TAIWANESE JOURNAL OF MATHEMATICS
Vol. 23, No. 2, pp. 293–305, April 2019
DOI: 10.11650/tjm/190107

A Survey of Threshold Regression for Time-to-event Analysis and Applications

Mei-Ling Ting Lee

In memory of Professor Hwai-Chiuan Wang

Abstract. In analyzing time-to-event data, proportional hazards (PH) regression is an ubiquitous model used in many fields. PH regression, however, requires a strong assumption that is not always appropriate. Threshold regression (TR) is one of the alternative models. A first-hitting-time (FHT) survival model postulates a health status process for a patient that gradually declines until the patient dies when the health level first reaches a critical threshold. In this article, we review the development of threshold regression models and their applications.

1. Introduction

Proportional hazards (PH) regression introduced by Sir David Cox in 1972 is an ubiquitous methodology for analyzing survival and time-to-event data. The PH regression model assumes that the time to the event or endpoint of interest is a positive random variable with a hazard function of the following form:

$$h(t) = h_0(t) \exp(\boldsymbol{z}\boldsymbol{\beta}).$$

Here $h_0(t)$ is a fixed baseline hazard function, $\boldsymbol{z}$ is a row vector of covariates, and $\boldsymbol{\beta}$ is a column vector of regression coefficients that are to be estimated. Although extensions with time-varying covariates of the form $\boldsymbol{z}(t)$ exist, we limit our attention to fixed covariates in this review.

PH regression is easy to use and has found applications in many disciplines ranging from engineering to medicine. The wide use of the PH model reflects more its mathematical convenience than its realism. The occurrence of proportional hazard functions in nature is actually rare. Realizing that the proportional hazards assumption of PH regression is

Received May 7, 2018; Accepted December 20, 2018.
Communicated by Tai-Chia Lin.
2010 *Mathematics Subject Classification.* 92B15, 62P10.
Key words and phrases. boundary, Brownian motion, first hitting time, inverse Gaussian distribution, running time, stochastic process, survival analysis, Wiener process.

not always appropriate, statistical researchers have explored many alternative models and methods over the past forty years. One of these alternatives is *threshold regression* (TR).

Aalen and Gjessing [2, p. 1] make the telling point that the "hazard rate is really an elusive concept, especially when one tries to interpret its shape considered as a function of time." These authors are highlighting the important point that the hazard function is only a derivative feature that may lie on the pathway to understanding but is not the end of the journey itself. A deeper understanding will be obtained if the risk mechanism behind the hazard pattern is probed by the investigator. See also Aalen, Borgan and Gjessing [1].

2. Threshold regression models

2.1. First-hitting-time models

Instead of focusing on hazard rate, threshold regression is based on the concept of a first hitting time. The approach provides an investigator with a general conceptual framework. The defining feature of threshold regression is that the event time is defined as the first time an underlying stochastic process hits a boundary threshold. In a medical context, for example, the event of interest might be death and the time of death is the moment when the patient's latent health status first reaches a critical threshold at zero.

For a stochastic process $\{Y(t), t \geq 0\}$, let $\mathcal{B}$ denote the boundary threshold and S, the first hitting time. Then their interconnection can be expressed mathematically as follows:

$$S = \inf\{t : Y(t) \in \mathcal{B}\},$$

where initial level $Y(0) \notin \mathcal{B}$. In medical applications, the stochastic process $\{Y(t)\}$ may describe the time trajectory of health or disease for a subject. The parameter t denotes time. The boundary $\mathcal{B}$ is a critical health state, disease state or other medical end point, such as death, a diagnosis of cancer, or hospital discharge. The first hitting time S is the time for the sample path of the stochastic process to first reach the boundary $\mathcal{B}$. It is this first hitting time, or FHT for short, that is the time-to-event or survival time of interest.

The stochastic process $\{Y(t)\}$ may take many forms, including a Wiener process, gamma process, Ornstein-Uhlenbeck process, or Markov chain. The nature of the boundary state may vary widely, for example, a fixed threshold in a Wiener process or an absorbing state in a Markov chain.

Previous work that has considered regression structures for FHT models includes Whitmore [22], Whitmore, Crowder and Lawless [23], Lee, DeGruttola and Schoenfeld [10] and Lee et al. [14]. Lee, DeGruttola and Schoenfeld [10] use a bivariate Wiener diffusion process as the basis of a regression model for the study of progression to death in AIDS, with CD4 cell count serving as a marker process.

2.2. Computational aspects of threshold regression models

Consider a Wiener process $\{Y(t), t \geq 0\}$ with mean parameter μ, variance parameter σ^2, and initial value $Y(0) = y_0 > 0$. The time required for the process to reach the zero level for the first time has an inverse Gaussian distribution if the process mean parameter μ is negative so the process tends to drift toward the zero level. See Lee and Whitmore [11].

Computational details of threshold regression methods in STATA and R packages can be found in user manuals by Xiao et al. [25, 26].

When we assume that the latent stochastic process is a Wiener process $Y(t)$ starting at y_0 with drift μ and variance σ^2, we know that it has the following properties:

1. $Y(t)$ has independent increments; for any non-overlapping time intervals (t_1, t_2), (t_3, t_4), $Y(t_2) - Y(t_1)$ and $Y(t_4) - Y(t_3)$ are independent.

2. $Y(t_2) - Y(t_1)$ is normally distributed with mean $\mu(t_2 - t_1)$ and variance $\sigma^2(t_2 - t_1)$ with $t_1 < t_2$.

When we regard this Wiener process as a latent health status process, we can let $Y(0) = y_0 > 0$ be the initial health status, and define S as the first time a sample path of the health status process reaches level 0, i.e., $S = \inf\{t : Y(t) = 0\}$.

By using either the backward or forward diffusion equations subject to the initial condition and the boundary condition for the absorbing barrier, it can be shown that S follows an inverse Gaussian distribution with the following probability density function (p.d.f.).

$$f(t \mid \mu, \sigma^2, y_0) = \frac{y_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left[-\frac{(y_0 + \mu t)^2}{2\sigma^2 t}\right],$$

where $\sigma^2 > 0$, $y_0 > 0$ and $-\infty < \mu < \infty$. The p.d.f. is proper if $\mu \leq 0$. The cumulative distribution function (c.d.f.) of the first hitting time is

$$F(t \mid \mu, \sigma^2, y_0) = \Phi\left[-\frac{(y_0 + \mu t)}{\sqrt{\sigma^2 t}}\right] + \exp\left(-\frac{2y_0\mu}{\sigma^2}\right) \Phi\left[\frac{\mu t - y_0}{\sqrt{\sigma^2 t}}\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Note that if $\mu > 0$, the Wiener process may never hit the boundary at zero and hence there is a probability that the FHT is ∞, with $P(\text{FHT} = \infty) = 1 - \exp(-2y_0\mu/\sigma^2)$. See Cox and Miller [6].

By carefully examining the above equations, it can be seen that both $f(t \mid \mu, \sigma^2, y_0)$ and $F(t \mid \mu, \sigma^2, y_0)$ actually depend on y_0/σ and μ/σ only. Hence we need to fix one of the three parameters (μ, y_0, σ) to avoid over-parameterization. Because the degradation process is latent with undefined measurement scale, we choose to set the variance parameter $\sigma^2 = 1$ in computations without loss of generality. Then we can regress the other two process

parameters, y_0 and μ, on covariates. Including a constant intercept term, the covariate vector can be written as $\boldsymbol{z}' = (1, z_1, \ldots, z_k)$, where $z_1, \ldots, z_k$ are covariates.

Note that the leading 1 in $\boldsymbol{z}'$ allows for a constant term in the regression relationship. We assume that μ and $\ln(y_0)$ are linear in regression coefficients, so they are linked to the covariates with the following regression forms:

$$\begin{aligned} \ln(y_0) &= \gamma_0 + \gamma_1 z_1 + \cdots + \gamma_k z_k = \boldsymbol{z}'\boldsymbol{\gamma}, \\ \mu &= \beta_0 + \beta_1 z_1 + \cdots + \beta_k z_k = \boldsymbol{z}'\boldsymbol{\beta}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_k)'$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$ are regression coefficient vectors. If some covariates are regarded as unimportant in predicting $\ln(y_0)$ or μ, then these covariates can be removed from the regression model by setting the corresponding elements in $\boldsymbol{\gamma}$ or $\boldsymbol{\beta}$ to zero. For example, if covariate z_1 in $\boldsymbol{z}'$ is not important to predict $\ln(y_0)$, we can set γ_1 to zero in (2.1).

Maximum likelihood estimation (MLE) is used to estimate the regression coefficients. A subject i in the sample data set with an observed exact death time $t^{(i)}$ contributes the FHT probability density $f(t^{(i)} \mid \mu^{(i)}, y_0^{(i)})$ to the sample likelihood function. A subject j in the sample data set who lives to the end of the study provides a right-censored event time. All we know about this subject is that the event time is larger than the on-study time. Therefore the contribution by a surviving subject j to the sample likelihood function is the survival function evaluated at the corresponding on-study time $t^{(j)}$, namely, $1 - F(t^{(j)} \mid \mu^{(j)}, y_0^{(j)})$. Among the n subjects in the sample, subjects with observed death times are indexed from 1 to n_1 and subjects with right-censored event times are indexed from $n_1 + 1$ to n. Hence, the log-likelihood function is

$$\ln L(\beta, \gamma) = \sum_{i=1}^{n_1} \ln f(t^{(i)} \mid \mu^{(i)}, y_0^{(i)}) + \sum_{j=n_1+1}^{n} \ln \left[1 - F(t^{(j)} \mid \mu^{(j)}, y_0^{(j)})\right].$$

2.3. Three building blocks of threshold regression models

It can be seen that a TR model has three building blocks: (1) a stochastic process that describes the evolution of a subject's underlying health state; (2) a boundary or threshold that defines a critical level or condition that triggers the event of interest when it is reached by the process for the first time; and (3) a time scale on which the process unfolds. Each of these building blocks may have parameters that depend on a covariate vector $\boldsymbol{z}$ through regression link functions. After a specification of the threshold regression model, the regression functions can be estimated and then various inferences made using conventional statistical theory. Lee and Whitmore [12] investigate theoretical connections between PH regression and threshold regression and show that PH regression can be

considered as a special case of TR using one of two methods of construction: one based on altering the TR time scale and the other based on varying the TR boundary. A case demonstration was presented in Lee and Whitmore [12] to highlight the understanding of scientific foundations that TR can offer in comparison to PH regression. Stogiannis et al. [20] and William and Law [21] also discussed comparisons between FHT and PH regression models.

2.4. Analytical running time versus calendar time

As pointed in Lee and Whitmore [11], the natural time scale of the parent stochastic process in many applications is not calendar or clock time. For example, a mechanical component may wear according to the amount of its usage. Mathematical research on different time scales has been carried out by many researchers. Cox and Oakes [7, Section 1.2, p. 34] pointed out that often the scale for measuring time is clock time, although other possibilities certainly arise, such as the use of operating time of a system, mileage of a car, or some measure of cumulative load encountered.

These accumulation measures are increasing with calendar time and thus are alternative progression scales for the stochastic process. Such measures are given a variety of names, depending on the context, such as operational time, disease progression, running time, or analytical time.

If $r(t)$ denotes the transformation of calendar time t to running time r, with $r(0) = 0$, and $Y(r)$ is the process of interest defined in terms of running time r, then the resulting process expressed in terms of calendar time is the subordinated process $Y^*(t) = Y[r(t)]$. The running time scale $r(t)$ is included in the FHT model in order to make the model a more valid representation of reality. With a correct specification of running time, one would expect health status or component strength to decline steadily and predictably against the scale that measures the accumulating ear and tear of running time.

2.5. Threshold regression for longitudinal data analysis

Assume that each individual has observation vectors of form $(t_j, f_j, y_j, \boldsymbol{z}_j)$, $j = 0, 1, \ldots, m$, where $t_0 = 0 \le t_1 \le \cdots \le t_m$. Here t_j is the time of the jth observation, f_j is an indicator variable for whether the time t_j is an FHT, y_j is the state of the process at time t_j, and $\boldsymbol{z}_j$ is the covariate vector of the jth observation for the individual.

Time points:	$0 = t_0 \le t_1 \le \cdots \le t_m$,
Failure codes:	$f_0 = 0, f_1 = 0, \ldots, f_{m-1} = 0, f_m = 0$ or 1,
Readings on the process:	$y_0, y_1, \ldots, y_m$,
Covariate vectors:	$\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$.

The data structures may have a variety of specialized features.

1. The data sets usually consist of a sample of individuals, $i = 1, \ldots, n$, with individual parent processes $\{Y_i(t)\}$ and boundary sets $\mathcal{B}^{(i)}$. The individual processes are often assumed to be mutually independent.

2. Where there are competing modes of failure, then the cause of failure d will be recorded for each individual.

3. The final observation time t_m for an individual is a random stopping time if $f_m = 1$.

4. Thus, $t_m = S$ and $y_m = Y(S)$ if $f_m = 1$. Here $Y(S) \in \mathcal{B}$ is the threshold state realized by the individual at the FHT. If $f_m = 0$ then time t_m is a right censoring time for the FHT, i.e., $t_m < S$. If $t_{m-1} < S \leq t_m$ then the survival time S is interval censored.

5. The data are longitudinal if there is more than one reading available for some individuals, i.e., if $m > 1$ for some individuals.

6. If the stochastic process Y is latent then the data set will have no observations y_j, although there may still be readings on the covariate vectors $\boldsymbol{z}_j$.

7. If the data set consists only of a single time t and failure indicator f for each individual then the data set constitutes censored survival data. With a baseline covariate vector $\boldsymbol{z}_0$ available, the data provide a basis for censored survival threshold regression.

8. Let $Y(t_j)$ be abbreviated y_j for any individual. The reading y_j on the parent process, for $j < m$, is a realization of the conditional random variable $Y_j \mid S > t_j$. The conditioning event is that the process has reached state y_j at time t_j without experiencing an FHT.

9. Where $\{Y(t)\}$ is a Markov process (which is the most common type of model), we have for any individual that

$$P(Y_j = y_j \mid y_{j-1}, \ldots, y_0, S > t_j) = P(Y_j = y_j \mid Y_{j-1} = y_{j-1}, S > t_j) \quad \text{for } j < m.$$

   In other words, the distribution of the next observation Y_j depends only on the value of preceding observation y_{j-1} and the fact that no FHT has yet occurred. The sample path by which y_{j-1} was attained is immaterial (Lee and Whitmore [11]).

Using previous notation, let $\{A_j\}$ denote the *longitudinal observation process*, defined on the time points t_j, $j = 0, 1, \ldots$. If the individual survives beyond time t_j then failure

code $f_j = 0$ and $A_j = \{S > t_j, y_j, \boldsymbol{z}_j\}$ for $j \leq m$. If the individual fails in the final interval $(t_{m-1}, t_m]$ then $f_m = 1$ and $A_m = \{S \in (t_{m-1}, t_m], x_m \in \mathcal{B}\}$. As defined earlier, S is the stopping time for the longitudinal observation process. We note that z_m is not defined when the individual has failed and, hence, is dropped from the expression for A_m. Moreover, the final reading y_m for the parent process lies inside the boundary set $\mathcal{B}$ when the individual has failed.

Longitudinal data of this kind pose an interesting challenge for first-hitting-time models, as for most time-to-event models. This method can be referred to as an *uncoupling* procedure because it effectively unlinks the longitudinal observations into a set of independent conditional observations. Lee, Whitmore, Rosner [16] used the uncoupling procedure to analyze longitudinal data from the Nurses Study. With the preceding notation, the probability of observing the longitudinal data record of an individual can be expanded as a product of conditional probabilities as follows:

$$P(A_m, A_{m-1}, \ldots, A_1, A_0) = P(A_0) \prod_{j=1}^{m} P(A_j \mid A_{j-1}, \ldots, A_0). \tag{2.2}$$

Now we come to the crucial assumption. If it can be assumed that $\{A_j, j = 0, 1, \ldots\}$ is a Markov process with initial state A_0, then (2.2) can be simplified as follows:

$$P(A_m, A_{m-1}, \ldots, A_1, A_0) = P(A_0) \prod_{j=1}^{m} P(A_j \mid A_{j-1}). \tag{2.3}$$

In other words, the probability of observing A_j depends only on its preceding state A_{j-1} and not on the earlier history of the observation process. The explicit forms of the probability elements on the right-hand side of (2.3) are as follows:

$$P(A_j \mid A_{j-1}) = P(S > t_j, y_j, \boldsymbol{z}_j \mid S > t_{j-1}, y_{j-1}, \boldsymbol{z}_{j-1}) \quad \text{if } f_j = 0,\ j \leq m, \tag{2.4}$$

$$P(A_m \mid A_{m-1}) = P(S \in (t_{m-1}, t_m], y_m \in \mathcal{B} \mid S > t_{m-1}, y_{m-1}, \boldsymbol{z}_{m-1}) \quad \text{if } f_m = 1. \tag{2.5}$$

If no observations are available on the parent process then y_j is dropped from the A_j notation, giving $A_j = \{S > t_j, \boldsymbol{z}_j\}$ if $f_j = 0$, $j \leq m$, and $A_m = \{S \in (t_{m-1}, t_m]\}$ if $f_m = 1$. Again, invoking the Markov assumption for the observation process, (2.4) and (2.5) take the revised forms:

$$\begin{aligned} P(A_j \mid A_{j-1}) &= P(S > t_j, \boldsymbol{z}_j \mid S > t_{j-1}, \boldsymbol{z}_{j-1}) \quad \text{if } f_j = 0 \text{ for } j \leq m, \\ P(A_m \mid A_{m-1}) &= P(S \in (t_{m-1}, t_m] \mid S > t_{m-1}, \boldsymbol{z}_{m-1}) \quad \text{if } f_m = 1. \end{aligned}$$

Statement (2.3) is the theoretical justification for the *uncoupling procedure*.

2.6. Cure rate

Some FHT models may offer a positive probability of no FHT taking place in finite time. Thus, for example, a medical treatment may offer a cure, some animals in a population may be immune to infection, some stock prices may never reach $1000, and some marriages may never end in divorce. The fact that $P(S = \infty) > 0$ in some FHT models is closely related to competing risks. Generally, if the FHT model takes account of *all* competing risks then eventual failure from some cause is assured. If, however, the FHT model takes account of only one or a few competing risks then there is a positive probability that the FHT will be infinite to accommodate those individuals who are not susceptible to the limited array of causes of failure that are considered in the model. To illustrate the natural way in which FHT models take account of a cure rate, consider a Wiener diffusion model with a fixed boundary at zero (the time axis). If the drift of the process is away from the boundary, i.e., $\mu > 0$, then a finite FHT is not assured and, in particular, $P(S < \infty) = \exp(-2y_0\mu/\sigma^2)$. Likewise, a modified gamma process with a cure rate might be defined as follows:

$$Y(t) = \begin{cases} y_0 & \text{with probability } 1-p, \\ y_0 - Z(t) & \text{with probability } p. \end{cases}$$

Here parameter p is a susceptibility fraction, with $0 \leq p \leq 1$, and $Z(t)$ is a gamma stochastic process. As an example of this last model, a subject may have a malignant or benign form of a disease with probabilities p and $1-p$, respectively. The malignant form progresses monotonically towards a medical endpoint (e.g., death).

See Lee and Whitmore [11] and Balka et al. [5] for a review on the implementation of cure models based on first hitting times for Wiener processes.

3. Applications

The TR model has recently begun to attract more attention. Bayesian random effects threshold regression was discussed in Pennell et al. [19]. Longitudinal analysis was examined in Lee, Whitmore and Rosner [16]. In this section, we review some case studies using TR regression models.

Lee, Chang, Whitmore [9] use the TR model to analyze data from a randomized clinical trial for treatment of multiple myeloma. The trial compares VELCADE and high-dose Dexamethasone, the former being a new therapy and the latter an established therapy for this disease. Patients are switched between the two drugs based on patient response. The novel contribution of this work is the modeling of this clinical trial design using a mixture of TR models. Specifically, they propose a mixture FHT model to fit the survival distribution. The model includes a composite time scale that differentiates the rate of

disease progression before and after switching. The analysis shows significant benefit from initial treatment by VELCADE. A comparison is made with a Cox proportional hazards regression analysis of the same data. Although the Cox regression results agree broadly with the TR results, TR provides more subtle insights into the source and nature of the comparative benefits of VELCADE than offered by the Cox methodology.

Lee et al. [15] considered a case-control study of lung cancer mortality in U.S. railroad workers. The case-control study data for workers in jobs with and without diesel exhaust exposure are reanalyzed using threshold regression methodology. The study included 1256 workers who died of lung cancer and 2385 controls who died primarily of circulatory system diseases. Diesel exhaust exposure was assessed using railroad job histories from the US Railroad Retirement Board and an industrial hygiene survey. Smoking habits were available from next-of-kin and potential asbestos exposure was assessed by job history review. The new analysis reassesses lung cancer mortality and examines circulatory system disease mortality. Jobs with regular exposure to diesel exhaust had a survival pattern characterized by an initial delay in mortality, followed by a rapid deterioration of health prior to death. The pattern is seen in subjects dying of lung cancer, circulatory system diseases, and other causes. The unique pattern is illustrated using a new type of KaplanMeier survival plot in which the analytical time scale represents a measure of disease progression rather than calendar time. The disease progression scale accounts for a healthy-worker effect when describing the effects of cumulative exposures on mortality.

Aaron et al. [3] compares a Poisson process TR model with a Wiener diffusion TR model for the occurrence of acute exacerbations in chronic obstructive pulmonary disease (COPD). They incorporate the causal determinants of disease operating in each patient. They test the methodology on COPD data from a randomized clinical trial. Results show that both models provide reasonably accurate fits to the clinical trial data. Analysis of the clinical trial data set using these TR models revealed that patients who experienced multiple exacerbations showed a progressive acceleration in the rate of exacerbations, and successive shortening of stable intervals between exacerbations.

Li and Lee [17] consider a semi-parametric modeling approach for TR and contribute details about theory and implementation for model fitting and statistical inferences with semi-parametric varying coefficients. Extensive simulations are carried out to examine the finite sample performance of the parametric and non-parametric estimates. A real example is analysed to illustrate the methods, along with a careful diagnosis of model assumptions.

Whitmore, Ramsay and Aaron [24] formulate a recurrent event process as a succession of independent and identically distributed first hitting times for a Wiener sample path as it passes through successive equally-spaced levels. They develop exact mathematical results

for statistical inferences based on several observation schemes that include observation initiated at a renewal point, observation of a stationary process over a finite window, and other variants. They demonstrate their results using data from a clinical trial for COPD in which the recurrent events are successive exacerbations of the condition.

Aaron et al. [4] construct a statistical model to assess the risk of death for cystic fibrosis (CF) patients between scheduled annual clinic visits. The model includes a CF health index that shows the influence of risk factors on CF chronic health and on the severity and frequency of CF exacerbations. Their model produces an accurate clinical scoring tool for prediction of one-year survival of CF patients. The tool can be used by clinicians to decide on optimal timing for lung transplant referral.

Mulatya et al. [18] propose utilizing a longitudinal threshold model to estimate the distribution of the elapsed time between two thresholds of the longitudinal process from repeated measurements. They extend this modeling framework to be used with multiple thresholds. A Wiener process under the first hitting time framework is used to represent a survival distribution. They demonstrate their model through simulation studies and an analysis of data from the Consortium on Safe Labor study.

Considering a composite of a chronic degradation process for skeletal health combined with a random stream of shocks from external traumas, He et al. [8] develop a shock-degradation TR model and study first and second fractures of elderly women using data from the Study of Osteoporotic Fractures.

Lee and Whitmore [13] consider a family of system failure models in which shock streams that follow a Fréchet process are superimposed on a degrading system described by a stochastic process with stationary independent increments.

Li and Whitmore [13] extended the TR model to the setting of complex sample survey designs that involve (a) differential selection probabilities of study subjects and (b) intracluster correlations induced by multistage cluster sampling. The pseudo-maximum likelihood estimation technique is applied to estimate the TR model parameters. Computationally efficient Taylor linearization variance estimators that consider both the intracluster correlation and the differential selection probabilities are developed. The proposed methods are evaluated by using simulation experiments with various complex designs and illustrated empirically by using mortality-linked Third National Health and Nutrition Examination Survey Phase II genetic data.

4. Future research

The earlier work on model validation has been largely restricted to the FHT model for a Wiener process and thus extensions to other FHT models need attention. Much remains to be done on model validation and diagnostic techniques in the general context of threshold

regression.

Because the TR model involves the simultaneous estimation of several regression functions, variable selection is an important aspect that needs attention. For a Wiener diffusion TR model, for example, at least two regression functions are estimated; one for the initial baseline process level and another for the mean drift. It is also important to work closely with subject-matter specialists to ensure that the FHT models have realistic features and that the findings emerging from the analysis make practical sense.

Acknowledgments

The work is support in part by R01EY02445.

References

[1] O. O. Aalen, Ø. Borgan and H. K. Gjessing, *Survival and Event History Analysis: A Process Point of View*, Statistics for Biology and Health, Springer, New York, 2008.

[2] O. O. Aalen and H. K. Gjessing, *Understanding the shape of the hazard rate: a process point of view*, Statist. Sci. **16**, (2001), no. 1, 1–22.

[3] S. D. Aaron, T. Ramsay, K. Vandemheen and G. A. Whitmore, *A threshold regression model for recurrent exacerbations in chronic obstructive pulmonary disease*, J. Clin. Epidemiol. **63** (2010), no. 12, 1324–1331.

[4] S. D. Aaron, A. L. Stephenson, D. W. Cameron and G. A. Whitmore, *A statistical model to predict one-year risk of death in patients with cystic fibrosis*, J. Clin. Epidemiol. **68** (2015), no. 11, 1336–1345.

[5] J. Balka, A. F. Desmond and P. D. McNicholas, *Review and implementation of cure models based on first hitting times for Wiener processes*, Lifetime Data Anal. **15** (2009), no. 2, 147–176.

[6] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, John Wiley & Sons, New York, 1965.

[7] D. R. Cox and D. Oakes, *Analysis of Survival Data*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1984.

[8] X. He, G. A. Whitmore, G. Y. Loo, M. C. Hochberg and M.-L. T. Lee, *A model for time to fracture with a shock stream superimposed on progressive degradation: the study of osteoporotic fractures*, Stat. Med. **34** (2015), no. 4, 652–663.

[9] M.-L. T. Lee, M. Chang and G. A. Whitmore, *A threshold regression mixture model for assessing treatment efficacy in a multiple myeloma clinical trial*, J. Biopharm. Statist. **18** (2008), no. 6, 1136–1149.

[10] M.-L. T. Lee, V. DeGruttola and D. Schoenfeld, *A model for markers and latent health status*, J. R. Stat. Soc. Ser. B Stat. Methodol. **62** (2000), no. 4, 747–762.

[11] M.-L. T. Lee and G. A. Whitmore, *Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary*, Statist. Sci. **21** (2006), no. 4, 501–513.

[12] ———, *Proportional hazards and threshold regression: their theoretical and practical connections*, Lifetime Data Anal. **16** (2010), no. 2, 196–214.

[13] ———, *Practical applications of a family of shock-degradation failure models*, in: *Statistical Modeling for Degradation Data*, 211–229, ICSA Book Ser. Stat., Springer, Singapore, 2017.

[14] M.-L. T. Lee, G. A. Whitmore, F. Laden, J. E. Hart and E. Garshick, *Assessing lung cancer risk in railroad workers using a first hitting time regression model*, Environmetrics **15** (2004), no. 5, 501–512.

[15] ———, *A case-control study relating railroad worker mortality to diesel exhaust exposure using a threshold regression model*, J. Statist. Plann. Inference **139** (2009), no. 5, 1633–1642.

[16] M.-L. T. Lee, G. A. Whitmore and B. A. Rosner, *Threshold regression for survival data with time-varying covariates*, Stat. Med. **29** (2010), no. 7-8, 896–905.

[17] J. Li and M.-L. T. Lee, *Analysis of failure time using threshold regression with semiparametric varying coefficients*, Stat. Neerl. **65** (2011), no. 2, 164–182.

[18] C. M. Mulatya, A. C. McLain, B. Cai, J. W. Hardin and P. S. Albert, *Estimating time to event characteristics via longitudinal threshold regression models–an application to cervical dilation progression*, Stat. Med. **35** (2016), no. 24, 4368–4379.

[19] M. L. Pennell, G. A. Whitmore and M.-L. T. Lee, *Bayesian random-effects threshold regression with application to survival data with nonproportional hazards*, Biostatistics **11** (2009), no. 1, 111–126.

[20] D. Stogiannis, C. Caroni, C. E. Anagnostopoulos and I. K. Toumpoulis, *Comparing first hitting time and proportional hazards regression models*, J. Appl. Stat. **38** (2011), no. 7, 1483–1492.

[21] C. L. William and C. Law, *Threshold regression and first hitting time models*, Research & Reviews: J. Stat. Math. Sci. **1** (2015), no. 1, 38–48.

[22] G. A. Whitmore, *A regression method for censored inverse-Gaussian data*, Canad. J. Statist. **11** (1983), no. 4, 305–315.

[23] G. A. Whitmore, M. J. Crowder and J. F. Lawless, *Failure inference from a marker process based on a bivariate Wiener model*, Lifetime Data Anal. **4** (1998), no. 3, 229–251.

[24] G. A. Whitmore, T. Ramsay and S. D. Aaron, *Recurrent first hitting times in Wiener diffusion under several observation schemes*, Lifetime Data Analysis **18** (2012), no. 2, 157–176.

[25] T. Xiao, G. A. Whitmore, X. He and M.-L. T. Lee, *Threg: a new command to implement threshold regression model in STATA*, The STATA Journal **12** (2012), 257–283.

[26] _____, *The R package threg to implement threshold regression model*, J. Stat. Softw. **66** (2015), no. 8, 16 pp.

Mei-Ling Ting Lee
University of Maryland, College Park, MD 20742, USA
E-mail address: mltlee@umd.edu