Mixed Hitting-Time Models*

Jaap H. Abbring†

First draft, July 2007

Preliminary second draft, April 7, 2009

Abstract

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

*I thank seminar participants at Tinbergen Institute, Toulouse, PSE, Chicago, Northwestern, Tilburg, IZA Bonn, and RWI Essen; and participants in the Cowles 75th Anniversary Conference at Yale and the Econometric Society European Meeting in Milan, and in particular Andrew Chesher, for helpful discussion.

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

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

JEL codes: C14, C41.

1 Introduction

Mixed hitting-time (MHT) models are mixture duration models that specify durations as the first time a latent stochastic process crosses a heterogeneous threshold. Such models are of substantial interest because they can be reduced from optimal-stopping models with heterogeneous agents that do not naturally lead to a Cox (1972)–Lancaster (1979) mixed proportional hazards (MPH) structure. In this paper, we explore the empirical content of an MHT model in which the latent process is a spectrally-negative Lévy process, a continuous-time process with stationary and independent increments and no positive jumps, and the threshold is proportional in the effects of observed regressors and unobserved heterogeneity. We show that existing strategies for analyzing the identifiability of the MPH model can be adapted to prove this model's identifiability. In particular, we show that the latent Lévy process, the regressor effect on the threshold, and the distribution of the unobserved heterogeneity in the threshold are uniquely determined by data on durations and regressors. Some assumptions on the long-run behavior of the latent process are 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.

Continuous-time models involving latent processes crossing thresholds are common in econometrics. They arise naturally from economic models in which heterogeneous agents choose optimally from a discrete set of alternatives (see Dixit and Pindyck, 1994; Kyprianou, 2006; Boyarchenko and Levendorskiĭ, 2007; Stokey, 2009, for reviews). Jovanovic's (1979; 1984) model of job tenure is an early example in labor economics and Alvarez and Shimer's (2007) model of search and rest unemployment is a recent one. In his classic text book on econometric duration analysis, Lancaster (1990, Sections 3.4.2, 5.7 and 6.5) reviews a canonical special case of our model, a reduced-form marginal duration model that specifies durations as the first-passage times of a Brownian motion with drift, and relates it to Jovanovic's job tenure model. In Lancaster (1972), he applies this model to strike durations, interpreting the gap between the Brownian motion and the threshold as the level of disagreement. Other applications include marriage and divorce, firm entry and exit, and lumpy investment.

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

Our analysis also complements the literature on discrete-time discrete choice models pioneered by Heckman (1981a,c). In particular, Heckman and Navarro (2007) discuss a general discrete-time mixture duration model based on a latent process crossing thresholds (see Abbring and Heckman, 2007, 2008, for reviews). They emphasize the distinction between this model and a discrete-time MPH model and its extensions, and study its identifiability and its relation to dynamic discrete choice. This paper complements theirs with an analysis in continuous time. The 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.

The paper is organized as follows. Section 2 introduces the MHT model. Section 3 develops the paper's main ideas for the well-understood, and therefore instructive, special case in which the latent Lévy process is a Brownian motion with drift. In particular, the MHT model structure is explored, and the key connection between the analysis of its empirical content and the MPH identification literature is highlighted. Section 4 presents the general MHT model's implications for the data and the main identification results. Section 5 briefly discusses estimation from complete and censored data. Section 6 presents some simple examples of economic models that can be analyzed using the MHT model. Section 7 suggests three extensions. Finally, Section 8 concludes with a discussion of the relative merits of the MHT and MPH models as complementary frameworks for econometric duration analysis.

2 The Model

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

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

Definition 1. *A* Lévy process *is a stochastic process* $\{Y\}$ *such that the increment* $Y(t+\Delta) - Y(t)$ *is independent of* $\{Y(\tau); 0 \leq \tau \leq t\}$ *and has the same distribution as* $Y(\Delta)$*, for every* $t, \Delta \geq 0$*.*

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

An important example of a Lévy process is the scalar Brownian motion with drift, in which case $Y(\Delta)$ is normally distributed with mean $\mu\Delta$ and variance $\sigma^2\Delta$, for some scalar parameters $\mu \in \mathsf{R}$ and $\sigma \in [0, \infty)$. Brownian motion is the single Lévy process with continuous sample paths. In general, Lévy processes may have jumps. The jump process $\{\Delta Y\}$ of a Lévy process $\{Y\}$ is a Poisson point process with characteristic measure Υ such that $\int \min\{1, x^2\}\Upsilon(dx) < \infty$, and any Lévy process $\{Y\}$ can be written as the sum of a Brownian motion with drift and an independent pure-jump process with jumps governed by such a point process (Bertoin, 1996, Chapter I, Theorem 1). The characteristic measure of $\{Y\}$'s jump process is called its ***Levy measure*** and, together with the drift and variance parameters of its Brownian motion component, fully characterizes $\{Y\}$'s distributional properties. Key examples of pure-jump Lévy processes are compound Poisson processes, which have independently and identically distributed jumps at Poisson times. In fact, in distribution, each Lévy process can be approximated arbitrary closely by a sequence of compound Poisson processes (Feller, 1971, Section IX.5, Theorem 2).

Let $T(y)$ denote the first time that the Lévy process $\{Y\}$ exceeds a threshold $y \in [0, \infty)$,

$$T(y) \equiv \inf\{t \geq 0 : Y(t) > y\}.$$

Here, we use the convention that $\inf \emptyset \equiv \infty$; that is, we set $T(y) = \infty$ if $\{Y\}$ never exceeds y. For completeness, we set $T(\infty) = \infty$. The (proportional) mixed hitting-time (MHT) model specifies that T is the first time that $Y(t)$ crosses $\phi(X)V$, or

$$T = T\left[\phi(X)V\right], \tag{1}$$

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

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

Note that this includes Brownian motion with drift as a special case.

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

For expositional convenience only, we assume that $\Pr(V=0)=0$. The model allows for an unobserved subpopulation $\{V=0\}$ of agents using a zero threshold. On this subpopulation, $T=T(0)=0$ almost surely, that is $\Pr(T=0,V=0)=\Pr(V=0)$, because $\{Y\}$ visits $(0,\infty)$ at arbitrarily small times almost surely (Bertoin, 1996, Chapter VII, Theorem 1). The case in which V, and therefore T, has a mass point at 0 may be of interest in some applications, but even then data on immediate transitions may not be available. In applications in which a mass point at 0 is relevant, our analysis under the assumption that $\Pr(V=0)=0$ can be applied to inference about 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 by $\Pr(T=0)$. Thus, our focus on the case in which $\Pr(V=0)=0$ is without loss of generality.

We could extend the model by also allowing for an ***observed*** subpopulation of stayers by taking ϕ to be a function into $(0,\infty]$. Because such a subpopulation can be trivially identified from complete data, this extension is of little interest for the purpose of this paper.

The same is true for an extension with a subpopulation with a zero threshold by including 0 in the range of ϕ.

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

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

3 Gaussian Example

We illustrate some of this paper's key ideas with the canonical example in which $\{Y\}$ is a Brownian motion with upward drift. In this case, we can write

$$Y(t) = \mu t + \sigma W(t)$$

for some $\mu \in (0, \infty)$ and $\sigma \in [0, \infty)$, with $W(t)$ a standard Brownian motion, or Wiener process, and $W(0) = 0$. Note that the Lévy measure $\Upsilon = 0$ in this example. Recall from (1) that the MHT model specifies T as the first time $\{Y\}$ crosses the heterogeneous threshold $\phi(X)V$. For expositional convenience, we assume, in this section only, that $V < \infty$ almost surely. With $\mu > 0$, $T(y) < \infty$ for $y \in [0, \infty)$, and this ensures that $T < \infty$ almost surely.

3.1 Characterization

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

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

$$\begin{aligned} \overline{F}(t|y) &\equiv \Pr\left[T(y) > t\right] \\ &= \Phi\left(\frac{y - \mu t}{\sigma\sqrt{t}}\right) - \exp\left(\frac{2\mu y}{\sigma^2}\right)\Phi\left(-\frac{y + \mu t}{\sigma\sqrt{t}}\right), \end{aligned} \tag{2}$$

and its Lebesgue density

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

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

In the polar case with $\sigma = 0$, we have that $Y(t) = \mu t$, and $T(y) = \mu^{-1} y$ is a deterministic linear function of the threshold y. Then, $T = \mu^{-1}\phi(X)V$, and the MHT model reduces to the accelerated failure time model for $T|X$: V takes the role of a "baseline" duration

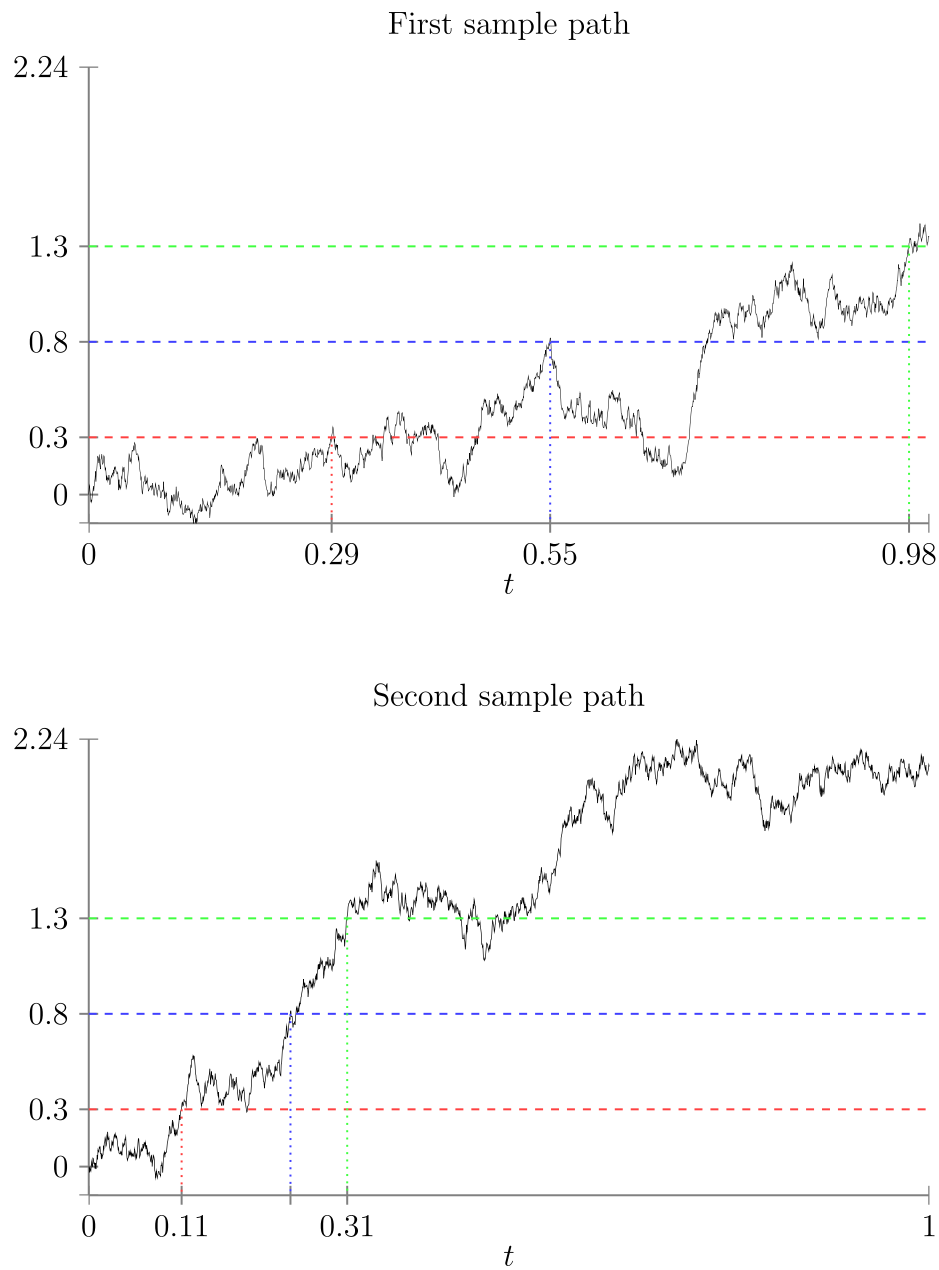
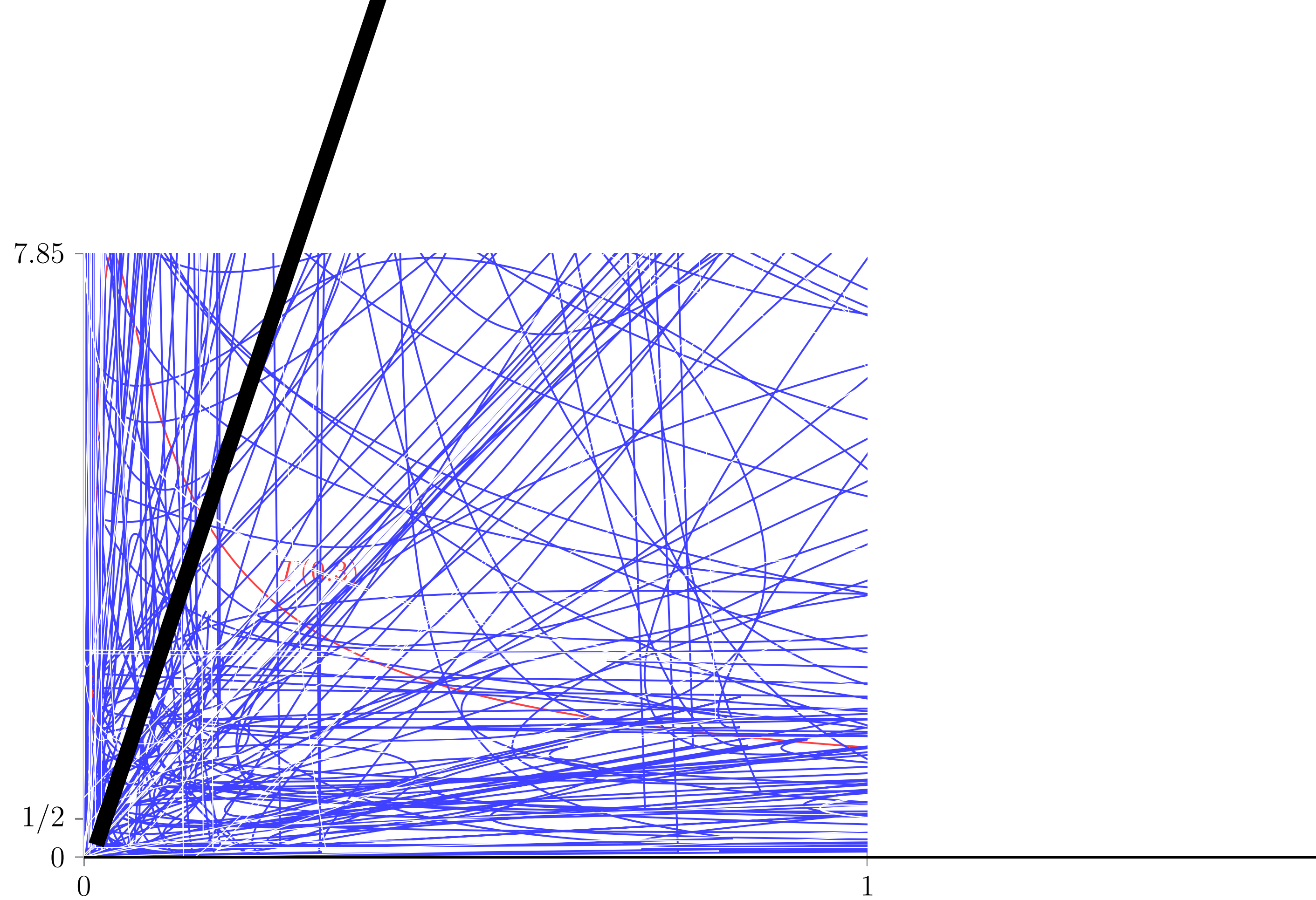


Figure 1: Two sample paths of $Y(t) = t + W(t)$, three possible thresholds, and the corresponding first hitting times.


7.85
1/2
0
0
1
T(0.3)

and 1.3, again for the case in which $\mu = 1$ and $\sigma = 1$. The hazard paths have a hump-shaped pattern: They start at 0, rise to a maximum that is attained between $y^2/(3\sigma^2) = y^2/3$ and $2y^2/(3\sigma^2) = 2y^2/3$, and then fall towards a limit $\mu^2/(2\sigma^2) = \frac{1}{2}$. The hazard rate corresponding to the lowest threshold ($y = 0.3$) is falling at most times, whereas that corresponding to the highest threshold ($y = 1.3$) is increasing for nearly all plotted times. Clearly, the hazard rates are not proportional; in this sense, the MHT model is substantially different from the MPH model.

By mixing over thresholds, a wide variety of distributions of $T|X$ can be generated. In the polar case with $\sigma = 0$, for example, for given X each distribution of T can be generated by picking the appropriate distribution G of V. Consequently, even in this special case in which $\{Y\}$ is degenerate upward drift, the model does not impose restrictions on the duration data if no variation with the regressors X is available or used (that is, if $\phi = 1$). It does however restrict the effect of any regressors to rescaling T.

3.2 Identifiability

This takes us to the question whether the model's structural determinants; μ, σ, ϕ, and G; can be uniquely determined ("identified") from large-sample data, the distribution of $T|X$. The latter is uniquely characterized by its Laplace transform,

$$\mathcal{L}_T(s|X) \equiv \mathsf{E}\left[\exp\left(-sT\right)|X\right], \quad s \in [0, \infty).$$

In turns out to be particularly convenient, both in this Gaussian example and in the general case, to study identification of the model's determinants in terms of $\mathcal{L}_T(\cdot|X)$.

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

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

so that

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

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

$$\mathcal{L}_T(s|X) = \mathcal{L}_G\left[\phi(X)\Lambda(s)\right]. \tag{5}$$

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

One trivial identification problem requires our attention upfront. Take the time T implied by (1) if $\{Y\}$ is a Brownian motion with parameters μ and σ, for a threshold $\phi(X)V$. Clearly, the process $\{\kappa\nu Y\}$, a Brownian motion with parameters $\kappa\nu\mu$ and $\kappa\nu\sigma$, and threshold $(\kappa\phi(X))(\nu V)$, with $\kappa, \nu \in (0,\infty)$, produce the same time T and are observationally equivalent, in terms of the distribution of $T|X$ or, equivalently, its Laplace transform $\mathcal{L}_T(\cdot|X)$. Like the latent error and index in static discrete-choice models, the latent process and threshold in the MHT model are only identified up to scale. At best, we can hope for identifiability of the distribution of $\{Y\}$, ϕ, and G up to two innocuous scale normalizations.

Key to this paper's identifiability analysis is an analogy with the analysis of the MPH model. To appreciate this, note that the right-hand side of (5) equals the survival function—rather than the Laplace transform— of $T|X$ in an MPH model with integrated baseline hazard

Λ, regressor effect $\phi(X)$, and unobserved-heterogeneity distribution G. It is easily checked that, in the MHT model, Λ is an increasing function such that $\lim_{s\to\infty}\Lambda(s)=\infty$ and that, in this example, $\Lambda(0)=0$. We can therefore borrow insights from the MPH identification literature pioneered by Elbers and Ridder (1982), Heckman and Singer (1984a), and Ridder (1990), exploiting the structure imposed by the MHT model on, in particular, Λ.

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

Note that, even in the Gaussian special case, regressor variation is crucial to identifiability. For example, take again the polar case with $\sigma=0$. Suppose that $\phi=1$ and $\mu=1$, so that $T=V$. Clearly, if V has an inverse Gaussian distribution with location parameter $\tilde{\mu}^{-1}$ and scale parameter $\tilde{\sigma}^{-2}$, with $\tilde{\mu},\tilde{\sigma}\in(0,\infty)$, then an alternative specification with a latent process $\{\tilde{Y}\}$ such that $\tilde{Y}(t)=\tilde{\mu}t+\tilde{\sigma}W(t)$ and a homogeneous unit threshold is observationally equivalent.

Finally, the analogy between the MHT and the MPH models should not be mistaken for a substantial similarity. 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, which, with the assumption that $\{Y\}$ is spectrally negative, ensures that $T(y)$ is infinitely divisible. ‘

4 Empirical Content

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

4.1 Characterization

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

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

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

Before we can derive $\mathcal{L}_{T(y)}$, we first have to introduce a common probabilistic characterization of the latent Lévy process. Recall from Section 2 that $\{Y\}$ can be decomposed in a Brownian motion with drift and an independent pure-jump process with jumps $\{\Delta Y\}$ following a Poisson point process. Therefore, $\{Y\}$ is fully characterized by the drift and dispersion coefficients μ and σ of its Brownian motion component and the characteristic (Lévy) measure Υ of $\{\Delta Y\}$. The latter satisfies $\int \min\{1, x^2\}\Upsilon(dx) < \infty$ and, because we exclude positive jumps, has negative support. It follows (Bertoin, 1996, Section VII.1) that $\mathsf{E}\left[\exp\left(sY(t)\right)\right] = \exp\left[\psi(s)t\right]$, for $s \in \mathsf{C} : \Re(s) \geq 0$, with the ***Laplace exponent*** ψ given by the

Lévy-Khintchine formula,

$$\psi(s) = \mu s + \frac{\sigma^2}{2}s^2 + \int_{(-\infty;0)} \left[e^{sx} - 1 - sxI(x > -1)\right] \Upsilon(dx). \tag{6}$$

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

$$\mathcal{L}_{T(y)}(s) = \exp\left[-\Lambda(s)y\right]. \tag{7}$$

In fact, $\{T\}$ is a ***killed subordinator*** with Laplace exponent Λ. That is, it is a nondecreasing Lévy process with Laplace exponent $\Lambda - \Lambda(0)$, forced to equal ∞ (the ***graveyard*** state) from some random threshold level $E_{\Lambda(0)}$ up if $\Lambda(0) > 0$. Here, $E_{\Lambda(0)}$ has an exponential distribution with parameter $\Lambda(0)$, and is independent from $(\{Y\}, X, V)$. Note that the probability $\Pr(E_{\Lambda(0)} \leq y) = 1 - \exp\left[-\Lambda(0)y\right]$ that $\{T\}$ has been ***killed*** at or below threshold level y equals the share $1 - \mathcal{L}_{T(y)}(0)$ of defecting movers at threshold level y. The fact that $\{T\}$ is a killed subordinator will allow us to draw from the extensive literature on subordinators in the next section's analysis of identifiability.

If, for example, $\{Y\}$ is a Brownian motion with general drift coefficient $\mu \in \mathsf{R}$ and dispersion coefficient $\sigma \in (0, \infty)$, we have that $\psi(s) = \mu s + \sigma^2 s^2/2$, so that $\Lambda(0) = \min\{0, -2\mu/\sigma^2\}$ and $\Lambda(s) = \left[\sqrt{\mu^2 + 2\sigma^2 s} - \mu\right]/\sigma^2$. If $\mu \geq 0$, then $\Lambda(0) = 0$, $T(y)$ is nondefective, and substituting in (7) gives the Laplace transform (4) of Section 3's Gaussian example. If $\mu < 0$, on the other hand, $\Lambda(0) = -2\mu/\sigma^2 > 0$ and the distribution of $T(y)$ has a defect of size $1 - \exp(2y\mu/\sigma^2)$. Note that in this case, $\sigma = 0$ is excluded to avoid the trivial outcome that $T(y) = \infty$ almost surely. Either way, $\{T\}$ is an inverse-Gaussian subordinator, killed at an independent exponential rate $\Lambda(0)$ if $\Lambda(0) > 0$.

Now define $\mathcal{L}_T(\cdot|X, V)$ and $\mathcal{L}_T(\cdot|X)$ analogously to $\mathcal{L}_{T(y)}$. From (1) and (7), it follows that

$$\mathcal{L}_T(s|X, V) = \exp\left[-\Lambda(s)\phi(X)V\right], \tag{8}$$

so that

$$\mathcal{L}_T(s|X) = \mathcal{L}_G\left[\Lambda(s)\phi(X)\right], \tag{9}$$

where $\mathcal{L}_G$ is again the Laplace transform of the unobservable's distribution G. Note that this expression for the Laplace transform of $T|X$ is the same as that for Section 3's Gaussian example in (5). However, in the general case here, we do not require that Λ has equation (4)'s inverse Gaussian two-parameter specification. Instead, we have semiparametrically specified Λ as the inverse of the latent process's Laplace exponent ψ in (6). This way, we now also allow for defecting movers, if $\Lambda(0) > 0$. Moreover, there can be a mass of stayers, if G has a mass point at ∞.

4.2 Identifiability

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

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

We focus on the "two-sample" case that $\mathcal{X} = \{0, 1\}$ and $\phi(x) = \beta^x$, for some $\beta \in (0, \infty)$. This assumes minimal regressor 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 regressors. This assumption can be tested, because $F_0 \neq F_1$ if and only if $\beta \neq 1$. Note that we have also fixed $\phi(0) = 1$, which is an innocuous normalization because the scale of V is unrestricted at this point.

Denote the distribution of $T|X = x$ by F_x. We have the following result on the identifiability of $(\Lambda, \beta, \mathcal{L}_G)$ from (F_0, F_1).

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

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

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

Proposition 1 establishes identification up to a power transformation, indexed by ρ, and an innocuous normalization, indexed by κ. Its proof, given in the Appendix, exploits an analogy with the analysis of the two-sample MPH model. Recall that the right-hand side of (5) equals the survival function— rather than the Laplace transform— of $T|X$ in a two-sample MPH model with integrated baseline Λ, regressor effect β^X, and unobserved-heterogeneity distribution G. We can therefore follow a strategy of proof pioneered by Elbers and Ridder (1982) and Ridder (1990). Doing so, we need to address the fact that defective duration distributions naturally arise in the context of an MHT model. Proposition 1 explicitly entertains the possibilities that there are stayers and defecting movers. The latter, a defect in the distribution of $T|(X,V)$, arises if $\Lambda(0) > 0$ and creates an identification problem similar to a left-censoring problem in the MPH model. To solve it, we use the analyticity of the Laplace transform.

Proposition 1 implies that $\mathcal{L}_G(0) = \widetilde{\mathcal{L}_G}(0)$ for any two observationally equivalent MHT triplets $(\Lambda, \beta, \mathcal{L}_G)$ and $(\tilde{\Lambda}, \tilde{\beta}, \widetilde{\mathcal{L}_G})$. Because $\Pr(V < \infty|X) = \Pr(V < \infty) = \mathcal{L}_G(0)$, $\Pr(T = \infty, V < \infty|X) = \mathcal{L}_T(0|X) - \mathcal{L}_G(0)$, and $\mathcal{L}_T(0|X)$ is data, this gives

Corollary 1 (Identifiability of the Mover-Stayer Structure). *The conditional probabilities* $\Pr(V = \infty|X = x)$ *of stayers and* $\Pr(T = \infty, V < \infty|X = x)$ *of defecting movers,* $x = 0, 1$*, are uniquely determined by* (F_0, F_1).

Intuitively, the two types of defect can be distinguished because 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. Abbring (2002) proves a similar result for the MPH model, but relying on an additional assumption on G.

The observational equivalence characterized by Proposition 1 can be given an appealing stochastic interpretation. The factor κ simply corresponds to a common rescaling of the threshold and latent process. For expositional clarity, we set $\kappa = 1$, and focus on the interpretation of ρ. Without loss of generality, let $\rho \in [1/2, 1)$. Let $\{S \}$ be a *stable subordinator* of index ρ; that is, $\{S \}$ is an increasing Lévy process such that $S (y)$ has Laplace transform $\mathsf{E}\left[\exp(-sS (y))\right] = \exp\left(-sy \right)$ (Bertoin, 1996, Section 3.1). Then, if $\{T\}$ is the hitting-time process characterized by Λ, the process $\{T\left[S (y)\right]; y \geq 0\}$, has Laplace exponent $\tilde{\Lambda}$ (Feller, 1971, Section XVII.4(e)). Consequently, for each given threshold level y, $\tilde{\Lambda}(y)$ corresponds to a positive-stable mixture $T[S\ (y)]$ over $\{T\}$. Thus, we can interpret $(\tilde{\Lambda}, \tilde{\beta}, \widetilde{\mathcal{L}_G})$ as reassigning some of the threshold heterogeneity in $(\Lambda, \beta, \mathcal{L}_G)$ to the individual hitting time process. Indeed, $|\tilde{\beta} - 1| = |\beta\ \ - 1| < |\beta - 1|$, so that there is less observed variation in the thresholds between the two samples. There are also various ways in which we can interpret $\tilde{G}$ as specifying less unobserved heterogeneity than G. Suppose, for example, that $\tilde{G}$ is an infinitely-divisible distribution, such as the inverse-Gaussian, Gamma, compound-Poisson, or positive-stable distributions. Then, G exhibits more variation than $\tilde{G}$ in the sense that it is the distribution of $\tilde{V} S$, where $\tilde{V}$ has distribution $\tilde{G}$, and S is an independent positive-stable random variable of index ρ (Feller, 1971, Section XVII.4(e)).

The restriction of ρ to $[1/2, 2]$ in Proposition 1 relies on the special structure of Λ. Suppose that Λ corresponds to the hitting-time process of a spectrally negative Lévy processes with Laplace exponent ψ. On the one hand, recall from the discussion of the Lévy-Khintchine formula (6) that ψ is convex and $\psi(s) \to \infty$ as $s \to \infty$. On the other hand, Bertoin (1996, Chapter I, Proposition 2) shows that the behavior of ψ at infinity is dominated by the quadratic term, $\sigma^2 s^2/2$. Now suppose that $\tilde{\Lambda} = \kappa\Lambda$ 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 $\tilde{\psi}(s) = \psi\left[(s/\kappa)^{1=}\ \right]$. Because $\tilde{\psi}$ should at least be of linear order and at most of quadratic order at ∞, just like ψ, it is necessary that $1/2 \leq \rho \leq 2$.

Note that Λ is a (possibly killed) subordinator if Λ is a subordinator, for all $\rho \in (0, 1)$, and not just for $\rho \in [1/2, 1]$. Thus, the requirement that Λ corresponds to the hitting-time process of a spectrally-negative Lévy process imposes structure beyond requiring that Λ is the Laplace exponent of a (killed) subordinator. On the one hand, this suggests that any strategy for point identification of Λ that does not exploit this structure, such as the application of Proposition 2, provides overidentifying restrictions that can be used in testing the MHT structure. On the other hand, when estimating the MHT model, it is more convenient to parameterize the model in terms of ψ, rather than specifying Λ directly. We will come back to this in Section 5.

Identification of the power transformation ρ requires further assumptions on either $\mathcal{L}_G$ or Λ. In their pioneering work on the MPH model, Elbers and Ridder (1982) have proved identifiability of the two-sample MPH model, up to scale, under the assumption that the unobserved factor has a finite mean. Within the context of an MPH model, this is an arbitrary normalization with substantive meaning (Ridder, 1990). In some cases, the corresponding assumption on the MHT model, $\mathsf{E}[VI(V < \infty)] < \infty$, may follow naturally from optimal stopping models in which threshold heterogeneity is reduced from primitive unobserved heterogeneity (see Section 6). In other cases, it will be a similarly arbitrary normalization. It yields identification, up to scale and without conditions on Λ, because two Laplace transforms $\widetilde{\mathcal{L}_G}$ and $\mathcal{L}_G$ such that $\widetilde{\mathcal{L}_G}(s) = \mathcal{L}_G((s/c)^{1=}\ )$ for all $s \in [0, \infty)$ can only both correspond to positive random variables V with $\mathsf{E}[VI(V < \infty)] < \infty$ if $\rho = 1$ (Ridder, 1990). We summarize in

Proposition 2 (**Identifiability of the MHT Model Under a Finite-Mean Assumption on** G). *Suppose that* $\mathsf{E}[VI(V < \infty)] < \infty$*. If two MHT triplets* $(\Lambda, \beta, \mathcal{L}_G)$ *and* $(\tilde{\Lambda}, \tilde{\beta}, \widetilde{\mathcal{L}_G})$ *imply the same pair of distributions* (F_0, F_1)*, then* $\tilde{\beta} = \beta$*,* $\tilde{\Lambda} = \kappa\Lambda$*, and* $\widetilde{\mathcal{L}_G}(\kappa s) = \mathcal{L}_G(s)$ *for all* $s \in [0, \infty)$*, for some* $\kappa \in (0, \infty)$*.*

In cases in which there is no substantial justification for a finite-mean assumption, the special structure of Λ offers a more attractive approach to point identification in the MHT model. In particular, note that $\lim_{s \downarrow 0} \Lambda'(s) > 0$. So, we can achieve identification for the (identified) case without defecting movers by requiring that $\lim_{s \downarrow 0} \Lambda'(s) < \infty$: If $\Lambda(0) = 0$ and $0 < \lim_{s \downarrow 0} \Lambda'(s) < \infty$, then $0 < \lim_{s \downarrow 0} d\Lambda(s)\ /ds < \infty$ if and only if $\rho = 1$. The assumption that $\lim_{s \downarrow 0} \Lambda'(s) < \infty$ is equivalent to the assumption that $\{Y\}$ does not *oscillate*, that is that it either drifts to ∞; that is, $\lim_{t \to \infty} Y(t) = \infty$ almost surely; or to $-\infty$; that is,

$\lim_{t\to\infty} Y(t) = -\infty$ almost surely (see Bertoin, 1996, Chapter VII, Corollary 2). An example of a Lévy process that oscillates is the (driftless) Wiener process $\{W\}$. For the case with defecting movers, that is $\Lambda(0) > 0$, a similar argument can be developed. The latent process always drifts to $-\infty$ in this case, but we need the additional assumption that $\mathsf{E}[Y(1)] > -\infty$ (note that we also have that $\mathsf{E}[Y(1)] < \infty$ by the assumption that $\{Y\}$ has no positive jumps). This involves some analysis of the Laplace exponent ψ underlying Λ. We relegate this analysis to the Appendix, where we prove

Proposition 3 (Identifiability of the MHT Model Based on Conditions on $\{Y\}$). *Assume that $\{Y\}$ does not oscillate and that $\mathsf{E}[Y(1)] > -\infty$. If two MHT triplets $(\Lambda, \beta, \mathcal{L}_G)$ and $(\tilde{\Lambda}, \tilde{\beta}, \widetilde{\mathcal{L}_G})$ imply the same pair of distributions (F_0, F_1), then $\tilde{\beta} = \beta$, $\tilde{\Lambda} = \kappa\Lambda$, and $\widetilde{\mathcal{L}_G}(\kappa s) = \mathcal{L}_G(s)$ for all $s \in [0, \infty)$, for some $\kappa \in (0, \infty)$.*

We pin down the power transformation ρ in Proposition 1 by restricting the class of inverse Laplace exponents so that it is not closed under power transformation. In their analysis of the semiparametric identifiability of the MPH model, Ridder and Woutersen (2003) use an analogous assumption on the baseline hazard. Unlike their assumption for the MPH model, however, ours can be related to more primitive assumptions, on the latent stochastic process $\{Y\}$. In particular, $\lim_{s\downarrow 0} \Lambda'(s) > 0$ follows without further assumptions on the MHT model; Ridder and Woutersen's analogous condition on the baseline hazard in the MPH model is an arbitrary restriction on this hazard's behavior near time 0.

Next, we revisit Section 3's Gaussian example $Y(t) = \mu t + \sigma W(t)$, but with $\mu \in \mathsf{R}$, $\sigma \in [0, \infty)$, $\sigma \in (0, \infty)$ if $\mu \leq 0$ (to avoid the trivial case in which $\{Y\}$ is nonincreasing), and general G. In this case,

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

so that we have

Corollary 2 (Identifiability of the Gaussian MHT Model). *If two Gaussian MHT triplets* $(\Lambda, \beta, \mathcal{L}_G)$ *and* $(\tilde{\Lambda}, \tilde{\beta}, \widetilde{\mathcal{L}_G})$ *imply the same pair of distributions* (F_0, F_1)*, then either one of the following is true:*

(i). $\tilde{\beta} = \beta$, $\tilde{\Lambda} = \kappa\Lambda$, *and* $\widetilde{\mathcal{L}_G}(s) = \mathcal{L}_G(\kappa s)$ *for all* $s \in [0, \infty)$*, for some* $\kappa \in (0, \infty)$*;*

(ii). $\tilde{\beta} = \beta^2$ *and, for all* $s \in [0, \infty)$. $\tilde{\Lambda}(s) = \kappa\Lambda(s)^2 = \nu s$ *and* $\widetilde{\mathcal{L}_G}(\kappa s^2) = \mathcal{L}_G(s)$*, for some* $\kappa, \nu \in (0, \infty)$*; or*

(iii). $\tilde{\beta} = \beta^{1/2}$ *and, for all* $s \in [0, \infty)$, $\tilde{\Lambda}(s) = \kappa\Lambda(s)^{1/2} = \nu\sqrt{s}$ *and* $\widetilde{\mathcal{L}_G}(\kappa s^{1/2}) = \mathcal{L}_G(s)$*, for some* $\kappa, \nu \in (0, \infty)$.

Thus, if two Gaussian MHT triplets are observationally equivalent, then they are either the same, up to an innocuous scale normalization, or one triplet corresponds to a degenerate upward drift and the other to a driftless nondegenerate Brownian motion. Note that identification was ensured in Section 3's example by excluding the latter specification. More generally, identification, up to scale, can be achieved by either requiring a nondegenerate latent process ($\sigma > 0$) or drift ($\mu \neq 0$).

Finally, it is important to note that the analogy with the MPH literature stretches beyond the set of basic results exploited so far. For example, consider the case in which we have multiple-spell data. In particular, suppose that we have stratified data, with one shared value of V and observations on two durations, T_1 and T_2, in each stratum. The two durations may concern a single agent's consecutive spells, or the single spells of two agents who are known to have the same value of V. Formally, suppose we observe the joint distribution of (T_1, T_2); for now, suppress covariates X. Let $T_1 = \inf\{t \geq 0 : Y_1(t) > V\}$ and $T_2 = \inf\{t \geq 0 : Y_2(t) > V\}$, with $\{Y_1\}$ and $\{Y_2\}$ independent spectrally-negative Lévy processes; and V a nonnegative random variable, distributed independently from $(\{Y_1\}, \{Y_2\})$ with distribution G. Denote the Laplace exponent of the hitting-time process corresponding to $\{Y_k\}$ with Λ_k; $k = 1, 2$. Then, analogously to Section 4.1's analysis for the single-spell case, it can be shown that

$$\mathcal{L}_{T_1;T_2}(s_2, s_2) \equiv \mathsf{E}\left[I(T_1 < \infty, T_2 < \infty)\exp\left(-s_1 T_1 - s_2 T_2\right)\right] = \mathcal{L}_G\left[\Lambda_1(s_1) + \Lambda_2(s_2)\right].$$

A similar expression again appears in the MPH literature, for the joint ***survival function*** of (T_1, T_2). Honoré's (1993) Theorem 1 for the MPH model translates directly into

Proposition 4 (Identifiability of the MHT Model from Stratified Data). *If two two-spell MHT triplets* $(\Lambda_1, \Lambda_2, \mathcal{L}_G)$ *and* $(\tilde{\Lambda}_1, \tilde{\Lambda}_2, \widetilde{\mathcal{L}_G})$ *imply the same joint distribution of* (T_1, T_2)*, then* $\tilde{\Lambda}_1 = \kappa\Lambda_1$*,* $\tilde{\Lambda}_2 = \kappa\Lambda_2$*, and* $\widetilde{\mathcal{L}_G}(\kappa s) = \mathcal{L}_G(s)$ *for all* $s \in [0, \infty)$*, for some* $\kappa \in (0, \infty)$*.*

Note that this identification result for multiple-spell data, unlike Propositions 2 and 3 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. Consequently, Proposition 4 also provides identification in a model extended with covariates X that interact in an unrestricted way with $\{Y_1\}$, $\{Y_2\}$, and V.

4.3 Censoring

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

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

5 Estimation

So far, we have ignored sampling variation. This section briefly discusses estimation of the MHT model, based on its characterization in Section 4.1, and standard likelihood and moment methods. Abbring and Salimans (2009) provide a full development of the estimators, their distributional properties, and their computation. They also develop and illustrate their implementation in Matlab/KNITRO.

Let Λ, ϕ and G be specified up to a finite vector of unknown parameters $\alpha \in \mathcal{A}$. We assume that this parameterization is one-to-one, so that α is uniquely determined by $(\Lambda, \phi, \mathcal{L}_G)$. In the two-sample specification $\phi(X) = \beta^X$, it is sufficient that $\mathcal{X} = \{0, 1\}$. More generally, if we have multivariate and continuous regressors, we can specify $\phi(X) = \exp(X'\beta)$. Then, we require the "rank condition" that the support $\mathcal{X}$ of X contains a nonempty open set in R^K. Note that this excludes an intercept from $\ln[\phi(X)] = X'\beta$, and thus embodies a scale normalization on ϕ similar to that in the two-sample specification.

Suppose that we have a complete random sample $((T_1, X_1), \ldots, (T_N, X_N))$ from the "true" distribution of (T, X), which corresponds to the distribution of $T|X$ induced by the parametric MHT model with "true" parameter vector $\alpha_0 \in \mathcal{A}$, and some marginal distribution of X. Our objective is to estimate the parameters α_0 of the distribution $T|X$. The marginal distribution of X is taken to be ancillary, so that we can ignore it and focus on the distribution of $T|X$ in our pursuit of an efficient estimator of α_0.

5.1 Maximum Likelihood

First, consider Section 3's Gaussian special case: $\{Y\}$ is a Brownian motion with drift $\mu > 0$, so that, by the analysis in the previous section, $T|X$ has a mixed (nondefective) inverse Gaussian distribution. Assume that $\phi(X)$ is nondegenerate; that is, the threshold varies with the observed covariates. Then, Corollary 2 ensures that α_0 is uniquely determined from the distribution of $T|X$, provided that we impose scale normalizations on two of the three functions Λ, ϕ, and $\mathcal{L}_G$. In this case, we can normalize ϕ as discussed before, and add a normalization on Λ such as $\mu = 1$. Alternatively, we may drop one of these normalizations and fix a scale parameter of $\mathcal{L}_G$.

In this special case, it is very easy to estimate α_0 by maximum likelihood. A conditional likelihood $L_N(\alpha)$ of $(T_1, \ldots, T_N)|(X_1, \ldots, X_N)$ can be constructed using the explicit expression for the density of $T(y)$ in (3): $L_N(\alpha) = \prod_{i=1}^{N} \int f\left[T_i|\phi(X_i)v\right] dG(v)$. Here, the dependence of f (through μ and σ), ϕ, and G (or $\mathcal{L}_G$) on the parameter vector α is kept implicit. Under standard regularity conditions, the maximizer of $L_N(\alpha)$ is a consistent and asymptotically normal estimator of α_0. It is also efficient under the assumption that the marginal distribution of X carries no information on α_0.

A finitely discrete specification of G is popular in duration analysis 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).

The Gaussian special case can be estimated by maximum likelihood because it comes with explicit expressions for the density and survival function of $T|X$. This feature it shares with many of the models studied in the statistics literature (Lee and Whitmore, 2006). In the general Lévy case, however, such expressions are not available, and maximum likelihood cannot be directly implemented. In these cases, a generalized method-of-moments (GMM) estimator can be based on the Laplace transform characterization of the distribution of $T|X$ in (9).

5.2 Generalized Method of Moments

The expression (9) of the Laplace transform of $T|X$ in terms of the model's primitives provides a continuum of conditional moment conditions, one for each point s at which the Laplace transform can be evaluated. Abbring and Salimans (2009) develop a GMM estimator based on these moment conditions.

Define $h(t, x; s, \alpha) \equiv \exp(-st) - \mathcal{L}_{\mathsf{G}}\left[\Lambda(s)\phi(x)\right]$. Recall that α_0 denotes the "true parameter", the parameter that gives the data generating process. Then, it follows from (9) that $\mathsf{E}[h(T, X; s, \alpha_0)|X] = 0$ for all $s \in (0, \infty)$. In our estimation procedure, we will specify an $(M \times 1)$-vector Z of instruments, and use the unconditional moment conditions

$$\mathsf{E}[h(T, X; s, \alpha_0)Z] = 0, \quad s \in (0, \infty). \tag{10}$$

The canonical example takes $M = K + 1$ and $Z' = [1\ X']$, which gives $K + 1$ unconditional moment conditions, $\mathsf{E}[h(T, X; s, \alpha_0)] = 0$ and $\mathsf{E}[h(T, X; s, \alpha_0)X] = 0$, for each s. We assume that the set of moment conditions (10) uniquely determines α_0.

We first construct a consistent GMM estimator with naive weighting of the moments. This estimator is easy to compute; it can serve as the first step in a more efficient two-step estimator, and may be of interest in its own right. Denote the empirical analog to the moment vector in the left-hand side of (10) with

$$h_N(s,\alpha) \equiv N^{-1} \sum_{i=1}^{N} h(T_i, X_i; s, \alpha) Z_i. \tag{11}$$

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

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

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

The analysis of Carrasco and Florens (2000) can be adapted to prove that, under appropriate regularity conditions, $\hat{\alpha}_N$ is $\sqrt{N}$-consistent and asymptotically normal. Moreover, Carrasco and Florens's (2002) method for efficient estimation based on empirical characteristic functions can be adapted to produce an GMM estimator of the MHT model that efficiently weights across evaluation points of $\mathcal{L}_{T|X}$, for given finite instrument vector Z. This estimator is a two-step estimator that uses $\hat{\alpha}_N$ as a first-stage estimator.

5.3 Censoring

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

The generalization of the GMM estimators to independently censored data is not covered by Abbring and Salimans (2009), but feasible. In the two-sample case; or more generally, in the case that the support $\mathcal{X}$ of X is finite; the GMM estimator can be readily adapted to allow for independent censoring, by nonparametrically correcting the empirical moments in (11) for censoring. To this end, first estimate the distribution of T in each sample using the Nelson-Aalen estimator or, in special cases, the Kaplan-Meier estimator (see e.g. Andersen et al., 1993, Section IV.1). Then, compute the empirical analogue of the moment condition (10) using these nonparametric estimators of the distribution of T; instead of the empirical distribution function, as in (11). Provided that the censoring mechanism is such that the distribution of T is identified in each sample, its nonparametric estimator is consistent and asymptotically Gaussian, and the properties of the censoring-corrected GMM estimator can be derived in a standard manner.

In the case that $\phi(X) = \exp(X'\beta)$, with $\mathcal{X}$ general, we cannot rely on repeated application of the Nelson-Aalen estimator to each sample. Instead, we need a semiparametric estimator of the distribution of $T|X$ to compute the empirical analogue of the moment condition (10).

6 Structural Examples

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

This section presents some simple examples of such models. With payoffs that are monotonic in a Lévy state variable, threshold rules routinely arise. We primarily focus on the way primitive heterogeneity generates ***heterogeneous*** threshold rules, and how this squares with the MHT model. We first study the optimal timing of an irreversible investment. This well-studied problem— it is closely related to the analysis of American options in finance— is a good vehicle to introduce the relation between optimal stopping models and the MHT framework. We then study two simple models of optimal transitions between unemployment and employment. By not only modeling the transitions out of the state of interest, but also the transitions into it, we specify the initial conditions for the analysis of durations in the state. This fixes some unwanted free parameters, and conveniently structures the dependence of the threshold rules on primitive heterogeneity.

6.1 Investment Timing

McDonald and Siegel (1986) study the optimal timing of an investment in an irreversible 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, a perpetual American call option— to real investments (Dixit and Pindyck, 1994). Here, we discuss an extension 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. Recall that this includes the Brownian motion case originally studied by McDonald and Siegel. The agent chooses a random investment time T that maximizes expected net payoffs, discounted at a rate R,

$$v_M(T) \equiv \mathsf{E}\left[\exp(-RT)\,(U(T)-K)_+\right],$$

where $(\cdot)_+ \equiv \max\{0, \cdot\}$ and the dependence of $v_M(T)$ on (K, U_0, R) is kept implicit. 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 $\{Y(\tau); 0 \leq \tau \leq t\}$, K, U_0, and R. Formally, if $\{\mathcal{F}\}$ is the filtration generated by these variables, then $\{T \leq t\}$ should be adapted to $\{\mathcal{F}\}$.

Suppose that $R > \ln \mathsf{E}\left[\exp(Y(1)\right] = \psi(1)$, so that $\Lambda(R) > 1$. For example, in the Brownian motion case, this requires that $R > \mu + \sigma^2/2$. Denote $\overline{Y}(t) \equiv \sup_{\in[0;t]} Y(t)$. Let E_R be an independent exponential time with parameter R. Then, because $\{Y\}$ is spectrally negative, $\overline{Y}(E_R)$ 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 if $\{Y\}$ crosses

$$\underline{y} = \ln\left[\frac{K}{U_0}\frac{\Lambda(R)}{\Lambda(R)-1}\right]_+, \tag{12}$$

where we again keep the dependence of $\underline{y}$ on (K, U_0, R) implicit.

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

$$v_n(T) \equiv \mathsf{E}\left[\exp(-RT)\,(U_0 + Y(T) - K)_+^n\right], \quad n \in \mathsf{Z}_+.$$

Here, we can interpret $U_0 + Y(t)$ as a project's value at time t, with K again the investment cost. Theorem 2 in Kyprianou and Surya (2005) gives optimal investment thresholds for all $n \in \mathsf{Z}_+$. Again applying the simplifications brought by the absence of positive shocks, these thresholds reduce to

$$\underline{y}_1 = \left[K - U_0 + \frac{1}{\Lambda(R)}\right]_+ \quad \text{and} \quad \underline{y}_2 = \left[K - U_0 + \frac{2}{\Lambda(R)}\right]_+,$$

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

In both specifications, primitive heterogeneity in investment costs K, initial project values U_0, and discount rates R generates heterogeneity in the investment thresholds. Suppose that we have data on investment times T and covariates X, and that (K, U_0, R) is fully determined by X and an unobserved heterogeneity factor V. Then, we can apply any of Propositions 1–3 if we assume that the threshold is proportional in the effects of X and 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 be made; log-linearity is a natural first choice. Typically, this implies that the primitive heterogeneity in (K, U_0, R) depends on the parameters of Λ, which is unattractive. For example, in Novikov and Shiryaev's specification, with $n = 1$ and $U_0 = K$, we get $\underline{y}_1 = \phi(X)V$ if $R = \Lambda^{-1}\left[\{\phi(X)V\}^{-1}\right]$. Note though that, following the discussion at the end of Section 4.2, we can invoke alternative identification results, yielding identification of more attractive specifications, if we impose more structure or use more information. For example, if data stratified on V are available, with multiple durations per stratum, Proposition 4 can be applied to establish identification of a model in which X enters in an unrestricted way. This accommodates any specification of the primitive 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 value dynamics, coded into Λ, and investment threshold heterogeneity. This provides an empirical distinction of state dependence and heterogeneity in the timing of investment that is consistent with theory. The results can further be used to explore the primitives underlying the investment decision rule. Obviously, without further information on these primitives, or strong assumptions, they can typically not be fully identified. For example, in Novikov and Shiryaev's specification with $U_0 = K$ and heterogeneous R, the cases with linear and quadratic utility give different estimates of the distribution of R, even if the distribution of the threshold and Λ are known. Nevertheless, the MHT identification results provide a useful first stage for exploring second-stage identification of the deeper parameters, and their implications. For example, in Novikov and Shiryaev's example with $U_0 = K$ and $n = 1$, the investment option's value is $v_1[T(\underline{y}_1)] = \exp(-1)\underline{y}_1$. Thus, from the MHT analysis, not only the distribution of R, but also the distribution of option values is identified up to scale if we assume linear utility.

From an empirical perspective, one 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. In theory, one would expect these quantities to depend on the way agents ended up with their investment option to begin with. This is an instance of the initial-conditions problem studied by Heckman (1981b). In the next section, we address this issue by explicitly modeling entry into the state of interest along with exit from this state.

6.2 Unemployment Durations and Heterogeneous Entry and Exit Costs

Consider a labor market in which workers continuously choose between unemployment and employment. A worker earns a flow B when unemployed, and

$$U(t) \equiv U_0 \exp\left[\mu t + \sigma W(t)\right]$$

when employed. Note that $U(t)$ is a geometric Brownian motion with drift, and that $\mathsf{E}[U(t)] = U_0 \exp[(\mu + \sigma^2/2)t]$. Workers incur a lump-sum cost $\underline{K} \geq 0$ when they leave their job; and pay $\overline{K} \geq 0$ when they enter a job. They maximize expected earnings, discounted at a rate $R > \mu + \sigma^2/2$.

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

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

Interestingly, if $\overline{K} < \infty$, then $\underline{y} < \infty$ even if $\underline{K} \to \infty$. This provides an example of a case in which unrestricted primitive heterogeneity leads to bounded threshold heterogeneity. In this case, threshold heterogeneity has a finite mean, $\mathsf{E}[V] < \infty$, and Proposition 2 provides point identification.

6.3 Job Separations and Heterogeneous Search

In Dixit's (1989) model, transaction costs are modeled as lump-sum entry and exit costs. Alternatively, we may assume that workers face heterogeneous search frictions when they leave employment.

Again consider a labor market in which workers are either employed or unemployed. Unemployed search sequentially for jobs, earning a flow utility B. For simplicity, assume that they encounter jobs at an exogenous Poisson rate A. Moreover, suppose that all jobs are identical, yielding utility $U(t) \equiv U_0 \exp[-\alpha Y(t)]$ at ***job tenure*** t, with $\{Y\}$ a Lévy process, $U_0 > 0$, and $\alpha > 0$; and that $B \leq U_0$, so that workers accept the first job they encounter. Finally, suppose that $\{Y\}$ is a compound Poisson process with negative jumps and positive drift. In particular, let $Y(t) = \mu t + \Delta Y(t)$, with $\mu > 0$ and $\Delta Y(t)$ shocks that arrive at a Poisson rate $\lambda > 0$ and have an independent exponential distribution on $(-\infty, 0)$ with parameter $\omega > 0$.

The expected discounted utility when unemployed is time-invariant; denote it with W. The value of employment in state Y satisfies the Bellman equation

$$(R + \lambda)\, v(Y) = U_0 \exp(-\alpha Y) + \lambda \int_0^\infty v(Y - e)\omega \exp(-\omega e) de + \mu v'(Y).$$

We now assume that $\omega > \alpha$, and that $R > \lambda\alpha/(\omega - \alpha) - \mu\alpha$. From a standard analysis; using a no-bubble condition, value matching ($v(\underline{y}) = W$), and smooth pasting ($\lim_{y \uparrow \underline{y}} v'(y) = 0$); we can solve for v and $\underline{y}$ given W:

$$v(Y) = \gamma \exp(-\alpha Y) + \delta(W) \exp(\tau Y) \quad \text{and} \quad \underline{y} = (\tau + \alpha)^{-1} \ln\left(\frac{\alpha\gamma}{\delta(W)\tau}\right),$$

where $\gamma > 0$ and $\tau > 0$ are parameters depending only on model primitives. The parameter $\delta(W)$ depends on the endogenous value W of unemployment, which is given by

$$W = \frac{B + AV(0)}{A + R} = \frac{B + A[\gamma + \delta(W)]}{A + R}.$$

It can be shown that that a unique solution $(V, W, \underline{y})$ exists; such that $|V(0) - W|$ and $\underline{y}$

decrease with A, and $|V(0) - W| \to 0$ and $\underline{y} \to 0$ as $A \to \infty$. As $A \to 0$, $W \to B/R$ and $\overline{y}$ may either diverge to ∞ or converge to a finite limit.

Heterogeneity in A generates heterogeneity in the job separation threshold $\underline{y}$. As before, under assumptions that ensure that $\underline{y} = \phi(X)V$, the MHT model can be applied to employment duration data to learn about job separations. The fact that $\overline{y}$ is, under some conditions, bounded may be exploited to justify the assumption that $\mathsf{E}[V] < \infty$, so that Proposition 2 can be applied.

As in Section 6.1, deeper parameters can possibly be identified if more data are available. In particular, note that the model specifies that unemployment durations conditional on A are exponential, so that the distribution of A is identified from a random sample of unemployment durations by the uniqueness of the Laplace transform (Feller, 1971, Section XIII.1, Theorem 1).

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 have 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, if that shock brings a sufficiently low payoff to employment. It directly implies a separation hazard, which is the product of the arrival rate of new productivity draws times the time-invariant probability that such a draw is below the separation threshold.

This can be contrasted with the specification studied here, which involves persistent idiosyncratic shocks that improve the payoffs in employment, combined with a common continuous drift towards separation. Because shocks can only improve payoffs to employment, separations do not take place at Poisson times, and a hazard specification is not directly implied. Because shocks are persistent, the model implies that individual workers, with given thresholds, have time-varying rates of leaving their jobs.

7 Extensions

This section suggests three extensions that are important, but beyond the scope of this paper.

7.1 Time-Varying Covariates

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

It is well known that time variation in observed covariates can be exploited to relax some of the more controversial identifying assumptions for the MPH model, such as Elbers and Ridder's (1982) finite-mean assumption (see e.g. Heckman and Taber, 1994). From this perspective, the case of time-invariant regressors, and in fact a single binary one, can be seen as informing us what can be learned with minimal regressor variation. Additional time-variation in the regressors can only aid identification, as with the MPH model.

7.2 Nonstationary Increments

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

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

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

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

7.3 Generalized Ornstein-Uhlenbeck Processes

Lévy processes are a key component in many process-based duration models in econometrics and statistics. Another frequent choice is the Ornstein-Uhlenbeck process (e.g. Aalen and Gjessing, 2004). This process allows for mean reversion and may be more appropriate in some applications. A specification for $\{Y\}$ that includes both as special cases is the Ornstein-Uhlenbeck process driven by a Lévy process. Such a process satisfies

$$dY(t) = -\varrho Y(t)dt + dZ(t),$$

with $\varrho \in [0, \infty)$ and $\{Z\}$ a Lévy process. The usual Ornstein-Uhlenbeck process arises if $\{Z\}$ is a Brownian motion and $\varrho > 0$. We explicitly include the boundary case $\varrho = 0$, in which $\{Y\}$ is a Lévy process. The Laplace transform of the distribution of $T|X$ in a MHT model generalized this way can be derived from Novikov (2004), who provides explicit expressions for the Laplace transform of the hitting-time distribution of an Ornstein-Uhlenbeck process driven by a spectrally-negative Lévy process. However, even though the generalized model adds only one parameter, ϱ, Novikov's results suggest that an analysis of its identifiability requires more than just a simple variation of the present paper's analysis.

8 Conclusion

This paper's main contribution is to provide fundamental insight in the empirical content of a framework for econometric duration analysis, the MHT model, that is connected to an important class of dynamic economic models with heterogeneous agents. It does so by highlighting and exploiting an analogy with the identification analysis of the MPH model, thus extending the relevance of the MPH literature to a much wider class of models.

The MHT model studied in this paper complements the MPH model; it by no means substitutes for it. The MPH model is arguably the most popular framework for econometric duration analysis (Van den Berg, 2001). In labor economics, it is most often justified as the reduced form of a job-search model. However, proportionality of the hazard rate between a duration factor on the one hand and heterogeneity factors on the other hand is hard to generate from nonstationary search models; for all we know, very special assumptions on agents' expectations and functional forms are needed (Van den Berg, 2001). Our analysis does not seek to resolve this issue; it rather offers a candidate reduced form for a class of dynamic economic models that is distinct from the search models usually associated with the MPH model. The MHT model's convenient proportional structure arises from assumptions on its primitives, notably the Lévy assumption on the latent process, and may be easier to defend in applications.

The MHT model is also a rich descriptive framework, which imposes restrictions only on the variation of durations with the regressors, not on marginal duration distributions. It includes the accelerated failure time model as a special case, and interprets this as a polar specification in which all variation in duration outcomes is due to *ex ante* heterogeneity. More generally, the Lévy structure on the latent process to great extent fixes agent-level time effects; heterogeneity is key to generating rich observed dynamics. We have discussed extensions of the framework that allow for more direct control of agent-level time effects, as through the baseline hazard in the MPH model. Justifying such time effects from dynamic economic theory will, however, not be any easier than justifying the MPH model's proportional time effects.

Appendix

Proof of Proposition 1. Denote $\mathcal{L}_x(\cdot) \equiv \mathcal{L}_T(\cdot|X = x)$ and note that $\mathcal{L}_0$ and $\mathcal{L}_1$ are uniquely determined by F_0 and F_1. Without loss of generality, let $\beta < 1$.

First, note that observational equivalence implies that

$$\mathcal{L}_G \circ (\beta \mathcal{L}_G^{-1}) = \mathcal{L}_1 \circ (\mathcal{L}_0^{-1}) = \widetilde{\mathcal{L}_G} \circ (\tilde{\beta} \widetilde{\mathcal{L}_G}^{-1})$$

on $(0, \mathcal{L}_0(0)]$. Without loss of generality, let $\mathcal{L}_G(0) \leq \widetilde{\mathcal{L}_G}(0)$. Because $\mathcal{L}_0(0) > 0$ and $\mathcal{L}_G \circ (\beta \mathcal{L}_G^{-1})$ and $\widetilde{\mathcal{L}_G} \circ (\tilde{\beta} \widetilde{\mathcal{L}_G}^{-1})$ are analytic on $(0, \mathcal{L}_G(0))$ (Kortram et al., 1995), this equality extends to $(0, \mathcal{L}_G(0))$. Iterating n times, this implies that

$$\mathcal{L}_G \circ (\beta^n \mathcal{L}_G^{-1}) = \widetilde{\mathcal{L}_G} \circ (\tilde{\beta}^n \widetilde{\mathcal{L}_G}^{-1})$$

on $(0, \mathcal{L}_G(0))$. With $K \equiv \widetilde{\mathcal{L}_G}^{-1} \circ \mathcal{L}_G$, this gives $K(\beta^n s) = \tilde{\beta}^n K(s)$ and therefore

$$\frac{K'(s)}{K(s)} = \frac{K'(\beta^n s)}{K(\beta^n s)/\beta^n}$$

for $s \in (0, \infty)$ and $n \in \mathsf{N}$. This implies that $K'(s)/K(s) = \rho/s$ for some $\rho \in (0, \infty)$, so that $K(s) = \kappa s$ and $\mathcal{L}_G(s) = \widetilde{\mathcal{L}_G}(\kappa s)$, for some $\kappa \in (0, \infty)$. With observational equivalence, in particular $\mathcal{L}_G \circ \Lambda = \widetilde{\mathcal{L}_G} \circ \tilde{\Lambda}$, this implies that $\tilde{\Lambda} = \kappa \Lambda$. And, with $\mathcal{L}_G \circ (\beta \Lambda) = \widetilde{\mathcal{L}_G} \circ (\tilde{\beta} \tilde{\Lambda})$, this implies $\tilde{\beta} = \beta$.

Finally, let ψ and $\tilde{\psi}$ be the Laplace exponents of the latent processes corresponding to Λ and $\tilde{\Lambda}$ respectively. Then, both ψ and $\tilde{\psi}$ should satisfy the Lévy-Khintchine formula (6). Because ψ is convex and $\psi(s) \to \infty$ as $s \to \infty$ (Bertoin, 1996, Chapter VII, Section 1), $s^{-1}\psi(s)$ either converges to a strictly positive constant or diverges to ∞ as $s \to \infty$. Moreover, $s^{-2}\psi(s) \to \sigma^2/2 < \infty$ (Bertoin, 1996, Chapter I, Proposition 2). Obviously, the same asymptotic behavior is displayed by $\tilde{\psi}$. From $\psi\left[\Lambda(s)\right] = s = \tilde{\psi}\left[\tilde{\Lambda}(s)\right]$, $s \in [0, \infty)$, it follows that $\tilde{\psi}(s) = \psi\left[(s/\kappa)^{1=}\right]$. Therefore, if $\rho > 2$, then

$$\lim_{s \to \infty} s^{-1}\tilde{\psi}(s) = \lim_{s \to \infty} (s/\kappa)^{-} \ \psi(s) = 0.$$

Consequently, $\rho \leq 2$ and, by symmetry, $\rho \geq 1/2$. □

Proof of Proposition 3. Let $(\Lambda, \beta, \mathcal{L}_G)$ and $(\tilde{\Lambda}, \tilde{\beta}, \widetilde{\mathcal{L}_G})$ be any two observationally-equivalent MHT triplets. Without loss of generality, let $\Lambda(0) \geq \tilde{\Lambda}(0)$. Let $\psi : [0, \infty) \to \mathsf{R}$ and $\tilde{\psi} : [0, \infty) \to \mathsf{R}$ be the Laplace exponents corresponding to both MHT triplets. Note that $\psi = \Lambda^{-1}$ and $\tilde{\psi} = \tilde{\Lambda}^{-1}$ on $[\Lambda(0), \infty)$. By Proposition 1, we have that $\tilde{\Lambda} = \kappa\Lambda$, so that

$$\tilde{\psi}(s) = \psi(\kappa^{-1=} \ s^{1=} \ ), \qquad s \in [\Lambda(0), \infty).$$

Because $\tilde{\psi}$ and $s \mapsto \psi(\kappa^{-1=} \ s^{1=} \ )$ are analytic on $(0, \infty)$ and $\Lambda(0) < \infty$, this equality extends to $(0, \infty)$. The assumptions that $\{Y\}$ does not oscillate and $\mathsf{E}[Y(1)] > -\infty$ imply that $0 < \lim_{s \downarrow 0} |\psi'(s)| < \infty$ and $0 < \lim_{s \downarrow 0} |\tilde{\psi}'(s)| < \infty$. Because

$$\lim_{s \downarrow 0} |\tilde{\psi}'(s)| = \rho^{-1}\kappa^{-1=} \ \lim_{s \downarrow 0} s^{(1- \ )=} \ |\psi'(\kappa^{-1=} \ s^{1=} \ )|,$$

these bounds only hold jointly if $\rho = 1$. □

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. 2, 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. 5, 17

Abbring, J. H. and J. R. Campbell (2005). A firm's first year. Discussion Paper 05-046/3, Tinbergen Institute, Amsterdam. 35

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. 2

Abbring, J. H. and J. J. Heckman (2008). Dynamic policy analysis. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data* (third ed.)., Chapter 24, pp. 797–865. Dordrecht: Kluwer Academic Publishers. 2

Abbring, J. H. and T. Salimans (2009). Estimating mixed hitting-time models. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. 23, 25, 27

Abbring, J. H. and G. J. Van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika 94*, 87–99. 24

Alvarez, F. and R. Shimer (2007). Search and rest unemployment. Mimeo, University of Chicago, Chicago IL. 1

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag. 22, 27

Bertoin, J. (1996). *Levy Processes*. Number 121 in Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press. 3, 4, 5, 13, 14, 16, 17, 18, 20, 29, 39

Boyarchenko, S. and S. Levendorskiĭ (2007). *Irreversible Decisions under Uncertainty: Optimal Stopping Made Easy*. Berlin: Springer-Verlag. 1, 28, 34

Carrasco, M. and J.-P. Florens (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory 16*, 797–834. 26

Carrasco, M. and J.-P. Florens (2002). Efficient GMM estimation using the empirical characteristic function. Mimeo, Department of Economics, University of Rochester. 26

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B 34*, 187–202. 1, 9

Cox, D. R. and H. D. Miller (1965). *The Theory of Stochastic Processes*. London: Methuen. 7

Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy 97*(3), 620–638. 32

Dixit, A. K. and R. S. Pindyck (1994). *Investment under Uncertainty*. Princeton University Press. 1, 27, 28

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. 12, 16, 17, 19, 35

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd ed.), Volume II. New York: Wiley. 4, 15, 17, 18, 34

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. 31

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. 12

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. 24

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. 22

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy 87*, 972–990. 1, 36

Jovanovic, B. (1984). Matching, turnover, and unemployment. *Journal of Political Economy 92*, 108–122. 1, 36

Kortram, R., A. Lenstra, G. Ridder, and A. van Rooij (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica 49*, 269–281. 39

Kyprianou, A. E. (2006). *Introductionary Lectures on Fluctuations of Levy Processes with Applications*. Berlin: Springer-Verlag. 1, 28

Kyprianou, A. E. and B. A. Surya (2005). On the Novikov-Shiryaev optimal stopping problems in continuous time. *Electronic Communications in Probability 10*, 146–154. 30

Lancaster, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society Series A 135*(2), 257–271. 2

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. 1, 7, 36

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. 2

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. 2, 25

McDonald, R. and D. Siegel (1986). The value of waiting to invest. *Quarterly Journal of Economics 101* (4), 707–728. 28, 29

Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica 58*, 757–782. 9

Moffitt, R. and W. Nicholson (1982). The effect of unemployment insurance on unemployment: The case of federal supplemental benefits. *Review of Economics and Statistics 64* (1), 1–11. 9

Mordecki, E. (2002). Optimal stopping and perpetual options for Lévy processes. *Finance and Stochastics 6* (4), 473–493. 28, 29

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. 34

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

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. 29, 30, 31

Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies 57*, 167–182. 12, 17, 19

Ridder, G. and T. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica 71*, 1579–1589. 12, 20

Singpurwalla, N. D. (1995). Survival in dynamic environments. *Statistical Science 10*(1), 86–103. 2

Stokey, N. L. (2009). *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton, NJ: Princeton University Press. 1, 27, 34

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. 38