

# Mixed Hitting-Time Models\*

Jaap H. Abbring<sup>†</sup>

June 2007<sup>‡</sup>

## 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 present extensions to time-varying covariates and latent processes with more general time and dependency structures. 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 for helpful discussion.

<sup>†</sup>Department of Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, and Tinbergen Institute. E-mail: [jabbring@feweb.vu.nl](mailto:jabbring@feweb.vu.nl). [Http://staff.feweb.vu.nl/jabbring](http://staff.feweb.vu.nl/jabbring)

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

*JEL codes:* C14, C41.

<sup>‡</sup>Preliminary draft prepared for the Cowles 75th Anniversary Conference on June 11–12, 2007. First draft: March 23, 2006.

# 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 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. Jovanovic's (1979; 1984) model of job tenure is an example in labor economics. They also appear in many other fields in economics and finance. 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.<sup>1</sup> In Lancaster (1972), he

---

<sup>1</sup>See also this paper's Section 4.3.1.

applies this model to strike durations, interpreting the gap between the Brownian motion and the threshold as the level of disagreement.

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](#)).<sup>2</sup> 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 recent work on dynamic discrete choice models. [Heckman and Navarro \(2007\)](#) discuss a general discrete-time mixture duration model based on a latent process crossing thresholds.<sup>3</sup> 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 and Section 3 analyzes its empirical content. We first develop the well-understood, and therefore instructive, special case in which the latent Lévy process is a Brownian motion with drift. Then, we present the general model’s implications for the data and the main identification results. Section 4 discusses inference from censored data and an extension to time-varying covariates. It also present ways to relax the Lévy assumptions of stationary and independent increments. Finally, Section 5 concludes with some discussion of the relative merits of the MHT and MPH

---

<sup>2</sup>They have also been studying many specifications that are outside this paper’s class of models. See Section 4.3 for discussion of some extensions, such as hitting-time models based on Ornstein-Uhlenbeck processes ([Aalen and Gjessing, 2004](#)).

<sup>3</sup>See also [Abbring and Heckman \(2008a,b\)](#) for reviews.

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. Formally, we have<sup>4</sup>

**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  $\mu \in \mathbb{R}$  and  $\sigma \in \mathbb{R}_+$ . Brownian motion is the single Lévy process with continuous sample paths. In general, Lévy processes may have jumps. Examples are compound Poisson processes, which have independently and identically distributed jumps at Poisson times. More generally, the jump process  $\{\Delta Y\}$  of a Lévy process  $\{Y\}$  is a Poisson point process with characteristic measure  $\Upsilon$  such that  $\int \min\{1, x^2\}\Upsilon(dx) < \infty$ , and any Lévy process  $\{Y\}$  can be written as the sum of a Brownian motion with drift and an independent pure-jump process with jumps governed by such a point process (Bertoin, 1996, Chapter I. Theorem 1). The characteristic measure of  $\{Y\}$ 's jump process is called its *Lévy measure* and, together

---

<sup>4</sup>See Bertoin (1996) for a comprehensive exposition of Lévy processes and their analysis.

with the drift and variance parameters of its Brownian motion component, fully characterizes  $\{Y\}$ 's distributional properties.

Throughout the paper, we will focus on spectrally negative Lévy processes. These are Lévy processes of which the characteristic measure  $\Upsilon$  has negative support, *i.e.* Lévy processes without positive jumps. Let  $\{Y\}$  be such a process. Then, the (proportional) mixed hitting-time (MHT) model specifies that  $T$  is the first time that  $Y(t)$  crosses  $\phi(X)V$ , or

$$T = \inf\{t \geq 0 : Y(t) > \phi(X)V\}, \tag{1}$$

for some observed covariates  $X$  with support  $\mathcal{X} \in \mathbb{R}^k$ , measurable function  $\phi : \mathcal{X} \mapsto \mathbb{R}_{++}$ , and positive random variable  $V$ . We use the convention that  $\inf \emptyset \equiv \infty$ ; that is, we set  $T = \infty$  if  $\{Y\}$  never crosses  $\phi(X)V$ . The assumption that there are no positive jumps greatly facilitates the analysis of hitting times, because it excludes that the process jumps across the threshold without crossing it.

The factor  $V$  is interpreted as an unobserved individual effect and is assumed to be distributed independently of  $X$  with distribution  $G$  on  $\mathbb{R}_{++} \cup \{\infty\}$ . This explicitly allows for an unobserved subpopulation  $\{V = \infty\}$  of *stayers*, on which  $T = \infty$ .<sup>5</sup> 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 nonpositive).<sup>6</sup>

---

<sup>5</sup>We could extend the model by also allowing for an *observed* subpopulation of stayers by taking  $\phi$  to be a function into  $\mathbb{R}_{++} \cup \infty$ . Because such a subpopulation can be trivially identified from complete data, this extension is of little interest for the purpose of this paper.

<sup>6</sup>If  $\{Y\}$  is nonpositive, then  $\{-Y\}$  is a *subordinator* and has increasing sample paths ([Bertoin, 1996](#), Chapter III).

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 and fix the threshold at a common value. In the Lévy-based MHT model, all that matters is the distance between the threshold and the initial condition that needs to be traveled, and we have specified this distance as  $\phi(X)V$ . In different applications, different interpretations in terms of heterogeneous initial conditions and/or heterogeneous thresholds may be appropriate.

### 3 Empirical Content

#### 3.1 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

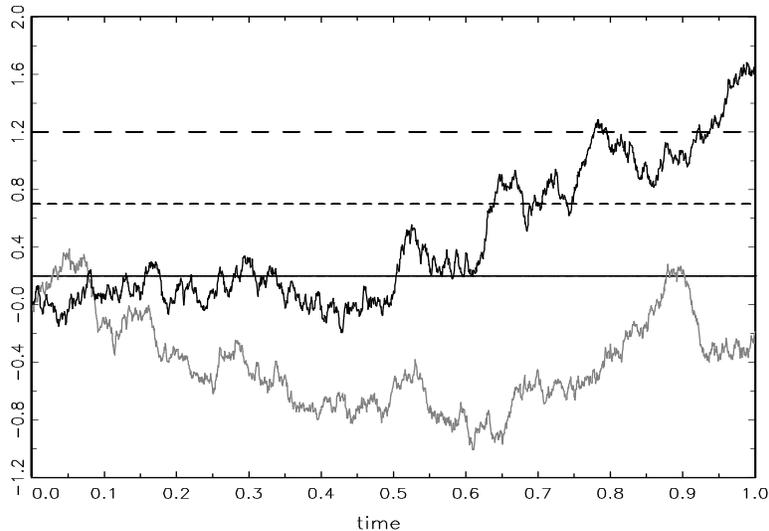
$$Y(t) = \mu t + \sigma W(t)$$

for some  $\mu \in \mathbb{R}_{++}$  and  $\sigma \in \mathbb{R}_+$ , 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 that the MHT model specifies  $T$  to be the first time that  $\{Y\}$  crosses a time-invariant threshold  $\phi(X)V$ .<sup>7</sup> For expositional convenience, we also assume, for the purpose of this example only, that

---

<sup>7</sup>Equivalently,  $T$  is the first time that the driftless Brownian motion  $\sigma W(t)$  crosses the affine threshold  $\phi(X)V - \mu t$ .

Figure 1: Two Sample Paths of a Wiener Process with Unit Drift ( $Y(t) = t + W(t)$ ) and Three Possible Thresholds



$V < \infty$  almost surely. As we will shortly see, this ensures that  $T < \infty$  almost surely.

Figure 1 plots two sample paths for the case in which  $\mu = \sigma = 1$ , with three possible exit thresholds, 0.2, 0.7 and 1.2. For a given threshold, the time that each path first crosses that threshold is a sample duration. The distribution of such times  $T$  for a given threshold, that is conditional on  $(X, V)$ , can be most conveniently characterized by its Laplace transform,

$$\mathcal{L}_T(s|X, V) \equiv \mathbb{E}[\exp(-sT) | X, V], \quad s \in \mathbb{R}_+.$$

From e.g. [Ross \(1983, Proposition 8.4.1\)](#), it follows that

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

If  $\sigma > 0$ , this is the Laplace transform of an inverse Gaussian distribution with param-

ters  $\mu/[\phi(X)V]$  and  $\sigma/[\phi(X)V]$ . Consequently, by the uniqueness of the Laplace transform,  $T|X, V$  has this inverse Gaussian distribution. This is the duration model reviewed by Lancaster (1990, Sections 4.2 and 5.7), which we extend by mixing over its parameters: Conditional on the observed regressors  $X$  only,  $T$  is distributed as a mixture of inverse Gaussian distributions.

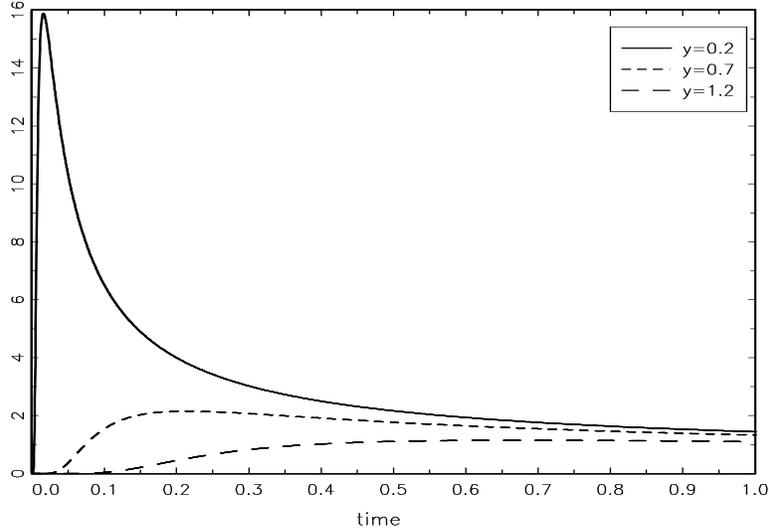
In the extreme case with  $\sigma = 0$ , we have that  $Y(t) = \mu t$  and  $T = \mu^{-1}\phi(X)V$ , so that the MHT model reduces to the accelerated failure time (AFT) model for  $T|X$  (e.g. Lancaster, 1990, Section 3.2.3). Then,  $V$  takes the role of a “baseline” duration variable, which is “accelerated” or “decelerated” by the regressor-dependent factor  $\mu^{-1}\phi(X)$ . An interpretation of the AFT model based on the MHT model is that it attributes all variation in durations for given  $X$  to *ex ante* unobserved heterogeneity. The fact that the MHT model can capture situations in which little or no uncertainty is resolved during the spell is appealing. Meyer (1990), for example, entertains this possibility as an alternative for a job-search model in his study of unemployment insurance and durations.

Although the hazard rate of  $T|X, V$  is not a primitive of the MHT model, it is useful to derive it for comparison with hazard-based models like the MPH model. Figure 2 plots the hazard-rate paths for the three threshold levels  $y$  plotted in the top graph, 0.2, 0.7 and 1.2, again for the case in which  $\mu = \sigma = 1$ .<sup>8</sup> 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 ( $y = 0.2$ ) is falling at most times, whereas that corresponding to the highest threshold ( $y = 1.2$ ) is increasing for all plotted times. Clearly, the hazard rates are not proportional; in this sense, the MHT model is substantially different from the MPH model.

---

<sup>8</sup>The density, cumulative distribution function, and hazard rate of the inverse Gaussian distribution can be given explicitly (Lancaster, 1990, Section 5.7), but this not carry over to the general case and we have no use for it in our analysis here.

Figure 2: Hazard Rates Corresponding to Figure 1’s Three Thresholds  $y$ , for a Wiener Process with Unit Drift ( $Y(t) = t + W(t)$ )



By mixing over thresholds, a wide variety of distributions of  $T|X$  can be generated. In the extreme 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  $T$ .

This takes us to the question whether the model’s structural determinants,  $\mu$ ,  $\sigma$ ,  $\phi$  and  $G$ , can be uniquely determined (“identified”) from large-sample data, the distribution of  $T|X$ . The latter is uniquely characterized by its Laplace transform  $\mathcal{L}_T(s|X) \equiv \mathbb{E}[\mathcal{L}_T(s|X, V)|X]$ , which is given by

$$\mathcal{L}_T(s|X) = \mathcal{L}_G[\phi(X)\Lambda(s)], \quad s \in \mathbb{R}_+. \quad (3)$$

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  $\mu$  and  $\sigma$ , 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  $(\kappa\phi(X))(\nu V)$  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  $\{Y\}$ ,  $\phi$ , 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 (3) equals the survival function—rather than the Laplace transform—of  $T|X$  in an MPH model with integrated baseline hazard  $\Lambda$ , regressor effect  $\phi(X)$ , and unobserved-heterogeneity distribution  $G$ .<sup>9</sup> We can therefore borrow insights from the MPH identification literature pioneered by [Elbers and Ridder \(1982\)](#), [Heckman and Singer \(1984\)](#), and [Ridder \(1990\)](#), exploiting the structure imposed by the MHT model on, in particular,  $\Lambda$ .

Consider the case that  $\phi(X) = \exp(X'\beta)$  for some parameter vector  $\beta \in \mathbb{R}^k$ . Note that  $\Lambda$  is differentiable on  $\mathbb{R}_{++}$  and that  $0 < \lim_{s \downarrow 0} \Lambda'(s) = \frac{1}{\mu} < \infty$ . Thus, [Ridder and Woutersen \(2003, Proposition 1\)](#) implies that  $\mu$ ,  $\sigma$ ,  $\beta$ , 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 Lévy processes, with nonparametric  $\Lambda$ . Doing so, we exploit and address various aspects of the more general MHT model.

Note that, even in this semi-parametric special case, regressor variation is crucial to

---

<sup>9</sup>It is easily checked that  $\Lambda$  is an increasing function such that  $\lim_{s \rightarrow \infty} \Lambda(s) = \infty$  and that, in this example,  $\Lambda(0) = 0$ .

identifiability. For example, take again the extreme case with  $\sigma = 0$ . Suppose that  $\phi = 1$  and  $\mu = 1$ , so that  $T = V$ . Clearly, if  $V$  has an inverse Gaussian distribution with parameters  $\tilde{\mu}$  and  $\tilde{\sigma}$ , 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.

### 3.2 Characterization

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. The distribution of  $T$  conditional on  $(X, V)$  is again fully determined, up to almost-sure equivalence, by its Laplace transform, which we now define as

$$\mathcal{L}_T(s|X, V) \equiv \mathbb{E}[\exp(-sT) I(T < \infty)|X, V], \quad s \in \mathbb{R}_+,$$

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

Before we can derive  $\mathcal{L}_T(\cdot|X, V)$ , we first have to introduce a common probabilistic characterization of the latent Lévy process. Recall from Section 2 that a Lévy process  $\{Y\}$  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  $\mu$  and  $\sigma$  of its Brownian motion component and the characteristic (Lévy) measure  $\Upsilon$  of  $\{\Delta Y\}$ . The latter satisfies  $\int \min\{1, x^2\}\Upsilon(dx) < \infty$  and, because we exclude positive jumps, has negative support. It follows (Bertoin, 1996, Section VII.1) that  $\mathbb{E}[\exp(sY(t))] = \exp[\psi(s)t]$ , for  $s \in \mathbb{C} : \Re(s) \geq 0$ , with the *Laplace exponent*  $\psi$

given by the Lévy-Khintchine formula,

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

The Laplace exponent, as a function on  $\mathbb{R}_+$ , 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  $\psi$  to  $[\Lambda(0), \infty)$ . Proposition 1 of [Bertoin \(1996, Chapter VII\)](#) implies that

$$\mathcal{L}_T(s|X, V) = \exp[-\Lambda(s)\phi(X)V]. \quad (4)$$

The Laplace transform of the distribution of  $T|X$  therefore is

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

with  $\mathcal{L}_G$  again the Laplace transform of the unobservable's distribution  $G$ .

If, for example,  $\{Y\}$  is a Brownian motion with general drift coefficient  $\mu \in \mathbb{R}$  and  $\sigma \in \mathbb{R}_{++}$ , 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  $\Lambda(0) = 0$ ,  $T$  is nondefective, and substituting in (4) gives the Laplace transform (2) of Section 3.1's Gaussian example. If  $\mu < 0$ , on the other hand,  $\Lambda(0) = -2\mu/\sigma^2 > 0$  and the distribution of  $T|X, V$  has a defect of size  $1 - \exp(2\phi(X)V\mu/\sigma^2)$ . Note that in this case,  $\sigma = 0$  is excluded to avoid the trivial outcome that  $T = \infty$  almost surely.

### 3.3 Identifiability

The distribution of  $T|X$  implied by the MHT model only depends on its primitives  $(\mu, \sigma^2, \Upsilon)$  and  $(\phi, G)$  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  $T|X$ .<sup>10,11</sup> Because there is a one-to-one relation between  $(\Lambda, \phi, \mathcal{L}_G)$  and the MHT model’s primitives, this identification analysis applies without change to these primitives.<sup>12</sup>

We focus on the “two-sample” case that  $\mathcal{X} = \{0, 1\}$  and  $\phi(x) = \beta^x$ , for some  $\beta \in \mathbb{R}_{++}$ . This assumes minimal regressor variation and thus poses the hardest identification problem.<sup>13</sup> 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  $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}}_G)$  imply the same pair of distributions  $(F_0, F_1)$ , then*

$$\begin{aligned}\tilde{\beta} &= \beta^\rho, \\ \tilde{\Lambda} &= c\Lambda^\rho, \text{ and} \\ \tilde{\mathcal{L}}_G(cs^\rho) &= \mathcal{L}_G(s) \text{ for all } s \in \mathbb{R}_+, \end{aligned}$$

---

<sup>10</sup>That is, we abstract from sampling variation in the analysis of identifiability

<sup>11</sup>The marginal distribution of  $X$  is ancillary to the model.

<sup>12</sup>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 parameters  $(\mu, \sigma^2, \Upsilon)$  of the latent Lévy process can be uniquely determined from  $\Lambda$  by the uniqueness of the Lévy-Khintchine representation (Bertoin, 1996, Chapter I, Theorem 1).

<sup>13</sup>This is similar to the challenge accepted by Elbers and Ridder (1982) in their analysis of the identifiability of the MPH model.

for some  $c, \rho \in \mathbb{R}_{++}$ .

Proposition 1 establishes identification up to a power transformation, indexed by  $\rho$ , and an innocuous normalization, indexed by  $c$ . 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 (3) equals the survival function—rather than the Laplace transform—of  $T|X$  in a two-sample MPH model with integrated baseline  $\Lambda$ , regressor effect  $\beta^X$ , 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  $(\widetilde{\Lambda}, \widetilde{\beta}, \widetilde{\mathcal{L}}_G)$ . 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  $G$ .

Identification of the power transformation  $\rho$  requires further assumptions on either  $\Lambda$  or  $\mathcal{L}_G$ . The MHT model facilitates a discussion of, in particular, assumptions on  $\Lambda$  in terms of

primitives. 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  $0 < \lim_{s \downarrow 0} d\Lambda(s)^\rho/ds < \infty$  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  $\infty$  ( $\lim_{t \rightarrow \infty} Y(t) = \infty$  almost surely) or to  $-\infty$  ( $\lim_{t \rightarrow \infty} Y(t) = -\infty$  almost surely).<sup>14,15</sup> In the case that  $\Lambda(0) = 0$  we have that  $\{Y\}$  either oscillates or drifts to  $\infty$ , so assuming that  $\{Y\}$  does not oscillate for the case without defecting movers selects the case that  $\{Y\}$  drifts to  $\infty$ . Thus, we have

**Corollary 2 (Identifiability of the MHT Model Based on Conditions on  $\{Y\}$ ).** *Let  $\Lambda(0) = 0$  and assume that  $\{Y\}$  drifts to  $\infty$ . If two MHT triplets  $(\Lambda, \beta, \mathcal{L}_G)$  and  $(\tilde{\Lambda}, \tilde{\beta}, \tilde{\mathcal{L}}_G)$  imply the same pair of distributions  $(F_0, F_1)$ , then  $\tilde{\beta} = \beta$ ,  $\tilde{\Lambda} = c\Lambda$ , and  $\tilde{\mathcal{L}}_G(cs) = \mathcal{L}_G(s)$  for all  $s \in \mathbb{R}_+$ , for some  $c \in \mathbb{R}_{++}$ .*

We pin down the power transformation  $\rho$  in Proposition 1 by restricting the class of inverse Laplace exponents  $\Lambda$  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.

By exploiting the Lévy structure on  $\Lambda$ , we avoid relying on a finite-mean assumption on  $V$  for identification. 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

---

<sup>14</sup>See [Bertoin \(1996, Chapter VII, Corollary 2\)](#).

<sup>15</sup>Note that  $\{W\}$  oscillates.

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). The corresponding assumption on the MHT model,  $\mathbb{E}[VI(V < \infty)] < \infty$ , is a similarly arbitrary normalization. It yields identification, up to scale and without conditions on  $\Lambda$ , because two Laplace transforms  $\widetilde{\mathcal{L}}_G$  and  $\mathcal{L}_G$  such that  $\widetilde{\mathcal{L}}_G(s) = \mathcal{L}_G((s/c)^{1/\rho})$  for all  $s \in \mathbb{R}_+$  can only both correspond to positive random variables  $V$  with  $\mathbb{E}[VI(V < \infty)] < \infty$  if  $\rho = 1$  (Ridder, 1990). For completeness, we summarize in

**Corollary 3 (Identifiability of the MHT Model Under a Finite-Mean Assumption on  $G$ ).** *Suppose that  $\mathbb{E}[VI(V < \infty)] < \infty$ . 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  $\tilde{\beta} = \beta$ ,  $\tilde{\Lambda} = c\Lambda$ , and  $\widetilde{\mathcal{L}}_G(cs) = \mathcal{L}_G(s)$  for all  $s \in \mathbb{R}_+$ , for some  $c \in \mathbb{R}_{++}$ .*

Finally, we revisit Section 3.1's Gaussian example  $Y(t) = \mu t + \sigma W(t)$ , but with  $\mu \in \mathbb{R}$ ,  $\sigma \in \mathbb{R}_+$ , and  $\sigma \in \mathbb{R}_{++}$  if  $\mu \leq 0$ .<sup>16</sup> In this case,

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

so that we have

**Corollary 4 (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).  $\tilde{\beta} = \beta$ ,  $\tilde{\Lambda} = c\Lambda$ , and  $\widetilde{\mathcal{L}}_G(s) = \mathcal{L}_G(cs)$  for all  $s \in \mathbb{R}_+$ , for some  $c \in \mathbb{R}_{++}$ ;

---

<sup>16</sup>If  $\mu \leq 0$  and  $\sigma = 0$ , then  $\{Y\}$  is nonincreasing, and  $T = \infty$  almost surely. Because  $\Pr(V < \infty) > 0$ , this case is (set-)identified, so that we can exclude it without loss of generality.

(ii).  $\tilde{\beta} = \beta^2$  and, for all  $s \in \mathbb{R}_+$ .  $\tilde{\Lambda}(s) = c\Lambda(s)^2 = d s$  and  $\tilde{\mathcal{L}}_G(cs^2) = \mathcal{L}_G(s)$ , for some  $c, d \in \mathbb{R}_{++}$ ; or

(iii).  $\tilde{\beta} = \beta^{1/2}$  and, for all  $s \in \mathbb{R}_+$ ,  $\tilde{\Lambda}(s) = c\Lambda(s)^{1/2} = d\sqrt{s}$  and  $\tilde{\mathcal{L}}_G(cs^{1/2}) = \mathcal{L}_G(s)$ , for some  $c, d \in \mathbb{R}_{++}$ .

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 3.1'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 ( $\mu \neq 0$ ).

## 4 Extensions

### 4.1 Censoring

Our identification analysis takes the distribution of  $(T, X)$ , or rather the Laplace transform of  $T|X$ , as known. If complete data on this distribution are available, the MHT model can be estimated in various ways. Because of the distribution's characterization in terms of its Laplace transform, empirical transform methods stand out as conceptually straightforward.<sup>17</sup>

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. Then, this paper's identification results carry over to the case of 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. In the two-sample case, empirical transforms,

---

<sup>17</sup>See e.g. Yao and Morgan (1999) for a review and a list of references.

and other statistics, can be readily adapted to such censoring patterns using the Nelson-Aalen estimator (see e.g. [Andersen et al., 1993](#), Section IV.1).

Our 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  $C \in \mathbb{R}_{++}$ , so that only the restriction of  $F$  to  $[0, C]$  is known. Then, only the restriction of  $G$  to  $[0, C]$  is identified.

## 4.2 Time-Varying Covariates

Following most of the duration-model identification literature, we have ignored time-varying covariates. Time-varying covariates are most easily and directly introduced in the MHT model as determinants of a time-varying threshold. 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. A complication specific to the MHT model is that the characterization of the duration distribution through its Laplace transform for the time-invariant regressor case does not extend directly to time-varying covariates. This complicates the identification analysis, but does not hamper inference using e.g. simulation methods.

### 4.3 The Latent Process

Finally, we consider relaxing the assumptions that  $\{Y\}$  has stationary and independent increments.

#### 4.3.1 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  $\xi : \mathbb{R}_+ \cup \{\infty\} \mapsto \mathbb{R}_+ \cup \{\infty\}$  and the distribution of  $U|X$  given by the MHT model. If  $\xi$  is linear, this simply gives the MHT model for  $T|X$ ; any nonlinearities correspond to additional duration dependence.

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

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

and  $U$  the first time of Brownian motion crosses a threshold that decreases linearly from a positive initial value.<sup>18</sup> 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

---

<sup>18</sup>This is equivalent to the first time a Brownian motion with upward drift crosses a positive threshold—this paper's adopted standard formulation— but the alternative phrasing is more natural in this context.

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.

### 4.3.2 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 (OU) 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 the stochastic differential equation

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

with  $\alpha \in \mathbb{R}_+$  and  $\{Z\}$  a Lévy process. The usual OU process arises if  $\{Z\}$  is a Brownian motion and  $\alpha > 0$ . We explicitly include the boundary case  $\alpha = 0$ , 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 OU process driven by a spectrally-negative Lévy process. However, even though the generalized model adds only one parameter,  $\alpha$ , [Novikov](#)'s results suggest that an analysis of its identifiability requires more than just a simple variation of the present paper's analysis. This is left for future work.

## 5 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 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 AFT 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

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

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

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

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

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

for  $s \in \mathbb{R}_{++}$  and  $n \in \mathbb{N}$ . This implies that  $K'(s)/K(s) = \rho/s$  for some  $\rho \in \mathbb{R}_{++}$ , so that  $K(s) = cs^\rho$  and  $\mathcal{L}_G(s) = \widetilde{\mathcal{L}}_G(cs^\rho)$ , for some  $c \in \mathbb{R}_{++}$ . With observational equivalence, in particular  $\mathcal{L}_G \circ \Lambda = \widetilde{\mathcal{L}}_G \circ \widetilde{\Lambda}$ , this implies that  $\widetilde{\Lambda} = c\Lambda^\rho$ . And, with  $\mathcal{L}_G \circ (\beta\Lambda) = \widetilde{\mathcal{L}}_G \circ (\widetilde{\beta}\widetilde{\Lambda})$ , this implies  $\widetilde{\beta} = \beta^\rho$ .  $\square$

## References

- Aalen, O. and H. Gjessing (2004). Survival models based on the Ornstein-Uhlenbeck process. *Lifetime Data Analysis 10*, 407–423. [2](#), [19](#)
- Aalen, O. O. and H. K. Gjessing (2001, February). Understanding the shape of the hazard rate: A process point of view. *Statistical Science 16*(1), 1–14. [2](#), [18](#)
- Abbring, J. H. (2002). Stayers versus defecting movers: A note on the identification of defective duration models. *Economics Letters 74*, 327–331. [4](#), [13](#)
- Abbring, J. H. and J. J. Heckman (2008a). Dynamic policy analysis. In L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data* (third ed.). Dordrecht: Kluwer. Forthcoming. [2](#)
- Abbring, J. H. and J. J. Heckman (2008b). 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 VI, Chapter 72. Amsterdam: Elsevier Science. Forthcoming. [2](#)
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag. [16](#), [17](#)
- Bertoin, J. (1996). *Lévy Processes*. Number 121 in Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press. [3](#), [4](#), [10](#), [11](#), [12](#), [14](#)
- 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. [9](#), [12](#), [13](#), [14](#), [17](#)

- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd ed.), Volume II. New York: Wiley. [12](#)
- Heckman, J. J. and S. Navarro (2007, February). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics* *136*(2), 341–396. [2](#)
- Heckman, J. J. and B. Singer (1984). The identifiability of the proportional hazard model. *Review of Economic Studies* *51*, 231–241. [9](#)
- 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. [17](#)
- Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy* *87*, 972–990. [1](#), [18](#)
- Jovanovic, B. (1984). Matching, turnover, and unemployment. *Journal of Political Economy* *92*, 108–122. [1](#), [18](#)
- Kortram, R., A. Lenstra, G. Ridder, and A. van Rooij (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica* *49*, 269–281. [22](#)
- Lancaster, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society Series A* *135*(2), 257–271. [1](#)
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press. [1](#), [7](#), [18](#)
- Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica* *58*, 757–782. [7](#)

- 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. [19](#)
- Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57, 167–182. [9](#), [13](#), [15](#)
- Ridder, G. and T. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica* 71, 1579–1589. [9](#), [14](#)
- Ross, S. M. (1983). *Stochastic Processes* (Second ed.). New York: Wiley. [6](#)
- Singpurwalla, N. D. (1995, February). Survival in dynamic environments. *Statistical Science* 10(1), 86–103. [2](#)
- 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 V. North Holland. [20](#)
- Yao, Q. and B. J. Morgan (1999). Empirical transform estimation for indexed stochastic models. *Journal of the Royal Statistical Society Series B* 61(1), 127–141. [16](#)