The Likelihood of Mixed Hitting Times*

Jaap H. Abbring† Tim Salimans‡

February 8, 2010

Abstract

We present a method for efficiently computing the likelihood of a mixed hitting-time model that specifies durations as the first time a latent Lévy process crosses a heterogeneous threshold. This likelihood is not generally known in closed form, but its Laplace transform is. Our approach to its computation relies on numerical methods for inverting Laplace transforms that exploit special properties of the first passage times of Lévy processes. We use our method to implement a maximum likelihood estimator of the mixed hitting-time model in MATLAB. We illustrate the application of this estimator with an analysis of Kennan's (1985) strike data.

*We thank Justin Dijk for excellent research assistance. A replication file for this paper and MATLAB code implementing the estimators developed in this paper will be made available from http://center.uvt.nl/staff/abbring/mhtm.

†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.

‡Tinbergen Institute, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: salimans@ese.eur.nl.

Keywords: duration analysis, first passage time, Laplace transform, Lévy process, maximum likelihood, Mellin's inverse formula, mixture, optimal stopping, strike duration.

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. They are of substantial interest because they can be applied to the analysis of optimal stopping decisions by heterogeneous agents. In particular, they can be applied to problems that do not lead to the mixed proportional hazards model, Lancaster's (1979) and Vaupel et al.'s (1979) popular extension of the Cox (1972) proportional hazards model. Examples include models of job durations, marriage durations, and the entry and exit of firms that are driven by Brownian motions and more general persistent processes. First hitting time duration models are also increasingly popular in statistics for their structural and descriptive appeal (Lee and Whitmore, 2006).

This paper considers likelihood-based empirical methods for 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. Spectrally-negative Lévy processes include Brownian motions with linear drifts and Poisson processes compounded with negative shocks as well-known special cases. Following empirical practice with mixture duration models such as the mixed proportional hazards model, we focus on parametric MHT models, and propose using flexible parameterizations that can approximate arbitrary functional forms by increasing the number of parameters. The main obstacle in applying standard parametric likelihood methods is that, in general, we have no explicit expression for the MHT model's likelihood. However, an explicit expression for its Laplace transform is generally available. Our approach to likelihood computation exploits this.

We adapt numerical methods for the inversion of the Laplace transforms of the first hitting times of Lévy processes to compute the conditional density and survival function implied by the MHT model. In turn, these are used to construct a likelihood for independently censored duration data. In the special case that the latent process is a Brownian motion, the likelihood can be explicitly expressed as a mixture of inverse Gaussian densities and survival functions. Therefore, we can use this special case as a benchmark for evaluating our procedure for computing the likelihood. We show that the numerical inversion that is required in the general case is sufficiently fast and precise to make maximum likelihood estimation feasible even if no explicit expression of the likelihood is available.

We implement a maximum likelihood estimator that uses this computational strategy with MATLAB, and illustrate its application with a reconsideration of Kennan's (1985) empirical analysis of US contract strike durations. Our strategy for computing the MHT model's likelihood can also be used to implement other likelihood-based empirical methods. For example, it can be combined with data augmentation and Markov chain Monte Carlo techniques to implement Bayesian estimators of the MHT model.

Abbring (2007) presented the MHT model studied in this paper, analyzed its empirical content, and highlighted its close relation to optimal stopping problems in economics. The present paper operationalizes this model by providing and analyzing feasible methods for computing its likelihood and its maximum likelihood estimator.

In a companion paper (Abbring and Salimans, 2010), we develop an alternative estimator that avoids computing the MHT model's likelihood altogether by directly matching the Laplace transform implied by the model to the empirical Laplace transform. To this end, we construct a generalized method of moments estimator based on the continuum of conditional moment conditions implied by the Laplace transform characterization of the data. We develop effective procedures for this estimator's computation, derive the asymptotic properties of this estimator and evaluate its finite-sample statistical performance against that of the asymptotically efficient maximum likelihood estimator developed here.

The remainder of this paper is organized as follows. Section 2 reviews the MHT model and the corresponding characterization of the data presented in Abbring (2007). Section 3 present a method for the computation of this model's likelihood and its derivatives.

Section 4 presents flexible model parameterizations and discusses the implementation of a maximum likelihood estimator. Section 5 applies this estimator to strike data. Section 6 concludes.

2 Mixed Hitting-Time Model

2.1 Specification

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 analysis of Lévy processes. 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 $\mu \in \mathbb{R}$ and $\sigma \in [0, \infty)$. Brownian motion is the single Lévy process with continuous sample paths. In general, Lévy processes may have jumps. 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 Υ such that $\int \min\{1, x^2\}\Upsilon(dx) < \infty$, and any Lévy process $\{Y\}$ can be written as the sum of a Brownian motion with drift and an independent pure-jump process with jumps governed by such a point process (Bertoin, 1996, Chapter I. Theorem 1). The characteristic measure of $\{Y\}$'s jump process is called its *Lévy measure* and, together with the drift and 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 Υ 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 $\phi : \mathcal{X} \mapsto (0, \infty)$, and nonnegative random variable V, with (X, V) independent of $\{Y\}$. 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.

The factor V is interpreted as 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 $T = \infty$. 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 nonpositive). For expositional convenience only, we also assume that $\Pr(V = 0) = 0$. Abbring (2007) provides further discussion.

2.2 Characterization

The distribution of T conditional on (X, V) is fully determined, up to almost-sure equivalence, by its Laplace transform,

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

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 $1 - \Pr(T < \infty|X, V) = 1 - \mathcal{L}_T(0|X, V)$.

Unlike the distribution of $T|(X, V)$, the Laplace transform $\mathcal{L}_T(\cdot|X, V)$ can be explicitly given for any specification of the latent process $\{Y\}$. This first requires a common probabilistic characterization of $\{Y\}$, in terms of its characteristic function. Bertoin (1996, Section VII.1) shows that

$$\mathbb{E}\left[\exp\left(sY(t)\right)\right] = \exp\left[\psi(s)t\right],$$

for all $s \in \mathbb{C}$ with nonnegative real parts, with the *Laplace exponent* ψ given by the Lévy-Khintchine formula,

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

Here, $\tilde{\mu} \in \mathbb{R}$ absorbs any linear drift of $\{Y\}$, $\sigma \geq 0$ is the dispersion parameter of its Brownian motion component; and Υ is the Lévy measure of its jump component, where Υ satisfies $\int \min\{1, x^2\}\Upsilon(dx) < \infty$ and has negative support. The Laplace exponent ψ of $\{Y\}$ fully characterizes its distributions, through its characteristic function $\mathbb{E}\left[\exp\left(iuY(t)\right)\right] = \exp\left[\psi(iu)t\right]$ for all $u \in \mathbb{R}$.

Equation (2) gives the most common parameterization of ψ. It corresponds to the Lévy-Itô decomposition of $\{Y\}$ in a Brownian motion with linear drift $\tilde{\mu}t$, a compound

Poisson process with jumps in $(-\infty,-1]$, and a pure-jump martingale with jumps in $(-1,0)$ (Bertoin, 1996, Section I.1). Alternative parameterizations arise if we decompose the jumps of $\{Y\}$ in small and large shocks in other ways. These parameterizations all have the same dispersion parameter σ and Lévy measure Υ, but have different drift parameters. For example, in the special case that $\int_0^1 x\Upsilon(dx)<\infty$, the *compensator* term for the small shocks in (2),

$$\int_{(-\infty,0)} sxI(x>-1)\Upsilon(dx)=\int_{(-1,0)} x\Upsilon(dx)s,$$

is a well-defined linear function of s. Therefore, in this case, we can alternatively parameterize ψ as

$$\psi(s)=\mu s+\frac{\sigma^2}{2}s^2+\int_{(-\infty,0)}\left(\mathrm{e}^{sx}-1\right)\Upsilon(dx), \tag{3}$$

where $\mu\equiv\tilde{\mu}+\int_{(-1,0)} x\Upsilon(dx)$. This includes the important special case that $\int_{(-\infty,0)}\Upsilon(dx)<\infty$, in which $\{Y\}$ is the sum of a Brownian motion with drift parameter μ and a compound Poisson process with jumps *of all sizes* in $(-\infty,0)$. In general, any of the equivalent parameterizations of ψ can be used in the MHT model's specification, but some are numerically and statistically more convenient than others; we return to this in Section 4.

With ψ determined, we are ready to analyze the Laplace transform $\mathcal{L}_T(\cdot|X,V)$. 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 (see Abbring, 2007)

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

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

$$\mathcal{L}_T(s|X) = \mathcal{L}\left[\Lambda(s)\phi(X)\right], \tag{4}$$

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

2.3 A Gaussian Example

Suppose that $\{Y\}$ is a Brownian motion with general drift coefficient $\mu \in \mathbb{R}$ and dispersion coefficient $\sigma \in (0, \infty)$. Then, we have that $\psi(s) = \mu s + \sigma^2 s^2/2$, so that $\Lambda(0)$ equals $\Lambda_{\mathrm{BM}}(0) \equiv \min\{0, -2\mu/\sigma^2\}$ and $\Lambda(s)$ equals

$$\Lambda_{\mathrm{BM}}(s) \equiv \frac{\sqrt{\mu^2 + 2\sigma^2 s} - \mu}{\sigma^2}. \tag{5}$$

Because there are no jumps, there is no ambiguity in the treatment of small and large jumps, and this parameterization of ψ is unique. In particular, the Lévy-Khintchine representations (2) and (3) of ψ coincide, and $\mu = \tilde{\mu}$.

In this special case, for positive $\phi(X)V$, the distribution of $T|X, V$ is inverse Gaussian (Cox and Miller, 1965, Section 5.4), with Lebesgue density

$$f_{\mathrm{BM}}(t|X, V) = \frac{\phi(X)V}{\sigma\sqrt{2\pi t^3}} \exp\left(-\frac{(\phi(X)V - \mu t)^2}{2\sigma^2 t}\right) \tag{6}$$

and survival function

$$\begin{aligned} \overline{F}_{\mathrm{BM}}(t|X, V) &\equiv \Pr\left(T > t|X, V\right) \\ &= \Phi\left(\frac{\phi(X)V - \mu t}{\sigma\sqrt{t}}\right) - \exp\left(\frac{2\mu\phi(X)V}{\sigma^2}\right)\Phi\left(-\frac{\phi(X)V + \mu t}{\sigma\sqrt{t}}\right). \end{aligned} \tag{7}$$

Here, Φ is the cumulative standard normal distribution function. If $\mu \geq 0$, then $\Lambda_{\mathrm{BM}}(0) = 0$ and the distribution of $T|X, V$ is nondefective for positive $\phi(X)V$. If $\mu < 0$, however, $\Lambda_{\mathrm{BM}}(0) = -2\mu/\sigma^2 > 0$ 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.

Either way, the MHT model (1) specifies a mixed inverse Gaussian distribution for $T|X$ in this special case. Mixed inverse Gaussian distributions have been used to model duration data in the statistical literature. For example, Aalen and Gjessing (2001) propose such a model with parametric mixing over the Brownian motion's drift coefficient μ. This paper extends and adapts this literature with estimators that allow for more general latent processes and mixing distributions.

3 Likelihood Computation

3.1 Parameterization

Let ψ, ϕ and $\mathcal{L}$ be specified up to a finite vector of unknown parameters $\alpha \in \mathcal{A}$. Assume that this parameterization is one-to-one, so that α is uniquely determined by (ψ, ϕ, G). In the case that $\ln \phi(X) = \delta + X'\beta$ for some scalar intercept δ and $K \times 1$ vector of slope parameters β, for example, this requires the "rank condition" that the support $\mathcal{X}$ of X contains a nonempty open set in $\mathbb{R}^K$.

With such a parameterization, under mild additional conditions, Abbring's (2007) results imply that α is uniquely determined ("identified") from the distribution of $T|X$. In particular, it is sufficient that

(i). the scales of $\{Y\}$, $\phi(X)$, and V are appropriately normalized;

(ii). $\phi(X)$ is nondegenerate; and

(iii). either V has a finite mean or the latent process $\{Y\}$ is such that $0 < |\psi'(0+)| < \infty$.

Throughout, we assume that the first two conditions hold, and explicitly note the assumptions on $\mathcal{L}$ and ψ required to ensure that the third condition holds as well.

The first condition's scale normalizations are innocuous, but need to be carefully implemented in any estimation procedure. They are needed because the durations T implied by the first hitting-time specification (1) are not affected by rescaling both the latent process $\{Y(t)\}$ and the threshold $\phi(X)V$ by the same factor, nor by rescaling the threshold factors $\phi(X)$ and V without changing the threshold itself. Specifically, any two specifications $(\psi, \phi, \mathcal{L})$ and $(\tilde{\psi}, \tilde{\phi}, \tilde{\mathcal{L}})$; with $\tilde{\psi}(s) = \psi(cs)$, $\tilde{\phi} = (c/d)\phi$, and $\tilde{\mathcal{L}}(v) = \mathcal{L}(dv)$ for some $c, d > 0$; are observationally equivalent. Stated differently, if $(\psi, \phi, \mathcal{L})$ corresponds to a latent process $\{Y\}$ and threshold $\phi(X)V$; and $(\tilde{\psi}, \tilde{\phi}, \tilde{\mathcal{L}})$ corresponds to a latent process $\{cY\}$, an observed threshold factor $c\phi(X)/d$, and an unobserved threshold factor dV; then the corresponding first hitting times are the same:

$$\inf\{t \geq 0 : Y(t) > \phi(X)V\} = \inf\{t \geq 0 : cY(t) > (c/d)\phi(X)dV\}$$

Identification therefore requires that the scale of two of $\{Y\}$, $\phi(X)$ and V are normalized. The most convenient way of implementing these two normalizations depends on the chosen parameterization, and will be discussed as we go.

The second condition ensures that the threshold varies with the regressors on their support. Such variation is key to the separate identification of the latent process and hetereogeneity. Abbring (2007) provides the following simple example of two MHT models without covariates ($\phi(X) \equiv 1$) that induce the same distribution of T. Both a model in which $\{Y\}$ is a Brownian motion with drift and V is degenerate at a single threshold value (that is, without heterogeneity) and a model in which $\{Y\}$ is degenerate linear drift ($\sigma = \Upsilon = 0$) and V has an inverse Gaussian distribution lead to an inverse Gaussian distribution of T.

The third condition is reminiscent of the conditions for identifiability of the mixed proportional hazards model. Abbring (2007) provides extensive discussion.

We also require that the parameterization of $(\psi, \phi, \mathcal{L})$ is sufficiently smooth to allow for the application of standard asymptotic theory. The choice of an appropriate parameterization of ψ is particularly important. We further discuss this in the context of specific parameterizations in Section 4.

3.2 Sampling

We explicitly deal with censoring, which is a common problem in applied duration analysis. Let $\{(T_1^*, X_1), \dots, (T_N^*, X_N)\}$ be a (complete) random sample from the distribution of (T, X) induced by the MHT model at the "true" parameter vector $\alpha_0 \in \mathcal{A}$ and some marginal distribution of X. We do not directly observe this complete sample, but only a censored version of it: $\{(T_1, D_1, X_1), \dots, (T_N, D_N, X_N)\}$. Here, $T_i \equiv \min\{T_i^*, C_i\}$ is the observed duration and $D_i \equiv I(T_i^* \leq C_i)$ a censoring indicator, for some random censoring time C_i; $i = 1, \dots, N$.

For expositional convenience, we focus on a simple type of independent right-censoring (Andersen et al., 1993). Assume that the complete observations (T_i^*, C_i, X_i) are independent across i and that, conditional on X_i, C_i is independent of T_i^*. That is, censoring times are not informative on the durations of interest. For example, if data are only collected for a deterministic time C_i, then C_i is trivially independent of T_i^*. The independent censoring assumption ensures that the likelihood of the observed durations T_i conditional on (C_i, X_i) only depends on the parameters α of the MHT model. We take the marginal distributions of the (C_i, X_i) to be ancillary, and focus on estimation of α_0 by maximizing this conditional likelihood.

With more general independent right censoring schemes, the resulting estimator remains a valid (but often, partial) likelihood estimator (Andersen et al., 1993). Moreover, the likelihood, and the corresponding estimator, can easily be adapted to other practically relevant sampling schemes, such as those involving interval censoring.

In the next section, we first consider the Gaussian special case. This allows us to discuss some practical details concerning normalizations in a well-understood framework in which the likelihood can be explicitly given. Section 3.4 then discusses likelihood computation in the general case.

3.3 Gaussian Special Case

Suppose that $\{Y\}$ is a Brownian motion with drift, so that, by the analysis in Section 2.3, $T|X$ has a mixed inverse Gaussian distribution. Because $|\psi'(0+)| = |\mu|$ in this case, identification of α_0 can be guaranteed by either assuming that G has finite mean or that $\mu \neq 0$ (Abbring, 2007).

In this special case, the log likelihood $\ell_N(\alpha)$ of α for $(T_1, \ldots, T_N)|\{(D_1, X_1), \ldots, (D_N, X_N)\}$ can be constructed using the explicit expression for the density and survival functions of $T|X, V$ in (6) and (7):

$$\ell_N(\alpha) = \sum_{i=1}^{N} \ln \int \theta_{\text{BM}}(T_i|X_i, v)^{D_i} \overline{F}_{\text{BM}}(T_i|X_i, v) dG(v), \tag{8}$$

with $\theta_{\text{BM}} \equiv f_{\text{BM}}/\overline{F}_{\text{BM}}$ the hazard rate corresponding to f_{BM}. Here, the dependence of θ_{BM} and $\overline{F}_{\text{BM}}$ (through μ, σ, and ϕ) and G on the parameter vector α is kept implicit. Under standard regularity conditions, the maximizer $\hat{\alpha}_N$ of $\ell_N(\alpha)$ is a consistent and asymptotically normal estimator of α_0. The estimator's asymptotic covariance matrix can be estimated in the standard way using either the score or Hessian characterization of the Fisher information matrix. It is asymptotically efficient under the assumption that the marginal distribution of X and the censoring times carry no information on α_0.

A typical parameterization would specify $\ln \phi(X) = \delta + X'\beta$, and a mixing distribution G that has finite support $\{v_1, \ldots, v_L\}$, for some fixed $L \in \mathbb{N}$, with parameters

$$\pi_l \equiv \Pr(V = v_l) = G(v_l) - G(v_l-); \qquad l = 1, \ldots, L. \tag{9}$$

A finitely discrete specification of G is popular because of its versatility and computational convenience; it also appears naturally in Heckman and Singer's (1984) influential work on semiparametric estimation of the MPH model. With it, the log likelihood in (8) reduces to

$$\ell_N(\alpha) = \sum_{i=1}^{N} \ln \sum_{l=1}^{L} \pi_l \theta_{\text{BM}}(T_i|X_i, v_l)^{D_i} \overline{F}_{\text{BM}}(T_i|X_i, v_l),$$

which is easy to compute using (6) and (7). In this parameterization, the two normalizations required can be implemented by setting $\delta = 0$, and setting $v_1 = 1$ with $\pi_1 > 0$. In the case that $\mu \neq 0$ is assumed, one of these normalizations can be replaced by a normalization of μ, such as $|\mu| = 1$.

The maximum likelihood estimator for the Gaussian special case of the MHT model and its asymptotic distribution are as easy to compute as, say, the maximum likelood estimator of the mixed proportional hazards model. In particular, with a computationally convenient specification of G like the discrete example above, explicit expressions for the likelihood and its derivatives are available; and computation can proceed directly by a search for a likelihood maximizer using standard numerical methods. The Gaussian special case shares this feature with many of the models studied in the statistics literature (Lee and Whitmore, 2006). In the general Lévy case or with general heterogeneity distributions, however, such explicit expressions are not available, and maximum likelihood cannot be implemented directly. The next section develops methods for computing the maximum likelihood estimator and its asymptotic distribution in this general case.

3.4 General Case

In general, the density and survival function of $T|X$ are not explicitly known, but can be computed by numerically inverting their Laplace transforms. We will develop fast and effective methods for computing the likelihood; its maximizer, the ML estimator; and its derivatives by adapting existing results for inverting the Laplace transform of the first hitting time of a Lévy process. We focus on the case with a nontrivial Gaussian component: $\sigma > 0$.

Our approach is based on the work of Rogers (2000), who applies a variant of Abate and Whitt's (1992) inversion method to the problem of calculating the first-passage-time distribution of a spectrally one-sided Lévy process. This approach builds on the fact that the Laplace transform $\mathcal{L}_T(\cdot|X) = \mathcal{L}[\Lambda(s)\phi(X)]$ of $T|X$ in (4) represents a one-to-one transformation of the probability density function $f(\cdot|X)$ of $T|X$,

$$\mathcal{L}[\Lambda(s)\phi(X)] = \int_0^\infty \exp(-st)f(t|X)dt. \tag{10}$$

The probability density function $f(\cdot|X)$ can be obtained by inverting this transformation using *Mellin's inverse formula* (see Davies, 2002),

$$f(t|X) = \frac{1}{2\pi \mathrm{i}} \lim_{R\to\infty} \int_{\gamma_R} \exp(st)\mathcal{L}[\Lambda(s)\phi(X)]\, ds. \tag{11}$$

Here, the integration is along the path $\gamma_R : u \in [-1, 1] \mapsto \gamma + \mathrm{i}Ru$, which traces out a straight line in $\mathbb{C}$, parallel to the imaginary axis from $\gamma - \mathrm{i}R$ to $\gamma + \mathrm{i}R$. Its parameter $\gamma \in \mathbb{R}$ should, in general, be chosen such that it is larger than the real part of any singularity in the Laplace transform $\mathcal{L}_T(\cdot|X)$. Because $\mathcal{L}_T(\cdot|X)$ is analytic for any s with nonnegative real part, we can choose any $\gamma \geq 0$.

The integral in (11) does not generally have an explicit solution, but can be efficiently approximated using numerical methods. A key complication is that our specification of $\mathcal{L}_T(\cdot|X)$ involves the inverse function Λ, which cannot generally be expressed in closed form. To circumvent this problem, we follow Rogers (2000) and integrate along the transformed path $\tilde{\gamma}_R = \psi \circ \Lambda_{\mathrm{BM}} \circ \gamma_R$ instead, which traces out a curve in $\mathbb{C}$ from $\psi[\Lambda_{\mathrm{BM}}(\gamma - iR)]$ to $\psi[\Lambda_{\mathrm{BM}}(\gamma + iR)]$ (where $\circ$ denotes function composition). Here, ψ is again the Laplace exponent of the latent process $\{Y\}$ and Λ_{BM} the inverse of the Laplace exponent of its Brownian motion component, for which (5) gives an explicit expression. Note that Λ_{BM} necessarily has the same dispersion parameter σ as ψ, but that its drift parameter is not uniquely pinned down (because the drift parameter of ψ depends on the way we deal with small shocks; see Section 2.2). Fortunately, the exact value of the drift parameter of Λ_{BM} plays no role in the argument that follows. It can generally be set to the drift parameter in the specific parameterization of ψ used; for example, $\tilde{\mu}$ in (2) or μ in (3). The MATLAB code accompanying this paper applies to specifications of ψ with compound Poisson jumps and sets the drift parameter of Λ_{BM} equal to μ in (3) (see Section 4).

Rogers (2000) shows that the transformed path $\tilde{\gamma}_R$ is close enough to γ_R, so that we can integrate along $\tilde{\gamma}_R$ in (11) instead. This gives an expression for $f(\cdot|X)$ that does not involve Λ:

$$\begin{aligned} f(t|X) &= \frac{1}{2\pi \mathrm{i}} \lim_{R\to\infty} \int_{\tilde{\gamma}_R} \exp(st)\mathcal{L}\left[\Lambda(s)\phi(X)\right] ds \\ &= \frac{1}{2\pi \mathrm{i}} \lim_{R\to\infty} \int_{\gamma_R} \exp\left[\psi\left\{\Lambda_{\text{BM}}(s)\right\} t\right] \mathcal{L}\left[\Lambda_{\text{BM}}(s)\phi(X)\right] d\psi[\Lambda_{\text{BM}}(s)]. \end{aligned} \tag{12}$$

This convenient change of integration path is valid because the differences of the end points of the curves mapped out by γ_R and $\tilde{\gamma}_R$ converge to zero as R grows large. In particular, because $\sigma > 0$,

$$\left|\frac{\gamma_R(1) - \tilde{\gamma}_R(1)}{\gamma_R(1)}\right| = \left|\frac{\gamma + \mathrm{i}R - \psi\left[\Lambda_{\text{BM}}\left(\gamma + \mathrm{i}R\right)\right]}{\gamma + \mathrm{i}R}\right| = \left|\frac{\psi_{\text{BM}}(z_R) - \psi(z_R)}{\psi_{\text{BM}}(z_R)}\right|,$$

with $z_R \equiv \Lambda_{\text{BM}}\left(\gamma + \mathrm{i}R\right)$, converges to zero as $R \to \infty$. The same result can be obtained for the other end point $\gamma_R(-1)$ if we instead take $z_R \equiv \Lambda_{\text{BM}}\left(\gamma - \mathrm{i}R\right)$.

Following Abate and Whitt (1995), we can apply the trapezoidal rule, and approximate the integral on the right hand side of (12) with the sum

$$\begin{aligned} &S_R(t|X) \equiv \frac{h}{2\pi} \sum_{r=-R}^{R} g(t, r|X), \text{ where} \\ &g(t, r|X) \equiv \exp\left\{\psi\left[\Lambda_{\text{BM}}\left(\gamma + \mathrm{i}rh\right)\right]t\right\} \mathcal{L}\left[\Lambda_{\text{BM}}(\gamma + \mathrm{i}rh)\phi(X)\right] \frac{d}{ds}\psi\left[\Lambda_{\text{BM}}(s)\right]\Big|_{\gamma+\mathrm{i}rh}. \end{aligned} \tag{13}$$

The error introduced by this discretization of the integral is bounded by

$$\frac{h^2 R}{6} \sup_{r \in (-R,R)} |g''(t, r|X)|$$

where $g''(t, r|X)$ denotes the second derivative of $g(t, r|X)$ with respect to r. This is a standard result for integration using the trapezoidal rule, and its application to the current problem is discussed in Abate and Whitt (1995). They argue that the actual error is likely to be much lower for our application as the integrand $g(t, r|X)$ oscillates and the approximation errors tend to cancel out. By reducing the step size h we can make the approximation arbitrarily precise.

Because $S_R(t|X)$ is a nearly alternating series in R, the limit $\lim_{R\to\infty} S_R(t|X)$ can be efficiently approximated using Euler summation:

$$f(t|X) \approx E(R, M, t|X) \equiv \sum_{m=0}^{M} 2^{-M} \binom{M}{m} S_{R+m}(t|X), \tag{14}$$

for some $M, R \in \mathbb{N}$. Abate and Whitt (1995) find that for most probability densities the error introduced by approximating the limit $R \to \infty$ by an Euler summation is well estimated by $E(R, M + 1, t|X) - E(R, M, t|X)$. In our case, this estimated error quickly tends to zero as M is increased, suggesting the approximation is accurate.

The log likelihood function of a sample of complete durations and covariates from an MHT model with parameters α can be computed by combining the individual approximate probabilities from (14) into the sum of their logarithms,

$$\ell_N(\alpha) = \sum_{i=1}^{N} \ln f(T_i|X_i) \approx \sum_{i=1}^{N} \ln(E(R, M, T_i|X_i)) \tag{15}$$

It is straightforward to extend this approach to independently censored data. The computation of the log likelihood contribution of a censored observation requires the computation of the survival function $\overline{F}(\cdot|X)$ at the censoring time and the corresponding covariate value. This survival function can be approximated along the lines above, using that the Laplace transform of $\overline{F}(\cdot|X)$ can be explicitly expressed in terms of the known transform $\mathcal{L}_T(\cdot|X)$ of $f(\cdot|X)$. In particular, using integration by parts, it is easy to show that

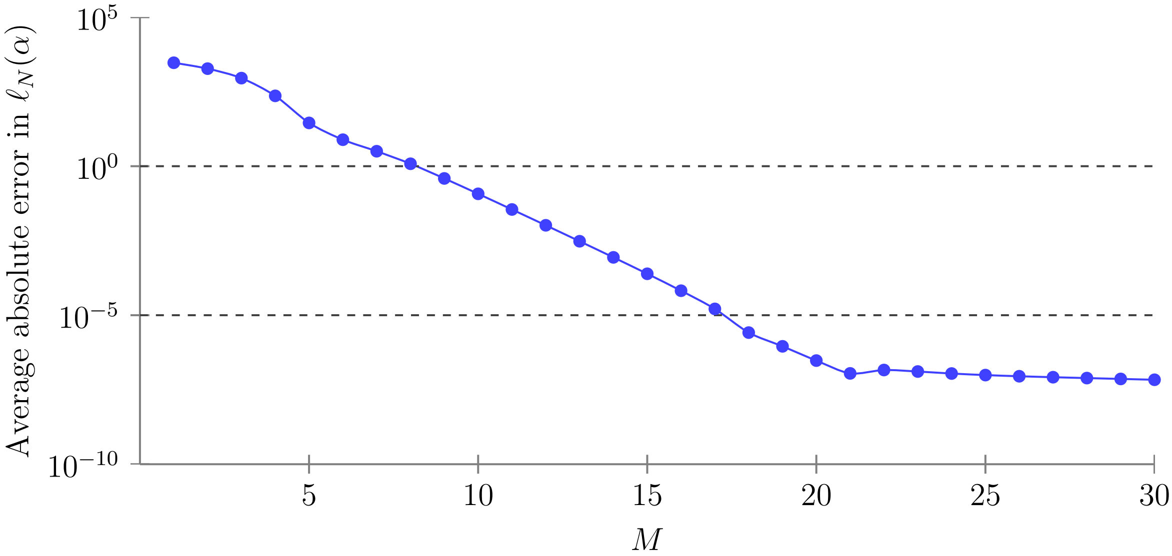
$$\frac{1-\mathcal{L}\left[\Lambda(s)\phi(X)\right]}{s} = \int_0^\infty \exp(-st)\overline{F}(t|X)dt = \frac{1-\mathcal{L}_T\left(s|X\right)}{s}.$$

With (4), this allows us to express a known function of the model's parameters as the Laplace transform of the survival function $\overline{F}(\cdot|X)$, analogously to the expression for the density in (10). This transformation can be numerically inverted to compute the survival function, and the likelihood contribution of each censored observation, using the strategy developed for the density. One minor difference is that the Laplace transform of the survival function may have a singularity at 0 if the durations do not have a (finite) mean; then, it is necessary to set $\gamma > 0$.

We approximate the score and Hessian of the log likelihood with the analytical first and second derivatives of the approximate log likelihood function. These exist and are well behaved because our approximation of the log likelihood function in (15) is smooth in the parameters.

The implementation of this method for computing the likelihood and its derivatives requires that we set the parameters that control the approximation in (14): γ, h, R, and M. Rogers (2000) provides guidance. We find that his suggestions for γ and h, $\gamma = 11/t$ and $h = 1/t$, yield good numerical performance in our case. We will adopt these as our default settings, together with $R = 6$ and $M = 15$, which Rogers claims provide a good accuracy to speed trade-off. As discussed below, additional accuracy can be obtained when needed by setting M higher.

Figure 1: Approximation Error of the Log Likelihood for Various M



Note: This figure is based on the log likelihood $\ell_N(\alpha)$ of an MHT model with a Brownian motion latent process and discrete unobserved heterogeneity with three support points for Kennan's (1985) complete strike duration data. It plots the average absolute difference between $\ell_N(\alpha)$ and its numerical approximation over 100 randomly drawn parameter values α, for a range of values of M. The errors are plotted on a logarithmic scale. The parameters are generated using our method of setting starting values for maximum likelihood estimation. This method sets the drift and variance parameters equal to their maximum likelihood estimates for a simple inverse Gaussian model with $\phi(X)V = 1$, which are known in closed form. Starting values for the support points v_l of the heterogeneity distribution are generated by exponentiating draws from a standard normal distribution. This ensures that the v_l vary in level, but are all approximately of the right scale. All three support points v_l receive probability mass 1/3. The parameter β multiplying the covariates is set to zero. For the current experiment, we found that setting the parameters to their final maximum likelihood estimates instead produced almost identical results.

3.5 Numerical Experiments

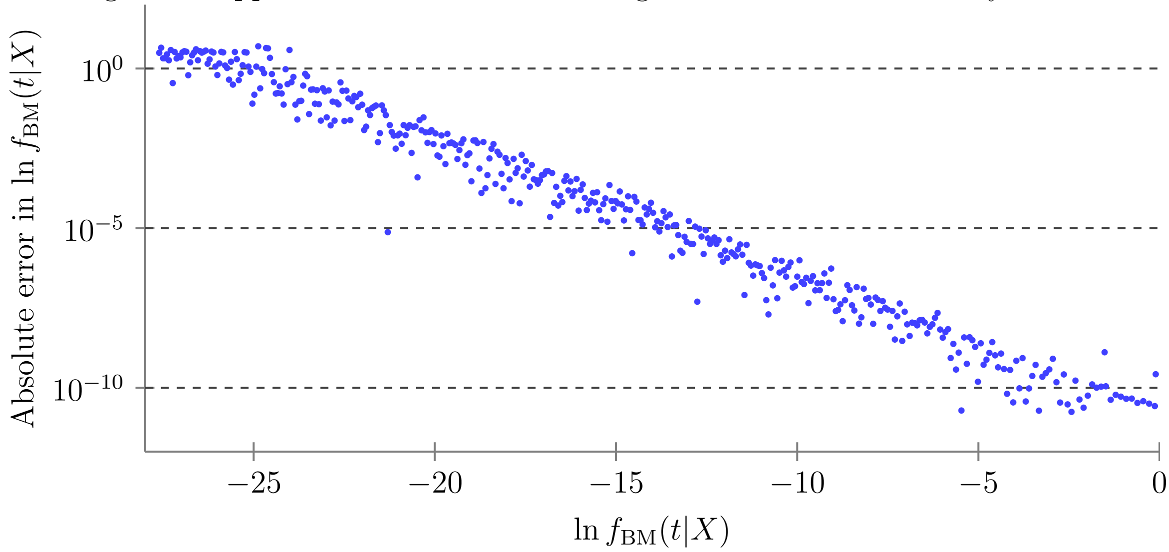
We have investigated the accuracy of the proposed likelihood approximation by conducting a range of numerical experiments. We discuss the results of two of these experiments here. Both experiments use the default settings for the parameters that control the approximation, unless explicitly stated otherwise.

The first experiment compares direct computations of the log likelihood function of the mixed inverse Gaussian model using the explicit expression for the density in (6) to its numerical approximations as we vary M. The log likelihood is calculated on the data set that we use in Section 5. This ensures that this experiment provides both a real life test case and a check on the results we present in that section. The data contain 566 complete strike durations. Because the approximation errors are close to unbiased, the error in the log likelihood scales with the root of the sample size.

Figure 1 plots the average of the absolute approximation error of the log likelihood, for different values of M, over a large set of model parameters randomly generated at the scale of their maximum likelihood estimates. We find that this average absolute error decreases exponentially with M; this result is robust across the various parameter values over which the plotted results are averaged. Consistently with Rogers (2000), we see that $M = 15$ already provides a decent approximation for most practical purposes. However, because the time required for the calculations grows only linearly in M, an extra thousandfold increase in precision can be obtained at a very low computational cost by setting $M = 20$ instead. Once $M > 20$, other factors, such as rounding errors, become important, and the approximation error levels off. We also find that, with $M = 20$, increasing R or decreasing the step size h adds very little to the precision of the inversion. The numerical approximation of the log likelihood takes about 15–20 times as long to calculate as the analytical expression. However, in absolute terms this is still very manageable: A maximization of the log likelihood function can be performed in under a minute on a regular computer for most model specifications.

The second experiment takes a closer look at the numerical approximation of the density f_{BM} of a basic inverse Gaussian model with parameters such that $\mu = \sigma^2 = \phi(X)V = 1$. We only present results for $M = 25$, but found very similar results for any $M \geq 15$. For the purpose of maximum likelihood estimation, we care most about the errors in the approximation of the *log* density, $\ln f_{\text{BM}}$. Figure 2 plots the absolute error of this approximation against the log density itself, on a logarithmic scale. The (log-)linear relation displayed by the graph implies that the absolute error in the approximation of $\ln f_{\text{BM}}(t|X)$ roughly equals $10^{-11}/f_{\text{BM}}(t|X)$. Consequently, the approximation error is generally small, but the approximation breaks down when the density gets very small (say,

Figure 2: Approximation Error of the Log Inverse Gaussian Density Function



Note: This figure plots the absolute difference between the log inverse Gaussian density $\ln f_{\mathrm{BM}}(t|X)$ with parameters $\mu = \sigma^2 = \phi(X)V = 1$ and its numerical approximation, on a logarithmic scale, against $\ln f_{\mathrm{BM}}(t|X)$, for a range of times t.

$f_{\mathrm{BM}}(t|X) < 10^{-10}$, or $\ln f_{\mathrm{BM}}(t|X) < -23$). When estimating the model with maximum likelihood, we can easily avoid this by setting reasonable starting values for the parameters. This ensures that the approximation is sufficiently precise for numerically robust maximum likelihood estimation.

4 Maximum Likelihood Estimation

This paper is accompanied with MATLAB code that implements a maximum likelihood estimator based on the previous section's approximate likelihood. We maximize this likelihood by means of a quasi-Newton algorithm with BFGS updates for the Hessian (see Nocedal and Wright, 2006). We use the analytical derivatives of the approximate likelihood to ensure quick and stable maximization, and to construct asymptotic standard errors.

We have implemented a range of computationally feasible, flexible parameterizations of the model. This section's remainder discusses these parameterizations.

4.1 Latent Process

We consider two parameterizations of the latent process $\{Y\}$. Both include a Brownian motion component with $\sigma > 0$.

The main specification specifies that $\{Y\}$ is a convolution of a nondegenerate Brownian motion with drift and a compound Poisson process with a finitely discrete shock distribution. Because $\int_{(-1,0)} x\Upsilon(dx) < \infty$ in this case, the Lévy-Khintchine formula (3) now offers the simplest way to parameterize ψ:

$$\psi(s) = \mu s + \frac{\sigma^2}{2}s^2 + \sum_{q=1}^{Q} \lambda_q \left(\mathrm{e}^{s\nu_q} - 1\right),$$

where μ and $\sigma^2 \geq 0$ are the Brownian drift and variance per time unit, and λ_q is the Poisson rate at which shocks of size $\nu_q < 0$ arrive; $q = 1, \ldots, Q$. Equivalently, in this specification, shocks arrive at a rate $\lambda \equiv \sum_{q=1}^{Q} \lambda_q$ and are drawn independently from a distribution with Q points of support $(\nu_1, \ldots, \nu_Q)$ with probabilities $(\lambda_1/\lambda, \ldots, \lambda_Q/\lambda)$.

An alternative is to specify $\{Y\}$ as a convolution of a nondegenerate Brownian motion with drift and a compound Poisson process with a gamma shock distribution. In this specification, shocks arrive at a Poisson rate λ, with their absolute sizes distributed according to a two-parameter gamma distribution $\Gamma_{\nu,\rho}$, with corresponding density

$$\frac{\nu^\rho}{\Gamma(\rho)} x^{\rho-1} \exp(-\nu x); \qquad \nu, \rho > 0;$$

and Laplace transform

$$\mathcal{L}_{\Gamma_{\nu,\rho}}(s) = \frac{1}{(s/\nu + 1)^\rho}. \tag{16}$$

We can again use (3), which now gives

$$\psi(s) = \mu s + \frac{\sigma^2}{2}s^2 + \lambda \left\{ \frac{1}{(s/\nu + 1)^\rho} - 1 \right\}.$$

4.2 Effect of the Observed Covariates

The threshold is naturally specified to be loglinear in the covariates:

$$\phi(X) = \exp(\delta + X\beta).$$

We assume that the $N \times (K+1)$ matrix with sampled observations of $(1\ X')$ in each row has full column rank.

4.3 Unobserved Heterogeneity

Finally, our procedure for computing the likelihood only depends on the unobserved heterogeneity distribution G through its Laplace transform $\mathcal{L}$. Therefore, any distribution with nonnegative support that admits an explicit expression for its Laplace transform is a convenient candidate for G. We consider two such specifications.

The main specification is Section 3's finitely discrete distribution. The corresponding Laplace transform is

$$\mathcal{L}(s) = \sum_{l=1}^{L} \pi_l \exp(-s v_l).$$

A simple and low-dimensional alternative is to specify a gamma distribution $\Gamma_{\omega,\tau}$ for G. Analogously to (16), this gives

$$\mathcal{L}(s) = \frac{1}{(s/\omega + 1)^{\tau}}.$$

4.4 Scale Normalizations

Recall from Section 3.1 that we need to normalize the scales of two out of ψ, ϕ, and $\mathcal{L}$. The MATLAB code currently normalizes the covariate effects $\phi(X)$ by setting $\delta = 0$, and ψ by setting $\mu = 1$. Note that this implicitly assumes that $\mu > 0$. It would be straightforward to adapt the code to allow more generally for $|\mu| = 1$.

One of these normalizations can be replaced by a normalization on $\mathcal{L}$. A discrete unobserved heterogeneity distribution can, for example, be normalized by requiring $v_1 = 1$ and $\pi_1 > 0$. A gamma distribution can be normalized by setting its scale parameter $\omega = 1$.

5 Strike Durations

The mere existence of nontrivial delays in labor agreements has puzzled economists; duration patterns in their resolution have been studied to learn more about underlying bargaining games and information structures.

Lancaster (1972) analyzes strike durations using a Gaussian MHT model with regressors, but without unobserved heterogeneity. He interprets the gap between the Brownian motion and the threshold as the level of disagreement, and concludes that this model fits his data for the United Kingdom well. Others have used proportional hazards models to study strike durations. Kennan (1985), in particular, shows that the US strike duration hazard is U-shaped and takes this as evidence against Lancaster's (homogeneous) MHT model. He notes that this aspect of the data can be interpreted in terms of heterogeneity in the conflicts underlying the strikes, but does not subsequently pursue this in his empirical analysis.

Here, we will investigate whether Kennan's strike data can be matched well by a more general MHT model that explicitly takes into account unobserved heterogeneity in strikes. Such a model comes with Lancaster's attractive interpretation in terms of a level of disagreement that may both vary over time and may initially be heterogeneous between strikes. We will explicitly discuss our estimation results in terms of this interpretation, with an implicit understanding that it is our modest objective to illustrate our methods and the descriptive and potential structural appeal of the MHT model, without providing a fully structural analysis of strike durations.

Kennan's data cover all contract strikes in US manufacturing in the period 1968–

1976 that involved at least a thousand workers, and that were classified to be primarily about "general wage changes". They include the durations in days of 566 strikes and, for each strike, a measure of the state of the business cycle in the month it started: The residuals of a regression of log industrial production in US manufacturing on linear and quadratic trend terms and seasonal dummies. We obtained the data in a fixed format text file `strkdur.asc` from Cameron and Trivedi's (2005) web page. We divided all strike durations by seven, so that they are measured in weeks.

Table 1 reports maximum likelihood estimates for a range of Section 4's flexible parameterizations. All reported estimates are computed using Section 3.4's numerical methods, with $M = 25$. To further check these methods and their MATLAB implementation, we have also computed the same estimates for lower values of $M \geq 15$ (not reported), and estimates for the first five specifications using the explicit expressions for the log likelihood that are available in these cases (not reported). These results are virtually identical to those reported in Table 1.

In all cases, we specify $\phi(X) = \exp(X\beta)$, with X the scalar business cycle indicator. Columns I–V presents estimates of models with Brownian motion latent processes and discrete unobserved heterogeneity. Throughout, the drift is normalized to 1 per week ($\mu = 1$), so that $\mathbb{E}\left[T|X,V\right] = -\mathcal{L}'_T(0+|X,V) = \exp(X\beta)V$. By its construction as a regression residual, X varies around zero and is close to zero on average in the sample. Consequently, V can be interpreted as the unobserved initial level of disagreement, measured as the mean number of strike weeks it commands.

The log likelihood substantially improves when adding a second, third and fourth support point to the distribution of V, between Columns I and IV, but a fifth support point (Column V) hardly changes the fit and the other parameters' estimates. The estimates indicate that there is both substantial heterogeneity in the strikes' initial levels of disagreement and uncertainty in their evolution over time. The numbers in Column IV imply that there are four unobserved types of labor conflict, on average commanding

Table 1: Maximum Likelihood Estimates for Kennan's (1985) Strike Duration Data

	I	II	III	IV	V	VI	VII
μ	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
σ^2	19.6592 (3.1752)	6.2185 (0.8702)	2.0675 (0.4433)	1.2272 (0.2423)	1.1966 (0.2224)	0.5423 (0.2808)	5.1469 (0.9768)
λ						0.0186 (0.0183)	
ν						−5.1321 (2.3211)	
β	−0.9306 (0.6010)	−1.7722 (0.6855)	−1.0846 (0.6572)	−0.8669 (0.6514)	−0.8623 (0.6338)	−0.5788 (0.6148)	−2.1198 (0.7881)
ω							0.4446 (0.0730)
τ							2.7911 (0.4373)
v_1	6.2603 (0.4688)	2.5431 (0.1993)	1.5369 (0.1508)	1.1045 (0.1213)	1.0312 (0.1644)	0.7546 (0.1602)	
v_2		8.7509 (0.5194)	5.8883 (0.3999)	3.2094 (0.4531)	1.7564 (1.0282)	2.0832 (0.5127)	
v_3			18.1612 (1.0108)	7.1654 (0.5598)	3.5180 (0.7618)	4.1380 (0.8364)	
v_4				18.5572 (0.7028)	7.3032 (0.6467)	7.4121 (0.5533)	
v_5					18.5749 (0.6945)	17.0035 (1.1016)	
pi_1	1 (0)	0.3991 (0.0439)	0.3534 (0.0335)	0.2519 (0.0380)	0.1986 (0.1160)	0.1978 (0.0398)	
pi_2		0.6009 (0.0439)	0.4923 (0.0347)	0.2826 (0.0507)	0.0981 (0.1300)	0.2009 (0.0688)	
pi_3			0.1543 (0.0231)	0.3146 (0.0541)	0.2561 (0.0825)	0.2230 (0.0617)	
pi_4				0.1508 (0.0191)	0.2969 (0.0646)	0.2379 (0.0609)	
pi_5					0.1503 (0.0190)	0.1403 (0.0200)	
ℓ_N	−1658.9	−1588.7	−1583.0	−1576.3	−1576.1	−1575.4	−1594.2

Note: The drift is normalized to 1 per week. All specifications include a single covariate, Kennan's (1985) deseasonalized and detrended log industrial production. Asymptotic standard errors are in parentheses.

respectively 1.10, 3.21, 7.17, and 18.56 strike weeks. Each type's level of disagreement evolves with a standard deviation per week just above the unit drift towards agreement.

It is instructive to note that the variance of the latent process drops substantially, from close to 20 to just over 1, when more heterogeneity is added between Columns I and IV. Clearly, Column I's specification falsely attributes heterogeneity in the strikes' initial levels of disagreement to uncertainty in their evolution over time.

The estimates of the coefficient β reflect the effect of the business cycle on strike durations. In line with Kennan's (1985) results, strikes that begin in months with low production last longer. In the MHT model, this is captured by a countercyclical threshold: In times with low production, in expectation, conflicts command more strike days. One interpretation is that strike days are less costly in times with low production. The precision of the estimates of β is low. This is consistent with Kennan's results. He obtains more precise results with a binary cyclical indicator constructed from the indicator used here. For simplicity, we do not follow this lead here.

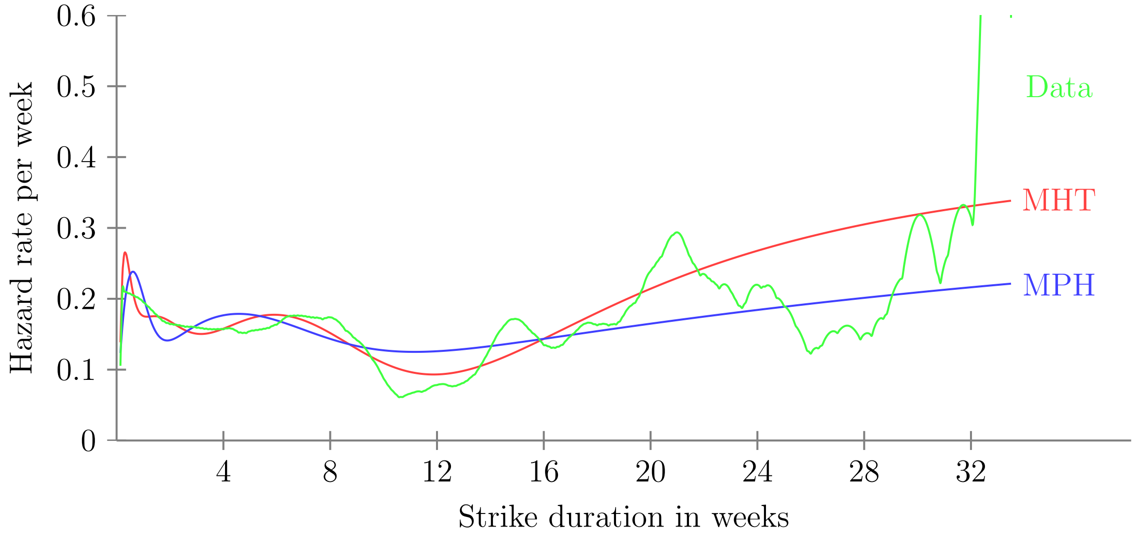
Column VI reports an estimate of a specification that includes discrete shocks of size ν at Poisson times. The estimates point to an infrequent shock that sets back just over five weeks of drift towards agreement. The shock only somewhat improves the likelihood; a specification without shock, such as those in Columns IV and V, seems to be sufficient.

A very similar result is found with a gamma shock at a Poisson time (not reported). With this specification, virtually the same estimate of the arrival rate of the shocks is obtained. Moreover, the estimated gamma shock distribution is close to degenerate at Column VI's estimate of the shock size (ν). Specifically, the estimates of the shape (ρ) and scale (ν) parameters of the gamma distribution are both very large, and their ratio equals Column VI's estimated shock size. As expected, the same log likelihood is found.

Finally, Column VII reports estimates of a specification with gamma heterogeneity. This specification is clearly inferior to that with any amount of discrete heterogeneity.

Figure 3 plots the aggregate hazard implied by the MHT model's estimates in Column

Figure 3: Aggregate Strike End Hazard Rates



Note: This graph plots the empirical strike end hazard rate (Data), computed with Epanechnikov kernel smoothing from Kennan's (1985) data, and the corresponding hazards implied by estimated MHT and MPH models. For the MHT model, the ML estimates in Table 1 for a specification with a latent Brownian motion and a discrete unobserved heterogeneity distribution with four support points are used. For the MPH model, we use ML estimates of a model with the same discrete heterogeneity distribution and a Weibull baseline. Estimated hazard rates of the unconditional distribution of T are plotted, based on the estimated distributions of $T|X$ implied by the models and the empirical distribution of the covariate X.

IV of Table 1. It also plots the hazard implied by estimates a MPH hazard model with a Weibull baseline and a discrete heterogeneity distribution with four support points. Note that this MPH specification has exactly the same number of parameters as Column IV's MHT specification.[1] In both cases, we computed the distribution of $T|X$ implied by these estimates, integrated over the empirical distribution of X, and computed and plotted the hazard rate of the resulting distribution. Figure 3 also plots the empirical hazard rate, computed by kernel smoothing the raw data.

The MHT model fits the empirical hazard well. The MPH model's fit seems to be slightly worse. This is confirmed by the MPH model's log likelihood, which, at -1583.4, is more than seven points lower. Because the Weibull baseline is monotonic, the Weibull MPH model can only fit the nonmonotonic strike hazard by compensating an increasing baseline hazard with negative duration dependence due to unobserved heterogeneity. Of

[1]However, estimates of two of the support points of the heterogeneity distribution converged to the same value.

course, usually MPH models with richer specifications of the baseline hazard are estimated and a sufficiently rich specification can fit the empirical hazard arbitrarily well.

6 Conclusion

The results in this paper enable applied researchers to analyze duration data with mixed hitting-time (MHT) models using standard likelihood-based estimation and inference methods. The MATLAB code for maximum likelihood estimation that accompanies this paper can directly be applied to either complete or independently right-censored duration data, and is easy to adapt to more general censoring schemes. Alternatively, the procedures for likelihood computation provided with this code can be used to implement other likelihood-based methods. For example, they can be combined with data augmentation and Markov chain Monte Carlo methods to implement a Bayesian estimator that can flexibly deal with unobserved heterogeneity.

Two types of empirical application of the MHT framework can be distinguished. First, it can be used as a descriptive framework, much like Cox's (1972) proportional hazards model and Lancaster's (1979) mixed proportional hazards model. Section 5's analysis of Kennan's (1985) strike data shows that estimates of the MHT model have descriptive appeal, with natural interpretations that nicely complement those that could be obtained from a proportional hazards analysis. Indeed, in statistics, there is increasing interest in the descriptive analysis of duration data with first hitting time models (Aalen and Gjessing, 2001; Lee and Whitmore, 2006; Singpurwalla, 1995; Yashin and Manton, 1997).

Second, it can be applied to the structural empirical analysis of heterogeneous agents' optimal stopping decisions. Abbring (2007) presents a range of examples, based on the type of optimal stopping models that are reviewed and analyzed in Boyarchenko and Levendorskiĭ (2007); Dixit and Pindyck (1994); Kyprianou (2006); Stokey (2009). These include McDonald and Siegel's (1986) model for the optimal timing of an irreversible investment; a model of unemployment durations based on Dixit's (1989) model of entry and exit, complemented with heterogeneity in transition costs; and a model of job separations with heterogeneous search. The identification results in Abbring (2007, 2010) show that data on durations and covariates are informative on the economic primitives of such models. The methods developed in this paper can be applied to measure those primitives.

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. 8, 27

Abate, J. and W. Whitt (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems 10*, 5–88. 13

Abate, J. and W. Whitt (1995). Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing 7*(1), 38–43. 14, 15

Abbring, J. H. (2002). Stayers versus defecting movers: A note on the identification of defective duration models. *Economics Letters 74*, 327–331. 4

Abbring, J. H. (2007). Mixed hitting-time models. Discussion Paper 07-57/3, Tinbergen Institute, Amsterdam. Revised, August 2009. 2, 4, 6, 8, 9, 11, 27, 28

Abbring, J. H. (2010). Identification of dynamic discrete choice models. *Annual Review of Economics 2*, Forthcoming. 28

Abbring, J. H. and T. Salimans (2010). Estimating mixed hitting-time models. Mimeo, CentER, Department of Econometrics & OR, Tilburg University, Tilburg. 2

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag. 10

Bertoin, J. (1996). *Lévy Processes*. Number 121 in Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press. 3, 4, 5, 6

Boyarchenko, S. and S. Levendorskiĭ (2007). *Irreversible Decisions under Uncertainty: Optimal Stopping Made Easy*. Berlin: Springer-Verlag. 27

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press. 23

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B 34*, 187–202. 1, 27

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

Davies, B. (2002). *Integral transforms and their applications*. Springer-Verlag. 13

Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy 97*(3), 620–638. 27

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

Heckman, J. J. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52*, 271–320. 11

Kennan, J. (1985). The duration of contract strikes in U.S. manufacturing. *Journal of Econometrics 28*, 5–28. 1, 2, 17, 22, 24, 25, 26, 27

Kyprianou, A. E. (2006). *Introductionary Lectures on Fluctuations of Lévy Processes with Applications*. Berlin: Springer-Verlag. 27

Lancaster, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society Series A 135*(2), 257–271. 22

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica 47*, 939–956. 1, 27

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. 1, 12, 27

McDonald, R. and D. Siegel (1986). The value of waiting to invest. *Quarterly Journal of Economics 101*(4), 707–728. 27

Nocedal, J. and S. J. Wright (2006). *Numerical Optimization*. Springer-Verlag. 19

Rogers, L. C. G. (2000). Evaluating first-passage probabilities for spectrally one-sided Lévy processes. *Journal of Applied Probability 37*(4), 1173–1180. 13, 14, 16, 18

Singpurwalla, N. D. (1995). Survival in dynamic environments. *Statistical Science 10*(1), 86–103. 27

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

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16*, 439–454. 1

Yashin, A. and K. Manton (1997). Effects of unobserved and partially observed covariate processes on system failure: A review of models and estimation strategies. *Statistical Science 12*, 20–34. 27