Article

Multivariate Threshold Regression Models with Cure Rates: Identification and Estimation in the Presence of the Esscher Property

Mei-Ling Ting Lee [1,*] and George A. Whitmore [2,3]

1   School of Public Health, University of Maryland, College Park, MD 20742, USA
2   Faculty of Management, McGill University, Montreal, QC H3A 0G4, Canada; george.whitmore@mcgill.ca
3   Ottawa Hospital Research Institute, Ottawa, ON K1Y 4E9, Canada
*   Correspondence: mltlee@umd.edu; Tel.: +1-301-405-4581; Fax: +1-301-314-6532

Abstract: The first hitting time of a boundary or threshold by the sample path of a stochastic process is the central concept of threshold regression models for survival data analysis. Regression functions for the process and threshold parameters in these models are multivariate combinations of explanatory variates. The stochastic process under investigation may be a univariate stochastic process or a multivariate stochastic process. The stochastic processes of interest to us in this report are those that possess stationary independent increments (i.e., Lévy processes) as well as the Esscher property. The Esscher transform is a transformation of probability density functions that has applications in actuarial science, financial engineering, and other fields. Lévy processes with this property are often encountered in practical applications. Frequently, these applications also involve a 'cure rate' fraction because some individuals are susceptible to failure and others not. Cure rates may arise endogenously from the model alone or exogenously from mixing of distinct statistical populations in the data set. We show, using both theoretical analysis and case demonstrations, that model estimates derived from typical survival data may not be able to distinguish between individuals in the cure rate fraction who are not susceptible to failure and those who may be susceptible to failure but escape the fate by chance. The ambiguity is aggravated by right censoring of survival times and by minor misspecifications of the model. Slightly incorrect specifications for regression functions or for the stochastic process can lead to problems with model identification and estimation. In this situation, additional guidance for estimating the fraction of non-susceptibles must come from subject matter expertise or from data types other than survival times, censored or otherwise. The identifiability issue is confronted directly in threshold regression but is also present when applying other kinds of models commonly used for survival data analysis. Other methods, however, usually do not provide a framework for recognizing or dealing with the issue and so the issue is often unintentionally ignored. The theoretical foundations of this work are set out, which presents new and somewhat surprising results for the first hitting time distributions of Lévy processes that have the Esscher property.

Keywords: bilateral gamma process; bilateral inverse gaussian process; cure rate; degradation process; esscher transform; exponential tilt; failure event; failure time; first hitting time; interval censoring; Lévy process; multidimensional; multivariate; poisson-bernoulli random walk; stationary independent increments; stochastic process; survival time; threshold regression; wiener process

1. Introduction

Event times for systems and their components are of research interest in many fields including medicine, engineering, the natural sciences, economics and the social sciences. These events are important milestones, happenings or outcomes such as deaths, hospitalizations, divorces, business bankruptcies, engineering failures, and so on. We will use the generic term 'failure' to describe the event and 'failure time' or 'survival time' to describe

the time to occurrence of the event. These failure times are often determined by an underlying stochastic process reaching a threshold, boundary or critical condition that triggers the failure event. The failure times are therefore first hitting times or first passage times. First-hitting-time models for events are ubiquitous in applications across diverse fields.

Study populations are usually a mixture of individuals, such as patients, marriages, business establishments, medical devices, plants, and the like. An individual in the study population may be at risk of eventual failure from the specific cause being investigated in the study. We refer to this type of individual as a susceptible. On the other hand, the study population may include an individual who is not at risk of failure; an individual we call a non-suspectible. Those individuals who are susceptible to failure may nevertheless escape eventual failure. They may avoid failure because their condition strengthens over time, their exposure to adverse conditions declines, or simple luck may be on their side. Susceptibles may also avoid failure of the type under study because they experience a failure event of a different type (often called a competing risk). If the non-susceptibles in a study population can be identified ahead they would usually not enter the study or at least be segregated for separate investigation. Often, however, it is not possible to know in advance (or ever) if a particular individual is truly non-susceptible. In this report, we are interested in the formulation, estimation and interpretation of survival models for study populations in diverse application fields that are a mixture of susceptibles and non-susceptibles who cannot be identified in advance.

We start our investigation by looking carefully at the idea of susceptibility to failure. We then present a setting for our investigation that involves a wide class of first hitting time models for study populations of mixed susceptibility. We examine several first hitting time models in some detail to reveal the underlying mechanisms at work. Our investigation includes the impact of regression structures on the interpretation of these first hitting time models, drawing upon the growing body of work on threshold regression. We also consider the important influences of interval- and right-censoring in survival data analysis. In the medical literature, we note that the proportion of a study population that will avoid eventual failure is often referred to as the cured subpopulation and the corresponding fraction of the study population as the cure rate. We will use this terminology even in settings where avoiding failure is not usually described as a 'cure'. Thus, a marriage is 'cured' if it does not end in divorce or a business establishment is 'cured' if it does not go bankrupt.

In our study we consider threshold regession applications that involve Lévy processes having the Esscher property (which is a common circumstance). In these contexts, the correct sign of the process drift parameter cannot be identified from censored failure data alone when the study population has some fraction of non-susceptibles. The subject matter under study or other kinds of data must provide additional information about the fraction of susceptibles, if any, who might escape failure by chance. Having censored survival data aggravates the situation. This susceptibility identification challenge can be faced squarely and analyzed in threshold regression methods. The ambiguity surrounding the susceptibility fraction originates in the subject matter of the study and therefore is present irrespective of the method chosen for survival data analysis. Aside from threshold regression, other methods for survival analysis generally provide no way to recognize or deal with the issue. As a consequence, the problem usually goes unaddressed when other methods are used.

2. Exogenous and Endogenous Cure Rates

Consider a univariate stochastic process $\{D(t), t \geq 0\}$ with $D(0) = 0$. Let S denote the first hitting time of a threshold at $d_0 > 0$ for this process. This first hitting event is the failure event and S is the failure time or survival time. We assume the study population has a proportion p of individuals who are at risk of experiencing the event dictated by this process. As already stated, we refer to these individuals as susceptibles because each individual may experience failure within the study's time horizon. The stochastic sample

paths of susceptibles are assumed to be probabilistically independent. The remaining proportion $1 - p$ of the study population is assumed to be not susceptible to this kind of failure because their individual experiences are not governed by stochastic process $\{D(t)\}$. We refer to proportion $1 - p$ as the exogenous cure rate. As an example, a medical study population may have a subpopulation of individuals who are at risk of contracting a progressive fatal disease (fraction p) while the remainder of the population are not at risk (fraction $1 - p$). As a second example, a study population may have a subpopulation of individuals who engage in a high-risk sport (say, mountain climbing) and are therefore exposed to a high-risk sport injury or death while the remainder of the population never participate in the high-risk sport. The setting just described implies that a randomly chosen individual from the study population has a mixed susceptibility survival function (s.f.), which we denote by $\overline{F}_R(s)$, where subscript R reminds us of the reference study population. This survival function has the following mixture form:

$$\overline{F}_R(s) = p\overline{F}(s) + 1 - p. \tag{1}$$

Here p denotes the susceptibility probability with $0 < p \leq 1$ and $\overline{F}(s)$ is the s.f. of susceptibles. Our notation suppresses the dependence of s.f. $\overline{F}(s)$ on model parameters.

As we are interested later in the behavior of $\overline{F}(s)$ as s goes to infinity, we use the following notation $1 - p_0 = \lim_{s \to \infty} \overline{F}(s) = \lim_{s \to \infty} P(S > s)$ for the relevant limit. Proportion p_0 is the fraction of susceptible individuals who would eventually experience failure if process $\{D(t)\}$ were allowed to operate indefinitely without competition from other causes of failure or encountering a study time horizon. Thus, a situation where $p_0 = 1$ is one where every susceptible individual would eventually experience failure if process $\{D(t)\}$ goes on without end. In real applications, of course, a failure time S is often unobservable because it is right censored by a competing failure event or the end of study follow-up. For some processes, there is no guarantee that a susceptibile individual will eventually experience failure, that is, $p_0 < 1$. In this situation, the individual is at risk of failure to a varying degree over time (i.e., is susceptible) but, in fact, may not experience the failure event, irrespective of the length of observation. This situation also results in a right-censored failure time with censoring produced by either a competing failure event or the end of study follow-up. The practical consequence is that susceptibles and non-susceptibles are lumped together and indistinguishable when we only know their censored failure times.

To explain and study this phenomenon, we investigate a large collection of stochastic processes that are widely found in practical applications of threshold regression. Specifically, we consider stochastic processes $\{D(t)\}$ known as Lévy processes. The only type of Lévy process that possesses continuous sample paths is a Wiener process. The sample paths of all other Lévy processes exhibit jump behavior. As a result, when we speak of the first hitting time S of a level d_0 by a process $\{D(t)\}$, we know that the level $D(S)$ at the moment of hitting the threshold will overshoot d_0 by some positive amount, except in two circumstances: (1) when $\{D(t)\}$ is a Wiener process and (2) when $\{D(t)\}$ is a process with fixed jump increments and d_0 equals a multiple of the increment. We distinguish between cases where $D(S) = d_0$ (no overshoot) and $D(S) > d_0$ (overshoot) in subsequent developments.

All Lévy processes possess the property of stationary independent increments. We consider that large subclass of Lévy processes $\{D(t)\}$ which are (1) bidirectional (i.e., their sample paths move up and down through time) and (2) possess a cumulant generating function (c.g.f.) $\kappa(\zeta)$, defined as follows:

$$\ln E\{e^{\zeta[D(t)]}\} = t\kappa(\zeta) \quad \text{for all } \zeta \in \mathcal{Z}, \tag{2}$$

where $\mathcal{Z}$ is an open interval containing 0. We let δ denote the mean parameter of the process. The c.g.f. assures the mean parameter exists. We will refer to such a process $\{D(t)\}$ with a positive mean parameter $\delta > 0$ as a primal process. Every primal process has a twin process $\{D_*(t)\}$, called its dual process, which follows the same probability law as $\{-D(t)\}$, the negative of the primal process. This dual process has c.g.f $\kappa_*(\zeta) = \kappa(-\zeta)$

and mean parameter $\delta_* = -\delta$. Notice that we add an asterisk to identify characteristics of the dual process. As $\delta > 0$ for the primal process $\{D(t)\}$, its sample paths tend to drift toward the failure threshold d_0 and the failure event will eventually be observed if it is not censored, that is, $p_0 = \lim_{s \to \infty} P(S \leq s) = 1$. In contrast, for the dual process $\{D_*(t)\}$ for which $\delta_* < 0$, the process tends to drift away from the threshold at d_0 and $p_{0*} = \lim_{s \to \infty} P_*(S \leq s) < 1$. The reversed sign of the dual process transforms its stochastic behavior from one in which eventual failure is guaranteed to one in which it is not guaranteed. As we will show, the existence of these twinned processes produces an identity crisis for interpreting threshold regression models with a cure rate.

We now define the conditional s.f. $\overline{K}(s) = P(S > s | S < \infty)$. Thus, $\overline{K}(s)$ is the survival function for those individuals who eventually will experience failure. For the primal process, we have $\overline{F}(s) = \overline{K}(s)$ because $p_0 = P(S < \infty) = 1$. For the dual process, we have $\overline{F}_*(s) = p_{0*}\overline{K}_*(s) + 1 - p_{0*}$ because $p_{0*} = P_*(S < \infty) < 1$. As shown in the Appendix A, the primal-dual pair of processes $\{D(t)\}$ and $\{D_*(t)\}$ have identical conditional survival functions $\overline{K}(s)$ and $\overline{K}_*(s)$ if (1) the primal process possesses the Esscher property and, in the case of processes that overshoot, (2) the primal process is in equilibrium. We now briefly define the two properties. The Appendix A provides more technical background and explanations of the properties.

Esscher Property. The Esscher property holds if the c.g.f. of the primal process has the form $\kappa(\zeta) = \kappa(-\zeta - U)$, where U is a positive constant defined by

$$U = -\sup\{\zeta : \zeta \in \mathbf{Z}, \zeta < 0, \kappa(\zeta) > 0\}. \tag{3}$$

This property is named for the Esscher transformation of a probability density function. Mathematically, the transformation is $f_*(x) = C(a) \exp(ax) f(x)$ for a constant a and normalizing constant $C(a)$. In our use of the transformation, we set $a = -U$ so $C(-U) = 1$. In this situation, the the original and transformed probability density functions $f(x)$ and $f_*(x)$ have c.g.f.s of form $\kappa(\zeta)$ and $\kappa(-\zeta - U)$, respectively, with $\kappa(\zeta) = \kappa(-\zeta - U)$.

Equilibrium Property. The equilibrium property holds if the process $\{D(t)\}$ has been operating for an extended earlier period and observation of the process starts at a random moment when the process has an upward crossing of level $D = 0$. This random moment is taken as the origin of the time scale ($t = 0$) and, hence, survival time S is measured from this moment. The precise mathematical definition of the equilibrium property, which requires a relaxation of the initial condition $D(0) = 0$, is set out in the Appendix A.

The presence of identical conditional survival functions under the Esscher and equilibrium properties gives rise to the identity crisis we anticipated earlier. The fact that $\overline{K}(s) \equiv \overline{K}_*(s)$ provides the following competing representations of the s.f. $\overline{F}_R(s)$ for the study population under the primal and dual processes.

$$
\begin{aligned}
\overline{F}_{R(primal)}(s) &= p\overline{F}(s) + (1 - p) \\
&= p\overline{K}(s) + (1 - p) \quad \text{for the primal process.}
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
\overline{F}_{R(dual)}(s) &= p_*\overline{F}_*(s) + (1 - p_*) \\
&= p_*[p_{0*}\overline{K}_*(s) + (1 - p_{0*})] + (1 - p_*) \\
&= p_*p_{0*}\overline{K}_*(s) + (1 - p_*p_{0*}) \\
&= p_*p_{0*}\overline{K}(s) + (1 - p_*p_{0*}) \quad \text{for the dual process.}
\end{aligned} \tag{5}
$$

We see from (4) and (5) that the s.f. for the study population is the same mathematical function of survival time s under the primal and dual processes with the exception that the apparent susceptibility rate for the dual process is p_*p_{0*} whereas its true susceptibility rate is p_*. The true susceptibility rates for the twinned processes, namely, p and p_*, are different but cannot be distinguished from survival data alone. Furthermore, with a limited study horizon and perhaps interruptions from competing events, the observed survival data will

be right-censored. So, it is only some fraction of the at-risk subpopulation whose failure times will actually be observed in the study. All others will have censored outcomes.

The proportion $1 - p$ in the primal process and $1 - p_*$ in the dual process are different exogenous cure rates. We now refer to probability $p_*(1 - p_{0*})$ as an endogenous cure rate for the dual process because it is this fraction of the susceptible population that will not experience failure under the dual model. The potential presence of both cure rates in the dual model leads to a combined cure rate for the study population of $1 - p_* + p_*(1 - p_{0*}) = 1 - p_* p_{0*}$. The challenge for survival data analysis that is raised by the mixed susceptibility models (4) and (5) is the difficulty of distinguishing non-susceptible cases from susceptible cases that happen to escape failure.

3. Examples of Mixed Susceptibility Models for Lévy Processes

Numerous families of Lévy processes have the Esscher propery and all of these have potential application as mixed survival models that include a cure rate component. The Appendix A presents mathematical details for four families of Lévy processes, namely, Wiener processes, Poisson-Bernoulli random walks, bilateral gamma processes and bilateral inverse Gaussian processes. We have chosen the Wiener family for detailed development as a case illustration in order to make the preceding ideas and issues concrete. We start with a brief overview of threshold regression to provide a context for the case illustration that follows.

Threshold Regression

Threshold regression refers to a category of regression models and methods for survival data analysis. The main idea of threshold regression is that a survival or failure time is often well-described mathematically as the first hitting time of a boundary or threshold by the sample path of a stochastic process. Regression functions for the process and threshold parameters in these models are multivariate combinations of explanatory variates. Estimation methods for threshold regression models are well developed and these have found extensive application. Wiener processes in particular have been widely used in threshold regression models. See, for example, [1–6]. The book by [7] provides an extensive overview of both theory and applications. Different variants of threshold regression models and notation are encountered. A common variant of what we presented earlier considers a process $\{Y(t)\}$ which starts at $Y(0) = y_0 > 0$ and has mean drift parameter μ and variance parameter $\sigma^2 > 0$. In this variant, the first hitting time S of a threshold at the zero level of the process defines the survival time. Note the following correspondence between this notation and our earlier notation: $D(t) = y_0 - Y(t)$, $d_0 = y_0$, $\delta = -\mu$.

For the Wiener process model, the dual process is one with drift away from the threshold, that is, one where $\delta_* = -\mu < 0$. For the dual model, the probability of eventual failure is given by:

$$p_{0*} = P_*(S < \infty) = \exp(2\delta_* d_0/\sigma^2) = \exp(-2\mu y_0/\sigma^2) \quad \text{for } \delta_* = -\mu < 0 \qquad (6)$$

Parameter σ^2 is usually set equal to 1 because the Wiener model is overparameterized for survival data. The Esscher property holds for all Wiener processes and, because the Wiener sample path is continuous, there is no overshoot. Our theory shows that the conditional survival functions of both the primal and dual processes are identical, i.e., $\overline{K}_*(s) \equiv \overline{K}(s)$. This conditional s.f. is an inverse Gaussian distribution [7–9].

Case Illustration

In an application of survival model (1), estimates of the model parameters might be calculated from interval-censored failure data or, more specifically, from right-censored failure data using maximum likelihood estimation. We assume the sample is a simple random sample of individuals from the study population. We also assume that the censoring mechanism is independent of whether the individual is susceptible or not and,

if susceptible, independent of the individual's failure time. The sample log-likelihood function for right-censored failure data takes the form:

$$\ln L(\boldsymbol{\theta}) = \sum_{i \in N_1} \ln f_R(t_i) + \sum_{i \in N_0} \ln \overline{F}_R(t_i) \tag{7}$$

where $\boldsymbol{\theta} = (d_0, \delta, p, \text{other parameters})$ is the vector of model parameters to be estimated and f_R is the probability density function (p.d.f.) corresponding to the population s.f. $\overline{F}_R$. Index set N_1 in (7) includes subjects i who failed at times $t_i, i \in N_1$, while index set N_0 includes subjects i who were censored by end of follow-up or a competing failure at times $t_i, i \in N_0$. Observe in this last case that the censoring time or time horizon may vary from one subject to the next. Finally, we note that in setting up the estimation problem we use the notation for the primal model.

We now give an illustration for the Wiener model. The estimation problem described by log-likelihood expression (7) seems to present nothing new in principle. Yet, a comparison of versions (4) and (5) of the population survival functions shows that there is an unresolved ambiguity. The sign of mean parameter $\mu = -\delta$ cannot be determined from the survival data alone. Table 1 shows alternate sets of estimates for model parameters from a single simulated data set. The simulated study population corresponds to a dual process with $\ln(y_0) = \ln(d_0) = 1$, $\mu = -\delta_* = 0.1$, and $p_* = 0.9$. The sample size is $n = 100,000$. All failure times are right censored at 70. The table shows the estimates in columns headed by the generic notation $\ln(y0)$, mu and $logit(p)$. Estimation of the model is done using a numerical gradient routine in Stata 12. The table shows the log-likelihood value and parameter estimates for two different sets of starting values.

Table 1. Maximum likelihood estimates of parameters of a mixed susceptibility Wiener model computed from two estimation runs. Estimates are based on a single simulated right-censored sample of 100,000 individuals but from estimation runs with different starting values for the numerical optimization routine. The true parameter values are $\ln(y_0) = 1$, $\mu = -\delta_* = 0.1$ and logit $p_* = \ln[0.9/(1-0.9)] = 2.1972$.

Estimation Run	Sample Log-Likelihood	Parameter Estimates		
		ln(y0)	mu	logit(p)
1	−238,975.75	1.003408	0.1035597	2.39924
2	−238,975.75	1.003406	−0.1035566	0.0844234

The two estimation runs in Table 1 produce identical sample log-likelihood values and almost identical estimates of $\ln(y0)$. The parameter estimates for mu and $logit(p)$ differ sharply. Closer study shows that the estimates match model variants (4) and (5). Estimation run 1 happens to provide estimates for the dual model in (5) with $\mu = -\delta_* = 0.1$ and logit $p_* = \ln[0.9/(1-0.9)] = 2.1972$. Here, the true value of p_{0*} is 0.5806, calculated from Formula (6). Estimation run 2 happens to provide estimates for the corresponding primal model. The mu estimate is reversed in sign relative to the dual model and the $logit(p)$ estimate is actually the estimate of $logit(p_* p_0)$, which equals $\ln\{0.9(0.5806)/[1 - 0.9(0.5806)]\} = 0.09030$. The example demonstrates that the sign of mean parameter μ is not resolvable from censored survival data for this Wiener model. A further implication is that one cannot know if the estimated cure rate represents a purely non-susceptible subpopulation (exogenous fraction $1 - p_*$) or a mixed subpopulation with some fraction of true non-susceptibles (exogenous fraction $1 - p_*$) combined with an added fraction of susceptibles who escaped failure by good fortune (endogenous fraction $p_* - p_* p_{0*}$). The latter case produces an apparent cure rate of $1 - p_* p_{0*}$. For the simulation scenario selected here, it is the first estimation run that gives correct estimates of the model. An outside observer, however, cannot know this fact from either the censored survival data set or the estimates derived from it.

4. Identifiability Challenge for the Mixed Susceptibility Model

The preceding case illustration, with the results displayed in Table 1, has demonstrated that the sign of the mean parameter for the threshold regression model cannot be decided unambiguously in the presence of possible exogenous and endogenous 'cures' in the study population. The sign of the estimated mean parameter depends, by chance, on the starting point for the maximum likelihood optimization. Because the conditional survival functions of the primal and dual processes are identical, i.e., $\overline{K}_*(s) \equiv \overline{K}(s)$, the survival data yield different but equally valid sets of estimates in terms of sample likelihood. If the study population were partitioned into subpopulations (say, subpopulations defined by age group, sex, marriage status, or other categorical variables) the same problem arises for each subpopulation, namely, the sign of the estimated mean parameter for each subpopulation would depend on the starting point for the maximum likelihood optimization. Independent subject matter knowledge would be needed to decide on the correct sign for the mean parameter.

In typical real-world applications of threshold regression, the mean parameter, threshold and exogenous cure rate fraction are made to depend on linear combinations of covariates through suitable regression link functions. The multivariate regression structures are rarely sufficiently detailed to define actual subpopulations and are usually misspecified to some degree. In these practical circumstances, estimation of the model from interval- or right-censored failure data would almost certainly lead to a unique maximum likelihood solution without any hint of the identifiability issue. But, our mathematical analysis for twinned stochastic processes anticipates that the computed threshold regression solution may have a competing local solution that is a close competitor to the global optimum. This close competitor would be a twinned process with an estimated mean having the opposite sign. Moreover, the solution and its twin would be associated with different cure rates.

As just mentioned, the drift parameter δ for our model often has a multivariate regression structure in threshold regression applications. A common formulation is an identity link function for δ connecting it to a linear combination of covariates. In this formulation, $\delta = \mathbf{z}\boldsymbol{\beta}$ where $\mathbf{z}$ is a k-component row vector of covariates and $\boldsymbol{\beta}$ is a matching column vector of regression coefficients. The leading covariate may be the unit value 1 so the leading regression coefficient is an intercept term. For a given covariate vector $\mathbf{z}$, the sign of the regression parameter vector must be reversed in order to reverse the sign of the drift parameter itself. In other words, $\delta_* = \mathbf{z}\boldsymbol{\beta}_* = \mathbf{z}(-\boldsymbol{\beta}) = -\delta$. This is an elementary mathematical change but it does suggest a way in which subject matter knowledge can check on the identifiability problem. If the regression vector is estimated from censored survival data, the investigator might examine the sign of the regression coefficient for each covariate (and the intercept) to see if the direction of effect is consistent with subject matter knowledge. Many applications, however, do not offer this opportunity to settle the ambiguity. Sampling error in estimates may make it unclear if the estimated sign of the coefficient is reliable. Moreover, the presence of multicollinearity in a regression model with multiple covariates may make it difficult to judge the directional effect of any single covariate. An investigator may be inclined to replace the identity link function in threshold regression with a log-linear link function of form $\ln \delta = \mathbf{z}\boldsymbol{\beta}$ so that the estimated drift is positive. This approach may be adopted because the investigator is confident that the drift should be towards failure. The strategy may be self-defeating. Although the mathematical device preserves the anticipated sign of the drift parameter, a substantively different statistical model is being estimated. The estimated model may be the wrong model and, in consequence, hides the ambiguity about the direction of drift.

The identifiability issue will not be present if it is known with certainty that the study population consists totally of susceptible individuals. In this circumstance, it is only the dual model which has some fraction of the study population that survives indefinitely. In the threshold regression context, the endogenous cure rate is determined by parameters of the stochastic process $\{D(t)\}$, as illustrated by Formula (6) for the Wiener process. Thus, no separate regression function should be employed for logit p_{0*} in this case.

The presence of censoring aggravates the identifiabilty issue because the true failure times for censored individuals are masked. Thus, precise estimates of the susceptibility rate and the other parameters are not available. In this situation, even available subject matter knowledge may not be sufficient to untangle p from $p_* p_{0*}$ in the primal and dual models. Of course, small study sample sizes also make the situation more difficult.

The identifiability issue might be resolved practically if the process level of a survivor is observed at the time of censoring. For example, the extent of deterioration of a knee joint for a patient with osteoarthritis might be measured at a clinic visit that precedes the knee's replacement. The visit censors the replacement, which will be triggered if and when deterioration reaches a prespecified critical level. In a situation where the process level of a survivor is observed, the sample log-likelihood function, such as the one presented in (7), would have the term $\sum_{i \in N_0} \ln \overline{F}_R(t_i)$ replaced by a sum of log-densities for known process levels $D(t_i) = d_i$ at censoring times t_i for subjects $i \in N_0$. This revised likelihood function will help to infer primal cases for which $\delta > 0$ from dual cases for which $\delta_* < 0$. See [10] for a discussion of statistical methods for this kind of data situation for Lévy processes.

5. Multidimensional and Other Complex Extensions

Our investigation of the identifiability challenge arising in mixed susceptibility models has been framed for a univariate stochastic process because this context provides the simplest demonstration of the issue. We now present a couple of extensions that illustrate where and how care may be needed to deal with the issue in multidimensional and other complex contexts.

Multidimensional stochastic processes are ubiquitous in many fields. Chemometrics provides an example of an important field that deals with multidimensional processes and has considerable interest in relation to time trends of complex chemical processes and their critical thresholds. See [11] for an extensive overview of the methodological challenges in chemometrics applied to environmental monitoring and [12,13] for specific application settings in chemometrics that reflect the complexity of data modelling in this field. Applications in chemometrics and many other fields may concern the first hitting time of a plane surface by a multidimensional process. When drawing statistical inferences from high-dimensional data about whether a multidimensional process is moving toward or away from a critical barrier in these applications, one can encounter the kind of identifiabily issue being considering in our study here. And when the issue is encountered, it is likely to be a significant scientific concern. The first hitting time of a plane in a high dimensional space is a prototypical problem for multivariate stochastic processes. This kind of application presents an exact analog of our univariate model setup and so our findings translate immediately to this context. There are many families of multidimensional processes. Our research has not attempted a general investigation of this interesting topic. To illustrate the extension, however, we consider an obvious case example, namely, a multidimensional Wiener process, which is frequently a valid model for real-world phenomena. To be specific, let $\mathbf{D}(t)$ denote a k-dimensional Wiener process operating in a k-dimensional linear vector space $\mathbf{X}$ with $\mathbf{D}(0) = \mathbf{0}$, where $\mathbf{0}$ is the origin of the space. Let superscript T denote the transpose of a vector. Consider a plane $\mathbf{a}^T \mathbf{x} = c$ in the $\mathbf{X}$ space, where we let $c > 0$ so the origin $\mathbf{0}$ lies outside the plane. Let the plane serve as a barrier or boundary for the process (i.e., a multidimensional threshold). We denote the projection of the sample path of the process $\mathbf{D}(t)$ onto a normal axis by $D(t)$. The normal axis is a line directed by a normal vector from the plane through the origin $\mathbf{0}$. $D(t)$ is a univariate process that has temporal movements toward or away from the origin in a direction that is orthogonal to the plane. In this situation, we know $D(t)$ is a Wiener process, with a mean δ and variance σ^2 that can be derived from the mean vector and covariance matrix of the parent Wiener process $\mathbf{D}(t)$. A first hitting time S of the plane occurs if and when the sample path of $D(t)$ first encounters the plane at distance $d_0 = c/\sqrt{\mathbf{a}^T \mathbf{a}}$ along the normal axis. $D(t)$ is a primal process in our terminology if its drift parameter is positive, i.e., if $\delta > 0$. $D(t)$ also possesses the Esscher property. This application includes the ambiguity of

mixed susceptibility in relation to first hitting times of the plane if the study population includes a proportion $1 - p$ of individual cases that are not susceptible to ever reaching the plane. For other families of multidimensional processes (such as multidimensional bilateral gamma processes), one would need to verify that the Esscher property holds for the univariate projection of the process onto a normal axis to the plane through the process origin.

For another extension of interest, we consider the conditions under which the Esscher property is preserved for stochastic processes formed from mathematical combinations of component stochastic processes. Again, this research question will not be investigated in depth here, but we sketch the following result to show the nature of the topic. Consider a set of k independent Lévy processes $D_j(t), j = 1, \ldots, k$, which each has the Esscher property. Let $D(t) = \sum_j a_j D_j(t)$ be a linear combination of the k processes, where the a_j are scalar coefficients. If $\kappa_j(\zeta)$ is the c.g.f. for a unit increment in process $D_j(t), j = 1, \ldots, k$, we know that the c.g.f. of $D(t)$ is given by $\phi(\zeta) = \sum_j \kappa_j(a_j\zeta)$. Moreover, as each component process satisfies the Esscher property, we have $\kappa_j(\zeta) = \kappa_j(-\zeta - U_j), j = 1, \ldots, k$, where U_j denotes the Esscher transform parameter for process $D_j(t)$, defined by (3). The Esscher property also implies that $\kappa_j(a_j\zeta) = \kappa_j(-a_j\zeta - a_jU_j)$. We now consider the sum $\sum_j \kappa_j(-a_j\zeta - a_jU_j)$. It is evident that if $U_1 = \cdots = U_k = U$ then we have:

$$
\begin{aligned}
\phi(\zeta) & = \sum_j \kappa_j(a_j\zeta) = \sum_j \kappa_j(-a_j\zeta - a_jU_j) = \sum_j \kappa_j(-a_j\zeta - a_jU) \\
& = \sum_j \kappa_j[a_j(-\zeta - U)] = \phi(-\zeta - U).
\end{aligned}
\tag{8}
$$

Thus, process $D(t)$ possesses the Esscher property when all component processes share the same Esscher transform parameter U_j. Reference to the case-demonstration process families considered in Appendix A.6 shows the kind of restrictions on the component process parameters that are implied by requiring that $U = U_j$ for all j. For example, if all component processes are Wiener processes then the ratio $2\delta_j/\sigma_j^2$ must be the same for all component processes. The preceding demonstration shows that the Esscher property is preserved in this linear combination of component processes even if the components come from different Lévy process families.

6. Final Remarks

The identifiability issue in this research arises under highly specialized circumstances and, hence, is an idealized construction. The exacting formulation is used to show the precise mathematical nature of the problem. Yet, it is clear that even when the exact conditions are not satisfied perfectly, the ambiguity of mixed susceptibility can still be present and problematic. Survival functions for first hitting times can be very similar for processes that might be drifting toward or away from a threshold, even if the generating process is not a true Levy process or the Esscher property is not quite valid. The situation reminds us that we may not be able to untangle susceptible from non-susceptible cases in the 'cure rate' fraction of a population based on censored survival data alone.

Our research focuses on explaining the ambiguous composition of the cure rate when both susceptible and non-susceptible fractions may be present. We show that the ambiguity arises precisely when the conditional survival functions for the primal and dual processes are identical. This research program can be broadened by asking what theoretical circumstances must hold for identical survival time distributions to arise from different generating models within the same model class. We created pairs of processes which were twinned by simple reversal of the sign of a drift parameter. Discovery of more theoretical circumstances that produce identical survival functions will require an extended and challenging research program, but the program may produce deep insights and results that advance mathematical statistics. Natural research questions for the program might include such questions as: (1) Can identical conditional survival functions be generated from nonlinear thresholds for some family of univariate processes or non-planar boundaries for

some family of multidimensional processes? (2) Do generalizations of the Esscher property exist whereby survival data or censored survival data necessarily produce ambiguous inferences for model parameters when the models have this property? Much interesting future work lies ahead.

Appendix A

Appendix A.1. Preliminary Elements

Consider a bidirectional stochastic process $\{D(t)\}$ with stationary, independent increments and initial value $D(0) = 0$. The bidirectional property requires that $D(1)$ have positive probability of taking both negative and positive values. We also require process $\{D(t)\}$ to have a cumulant generating function (c.g.f.) defined on an open interval $\mathcal{Z}$ which includes zero. We denote this c.g.f. by $\kappa(\zeta)$ so, by definition, $\ln E\{\exp[\zeta D(t)]\} = t\kappa(\zeta)$ for $t \geq 0$. These assumptions for process $\{D(t)\}$ imply that we are considering a subfamily of Lévy processes.

We let S denote the first hitting time of threshold $d_0 > 0$ by process $\{D(t)\}$. Thus, $D(S) \geq d_0$ is the process level reached at the hitting time S. The excess $V(S) = D(S) - d_0$ is referred to as the overshoot of the process at the first hitting time. For notational convenience, we let D denote $D(t)$ for some fixed time $t > 0$ and V denote $V(S)$ for $S = s$. Furthermore, we denote the probability density function (p.d.f.) of D by $f(D|t)$. This p.d.f. notation reminds us of the conditioning on time t. We denote the conditional p.d.f. of D given $S > t$ by $g(D|t)$, the p.d.f. of the first hitting time S given $S < \infty$ by $k(s)$, and the p.d.f. of V by $h(v|s)$. If the process has no overshoot then $V = 0$ and we take $h(v|s)$ as a limiting density function concentrated on 0 (a Dirac delta function).

We will refer to $\{D(t)\}$ as our primal process and its dual process by $\{D_*(t)\}$, where an asterisk subscript is used to designate properties of the dual process. The dual process $\{D_*(t)\}$ has the same probability law as $\{-D(t)\}$, that is, $\{D_*(t)\} \sim \{-D(t)\}$. So $\{D_*(t)\}$ has c.g.f. $\kappa_*(\nu) = \kappa(-\nu)$ defined for all $\nu \in \mathcal{Z}^* = -\mathcal{Z}$. We are interested in the situation where the primal process $\{D(t)\}$ has a positive mean parameter $\delta > 0$ so that its dual process has a negative mean parameter $\delta_* = -\delta < 0$. Sample paths of the primal process must eventually hit threshold d_0 so $P(S < \infty) = 1$ for its first hitting time S when $\delta > 0$. Sample paths of the dual process $\{D_*(t)\}$ are not guaranteed to reach the threshold d_0 so $P_*(S < \infty) < 1$ for its first hitting time S.

A sample path of the primal process $\{D(t)\}$ either reaches level D at time t without hitting threshold d_0 during interval $(0, t]$ or it does hit d_0 at some time $S = s \leq t$ and subsequently travels from $D(s)$ to D in the interval $(s, t]$. This line of reasoning gives us the following probability identity connecting the mathematical constituents of this setting:

$$f(D|t) = P(S > t)g(D|t) + P(S \leq t) \int_0^t k(s) \int_{d_0}^{\infty} h(v|s)f(D - d_0 - v|t - s)dvds. \quad \text{(A1)}$$

The corresponding probability identity for the dual process is:

$$f_*(D|t) = P_*(S > t)g_*(D|t) + P_*(S \leq t) \int_0^t k_*(s) \int_0^{\infty} h_*(v|s)f_*(D - d_0 - v|t - s)dvds. \quad \text{(A2)}$$

Here the functions with asterisks are the corresponding entities for the dual process that are found in (A1) for the primal process.

Appendix A.2. The Esscher Property

We now narrow the subfamily of stochastic processes of interest and consider primal processes $\{D(t)\}$ for which the c.g.f. $\kappa(\zeta)$ has the property:

$$\kappa(\zeta) = \kappa(-\zeta - U), \tag{A3}$$

where $U > 0$ is defined as follows:

$$U = -\sup\{\zeta : \zeta \in \mathbf{Z}, \zeta < 0, \kappa(\zeta) > 0\} = \inf\{v : v \in \mathbf{Z}_* = -\mathbf{Z}, v > 0, \kappa(-v) > 0\}. \tag{A4}$$

Later case demonstrations show that Lévy processes possessing property (A3) form a large family.

The property $\kappa(\zeta) = \kappa(-\zeta - U)$ implies that:

$$
\begin{aligned}
\exp[t\kappa(\zeta)] &= \int_{-\infty}^{\infty} \exp(\zeta D) f(D|t) dD = \int_{-\infty}^{\infty} \exp[(-\zeta - U)D] f(D|t) dD \\
&= \int_{-\infty}^{\infty} \exp(-\zeta D) \exp(-UD) f(D|t) dD,
\end{aligned}
\tag{A5}
$$

where we recall that $f(D|t)$ denotes the p.d.f. of $D = D(t)$. For the dual process $\{D_*(t)\}$, we use (A5) to obtain:

$$
\begin{aligned}
\exp[t\kappa_*(v)] &= \int_{-\infty}^{\infty} \exp(vD) f_*(D|t) dD = \exp[t\kappa(-v)] \\
&= \int_{-\infty}^{\infty} \exp(vD) \exp(-UD) f(D|t) dD,
\end{aligned}
\tag{A6}
$$

Equation (A6) shows that the dual p.d.f. is the following exponentially tilted version of the p.d.f. for the primal process:

$$f_*(D|t) = \exp(-UD) f(D|t) \tag{A7}$$

A transformation of a p.d.f. $p(x)$ into a second p.d.f. of form $p(x|a) = C(a) \exp(ax) p(x)$, where $C(a)$ is a normalizing constant, is known as an Esscher transform. The transformation found early application in actuarial science and more recently in financial engineering and simulation sampling methods [14–16]. The property $\kappa(\zeta) = \kappa(-\zeta - U)$ can be seen from (A7) to imply an Esscher transform of $f(D|t)$ with transform parameter $a = -U$. This particular choice of a ensures that normalizing constant $C(a)$ equals 1 so the transformed function $f_*(D|t)$ is itself a probability density function. Because of this connection to the Esscher transform, we refer to this property for our stochastic processes as the Esscher property.

Appendix A.3. Evaluating Bundles of Sample Paths

The Esscher property provides Lévy processes with a remarkable feature. This feature is apparent if we consider discrete versions of the sample paths of our processes. Let $D_0, D_1, \ldots, D_k$ be a sequence of levels of a sample path corresponding to ordered time points $\tau_0 = 0 < \tau_1 < \cdots < \tau_k = \tau$ of the process $\{D(t)\}$. Here $D_i = D(\tau_i)$ with $D_0 = 0$ and $D_k = D(\tau)$. Thus, the path experiences probabilistically independent increments of $\Delta D_i = D_i - D_{i-1}$ during time increments $\Delta \tau_i = \tau_i - \tau_{i-1}$, for $i = 1, \ldots, k$. We note that the partition of the path from the origin to $D(\tau)$ at time τ can be as fine as desired for the analysis at hand.

With the Esscher property, the probability densities of the primal and dual processes for any path increment i are related as follows:

$$f^{(i)}(\Delta D_i | \Delta \tau_i) = e^{U \Delta D_i} f_*^{(i)}(\Delta D_i | \Delta \tau_i), \tag{A8}$$

where $f^{(i)}(\Delta D_i | \Delta \tau_i)$ and $f_*^{(i)}(\Delta D_i | \Delta \tau_i)$ denote the respective probability densities. It follows then that the probability densities of any given sample path J from 0 at time 0 to $D(\tau)$ at time τ for the primal and dual processes are related by:

$$
\begin{aligned}
f(\mathsf{J}|\tau) &= \prod_{i=1}^{k} f^{(i)}(\Delta D_i | \Delta \tau_i) = \prod_{i=1}^{k} e^{U \Delta D_i} f_*^{(i)}(\Delta D_i | \Delta \tau_i) \\
&= e^{U \sum_i \Delta D_i} \prod_{i=1}^{k} f_*^{(i)}(\Delta D_i | \Delta \tau_i) = e^{U D(\tau)} f_*(\mathsf{J}|\tau),
\end{aligned} \tag{A9}
$$

where $f(\mathsf{J}|\tau)$ and $f_*(\mathsf{J}|\tau)$ denote the respective probability densities for path J. The relationship in (A9) shows that the probability densities of the primal and dual processes for any path J under the Esscher property are proportional, with the multiplier $\exp[UD(\tau)]$ being determined only by the cumulative vertical distance $D(\tau)$ travelled by the sample path J.

The result in (A9) provides a mechanism for evaluating probabilities of different bundles or sets of sample paths for a stochastic process observed over a given time interval $(0, t]$. The mechanism visualizes discrete simulations of many sample paths for the process using a fixed partition of the time interval $(0, t]$. In particular, we consider the simulation of a set $\mathcal{J}$ of independent sample paths from the primal process where $f(\mathsf{J}|t)$ denotes the probability density of a path $\mathsf{J} \in \mathcal{J}$. If we are interested in estimating the probability of experiencing a simulated sample path that meets a specified condition C then we sort through set $\mathcal{J}$ and keep only sample paths that meet the condition, resulting in a subset that we can denote by $\mathcal{J}_C \subseteq \mathcal{J}$. The probability density for this subset is therefore $\sum_{\mathsf{J} \in \mathcal{J}_C} f(\mathsf{J}|t)$.

We can exploit the correspondence of path densities in (A9) to evaluate the relationships between corresponding terms in the probability identities (A1) and (A2). P.d.f. $f(D|t)$, for example, is the probability density for the primal process of the set of sample paths that connect the origin to $D = D(t)$ at time t, which we now denote by $\mathcal{P}$. Using (A9), we then have the following link between $f(D|t)$ and $f_*(D|t)$ for the primal and dual processes.

$$
f(D|t) = \sum_{\mathsf{P} \in \mathcal{P}} f(\mathsf{P}|t) = \sum_{\mathsf{P} \in \mathcal{P}} e^{UD} f_*(\mathsf{P}|t) = e^{UD} \sum_{\mathsf{P} \in \mathcal{P}} f_*(\mathsf{P}|t) = e^{UD} f_*(D|t) \tag{A10}
$$

Next we consider the subset of $\mathcal{P}$ that contains sample paths that do not cross the threshold d_0 in the interval $(0, t]$. We denote this subset by $\mathcal{P}_0 \subseteq \mathcal{P}$. Using (A9), we have the following link between the conditional p.d.f.s $g(D|t)$ and $g_*(D|t)$ for the primal and dual processes.

$$
\begin{aligned}
P(S > t)g(D|t) &= \sum_{\mathsf{P} \in \mathcal{P}_0} f(\mathsf{P}|t) = \sum_{\mathsf{P} \in \mathcal{P}_0} e^{UD} f_*(\mathsf{P}|t) \\
&= e^{UD} \sum_{\mathsf{P} \in \mathcal{P}_0} f_*(\mathsf{P}|t) = P_*(S > t)e^{UD} g_*(D|t)
\end{aligned} \tag{A11}
$$

Note that probabilities $P(S > t)$ and $P_*(S > t)$ must be included to adjust for the condition that the first hitting time lies beyond t.

The density functions in the rightmost terms of (A1) and (A2) arise from sample paths that cross the threshold d_0 in the interval $(0, t]$. The correspondence of these primal and dual densities is established as follows. Let subset $\mathcal{P}_{(s,v)} \subseteq \mathcal{P} - \mathcal{P}_0$ consist of paths that first cross the threshold d_0 at a specified time s in the interval $(0, t]$, experience an overshoot of v, and then proceed to level $D = D(t)$ at time t. So, again using (A9), we have the following equation:

$$P(S \le t)k(s)h(v|s)f(D - d_0 - v|t - s) = \sum_{\mathsf{P} \in \mathcal{P}_{(s,v)}} f(\mathsf{P}|t) = \sum_{\mathsf{P} \in \mathcal{P}_{(s,v)}} e^{UD} f_*(\mathsf{P}|t)$$

$$= e^{UD} \sum_{\mathsf{P} \in \mathcal{P}_{(s,v)}} f_*(\mathsf{P}|t) = P_*(S \le t)k_*(s)e^{UD}h_*(v|s)f_*(D - d_0 - v|t - s). \tag{A12}$$

The distribution of process increment $D - D(s) = D - d_0 - v$ does not depend on $D(s)$ for either the primal or dual processes. Therefore, when we integrate both sides of (A12) with respect to $D - D(s)$, the following integrals evaluate to 1:

$$\int_{-\infty}^{\infty} f[D - D(s)|t - s]d[D - D(s)] = \int_{-\infty}^{\infty} e^{U[D - D(s)]} f_*[D - D(s)|t - s]d[D - D(s)] = 1. \tag{A13}$$

Using this result, Equation (A12) reduces to:

$$P(S \le t)k(s)h(v|s) = P_*(S \le t)k_*(s)e^{U(d_0+v)}h_*(v|s). \tag{A14}$$

Of particular interest is the limit of (A14) if we now let t go to infinity. As a first hitting is inevitable for the primal process, we have $P(S < \infty) = 1$. For the dual process, a first hitting may not occur in finite time so the limiting probability $P_*(S < \infty)$ will be less than 1. Equation (A14) therefore reduces to:

$$k(s)h(v|s) = P_*(S < \infty)k_*(s)e^{U(d_0+v)}h_*(v|s). \tag{A15}$$

To proceed further, we will now separate processes according to whether they overshoot threshold d_0 or not when their sample paths first exit the threshold.

Appendix A.4. First Hitting Times without Overshoot

Lévy processes that do not overshoot a threshold are of two types, namely, Wiener processes or processes with fixed jump increments where d_0 is a multiple of that fixed increment. With no overshoot, Equation (A15) simplifies as follows because random variable $V(S)$ equals 0 with probability 1.

$$k(s) = P_*(S < \infty)k_*(s)e^{Ud_0} \quad \text{for all } s > 0. \tag{A16}$$

As $k(s)$ and $k_*(s)$ are both p.d.f.s and $P_*(S < \infty)$ and e^{Ud_0} are both constants, it follows necessarily that $k(s) \equiv k_*(s)$ and $P_*(S < \infty) = e^{-Ud_0}$ in order to maintain the equality in (A16). Thus, the identity $\overline{K}(s) \equiv \overline{K}_*(s)$ is assured for all processes without overshoot.

Appendix A.5. First Hitting Times with Overshoot

We next consider bidirectional Lévy processes with jump components that do overshoot thresholds. To deal with these processes, we reformulate the first hitting time of the process. Instead of having primal process $\{D(t)\}$ start at the origin with $D(0) = 0$, we let the process operate until it encounters its first positive level and shift the time origin to that position. We denote this initial positive level by V_1 so $D(0) = V_1$. To proceed, we return to a discrete version of the sample path for the re-scaled process $\{D(t)\}$ and consider a partition of time interval $[0, \tau]$, with equally spaced time points $\tau_0 = 0 < \tau_1 < \cdots < \tau_k = \tau$ where $\tau_i = i\Delta\tau$ and $\Delta\tau = \tau/k$. Let the successive levels $D_i, i = 0, 1, \ldots, k$, correspond to $D_i = D(\tau_i)$, so $D_0 = D(0) = V_1$ and $D_k = D(\tau)$. We then consider the successive maxima $M_j = \max_{i \le j} D_i$ of the process, which starts with $M_1 = V_1$. We define the successive increments in the maxima by $\Delta M_j = M_j - M_{j-1}, j = 2, \ldots, k$. Now we keep only positive increments ΔM_j, discarding all ΔM_j that are zero. These positive increments are necessarily independent and identically distributed. The increments mark out the vertical advance of the partitioned primal process from starting level V_1 toward the failure threshold at d_0. We re-index the successive positive increments and denote them by $R_n = \Delta M_j$, $n = 2, \ldots, m$, where ΔM_j is the nth positive increment in the sequence. It is evident that the

set $\{V_1, R_2, \ldots, R_m\}$ is a sequence of m renewal intervals for a delayed renewal process [17]. The delay V_1 in the process is, in effect, the overshoot of the zero-axis by the process in its first passage to that level.

Building on this formulation as a delayed renewal process, we denote the s.f. of the successive renewal intervals $R_2, \ldots, R_m$ by $\overline{H}(r)$. Next, we assume that the delayed renewal process is, in fact, in equilibrium so $M_1 = V_1$ is a random draw from the equilibrium c.d.f.:

$$H_E(v) = \int_0^v \overline{H}(r)dr/\mu \tag{A17}$$

where μ denotes the mean of s.f. $\overline{H}(r)$. Parameter μ is the mean vertical advance of the process whenever a positive advance occurs. With the assumption that the renewal process is in equilibrium, it follows therefore that $H_E(v)$ is also the c.d.f. of the overshoot that will be experienced when the primal process crosses threshold d_0. Moreover, the equilibrium condition implies that the overshoot c.d.f. $H_E(v)$ will be independent of the first hitting time S of threshold d_0.

Finally, we consider the corresponding dual process $\{D_*(t)\}$ which, of course, is also bidirectional with a jump component. The preceding line of development for constructing an equilibrium renewal process for advances toward the failure threshold at d_0 can be applied to the dual process. For the dual process, we denote the s.f. of the renewal interval by $\overline{H}_*(r)$ and the equilibrium c.d.f. for overshoot by $H_{E*}(v)$.

We return now to Equation (A14) and rearrange it into a product of ratios for matched terms as follows:

$$\left\{ \frac{1}{e^{Ud_0}P_*(S < \infty)} \right\} \left\{ \frac{k(s)}{k_*(s)} \right\} \left\{ \frac{h(v|s)}{e^{Uv}h_*(v|s)} \right\} = 1, \tag{A18}$$

It can be seen that only the rightmost ratio on the left side of Equation (A18) contains terms that are functions of overshoot v. It follows therefore that this ratio cannot depend on v although it may depend on survival time s. Denoting this rightmost ratio by $Q(s)$, we have:

$$Q(s) = \frac{h(v|s)}{e^{Uv}h_*(v|s)}. \tag{A19}$$

Now, however, we can add that if the primal process and dual processes are in equilibrium then the densities in (A19) do not depend on s and, hence, $Q(s)$ equals some constant Q for all s. In this case, the two density functions in the ratio are those corresponding to c.d.f.s $H_E(v)$ and $H_{E*}(v)$, which we denote by $h_E(v)$ and $h_{E*}(v)$, respectively. Moreover, in this same equilibrium condition, it follows that the density functions $k(s)$ and $k_*(s)$ for the survival time S of the primal and dual processes must be identical. Thus, the identity $\overline{K}(s) \equiv \overline{K}_*(s)$ is assured for all bidirectional Lévy processes with overshoot, provided survival time is measured from a time point where the process is in equilibrium. As a final consequence, we see that constant $Q = e^{Ud_0}P_*(S < \infty)$ so the probability that S is finite is given by $P_*(S < \infty) = Qe^{-Ud_0}$. When the primal process has no overshoot, we have from (A19) that $Q(s) = Q = 1$ for all s and, hence, the probability that S is finite is given by $P_*(S < \infty) = e^{-Ud_0}$, a result we saw earlier.

An implication of adopting the equilibrium formulation for the overshoot case is that the first hitting time of d_0 will be immediate if $V_1 \geq d_0$. In other words, if the initial process maximum $M_1 = V_1$ falls above threshold d_0 then $S = 0$ and failure occurs at the outset. The survival function therefore has a probability mass at zero equal to $P(S = 0) = P(V_1 \geq d_0)$.

Appendix A.6. Case Demonstrations

We now present case demonstrations of stochastic process families in which the primal and dual processes have the same survival distributions. We begin by looking at two families whose sample paths do not exhibit overshoot. We then consider an additional two families that do exhibit overshoot.

Cases with No Overshoot

<u>Wiener Process</u>

As the primal process, consider a Wiener process $\{D(t)\}$ with mean parameter $\delta > 0$ and variance parameter σ^2. Let $t = 1$ so $D(1) = D$. The p.d.f. of D is:

$$f(D) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(D - \delta)^2/2\sigma^2] \tag{A20}$$

The c.g.f. for the process is:

$$\kappa(\zeta) = \delta\zeta + \sigma^2\zeta^2/2 \tag{A21}$$

which is defined for all real numbers ζ. As Wiener sample paths are continuous, we have no overshoot.

For dual process $\{D_*(t)\}$, $\kappa_*(\nu) = \kappa(-\nu)$ and $\delta_* = -\delta$. Also, $U = \inf\{\nu : \nu > 0, \kappa_*(\nu) > 0\} = -\sup\{\zeta : \zeta < 0, \kappa(\zeta) > 0\} = 2\delta/\sigma^2$ so:

$$P_*(S < \infty) = \exp(-Ud_0) = \exp(-2d_0\delta/\sigma^2). \tag{A22}$$

It also can be verified that $\kappa(\zeta) = \kappa(-\zeta - U)$ so the Esscher property holds.

The first hitting time S of the primal and dual processes has the following inverse Gaussian c.g.f.:

$$\phi(\xi) = -d_0\kappa^{-1}(-\xi) = \frac{d_0\delta}{\sigma^2}\left[1 - \left(1 - \frac{2\sigma^2\xi}{\delta^2}\right)^{1/2}\right] \tag{A23}$$

<u>Poisson-Bernoulli Random Walk</u>

As the primal process, consider a Poisson-Bernoulli random walk $\{D(t)\}$ defined by the following doubly stochastic process. Bernoulli events occur according to a Poisson process with rate parameter λ and each Bernoulli event is either an up-step of $+1$ with probability p or a down-step of -1 with probability $1 - p$. Assume the event outcomes occur in a mutually independent fashion. We take the threshold d_0 to be a natural number so there is no overshoot. The process therefore takes the following form:

$$D(t) = 2B[N(t)] - N(t). \tag{A24}$$

Here $\{N(t)\}$ denotes the Poisson process and $\{B[N(t)]\}$ denotes the binomial process for the number of up-step outcomes among the $N(t)$ Bernoulli trials occurring during time interval $(0, t]$.

The probability function of $D(1) = D$ for this process is evaluated by expanding the probability product $P[B = b|N(1) = n]P[N(1) = n]$ and then summing over all pairs (b, n) for which $d = 2b - n$, taking account of the condition that $b = 0, \ldots, n$ for $n = 0, 1, \ldots$. The result is the following form, which involves a tractable infinite series:

$$P(D = d) = \left[(p^{I(d \geq 0)}(1 - p)^{I(d < 0)}\right]^{|d|} \lambda^{|d|} \exp(-\lambda) \sum_{j=0}^{\infty} \frac{\lambda^{2j}[2p(1 - p)]^j}{j!(j + |d|)!}, \tag{A25}$$

where $I(\cdot)$ denotes an indicator function. A little mathematics gives the following c.g.f. for the process $\{D(t)\}$:

$$\kappa(\zeta) = \lambda\left[pe^{\zeta} + (1 - p)e^{-\zeta} - 1\right], \tag{A26}$$

which is defined for all real numbers ζ. The mean parameter is easily evaluated from this c.g.f. as $\delta = (2p - 1)\lambda$. To ensure a positive mean parameter, we set $p > 1/2$.

The dual process $\{D_*(t)\}$ is obtained by simply interchanging p and $1 - p$, that is, by letting $p_* = 1 - p < 1/2$. Again, for $\{D_*(t)\}$, $\kappa_*(\nu) = \kappa(-\nu)$, the process mean is $\delta_* = -\delta < 0$, and $U = \inf\{\nu : \nu > 0, \kappa_*(\nu) > 0\} = -\sup\{\zeta : \zeta < 0, \kappa(\zeta) > 0\} = \ln[p/(1-p)]$ so:

$$P_*(S < \infty) = \exp(-Ud_0) = [(1-p)/p]^{d_0} \quad \text{where } p > 1/2. \tag{A27}$$

It also can be verified with a little algebra that the stochastic process $\{D(t)\}$ has the Esscher property.

The first hitting time S of the primal and dual processes has the following c.g.f.:

$$\phi(\xi) = -d_0 \kappa^{-1}(-\xi) = -d_0 \ln\left[\frac{(\lambda - \xi) - \sqrt{(\lambda - \xi)^2 - 4\lambda^2 p(1-p)}}{2\lambda p}\right], \tag{A28}$$

Cases with Overshoot

For both cases with overshoot, we consider bilateral processes that are created by taking the difference of independent monotonically non-decreasing jump processes with stationary independent increments.

Bilateral Gamma Process

As the primal process, consider a bilateral gamma process $\{D(t)\}$, defined as the difference between two independent gamma processes $\{G_1(t)\}$ and $\{G_2(t)\}$, as follows:

$$D(t) = G_1(t) - G_2(t). \tag{A29}$$

In the most general case, the processes may have different positive shape parameters $\alpha_i, i = 1, 2$, and different positive scale parameters $\beta_i, i = 1, 2$. This process has attracted some interest in financial mathematics [18]. The general bilateral gamma process has c.g.f.:

$$\kappa(\zeta) = \alpha_1 \ln\left(\frac{\beta_1}{\beta_1 - \zeta}\right) + \alpha_2 \ln\left(\frac{\beta_2}{\beta_2 + \zeta}\right), \quad \text{where } \zeta \in \mathcal{Z} = (-\beta_2, \beta_1). \tag{A30}$$

The process mean is $\delta = (\alpha_1/\beta_1) - (\alpha_2/\beta_2)$.

Equal shape parameters. The special case in which the component processes $\{G_1(t)\}$ and $\{G_2(t)\}$ share the same shape parameter, say, $\alpha_1 = \alpha_2 = \alpha$, provides a family of processes $\{D(t)\}$ for which $\delta = \alpha(\beta_2 - \beta_1)/\beta_1\beta_2 > 0$ if $\beta_2 > \beta_1$. The c.g.f. in (A30) simplifies to:

$$\kappa(\zeta) = \alpha \ln\left[\frac{\beta_1\beta_2}{(\beta_1 - \zeta)(\beta_2 + \zeta)}\right] \tag{A31}$$

It can be verified that the Esscher property $\kappa(\zeta) = \kappa(-\zeta - U)$ holds in this case with $U = \beta_2 - \beta_1$. For this equal-shape case, a little algebra shows that the p.d.f. of $D(1) = D$ takes the following form:

$$f(D) = \frac{(\beta_1\beta_2)^\alpha}{\Gamma(\alpha)^2} \exp(-\beta_1 D) \int_{\max(0,-D)}^\infty [(u(D+u)]^{\alpha-1} \exp[-(\beta_1 + \beta_2)u]du. \tag{A32}$$

It is noteworthy that the integral in (A32) depends parametrically only on α and $\beta_1 + \beta_2$.

Equal scale parameters. The counterpart of the equal-shape case is the equal-scale case in which the component processes $\{G_1(t)\}$ and $\{G_2(t)\}$ share the same scale parameter, say, $\beta_1 = \beta_2 = \beta$. In this case, the family of processes $\{D(t)\}$ have mean $\delta = (\alpha_1 - \alpha_2)/\beta > 0$ if $\alpha_1 > \alpha_2$. The c.g.f. in (A30) simplifies to:

$$\kappa(\zeta) = \alpha_1 \ln\left(\frac{\beta}{\beta - \zeta}\right) + \alpha_2 \ln\left(\frac{\beta}{\beta + \zeta}\right), \tag{A33}$$

Quantity U is the positive root $(0 < U < \beta)$ of the following equation:

$$\kappa(-U) = \alpha_1 \ln\left(\frac{\beta}{\beta + U}\right) + \alpha_2 \ln\left(\frac{\beta}{\beta - U}\right) = 0. \tag{A34}$$

For this equal-scale case, however, the Esscher property $\kappa(\zeta) = \kappa(-\zeta - U)$ does not hold. We therefore do not pursue it further.

Bilateral Inverse Gaussian Process

As the primal process, consider a bilateral inverse Gaussian process $\{D(t)\}$, defined as the difference between two independent inverse Gaussian processes $\{H_1(t)\}$ and $\{H_2(t)\}$, as follows:

$$D(t) = H_1(t) - H_2(t). \tag{A35}$$

In the most general case, the processes may have different positive scale parameters $\eta_i, i = 1, 2$, and different positive shape parameters $\nu_i, i = 1, 2$. The general bilateral inverse Gaussian process has c.g.f.:

$$\kappa(\zeta) = \frac{1}{\nu_1}\left[\eta_1 - (\eta_1^2 - 2\nu_1\zeta)^{1/2}\right] + \frac{1}{\nu_2}\left[\eta_2 - (\eta_2^2 + 2\nu_2\zeta)^{1/2}\right] \quad \text{for} \quad \zeta \in \mathcal{Z} = (-\frac{\eta_2^2}{2\nu_2}, \frac{\eta_1^2}{2\nu_1}). \tag{A36}$$

Equal shape parameters. The special case in which the component processes $\{H_1(t)\}$ and $\{H_2(t)\}$ share the same shape parameter, say, $\nu_1 = \nu_2 = \nu$, provides a family of processes $\{D(t)\}$ for which $\delta = (\eta_2 - \eta_1)/(\eta_1\eta_2) > 0$ if $\eta_2 > \eta_1$. The c.g.f. in (A36) simplifies to:

$$\kappa(\zeta) = \frac{1}{\nu}\left\{\left[\eta_1 - (\eta_1^2 - 2\nu\zeta)^{1/2}\right] + \left[\eta_2 - (\eta_2^2 + 2\nu\zeta)^{1/2}\right]\right\}. \tag{A37}$$

It can be verified that the Esscher property $\kappa(\zeta) = \kappa(-\zeta - U)$ holds in this case with $U = (\eta_2^2 - \eta_1^2)/(2\nu)$.

Equal scale parameters. The counterpart of the equal-shape case is the equal-scale case in which the component processes $\{H_1(t)\}$ and $\{H_2(t)\}$ share the same scale parameter, say, $\eta_1 = \eta_2 = \eta$. In this case, the family of processes $\{D(t)\}$ have mean $\delta = 0$ and, hence, do not fall into the class of processes that interest us here. Equation $\kappa(\zeta) = 0$ has only a single root at 0 in the interval $\zeta \in \mathcal{Z}$ so the Esscher property does not hold.

References

1. Hellier, J.; Emsley, R.; Pickles, A. Estimating dose-response for time to remission with instrumental variable adjustment: The obscuring effects of drug titration in Genome Based Therapeutic Drugs for Depression Trial (GENDEP): Clinical trial data. *Trials* **2020**, *21*, 1–11. [CrossRef] [PubMed]
2. Lee, M.-L.T.; Whitmore, G.A. Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Stat. Sci.* **2006**, *21*, 501–513. [CrossRef]
3. Lee, M.-L.T.; Whitmore, G.A. Proportional hazards and threshold regression: Their theoretical and practical connections. *Lifetime Data Anal.* **2010**, *16*, 196–214. [CrossRef] [PubMed]
4. Lee, M.-L.T.; Whitmore, G.A.; Rosner, B.A. Threshold regression for survival data with time-varying covariates. *Stat. Med.* **2010**, *29*, 896–905. [CrossRef] [PubMed]
5. Sæbø, S.; Almøy, T.; Aastveit, A.H. Disease resistance modelled as first-passage times of genetically dependent stochastic processes. *Appl. Stat.* **2005**, *54*, 273–285.
6. Whitmore, G.A.; Schenkelberg, F. Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Anal.* **1997**, *3*, 1–19. [CrossRef] [PubMed]
7. Caroni, C. *First Hitting Time Regression Models: Lifetime Data Analysis Based on Underlying Stochastic Processes*; Wiley: Hoboken, NJ, USA, 2017.
8. Aalen, O.O.; Borgan, O.; Gjessing, H.K. *Survival and Event History Analysis: A Process Point of View (Statistics for Biology and Health)*; Springer: New York, NY, USA, 2008.
9. Cox, D.R.; Miller, H.D. *The Theory of Stochastic Processes*; Chapman and Hall: London, UK, 1965.
10. Lee, M.-L.T.; Whitmore, G.A. Distribution-free predictive inference for failure data using threshold regression. *submitted*.
11. Dupont, M.F.; Elbourne, A.; Cozzolino, D.; Chapman, J.; Truong, V.K.; Crawford, R.J.; Latham, K. Chemometrics for environmental monitoring: A review. *Anal. Methods* **2020**, *12*, 4597–4620. [CrossRef] [PubMed]

12. Medinger, J.; Nedyalkova, M.; Furlan, M.; Lüthi, T.; Hofmann, J.; Neels, A.; Lattuada, M. Preparation and machine-learning methods of nacre-like composites from the self-assembly of magnetic colloids exposed to rotating magnetic fields. *ACS Appl. Mater. Interfaces* **2021**, *13*, 48040–48052. [CrossRef] [PubMed]
13. Vakarelska, E.; Nedyalkova, M.; Vasighi, M.; Simeonov, V. Persistent organic pollutants (POPs)-QSPR classification models by means of machine learning strategies. *Chemosphere* **2022**, *287*, 132189. [CrossRef] [PubMed]
14. Esscher, F. On the probability function in the collective theory of risk. *Skand. Aktuarietidskr.* **1932**, *15*, 175–195.
15. Gerber, H.U.; Shiu, E.S.W. Option pricing by Esscher transforms (with discussion). *Trans. Soc. Actuar.* **1994**, *46*, 99–191.
16. Kawai, R. An importance sampling method based on the density transformation of Lévy processes. *Monte Carlo Methods Appl.* **2006**, *12*, 171–186. [CrossRef]
17. Ross, S.M. *Stochastic Processes*, 2nd ed.; Wiley: New York, NY, USA, 1996.
18. Küchler, U.; Tappe, S. Bilateral gamma distributions and processes in financial mathematics. *Stoch. Process. Appl.* **2008**, *118*, 261–283. [CrossRef]