

Mixed Hitting-Time Models*

Jaap H. Abbring[†]

September 2011

Forthcoming in *Econometrica*

Abstract

We study mixed hitting-time models that 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 deduced 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, a co-editor, and three referees; seminar participants at Tinbergen Institute Amsterdam, Toulouse, Paris School of Economics, Chicago, Northwestern, Tilburg, IZA Bonn, RWI Essen, and UPenn; 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. A project web page is maintained at mht.abbring.org.

[†]CentER, Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: jaap@abbring.org. Web: jaap.abbring.org.

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 multiplicative 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 covariates' 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 (Boyarchenko and Levendorskiĭ, 2007; Kyprianou, 2006). 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 multiplicative in the effects of elapsed duration and those of observed and unobserved heterogeneity. Because such multiplicativity 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](#); and [Abbring, 2010](#); 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.

An early application in labor economics is [Lancaster's \(1972\)](#) analysis of strikes. [Lancaster](#) models strike duration as the time that a Brownian motion with drift first hits a threshold that depends on the state of the business cycle. He interprets the gap between the Brownian motion and the threshold as the level of disagreement. [Whitmore \(1979\)](#) uses a similar model to study job tenure, with some discussion of parametric unobserved heterogeneity. [Lancaster \(1990, Sections 3.4.2, 5.7 and 6.5\)](#) reviews hitting-time models based on Brownian motion, and relates them to [Jovanovic's \(1979;1984\)](#) job tenure model. [Shimer \(2008\)](#) recently analyzed unemployment durations using [Alvarez and Shimer's \(2011\)](#) 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. [Aalen and Gjessing, 2001](#); [Lee and Whitmore, 2004, 2006](#); [Singpurwalla, 1995](#)). 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](#) gives examples of economic models that can be analyzed using the MHT model. Sec-

tion 4 presents the MHT model’s implications for the data, highlights the key connection between the analysis of its empirical content and the MPH identification literature, and develops the main identification results. Section 5 briefly considers estimation. Section 6 discusses extensions with time-varying covariates, and to latent processes with non-stationary and dependent increments. Finally, Section 7 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 .

2.1 Lévy Processes

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 right-continuous stochastic process $\{Y\}$ with left limits such that; for every $t, \Delta \geq 0$; the increment $Y(t + \Delta) - Y(t)$ is independent of $\{Y(t'); 0 \leq t' \leq t\}$ and has the same distribution as $Y(\Delta)$.

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 $Y(t)$ is normally distributed with mean μt and variance $\sigma^2 t$, for some scalar drift parameter $\mu \in \mathbb{R}$ and dispersion parameter $\sigma \in [0, \infty)$. Brownian motion is the single Lévy process with continuous sample paths. In general, Lévy processes may have jumps. The jump process $\{Y(t) - \lim_{t' \uparrow t} Y(t'); t > 0\}$ of a Lévy process $\{Y\}$ is a Poisson

point process with characteristic measure Υ such that $\int \min\{1, y^2\} \Upsilon(dy) < \infty$, 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, Theorem I.1). The characteristic measure of $\{Y\}$'s jump process is called its *Lévy measure* and, together with a drift parameter and the dispersion parameter of its Brownian motion component, fully characterizes $\{Y\}$'s distributional properties (see Section 4.1). 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).

2.2 Mixed Hitting Times

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 $T(\infty) \equiv \infty$. We specify T to be the first time that $\{Y\}$ crosses $\phi(X)V$, or

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

for some observed covariates X with support $\mathcal{X} \subseteq \mathbb{R}^k$, measurable function $\phi : \mathcal{X} \rightarrow (0, \infty)$, and positive random variable V , with (X, V) independent of $\{Y\}$. We will refer to (1) as the *mixed hitting-time (MHT) model*. We will pay some specific attention to a version of this model without covariates; that is, with $\phi = 1$. Such a model can be applied to strata defined by the covariates, without restrictions across the strata, and can thus be interpreted as a generalization of (1).

The hitting times $T(y)$ characterize durations for given thresholds y , 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. 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). In the special case of a Brownian motion with drift, the latter requires that $\mu > 0$ if $\sigma = 0$.

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.

2.3 A Gaussian Example

Before further motivating the MHT model’s specification with possible applications, we first highlight some of its salient features with the canonical example that $\{Y\}$ is a non-degenerate Brownian motion ($\sigma > 0$) with upward drift ($\mu > 0$). In this special case, the distribution of $T(y)$ is inverse Gaussian with location parameter y/μ and scale parameter $(y/\sigma)^2$ ([Cox and Miller, 1965](#)). Figure 1 displays two sample paths of $\{Y\}$ for the case

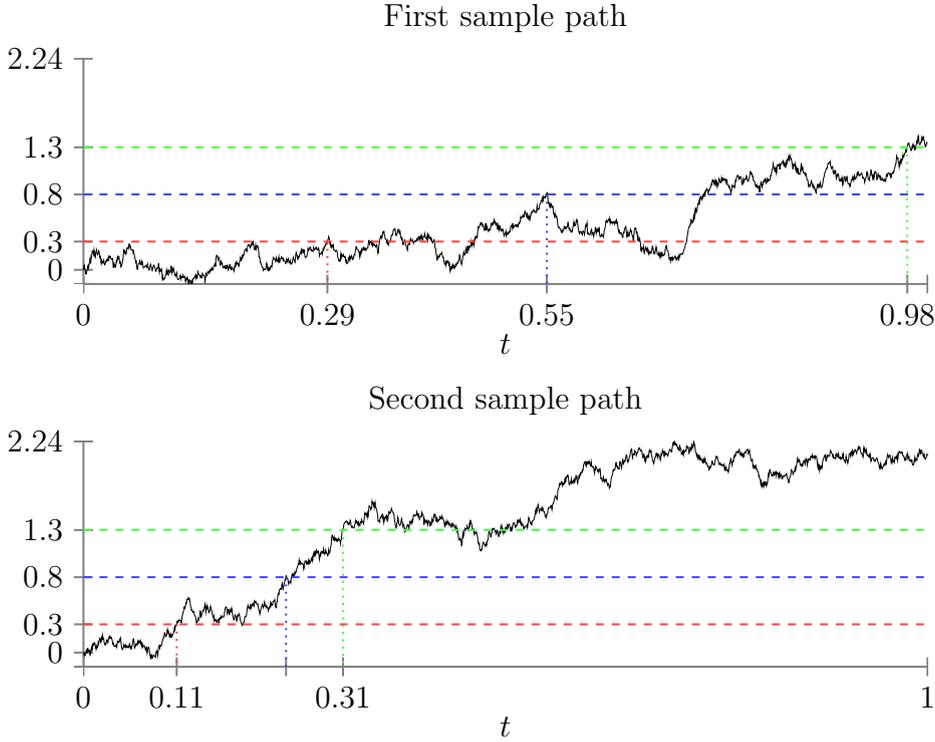


Figure 1: Two sample paths of a standard Brownian motion with unit drift, three possible thresholds, and the corresponding first hitting times.

that $\mu = \sigma = 1$, with three possible exit thresholds; 0.3, 0.8, and 1.3. For a given threshold y , the time that each sample path first crosses that threshold is a realization of $T(y)$. Figure 2 characterizes the distribution of $T(y)$ for each of Figure 1’s threshold values y by plotting the corresponding hazard rates. The three hazard paths have the same hump shape, but are clearly not proportional. Consequently, the hazard rate of $T = T[\phi(X)V]$ conditional on (X, V) is not multiplicative in a function of time and a function of (X, V) ; in this sense, the MHT model is structurally different from the MPH model.

The MHT model with $Y(t) = \mu t$, $\mu \in (0, \infty)$, is of particular interest in statistics. In this boundary case of Brownian motion, $T(y) = \mu^{-1}y$ is a deterministic linear function of the threshold y . Consequently, $T = \mu^{-1}\phi(X)V$, 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

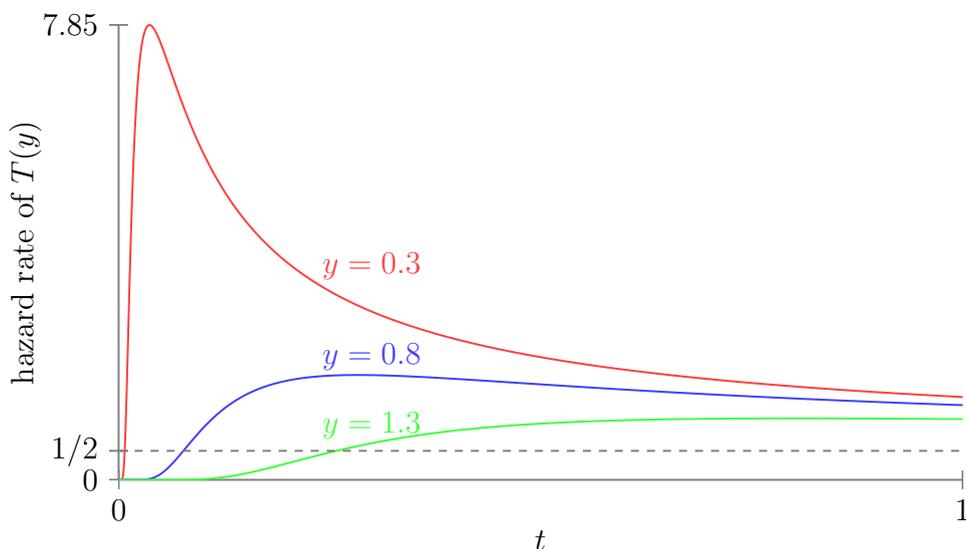


Figure 2: Hazard rates of $T(y)$ for Figure 1's three thresholds y in the case that $\{Y\}$ is a standard Brownian motion with unit drift.

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

Even in this boundary case, a wide variety of duration distributions can be generated by mixing over thresholds. 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 boundary 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 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 motion 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.

3.1 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 $\mathbb{E}[\exp(-RT) \{U(T) - K\}]$, the expected net payoff discounted at a rate R . 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, $\{T \leq t\}$ should be adapted to the filtration generated by these variables.

This investment problem's analysis requires some notation. Define

$$\Lambda(s) \equiv -\ln \mathbb{E} \left[\exp \{-sT(1)\} \cdot I \{T(1) < \infty\} \right], \quad s \in [0, \infty).$$

Here, $I(\cdot) \equiv 1$ if \cdot is true, and 0 otherwise. The factor $I \{T(1) < \infty\}$ makes explicit that the distribution of the hitting time $T(1)$ may be defective. The function Λ is called the *Laplace exponent* of the hitting times of $\{Y\}$ and is central to Section 4's identification analysis. It is fully determined by the distributions of $\{Y\}$; conversely, it also fully determines these distributions (see Section 4.2). The Laplace exponent Λ is strictly increasing, with $\Lambda(0) \geq 0$ and $\lim_{s \rightarrow \infty} \Lambda(s) = \infty$.

Suppose that R is such that $\Lambda(R) > 1$. For example, if $\{Y\}$ is a Brownian motion with general drift coefficient $\mu \in \mathbb{R}$ and dispersion coefficient $\sigma \in (0, \infty)$, then $\Lambda(s) = \left[\sqrt{\mu^2 + 2\sigma^2 s} - \mu \right] / \sigma^2$ and 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

$$\underline{Y}_M \equiv \max \left\{ \ln \left[\frac{K}{U_0} \cdot \frac{\Lambda(R)}{\Lambda(R) - 1} \right], 0 \right\}, \quad (2)$$

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 $\mathbb{E} \left[\exp(-RT) \max \{U_0 + Y(T) - K, 0\}^l \right]$ for some $l \in \{1, 2, \dots\}$. 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 $l \in \{1, 2, \dots\}$. 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 $l = 1$ and $l = 2$, 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 the MHT model to the empirical analysis of the latent process $\{Y\}$, the effect of the covariates X on the threshold, and the distribution of the unobserved heterogeneity in the threshold (if necessary, using [Appendix A](#)'s extension with zero thresholds). In particular, [Section 4.2](#)'s [Theorem 1](#) establishes conditions for the identification of these quantities under the assumption that the threshold is multiplicative in the effect of X and that of V .

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 to be made. Because thresholds are nonnegative, a multiplicative specification is a natural first choice. Typically, this implies that the primitive heterogeneity in (K, U_0, R) depends on Λ , which is unattractive. For example, in [Novikov and Shiryaev](#)’s specification, with $l = 1$ and $U_0 = K$, we get $\underline{Y}_1 = \phi(X)V$ if $R = \Lambda^{-1} [\{\phi(X)V\}^{-1}]$.

With more information or additional structure, this paper’s results allow for the analysis of more attractive specifications of the MHT model. For example, if data stratified on V are available, with multiple durations per stratum, [Section 4.6](#)’s [Theorem 2](#) can be applied to establish identification of a model in which X enters in an unrestricted way (that is, a version of the MHT model with $\phi = 1$ applied to strata defined by X). 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 explore 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 $l = 1$, the investment option’s value is $\mathbb{E}[\exp\{-RT(\underline{Y}_1)\} \underline{Y}_1] = \exp(-1)\underline{Y}_1$. 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 [\(2\)](#) 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.

3.2 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 $U(t^*) \equiv U_0 \exp[Y^*(t^*)]$ when employed at calendar time t^* , with $\{Y^*\}$ a Brownian motion with drift parameter μ and dispersion parameter σ . Note that $\{U\}$ 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 $\bar{K} \geq 0$ 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 $\{U\}$ increases above \bar{U} , and resigns when $\{U\}$ 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. Denote the calendar time at which the worker enters unemployment with T_0 . Let $t \equiv t^* - T_0$ be the duration since entering the sample at calendar time t^* . Then, unemployed start the sampled spell with earnings $U(t + T_0) = \underline{U}$, and end their spell when earnings hit the exit threshold $\bar{U} \geq \underline{U}$. Define $Y(t) \equiv \ln U(t + T_0) - \ln \underline{U}$, and note that $\{Y\}$ is a Brownian

motion with drift parameter μ and dispersion parameter σ . 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 multiplicative specification $\underline{Y}_D = \phi(X)V$ is natural.

If we generalize $\{Y^*\}$ to be a spectrally-negative Lévy process, unemployment durations continue to have an MHT structure of the type introduced in this paper. However, unlike in the Gaussian case, employment durations would not have such a structure: Employment may be terminated at the time of a negative earnings jump. This exemplifies the complementary nature of this paper's MHT model and hazard models.

3.3 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 $U_0 \exp[-\varsigma Y^j(t)]$. 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 $\varsigma > 0$ 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 $A > 0$, and earn a nonnegative 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 $W(y)$, and the expected discounted utility of unemployment with \underline{W} . We first provide some explicit results for the special case in which $\{Y^j\}$ is a compound Poisson process with shocks that arrive at a Poisson rate $\lambda > 0$ and have an independent exponential distribution on $(-\infty, 0)$ with parameter $\omega > 0$, and drift parameter $\mu > 0$. To ensure nontrivial job separation strategies; we assume that $\omega > \varsigma$, so that $\mathbb{E}[\exp\{-\varsigma Y^j(t)\}] < \infty$ for finite t ; and we assume that the discount rate $R > 0$ strictly exceeds the expected utility growth rate in employment, $\varsigma [\frac{\lambda}{\omega - \varsigma} - \mu]$, so that the expected discounted utility $W^*(y)$ of being employed forever in a job currently in state y exists and equals $\gamma \exp(-\varsigma y)$, with $\gamma \equiv U_0 / \{R - \varsigma [\frac{\lambda}{\omega - \varsigma} - \mu]\} > 0$.

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

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

with value matching, $\lim_{y \uparrow \underline{Y}_S} W(y) = \underline{W}$; smooth pasting, $\lim_{y \uparrow \underline{Y}_S} W'(y) = 0$; and a no-bubble condition, $\lim_{y \rightarrow -\infty} [W(y) - W^*(y)] = 0$. It is straightforward to verify that this implies that $W(y) = W^*(y) + \delta(\underline{W}) \exp(\varsigma y)$ and $\underline{Y}_S = (\varsigma + \varsigma)^{-1} \ln\left(\frac{\varsigma \gamma}{\delta(\underline{W}) \varsigma}\right)$; where $\zeta \equiv \left[R + \lambda - \mu\omega + \sqrt{(R + \lambda - \mu\omega)^2 + 4R\omega\mu} \right] / (2\mu) > 0$, and $\delta(\underline{W})$ is implicitly determined by $\delta(\underline{W}) = \exp(-\zeta \underline{Y}_S) [\underline{W} - W^*(\underline{Y}_S)]$. With $\underline{W} = \{B + A[\gamma + \delta(\underline{W})]\} / (A + R)$, this gives a unique solution $(W, \underline{W}, \underline{Y}_S)$.

The job separation threshold \underline{Y}_S decreases with A , and $\underline{Y}_S \downarrow 0$ 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. If A varies over $(0, \infty)$ in the population, then the job separation threshold \underline{Y}_S has support in $(0, \infty)$. As before, under assumptions that ensure that $\underline{Y}_S = \phi(X)V$, the MHT model can be applied to employment duration data to learn about job separations.

As in Section 3.1, 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.

4 Empirical Content

The distribution of $T|X$ cannot be explicitly expressed in terms of the MHT model's primitives, because the distributions of the hitting times $T(y)$ are not explicitly known, except in special cases. It turns out to be more convenient to analyze its Laplace transform,

$$\mathcal{L}_T(s|X) \equiv \mathbb{E} \left[\exp(-sT) \cdot I(T < \infty) \mid X \right], \quad s \in [0, \infty). \quad (3)$$

The factor $I(T < \infty)$ makes explicit that the distribution of $T|X$ may be defective. Note that the defect has probability mass $\Pr(T = \infty|X) = 1 - \mathcal{L}_T(0|X)$. The Laplace transform $\mathcal{L}_T(\cdot|X)$ uniquely characterizes the distribution of $T|X$ (up to almost sure equivalence; see [Feller, 1971](#), Section XIII.1, Theorem 1). It can be given explicitly, using results on the first hitting times of spectrally-negative Lévy processes.

4.1 Characterization

We first characterize the hitting-time process $\{T\} \equiv \{T(y); y \geq 0\}$ implied by $\{Y\}$. This requires a common probabilistic characterization of $\{Y\}$. [Bertoin \(1996, Section VII.1\)](#) shows 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,

$$\psi(s) = \mu s + \frac{\sigma^2}{2} s^2 + \int_{(-\infty, 0)} [e^{sy} - 1 - syI(y > -1)] \Upsilon(dy). \quad (4)$$

Here, $\mu \in \mathbb{R}$ absorbs any linear drift of $\{Y\}$, $\sigma \in [0, \infty)$ is the dispersion parameter of its Brownian motion component; and Υ is the Lévy measure of its jump component, where Υ satisfies $\int \min\{1, y^2\} \Upsilon(dy) < \infty$ and has negative support. The Laplace exponent ψ

of $\{Y\}$ fully characterizes its distributions, through its characteristic function $s \in \mathbb{R} \mapsto \mathbb{E}[\exp\{isY(t)\}] = \exp[\psi(is)t]$.

Equation (4) gives the most common parameterization of ψ . It corresponds to the Lévy-Itô decomposition of $\{Y\}$ in a Brownian motion with linear drift μt , a compound Poisson process with jumps in $(-\infty, -1]$, and a pure-jump martingale with jumps in $(-1, 0)$ (Bertoin, 1996, Section I.1). Alternative parameterizations arise if we decompose the jumps of $\{Y\}$ in small and large shocks in other ways. These parameterizations all have the same dispersion parameter σ and Lévy measure Υ , but have different drift parameters. They may be more convenient when estimating the MHT model (see Section 5). Here, the standard parameterization in (4) suffices.

The Laplace exponent ψ , 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 VII.1 of Bertoin (1996) implies that $\{T\}$ is a *killed subordinator* with Laplace exponent Λ :

$$\mathcal{L}_{T(y)}(s) \equiv \mathbb{E}\left[\exp\{-sT(y)\} \cdot I\{T(y) < \infty\}\right] = \exp[-\Lambda(s)y], \quad s \in [0, \infty). \quad (5)$$

That is, $\{T\}$ is a nondecreasing Lévy process (*subordinator*) with Laplace exponent $\Lambda - \Lambda(0)$, forced to equal ∞ (*killed*) 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 $(\{Y\}, X, V)$. 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 $\Pr[T(y) = \infty] = 1 - \mathcal{L}_{T(y)}(0)$ of defecting movers at this threshold level.

The result that $\{T\}$ is a killed subordinator is intuitive. First, note that $\{Y\}$ can only cross a threshold y after crossing all lower thresholds. Consequently, $\{T\}$ is nondecreasing. Next, recall from Section 2 that the assumption that $\{Y\}$ has no positive jumps ensures that $Y[T(y)] = y$ on $\{T(y) < \infty\}$. Therefore, on $\{T(y) < \infty\}$, $T(y + \Delta) - T(y)$ is the

time it takes $\{Y\}$ to move from y at time $T(y)$ to $y + \Delta$. Because $\{Y\}$ is a strong Markov process (Bertoin, 1996, Proposition I.6), this time is independent of $\{Y(t); 0 \leq t \leq T(y)\}$, and therefore of $\{T(y'); 0 \leq y' \leq y\}$, and has the same distribution as $T(\Delta)$. If $\Lambda(0) = 0$, $T(y) < \infty$ almost surely, and this implies that $\{T\}$ is a Lévy process. Because $\{T\}$ is nondecreasing, it follows that it is, more specifically, a subordinator. If $\Lambda(0) > 0$, then $T(y) = \infty$ with probability $1 - \exp[-\Lambda(0)y]$, and $\{T\}$ is a killed subordinator.

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 $\Lambda(0) = 0$ and $T(y)$ is nondefective. 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 $1 - \exp(2y\mu/\sigma^2)$. 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). Bertoin's (1996) Proposition I.2(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 (Bertoin, 1996, Section III.1). We will come back to this in Section 5.

Now define the Laplace transform $\mathcal{L}_T(\cdot|X, V)$ of the distribution of $T|X, V$ analogously to $\mathcal{L}_T(\cdot|X)$ in (3). Similarly, define \mathcal{L} to be the Laplace transform of the distribution G of V (for expositional simplicity, we suppress the subscript V from \mathcal{L}). From (1) and (5), it follows that $\mathcal{L}_T(s|X, V) = \exp[-\Lambda(s)\phi(X)V]$, so that

$$\mathcal{L}_T(s|X) = \mathcal{L}[\Lambda(s)\phi(X)]; \quad s \in [0, \infty), \quad (6)$$

almost surely.

4.2 Main Identification Result

Equation (6) characterizes the distribution of $T|X$ in terms of the *MHT triplet* $(\Lambda, \phi, \mathcal{L})$. In Section 4’s remainder, we study the identification question whether, conversely, $(\Lambda, \phi, \mathcal{L})$ is uniquely determined from the distribution of $T|X$.

Because there is a one-to-one relation between $(\Lambda, \phi, \mathcal{L})$ and the MHT model’s primitives $(\mu, \sigma^2, \Upsilon)$ and (ϕ, G) , 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, Theorem I.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.

Key to our analysis is an analogy with the identification analysis of the MPH model. To appreciate this, note that the right-hand side of (6) 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)

analyses of this MPH model. Because of possible defects in the MHT model, their analyses do 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) .

Lemma 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 $\tilde{\mathcal{L}}(0) = \mathcal{L}(0)$.*

Lemma 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

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

Lemma 1's proof is given in Appendix B, with the proofs of this section's other results. 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 (6) 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 $\nu : (0, \infty) \rightarrow (0, \infty)$ varies regularly with exponent $\tau \in \mathbb{R}$ at 0 (at ∞) if $\nu(bs)/\nu(s) \rightarrow b^\tau$ as $s \rightarrow 0$ ($s \rightarrow \infty$) for every $b \in (0, \infty)$.

A function that varies regularly with exponent 0 is also said to be *slowly varying*. Any function that has a positive (and finite) limit varies slowly; but slowly varying functions may converge to 0 or diverge, such as $s \mapsto |\ln(s)|$ and $s \mapsto 1/|\ln(s)|$. If ν varies regularly with exponent τ , then $\nu(s) = s^\tau \nu_0(s)$ for some slowly varying function ν_0 . By [Feller \(1971, Section VIII.8, Lemma 2\)](#), a function ν that varies regularly with exponent τ at ∞ (at 0) asymptotically satisfies $s^{\tau-\varepsilon} < \nu(s) < s^{\tau+\varepsilon}$, for any given $\varepsilon > 0$ ($\varepsilon < 0$).

Theorem 1 (Identifiability of the MHT Model). *Let $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ be MHT triplets that imply the same pair of distributions (F_0, F_1) and that satisfy at least one of the following:*

- (R1). $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at 0, with exponents $\tau \in (-1, 0]$ and $\tilde{\tau} \in (-1, 0]$;
- (R2). $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at ∞ , with exponents $\tau \in (-\infty, -1)$ and $\tilde{\tau} \in (-\infty, -1)$;
- (R3). $|\psi'|$ and $|\tilde{\psi}'|$ vary regularly at 0, with exponents $\tau \in (-1, 1]$ and $\tilde{\tau} \in (-1, 1]$; or
- (R4). $|\psi'|$ and $|\tilde{\psi}'|$ vary regularly at ∞ , with exponents $\tau \in [0, 1]$ and $\tilde{\tau} \in [0, 1]$.

Then $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in [1/2, 2]$ and

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

for some $\kappa \in (0, \infty)$.

Theorem 1 establishes identification up to a power transformation, indexed by ρ , and an innocuous normalization, indexed by κ . It is analogous to [Ridder's \(1990\) Theorem 1](#) for the generalized accelerated failure time (GAFT) model, which [Ridder](#) applied to the MPH model. Our analysis deviates in three important ways from [Ridder's](#). First, our proof makes explicit use of the assumption that the MHT triplets satisfy at least

one of (R1)–(R4). [Abbring and Ridder \(2011\)](#) show that [Ridder’s](#) Theorem 1 implicitly requires a similar regularity condition. We discuss (R1)–(R4) in Subsection 4.3. 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 Theorem 1 can be given an appealing stochastic interpretation. First, note that different values of κ correspond to different scale normalizations of the latent process and the threshold. Take the time T implied by (1) for some latent process $\{Y\}$, with inverse Laplace exponent Λ , and some threshold $\beta^X V$, with \mathcal{L} the Laplace transform of the distribution of V . Then, the process $\{\kappa^{-1}Y\}$ and threshold $\beta^X \cdot \kappa^{-1}V$, with $\kappa \in (0, \infty)$, produce the same time T and thus the same distributions F_0 and F_1 . Moreover, as in Theorem 1 with $\rho = 1$, $\{\kappa^{-1}Y\}$ has inverse Laplace exponent $\tilde{\Lambda} \equiv \kappa\Lambda$ and the distribution of $\kappa^{-1}V$ has Laplace transform $\tilde{\mathcal{L}}(s) \equiv \mathcal{L}(\kappa^{-1}s)$.

Next, 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 and $\{T\}$ a hitting-time process characterized by Λ . Then, the process $\{T[S_\rho(y)]; y \geq 0\}$ has Laplace exponent $\tilde{\Lambda} \equiv \Lambda^\rho$ ([Kyprianou, 2006](#), Lemma 2.15). 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 \(1971, Section VIII.9\)](#), Theorem 1, 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 Theorem 1 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 $\tilde{\Lambda} \equiv \kappa\Lambda^\rho$ 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 $\rho \in [1/2, 1]$. Theorem 1 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.

4.3 Tail Conditions

Theorem 1 requires that the MHT triplets satisfy one of (R1)–(R4). From the literature’s perspective, this is not a strong requirement, because it encompasses all approaches to the identification of the MHT model suggested by the MPH literature and includes the canonical special case that $\{Y\}$ is a Brownian motion with drift. This section reviews Theorem 1’s application to these cases.

First, note that (R1)–(R4) only require regular variation of $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ or $|\psi'|$ and $|\tilde{\psi}'|$; the ranges of the exponents of regular variation follow from the properties of the functions involved, and do not constitute additional restrictions, except for the exclusion of the boundary cases that $\tau = -1$ and/or $\tilde{\tau} = -1$ in (R1)–(R3). In particular, the ranges in (R1) and (R2) are determined by the restrictions that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ are decreasing and integrable, and the Lemma in Feller (1971, Section VIII.9). The ranges in (R3) and (R4) follow from the Lévy-Khintchine formula (4) and that same Lemma. In the boundary cases, the tails of F_0 and F_1 are not sufficiently informative on the MHT model’s

primitives and Theorem 1's proof breaks down (Abbring and Ridder, 2011, provide a detailed discussion for the GAFT model that can be adapted to the MHT framework).

Second, each of the assumptions suggested by the MPH literature to achieve identification up to scale implies that one of (R1)–(R4) holds with τ and $\tilde{\tau}$ equal to the same a priori known value. For example, Elbers and Ridder (1982) proved identifiability of the two-sample MPH model under the assumption that the unobserved factor has a finite mean. In the MHT model, the weaker assumption that all triplets $(\Lambda, \beta, \mathcal{L})$ satisfy $\lim_{s \downarrow 0} |\mathcal{L}'(s)| = \mathbb{E}[V \cdot I(V < \infty)] < \infty$ is appropriate, because this allows V to be defective. Because $\Pr(0 < V < \infty) > 0$, this implies that $0 < \lim_{s \downarrow 0} |\mathcal{L}'(s)| = \mathbb{E}[V \cdot I(V < \infty)] < \infty$, so that $|\mathcal{L}'|$ varies slowly at 0. Consequently, the requirement that both $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ in Theorem 1 satisfy this assumption implies (R1) with $\tau = \tilde{\tau} = 0$. In turn, this fixes $\rho = 1$ and yields identification up to scale.

Within the context of the MPH model, the assumption that the unobserved factor has a finite mean is an arbitrary normalization with substantive effects on the model interpretation (Ridder, 1990). The corresponding assumption on the MHT model in some cases follows naturally from optimal stopping models in which threshold heterogeneity is deduced from unbounded primitive unobserved heterogeneity. For example, in Section 3.2's model of unemployment durations with a given job entry cost $\bar{K} < \infty$, workers will enter employment at a finite wage, even if they would have to work at that wage forever: $\underline{Y}_D < \infty$ even if $\underline{K} \rightarrow \infty$. Consequently, given $\bar{K} < \infty$, unbounded heterogeneity in \underline{K} leads to bounded threshold heterogeneity. Similarly, in Section 3.3's model of job tenure with given flow utility $B > 0$ from unemployment, the job separation threshold \underline{Y}_S converges to a finite upper limit as job search becomes impossible ($A \rightarrow 0$). Thus, given $B > 0$, unbounded heterogeneity in search frictions implies bounded threshold heterogeneity. Both examples imply that, if the threshold is specified as $\phi(X)V$, V is bounded and therefore has a finite mean.

Following Heckman and Singer (1984a), we could alternatively fix $\rho = 1$ in Theorem 1

by assuming that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at 0 with the same exponent $\tau = \tilde{\tau} \in (-1, 0)$. We could also 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.

More recently, [Ridder and Woutersen \(2003\)](#) obtained identification of the MPH model by assuming that the baseline hazard is bounded away from zero and infinity near time 0. The analogous assumption on the MHT model without defecting movers; that is, with $\Lambda(0) = \tilde{\Lambda}(0) = 0$; requires that $\lim_{s \downarrow 0} \Lambda'(s) < \infty$ and $\lim_{s \downarrow 0} \tilde{\Lambda}'(s) < \infty$. By the inverse function theorem, and the convexity of ψ and $\tilde{\psi}$, this implies that $0 < \lim_{s \downarrow 0} \psi'(s) < \infty$ and $0 < \lim_{s \downarrow 0} \tilde{\psi}'(s) < \infty$, so that $|\psi'|$ and $|\tilde{\psi}'|$ vary slowly at 0. Consequently, [\(R3\)](#) is satisfied with $\tau = \tilde{\tau} = 0$, and [Theorem 1](#)'s result holds with $\rho = 1$. Trivially, this result directly extends to the general case in which movers may defect and $\Lambda(0) > 0$ and $\tilde{\Lambda}(0) > 0$ if, instead of making an assumption on the Laplace exponent of $\{T\}$, we directly require that $0 < \lim_{s \downarrow 0} |\psi'(s)| < \infty$ and $0 < \lim_{s \downarrow 0} |\tilde{\psi}'(s)| < \infty$.

The condition that $0 < \lim_{s \downarrow 0} |\psi'(s)| < \infty$ can be related to the long run behavior of the latent process $\{Y\}$: It 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 3.3](#)'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.

Finally, both [\(R3\)](#) and [\(R4\)](#) are satisfied in the canonical example that $\{Y\}$ is a Brownian motion with drift parameter $\mu \in \mathbb{R}$ and dispersion parameter $\sigma \in [0, \infty)$. In this case, $\psi'(s) = \mu + \sigma^2 s$, and $\sigma > 0$ if $\mu \leq 0$, so that $|\psi'|$ varies regularly at both 0, with exponent 1 if $\mu = 0$ and 0 otherwise, and ∞ , with exponent 1 if $\sigma > 0$ and 0 if $\sigma = 0$.

Consequently, if both ψ and $\tilde{\psi}$ are Laplace exponents of Brownian motion with drift, (R3) holds with $\tau \in \{0, 1\}$ and $\tilde{\tau} \in \{0, 1\}$ and (R4) holds with (possibly different) $\tau \in \{0, 1\}$ and $\tilde{\tau} \in \{0, 1\}$. Either way, Theorem 1's conclusion follows with $\rho \in \{1/2, 1, 2\}$. 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. Identification up to scale can be achieved by requiring either $\sigma > 0$ or $\mu > 0$.

4.4 Censoring and Competing Risks

The identification analysis so far assumes that F_1 and F_2 are known. In practice, duration data are often censored. With independent censoring (Andersen et al., 1993, Section II.1), F_1 and F_2 are 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 F_1 and F_2 . 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 time $c \in (0, \infty)$, only the restriction of G to $[0, c]$ can possibly be identified. Nevertheless, the specific structure implied by the Lévy assumption suggests that, subject to such support qualifications, results like Theorem 1 can be derived under independent right-censoring. We do not further explore this here.

The analysis of the independent competing-risks model is similar to that for independently censored data. Consider two durations $T^1 \equiv \inf\{t \geq 0 : Y^1(t) > \beta_1^X V^1\}$ and $T^2 \equiv \inf\{t \geq 0 : Y^2(t) > \beta_2^X V^2\}$, with $\{Y^1\}$ and $\{Y^2\}$ spectrally-negative Lévy processes, V^1 and V^2 positive random variables, β_1 and β_2 positive scalars, and $\mathcal{X} = \{0, 1\}$. Suppose that only the *identified minimum* of T^1 and T^2 , $(\min_j T^j, \arg \min_j T^j)$, is observed. As-

sume that $\{Y^1\}$, $\{Y^2\}$, V^1 , V^2 , and X are mutually independent so that, conditional on X , T^1 and T^2 are independent with distributions given by two-sample MHT models. Let T^1 and T^2 have unbounded supports; a sufficient condition for this is that $\{Y^1\}$ and $\{Y^2\}$ have nondegenerate Brownian motion components. Then, the distributions of $T^1|X$ and $T^2|X$ are uniquely determined from the distribution of $(\min_j T^j, \arg \min_j T^j)|X$ (Cox, 1962). Thus, Theorem 1 can be applied to the identification of the MHT triplets characterizing both distributions. However, if one or both durations have bounded support, then we face an identification problem similar to that with bounded censoring times.

4.5 Covariates and the Latent Process

Because the increments of the latent Lévy process are independent of its history, in particular its initial condition, an alternative model that specifies the initial condition $Y(0)$ to be heterogeneous, say equal to $-\phi(X)V$, and fixes the threshold at a common value of zero generates the same durations T . Similarly, we can redistribute a linear drift μt from $\{Y\}$ to the threshold, or rescale both with the same positive function of (X, V) , without changing the implications for T . Thus, a choice between these alternative specifications cannot be based on the observed distribution of $T|X$, and should be motivated by application-specific substantial considerations. Section 3's structural examples illustrate this.

With exclusion restrictions, it is possible to distinguish between the effects of covariates on the latent process and their effects on the threshold. To illustrate this, we augment Section 4.2's two-sample framework with regressors that may affect both the latent process and the threshold. To this end, suppose that $X \equiv (X^\phi, X^\Lambda)$, with X^ϕ binary and X^Λ taking values in $\mathcal{X}^\Lambda \subseteq \mathbb{R}^{k-1}$, so that $\mathcal{X} \equiv \{0, 1\} \times \mathcal{X}^\Lambda$. The threshold now depends on both X^ϕ and X^Λ , through $\phi(X)$. We assume that $\phi(0, X^\Lambda) \neq \phi(1, X^\Lambda)$ with positive probability. In addition, $\{Y\}$ may depend on X^Λ but, conditional on X^Λ , not on X^ϕ . To this end, we specify its parameters (μ, σ, Υ) as functions of X^Λ , so that the Laplace

exponent $\Lambda(\cdot, X^\Lambda)$ of $\{T\}$ depends on X^Λ .

We ensure that X^Λ cannot simply enter a scale factor in the Laplace exponent, by requiring that $\tilde{\Lambda}(s, X^\Lambda)/\Lambda(s, X^\Lambda)$ varies with s with positive probability whenever $\tilde{\Lambda}(\cdot, X^\Lambda) \neq \Lambda(\cdot, X^\Lambda)$ with positive probability. In the Gaussian case, for example, this condition is satisfied if we allow the drift parameter to depend on X^Λ , but take the dispersion parameter to be a positive scalar. It would be violated if we would allow the dispersion parameter to depend on X^Λ as well.

Take two MHT triplets $(\Lambda, \beta, \mathcal{L})$ and $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}})$ that imply the same distribution of $T|X$ almost surely. Assume that $\mathbb{E}[V \cdot I(V < \infty)] < \infty$ and normalize $\mathbb{E}[V \cdot I(V < \infty)] = 1$. Then, Theorem 1, applied conditional on X^Λ and on $\{\phi(0, X^\Lambda) \neq \phi(1, X^\Lambda)\}$, implies that $\tilde{\mathcal{L}} = \mathcal{L}$. In turn, this ensures that $\tilde{\phi}(X)\tilde{\Lambda}(\cdot, X^\Lambda) = \phi(X)\Lambda(\cdot, X^\Lambda)$ almost surely. Because this implies that $\tilde{\Lambda}(s, X^\Lambda)/\Lambda(s, X^\Lambda) = \phi(X)/\tilde{\phi}(X)$ almost surely does not vary with s , $\Lambda(\cdot, X^\Lambda) = \tilde{\Lambda}(\cdot, X^\Lambda)$ almost surely. It follows that $\phi(X) = \tilde{\phi}(X)$ almost surely.

4.6 Stratified Data

We can allow for general dependence of the latent process and the unobserved heterogeneity on the covariates if 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 $T^1 \equiv \inf\{t \geq 0 : Y^1(t) > V\}$ and $T^2 \equiv \inf\{t \geq 0 : Y^2(t) > V\}$, with $\{Y^1\}$ and $\{Y^2\}$ independent spectrally-negative Lévy processes; and V a positive 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, [Lemma 1](#) does not apply here. So, the following result is of independent value.

Theorem 2 (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 [Subsection 4.3's](#) results 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 .

5 Estimation

So far, we have ignored sampling variation. This section briefly discusses estimation of an MHT model, based on its characterization in [Section 4.1](#), and standard moment and likelihood methods.

5.1 Parameterization

Let Λ , ϕ and \mathcal{L} be specified up to a finite vector of unknown parameters $\alpha \in \mathcal{A}$. We assume that this parameterization is one-to-one and implements one of Section 4.3's identifying conditions and two scale normalizations on $(\Lambda, \phi, \mathcal{L})$. Then, Theorem 1 applies with $\kappa = \rho = 1$, and α is uniquely determined by the distribution of $T|X$. We also require that the parameterization is sufficiently smooth to allow for the application of standard asymptotic theory.

In the two-sample specification $\phi(x) = \beta^x$, we need that $\mathcal{X} = \{0, 1\}$. In general, we can specify $\phi(x) = \exp(x'\beta)$ and make assumptions on the support \mathcal{X} of X that ensure that β is uniquely determined from $x \in \mathcal{X} \mapsto \exp(x'\beta)$. With continuous covariates, for example, we could assume that \mathcal{X} contains a nonempty open set in \mathbb{R}^k .

The Lévy-Khintchine formula can be used to specify ψ ; Λ then follows by inversion. This ensures that Λ satisfies the model's restrictions (see Section 4.1). We can generate a smooth parameterization of Λ by using a version of the Lévy-Khintchine formula that, unlike (4), employs a gradual distinction between small and large shocks. As we noted in Section 4.1, this only affects the interpretation of the drift parameter μ . The 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 , and thus \mathcal{L} , 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).

5.2 Sampling

For expositional convenience, we focus on a simple type of independent right-censoring (Andersen et al., 1993). Let $\{(T_1^*, X_1), \dots, (T_n^*, X_n)\}$ be a (complete) random sample from the distribution of (T, X) induced by the MHT model at the “true” parameter vector $\alpha_0 \in \mathcal{A}$ and some marginal distribution of X . We do not directly observe this complete sample, but only a censored version of it: $\{(T_1, D_1, X_1), \dots, (T_n, D_n, X_n)\}$. Here, $T_i \equiv \min\{T_i^*, C_i\}$ is the observed duration and $D_i \equiv I(T_i^* \leq C_i)$ a censoring indicator, for some random censoring time C_i ; $i = 1, \dots, n$. We assume that the complete observations (T_i^*, C_i, X_i) are independent across i and that, conditional on X_i , C_i is independent of T_i^* . That is, censoring times are not informative on the durations of interest.

We take the marginal distributions of the (C_i, X_i) to be ancillary for α , and focus on estimating α_0 using the conditional moment restrictions and likelihood implied by the MHT model for $T|X$.

5.3 Generalized Method of Moments

Section 4.1’s characterization of the distribution of $T|X$ in terms of its Laplace transform provides a continuum of conditional moment conditions, one for each point s at which the Laplace transform can be evaluated. This suggests a generalized method of moments (GMM) estimator.

Define $h(t, x; s, \alpha) \equiv \exp(-st)I(t < \infty) - \mathcal{L}[\Lambda(s)\phi(x)]$. Then, it follows from (6) that $\mathbb{E}[h(T, X; s, \alpha_0)|X] = 0$ almost surely, for all $s \in (0, \infty)$. In our estimation procedure, we will specify an $(m \times 1)$ -vector Z of instruments based on X , and use the unconditional

moment conditions

$$\mathbb{E}[h(T, X; s, \alpha_0)Z] = 0, \quad s \in [0, \infty). \quad (7)$$

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 (7) uniquely determines α_0 .

Suppose, for now, that there is no censoring, so that $T_i = T_i^*$ for all i . 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 (7) with

$$h_n(s, \alpha) \equiv n^{-1} \sum_{i=1}^n h(T_i, X_i; s, \alpha)Z_i. \quad (8)$$

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 $\int_0^\infty h_n(s, \alpha)'Q_n h_n(s, \alpha)q_n(ds)$. 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 weighs the various moments corresponding to the m instruments, with weights independent of s . Examples include the $m \times m$ identity matrix and $(n^{-1} \sum_{i=1}^n Z_i Z_i')^{-1}$. The random probability measure q_n weighs the various moment conditions corresponding to the evaluation points s of the Laplace transform, identically across the instruments in Z_i . It has support in $[0, \infty)$ and converges to a nonrandom measure q . It could be finitely discrete, and selecting only a finite number of Laplace evaluation points, or absolutely continuous. Examples of the latter include $q_n(s) = \exp(-\varpi_n s)$ for either a fixed or a 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. More-

over, Carrasco and Florens's (2002) method for efficient estimation based on empirical characteristic functions can be adapted to produce an GMM estimator of the MHT model that efficiently weighs 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. A full development is left for Abbring and Salimans (2011a).

In the two-sample case or, more generally, the case that the support \mathcal{X} of X is finite, the GMM estimator can be readily adapted to allow for independent censoring, by non-parametrically correcting the empirical moments in (8) for censoring. To this end, first estimate the distribution of T in each sample using the Nelson-Aalen estimator or, in the special case of simple random right-censoring, the Kaplan-Meier estimator (see e.g. Andersen et al., 1993, Section IV.1). Then, compute the empirical analogue of the moment condition (7) using these nonparametric estimators of the distribution of T ; instead of the empirical distribution function, as in (8). Provided that the censoring mechanism is such that the distribution of $T|X$ 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 (7). In these cases, likelihood-based methods are a convenient alternative to GMM estimation.

5.4 Likelihood-Based Methods

The log (conditional) likelihood $\ln L_n(\alpha)$ of α for $(T_1, \dots, T_n) | \{(D_1, X_1), \dots, (D_n, X_n)\}$ is given by

$$\ln L_n(\alpha) = \sum_{i=1}^n \ln \int \theta(T_i | X_i, v)^{D_i} \bar{F}(T_i | X_i, v) dG(v), \quad (9)$$

with $\theta(\cdot|X, V)$ and $\bar{F}(\cdot|X, V)$ the hazard rate and survival function of $T|X, V$. Here, the dependence of θ and \bar{F} (through Λ and ϕ) and G on the parameter vector α is kept implicit.

Under standard regularity conditions, the maximizer $\hat{\alpha}_n$ of $\ln L_n(\alpha)$ is a consistent and asymptotically normal estimator of α_0 . The estimator's asymptotic covariance matrix can be estimated in the standard way using either the score or Hessian characterization of the Fisher information matrix. It is asymptotically efficient under the assumption that the marginal distribution of X and the censoring times carry no information on α_0 .

In the Gaussian special case; $T|X, V$ has an inverse Gaussian distribution, and we have explicit expressions for $\theta(T_i|X_i, v)$ and $\bar{F}(T_i|X_i, v)$ in (9). In the general case with shocks, such explicit expressions are not available, and the likelihood cannot be computed directly. [Abbring and Salimans \(2011b\)](#) however show that the log likelihood can, in general, be efficiently computed with numerical methods for inverting Laplace transforms that exploit special properties of the first hitting times of Lévy processes. This enables the computation of the maximum likelihood estimator and facilitates other likelihood-based methods.

6 Extensions

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

6.1 Time-Varying Covariates

Following most of the duration-model identification literature, we have ignored time-varying covariates. 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.

Specifically, time-varying covariates can be introduced in the MHT model as determinants of a time-varying threshold or, following Section 4.5’s extension, the latent process $\{Y\}$. For example, in a structural model, time-varying covariates may directly shift the drift parameter of the latent process and indirectly affect the threshold through the agent’s behavioral response. If the support of the covariate process includes constant sample paths then, under some regularity conditions, this paper’s identification results can be applied.

A drawback of a model with time-varying parameters is that it is hard to characterize its hitting-time process. As a consequence, we cannot directly exploit the time-varying covariates to derive more powerful identification results, as in the MPH literature. Moreover, the model may be hard to implement empirically. 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.

6.2 Nonstationary Increments

[Aalen and Gjessing \(2001\)](#) show that hitting-time models based on Brownian motions with drift towards the threshold 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 $T \equiv \xi(T^*)$, for an increasing time transformation $\xi : [0, \infty] \rightarrow [0, \infty]$ and the distribution of $T^*|X$ 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 $\xi(T^*)$, with

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

Here, T^* is 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 $\Pr(T^* \geq \eta^{-1})$ 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 T^* by a Gaussian MHT model for the distribution of $T^*|X$. The resulting model is a simple, one-parameter extension of the MHT model that allows for nonstationary increments.

6.3 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 $dY(t) = -\varrho Y(t)dt + dY^*(t)$, with $\varrho \in [0, \infty)$ and $\{Y^*\}$ a Lévy process. The usual Ornstein-Uhlenbeck process arises if $\{Y^*\}$ 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 an 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.

7 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 the MHT model with $Y(t) = \mu t$, and V distributed as a mixture of exponentials: $\Pr(V > v) = \int_0^\infty \exp(-wv)dG^*(w)$, 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 3.

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 decisions that are taken at Poisson times, such as sequential job search or insurance claim decisions. In some cases, hazards may also be the easiest way to specify the hitting-time process implied by a latent process that jumps across the threshold; see e.g. Section 3.2's discussion of employment durations. The fact that the resulting hazard rates are usually not multiplicative in the effects of elapsed duration and those of heterogeneity (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 this paper's MHT model. The fact that Section 3.3'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 the MHT model with $Y(t) = \mu t$, 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 examples with $Y(t) = \mu t$ are both special cases of this standard model from statistics. As discussed in Section 2.3, the AFT model can be interpreted as a boundary 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 an optimal stopping problem's state and a threshold rule that naturally follows as its optimal decision rule, 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 6.1 proposes that we either extend the MHT model by including time-varying covariates in its primitives or respect its basic structure 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, Theorem VII.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 Lemma 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

$$\mathcal{L} \circ (\beta\mathcal{L}^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}\tilde{\mathcal{L}}^{-1}) \quad \text{on } (0, \mathcal{L}(0)). \quad (10)$$

Note that both sides of (10) map $(0, \mathcal{L}(0))$ into itself. Thus, we can compose each side l times with itself, and find that

$$\mathcal{L} \circ (\beta^l\mathcal{L}^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}^l\tilde{\mathcal{L}}^{-1}) \quad \text{on } (0, \mathcal{L}(0)), \quad l \in \mathbb{N}. \quad (11)$$

Because \mathcal{L} and $\tilde{\mathcal{L}}$ are continuous at 0; evaluating both sides of (11) at a fixed $s \in (0, \mathcal{L}(0))$, and letting $l \rightarrow \infty$, gives $\mathcal{L}(0) = \tilde{\mathcal{L}}(0)$. \square

Proof of Theorem 1. 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 $\tilde{\beta} < 1$.

(R1). Suppose that $|\mathcal{L}'|$ and $|\tilde{\mathcal{L}}'|$ vary regularly at 0, with exponents $\tau, \tilde{\tau} \in (-1, 0]$.

By Lemma 1, $\mathcal{L}(0) = \tilde{\mathcal{L}}(0)$. With (11), this implies that $\mathcal{L}(0) - \mathcal{L} \circ (\beta^l\mathcal{L}^{-1}) = \tilde{\mathcal{L}}(0) - \tilde{\mathcal{L}} \circ (\tilde{\beta}^l\tilde{\mathcal{L}}^{-1})$ on $(0, \mathcal{L}(0))$, $l \in \mathbb{N}$. Taking logs, and then derivatives, yields

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

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

$$\frac{\tilde{\mathcal{K}}'(s)/\tilde{\mathcal{K}}(s)}{\mathcal{K}'(s)/\mathcal{K}(s)} = \frac{\beta^l\mathcal{K}(s) |\mathcal{L}'[\beta^l\mathcal{K}(s)]| / \{\mathcal{L}(0) - \mathcal{L}[\beta^l\mathcal{K}(s)]\}}{\tilde{\beta}^l\tilde{\mathcal{K}}(s) |\tilde{\mathcal{L}}'[\tilde{\beta}^l\tilde{\mathcal{K}}(s)]| / \{\tilde{\mathcal{L}}(0) - \tilde{\mathcal{L}}[\tilde{\beta}^l\tilde{\mathcal{K}}(s)]\}}, \quad (12)$$

for $s \in (0, \mathcal{L}(0))$ and $l \in \mathbb{N}$. By [Feller \(1971, Section VIII.9, Theorem 1\(a\)\)](#); the numerator in the right-hand side of (12) 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 $l \rightarrow \infty$. 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 definitions of \mathcal{K} and $\tilde{\mathcal{K}}$, 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 $\mathcal{L} \circ (\beta\Lambda) = \tilde{\mathcal{L}} \circ (\tilde{\beta}\tilde{\Lambda})$, we find that $\tilde{\beta} = \beta^\rho$.

(R2). 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 [Lemma 1's](#) proof, this gives $\mathcal{L} \circ (\beta^{-l}\mathcal{L}^{-1}) = \tilde{\mathcal{L}} \circ (\tilde{\beta}^{-l}\tilde{\mathcal{L}}^{-1})$ on $(0, \mathcal{L}(0))$, $l \in \mathbb{N}$. Taking logs and derivatives, and rearranging, yields

$$\frac{\tilde{\mathcal{K}}'(s)/\tilde{\mathcal{K}}(s)}{\mathcal{K}'(s)/\mathcal{K}(s)} = \frac{\beta^{-l}\mathcal{K}(s) |\mathcal{L}'[\beta^{-l}\mathcal{K}(s)]| / \{\mathcal{L}[\beta^{-l}\mathcal{K}(s)]\}}{\tilde{\beta}^{-l}\tilde{\mathcal{K}}(s) |\tilde{\mathcal{L}}'[\tilde{\beta}^{-l}\tilde{\mathcal{K}}(s)]| / \{\tilde{\mathcal{L}}[\tilde{\beta}^{-l}\tilde{\mathcal{K}}(s)]\}}, \quad (13)$$

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 (13) 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 $l \rightarrow \infty$. As in [Case \(R1\)](#), this gives $\tilde{\mathcal{L}}(\kappa s^\rho) = \mathcal{L}(s)$ for all s , $\tilde{\Lambda} = \kappa\Lambda^\rho$, and $\tilde{\beta} = \beta^\rho$.

(R3). 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 $\psi(0) = \psi[\Lambda(0)] = 0$ 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 at some $\Lambda_{min} \in [0, \Lambda(0)]$. Denote this minimum with $s_{min} \in (-\infty, 0]$. Note that the inverse of the restriction of ψ to $[\Lambda_{min}, \infty) \supseteq [\Lambda(0), \infty)$ exists. Extend Λ from $[0, \infty)$ to $[s_{min}, \infty)$ so that it equals

this inverse. Similarly, denote the unique minimum of $\tilde{\psi}$ with $\tilde{s}_{min} \in (-\infty, 0]$ and extend $\tilde{\Lambda}$ to $[\tilde{s}_{min}, \infty)$. Without loss of generality, suppose that $s_{min} \geq \tilde{s}_{min}$. Because ψ and $\tilde{\psi}$ are real analytic (Bertoin, 1996, Section VII.1), Λ and $\tilde{\Lambda}$ are real analytic on (s_{min}, ∞) 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_{min}, ∞) . With $\underline{s} < \infty$, using analytic extension, this implies that

$$\psi \circ (\beta\Lambda) = \tilde{\psi} \circ (\tilde{\beta}\tilde{\Lambda}) \quad \text{on } (s_{min}, \infty). \quad (14)$$

Note that both sides of (14) map (s_{min}, ∞) into itself. Thus, we can compose each side l times with itself, which gives $\psi \circ (\beta^l\Lambda) = \tilde{\psi} \circ (\tilde{\beta}^l\tilde{\Lambda})$ on (s_{min}, ∞) , $l \in \mathbb{N}$. Applying calculations that parallel those for Cases (R1) and (R2), we find that

$$\frac{\tilde{\Lambda}'(s)/\tilde{\Lambda}(s)}{\Lambda'(s)/\Lambda(s)} = \frac{\beta^l\Lambda(s) \left| \psi' [\beta^l\Lambda(s)] \right| / \psi [\beta^l\Lambda(s)]}{\tilde{\beta}^l\tilde{\Lambda}(s) \left| \tilde{\psi}' [\tilde{\beta}^l\tilde{\Lambda}(s)] \right| / \tilde{\psi} [\tilde{\beta}^l\tilde{\Lambda}(s)]}, \quad (15)$$

for all $s \in (0, \infty) \subseteq (s_{min}, \infty)$ and $l \in \mathbb{N}$. Feller (1971, Section VIII.9, Theorem 1(a)) implies that the right-hand side of (15) converges to $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in (0, \infty)$; for each given $s \in (0, \infty)$, as $l \rightarrow \infty$. 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 $\tilde{\beta} = \beta^\rho$.

(R4). 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 (R3), this can be used to show that (15) extends from $l \in \mathbb{N}$ to all $l \in \mathbb{Z}$. Feller (1971, Section VIII.9, Theorem 1(b)) implies that the right-hand side of (15) converges to $\rho \equiv (\tau + 1)/(\tilde{\tau} + 1) \in [1/2, 2]$; for each given $s \in (0, \infty)$, as $l \rightarrow -\infty$. Consequently, the conclusion of Case (R3) 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 (R4)'s tighter bound on ρ holds generally. To this end, note that both ψ and $\tilde{\psi}$ should satisfy the Lévy-Khintchine formula (4). Because ψ is convex and $\psi(s) \rightarrow \infty$ as $s \rightarrow \infty$ (Bertoin, 1996, Section VII.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, Proposition I.2). Obviously, the same asymptotic behavior is displayed by $\tilde{\psi}$. From $\psi[\Lambda(s)] = s = \tilde{\psi}[\tilde{\Lambda}(s)]$, it follows that $\tilde{\psi}(s) = \psi[(s/\kappa)^{1/\rho}]$, $s \in [\tilde{\Lambda}(0), \infty)$. 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 Theorem 2. 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,

$$\tilde{\Lambda}_j - \tilde{\Lambda}_j(0) = \kappa [\Lambda_j - \Lambda_j(0)]; \quad j = 1, 2; \quad (16)$$

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 (16) 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

$$\Lambda_j^{-1} \circ \Lambda_{12} = (\mathcal{L}^j)^{-1} \circ \mathcal{L}^{12} = \tilde{\Lambda}_j^{-1} \circ \tilde{\Lambda}_{12} \quad \text{on } [0, \infty); \quad j = 1, 2. \quad (17)$$

Substituting (17) in (16) gives

$$\tilde{\Lambda}_{12} - \tilde{\Lambda}_j(0) = \kappa [\Lambda_{12} - \Lambda_j(0)]; \quad j = 1, 2. \quad (18)$$

Moreover, (16) implies that

$$\tilde{\Lambda}_{12} - \tilde{\Lambda}_1(0) - \tilde{\Lambda}_2(0) = \kappa [\Lambda_{12} - \Lambda_1(0) - \Lambda_2(0)]. \quad (19)$$

Together, (18) and (19) imply that $\tilde{\Lambda}_j(0) = \kappa\Lambda_j(0)$; $j = 1, 2$. With (16), 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](#), [36](#)
- Aalen, O. O. and H. K. Gjessing (2004). Survival models based on the Ornstein-Uhlenbeck process. *Lifetime Data Analysis* 10, 407–423. [37](#)
- Abbring, J. H. (2002). Stayers versus defecting movers: A note on the identification of defective duration models. *Economics Letters* 74, 327–331. [6](#), [21](#)
- Abbring, J. H. (2010). Identification of dynamic discrete choice models. *Annual Review of Economics* 2, 367–394. [3](#)
- Abbring, J. H. and J. R. Campbell (2005). A firm’s first year. Discussion Paper 05-046/3, Tinbergen Institute, Amsterdam. [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 G. Ridder (2011). Regular variation and the identification of generalized accelerated failure-time models. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. [23](#), [25](#)
- Abbring, J. H. and T. Salimans (2011a). Estimating mixed hitting-time models. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. [34](#)
- Abbring, J. H. and T. Salimans (2011b). The likelihood of mixed hitting times. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. [35](#)
- Abbring, J. H. and G. J. Van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94, 87–99. [32](#)

- Alvarez, F. and R. Shimer (2011). Search and rest unemployment. *Econometrica* 79(1), 75–122. [3](#)
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag. [27](#), [32](#), [34](#)
- Bertoin, J. (1996). *Lévy Processes*. Number 121 in Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press. [4](#), [5](#), [10](#), [17](#), [18](#), [19](#), [20](#), [41](#), [44](#), [45](#)
- Boyarchenko, S. and S. Levendorskiĭ (2007). *Irreversible Decisions under Uncertainty: Optimal Stopping Made Easy*. Berlin: Springer-Verlag. [1](#), [9](#), [16](#)
- Carrasco, M. and J.-P. Florens (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16, 797–834. [33](#)
- Carrasco, M. and J.-P. Florens (2002). Efficient GMM estimation using the empirical characteristic function. Mimeo, Department of Economics, University of Rochester. [34](#)
- Cox, D. R. (1962). *Renewal Theory*. London: Methuen. [28](#)
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B* 34, 187–202. [7](#)
- Cox, D. R. and H. D. Miller (1965). *The Theory of Stochastic Processes*. London: Methuen. [6](#)
- Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy* 97(3), 620–638. [9](#), [13](#), [14](#)
- Dixit, A. K. and R. S. Pindyck (1994). *Investment under Uncertainty*. Princeton University Press. [1](#), [9](#), [10](#)
- 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. [20](#), [25](#), [35](#)
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (Second

- ed.), Volume II. New York: Wiley. [5](#), [16](#), [17](#), [20](#), [21](#), [22](#), [23](#), [24](#), [31](#), [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. [13](#)
- 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. [23](#), [25](#)
- 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. [31](#), [32](#)
- 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. [35](#)
- Honoré, B. E. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies* *60*, 241–246. [30](#), [45](#)
- 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](#), [9](#), [23](#)
- 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. [11](#)
- 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](#), [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](#)
- McDonald, R. and D. Siegel (1986). The value of waiting to invest. *Quarterly Journal of Economics* 101(4), 707–728. [9](#)
- 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. [8](#)
- Mordecki, E. (2002). Optimal stopping and perpetual options for Lévy processes. *Finance and Stochastics* 6(4), 473–493. [10](#)
- 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. [14](#), [16](#)
- 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. [37](#), [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. [11](#), [12](#)
- Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57, 167–182. [20](#), [22](#), [23](#), [25](#)
- Ridder, G. and T. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica* 71, 1579–1589. [26](#)
- 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](#), [9](#), [16](#)
- 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](#)

Whitmore, G. A. (1979). An inverse Gaussian model for labour turnover. *Journal of the Royal Statistical Society Series A* 142(4), 468–478. [3](#)

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

[41](#)