In N. Balakrishnan and C. Rao (Eds.), *Handbook of Statistics*, Volume 23, Chapter 30, pp. 537–543. Amsterdam: Elsevier. [3](#)
- Lee, M.-L. T. and G. A. Whitmore (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Statistical Science* *21*(4), 501–513. [3](#), [26](#)
- McDonald, R. and D. Siegel (1986). The value of waiting to invest. *Quarterly Journal of Economics* *101*(4), 707–728. [29](#)
- Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica* *58*, 757–782. [8](#)
- Moffitt, R. and W. Nicholson (1982). The effect of unemployment insurance on unemployment: The case of federal supplemental benefits. *Review of Economics and Statistics* *64*(1), 1–11. [9](#)

- Mordecki, E. (2002). Optimal stopping and perpetual options for Lévy processes. *Finance and Stochastics* 6(4), 473–493. [29](#), [30](#)
- Mortensen, D. T. and C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies* 61(3), 397–415. [33](#), [35](#)
- Novikov, A. (2004). Martingales and first-passage times for Ornstein-Uhlenbeck processes with a jump component. *Theory of Probability and its Applications* 48, 288–303. [38](#)
- Novikov, A. A. and A. N. Shiryaev (2005). On an effective solution of the optimal stopping problem for random walks. *Theory of Probability and its Applications* 49(2), 344–354. [30](#), [31](#), [32](#)
- Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57, 167–182. [11](#), [15](#), [18](#), [19](#)
- Ridder, G. and T. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica* 71, 1579–1589. [11](#), [17](#), [21](#)
- Shimer, R. (2008). The probability of finding a job. *American Economic Review* 98(2), 268–273. [3](#)
- Singpurwalla, N. D. (1995). Survival in dynamic environments. *Statistical Science* 10(1), 86–103. [3](#)
- Stokey, N. L. (2009). *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton, NJ: Princeton University Press. [1](#), [29](#), [35](#)
- Van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple durations. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 55, pp. 3381–3460. Amsterdam: Elsevier Science. [2](#), [39](#)
- Widder, D. V. (1946). *The Laplace Transform*. Princeton: Princeton University Press. [41](#)