Reliability Engineering and System Safety 124 (2014) 201–206

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Proportional hazards models of infrastructure system recovery

Kash Barker *, Hiba Baroud

School of Industrial and Systems Engineering, University of Oklahoma, USA

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form
18 December 2013
Accepted 21 December 2013
Available online 30 December 2013

Keywords:
Infrastructure systems
Recovery
Proportional hazards

ABSTRACT

As emphasis is being placed on a system's ability to withstand and to recover from a disruptive event, collectively referred to as dynamic resilience, there exists a need to quantify a system's ability to bounce back after a disruptive event. This work applies a statistical technique from biostatistics, the proportional hazards model, to describe (i) the instantaneous rate of recovery of an infrastructure system and (ii) the likelihood that recovery occurs prior to a given point in time. A major benefit of the proportional hazards model is its ability to describe a recovery event as a function of time as well as covariates describing the infrastructure system or disruptive event, among others, which can also vary with time. The proportional hazards approach is illustrated with a publicly available electric power outage data set.

1. Introduction

The resilience of infrastructure systems, both in the US and globally, is of significant concern. Following the events of September 11, 2001, the primary planning concern revolved around the protection from and prevention of terrorist attacks. However, as accidents and natural disasters become more prevalent and more impactful (e.g., Hurricanes Katrina and Rita in 2005, the *Deepwater Horizon* oil spill in 2010, the Japanese earthquake and tsunami in 2011), recent efforts have been placed on resilience, or the ability to "bounce back," from such disruptions. Demonstrated in several works on resilience [1–6] is the significance of accepting that disruptions will indeed occur and focusing on two aspects: (i) lessening the impact of such disruptions, and (ii) improving the speed with which recovery occurs. As such, the US Department of Homeland Security [7] emphasizes "strengthen national preparedness, timely response, and rapid recovery" of US infrastructure, particularly due to their interconnectedness with other infrastructure, industries, and the workforce. Resilience is often described as a function of *robustness*, or the ability of a system to resist the initial adverse effects of a disruptive event, and *rapidity*, or the rate or speed at which a system is able to return to an appropriate operability following the disruption [8]. Modeling the rapidity aspect of resilience, particularly as a rate of recovery, is addressed in this work.

Introduced here is a data-driven statistical technique to model the recovery of infrastructure systems following a disruptive event. Guikema [9] provides an introduction to the use of statistical methods (e.g., generalized linear models) for performing probabilistic risk analysis, with several applications in modeling and estimating the initial impacts of natural disasters to electric power service [10,11], among other infrastructures [12,13]. MacKenzie and Barker [14] apply regression to the study of interdependent recovery following an electric power outage to populate parameters in a multi-industry interdependency model.

This work describes the proportional hazards model (PHM) [15], a standard survival analysis technique in biostatistics with applications found in reliability engineering, as a means to derive a temporal, condition-based rate of recovery for infrastructures impacted by a disruptive event. Reliability growth can be described as a function of the dynamic assumptions of the underlying system [16]. Reliability applications of PHM include analyzing factors that impact the reliability behavior of systems and components [17] and planning for preventive maintenance repairs due to degrading condition variables [18].

Similar to the reliability analysis, policymakers should strive for improved *resilience analysis* encompassing the recovery dynamics of a disrupted system. Preparedness activities should not only enhance the reliability of the system but also its resilience (e.g., its strength in responding to and recovering from a disruption [19]). Further, decisions should be made prior to and in the aftermath of a disruption to effectively allocate resources to reduce impacts [20] and enhance recovery activities [21]. Resilience modeling so far has been concerned with the optimization of preparedness and resource allocation strategies under different circumstances, but no research exists that models the trajectory of recovery over time given external and internal impacts. Different systems have different recovery paths due to their structure, the nature of the disruption, the surrounding environment, among many other factors.

A model that can capture this relationship and translate the evolution of recovery over time is the PHM. We extend the use of the PHM, primarily used herein to model failure rate or infection rate (and sparse applications in modeling repair rates [22,23]), to

* Correspondence to: School of Industrial and Systems Engineering, University of Oklahoma, 202 West Boyd Street, Room 124, Norman, OK 73019, USA.
Tel.: +1 405 325 3721; fax: +1 405 325 7555.
E-mail address: kashbarker@ou.edu (K. Barker).

0951-8320/$ - see front matter
http://dx.doi.org/10.1016/j.ress.2013.12.004

model recovery rate and supplement the *rapidity* component of existing resilience paradigms. As will be depicted later, we model time-to-recover and its associated covariates rather than the tradition time-to-failure in traditional reliability analysis.

Section 2 provides background on the PHM, while Section 3 discusses its use in the infrastructure recovery context, with an application in recovery from electric power outages from publicly available data. Concluding remarks are provided in Section 4.

2. Methodological background

Survival analysis is a technique used to describe the duration between events. Many phenomena in medical research, engineering, and economics can be described using survival analysis techniques. For example, medical histories [24–26] and the failure of engineered systems [27–29] have been described using survival analysis. Survival distributions can be described by four functions at time t: (i) the probability density $f(t)$, (ii) the cumulative distribution function $F(t)$, (iii) the survivor (or reliability [30]) $S(t)=1-F(t)$, and (iv) the hazard $h(t)=f(t)/S(t)$ functions. One means to estimate the survival function with the incorporation of covariate effects, or the risk factors influencing the occurrence of and duration between events, is the proportional hazards model (PHM) [15].

Provided in Eq. (1), the PHM hazard function describing the rate at which failures occur is a function of a time-driven baseline hazard function, $h_0(t)$, and the state/condition of the system (or covariates influencing the hazard), vector $\mathbf{x}(t)$. $\boldsymbol{\beta}$ is a vector of regression coefficients reflecting the effect of the state of the system on the hazard function.

$$h(t,\mathbf{x}(t)) = h_0(t)\exp(\beta^{\mathrm{T}}\mathbf{x}(t)) \tag{1}$$

Generally speaking in a reliability engineering context, $h_0(t)$ can be derived from a pdf fit to time-to-failure data. A system is viewed periodically during inspection, and at each inspection time t_i, the covariates $\mathbf{x}(t_i)$ are recorded, as well as a 0/1 indicator of "no failure" or "failure." Regression parameters are then fit using the method of maximum likelihood. Likewise for a biostatistics application, patients are observed at t_i, physical characteristics $\mathbf{x}(t_i)$ of the patient are recorded, and the presence or absence of an ailment is noted.

A primary reason for its popularity is that it allows the ability to assess the effect of covariates on the hazard function, and ultimately the likelihood of the event, with a semi-parametric approach, (i.e., solving for $\boldsymbol{\beta}$ without specifying a baseline hazard function, $h_0(t)$). As such, the typical use of the PHM is used for descriptive purposes (i.e., identifying factors that significantly impact hazard/survival and interpreting the elements of $\boldsymbol{\beta}$), not prescriptive purposes (i.e., actually estimating hazard/survival given a set of covariates) (e.g., [31]). Another reason for the popularity of PHM in certain contexts is that observations can be censored, a term to describe when an event is incomplete for an observation during the observed period (e.g., when applying PHM to model the occurrence of cancer in patients, some patients may leave the study without cancer appearing). Such censored observations are no less important than those for whom observation is complete, and they would not be included in other similar statistical methods (e.g., logistic regression). The survival function is estimated using the Breslow estimator, where the function takes the form of a step function due to the assumption that hazard between distinct failure points is constant [26].

Other approaches can incorporate the effects of covariates on the hazard function or the probability of event occurrence. Generalized linear models can determine the likelihood or rate of occurrence of an event given covariates with a wide range of distributions for the link function, though their ability to capture time-varying conditions is lacking [32]. The accelerated failure time model is used especially when an underlying hazard function is known, or when the baseline hazard function would vary from observation to observation [33]. A more general representation of the relationship among the survivor function $S(t,\mathbf{x})$ and the baseline survival function $S_0(t)$ is the Royston–Parmar family of models which rely on the transformation $g(\bullet)$ such that $g(S(t,\mathbf{x}))= g(S_0(t))+\beta^{\mathrm{T}}\mathbf{x}$, where $g(\bullet)$ can represent the proportional hazard, proportional odds, or probit families [34]. However, the PHM can account for the effect of time-varying covariates, $\mathbf{x}(t)$, particularly useful for post-disruption decision making as the state of the system can vary over time. Allison [35] points out that the approach given time-varying covariates is technically a non-proportional hazards model. More specifically, the PHM assesses time independent covariates while an extension of the PHM (e.g., Cox regression, extended Cox regression) model time-varying covariates or a combination thereof [22,36,37,35].

3. Infrastructure recovery with proportional hazards models

The use of PHMs has mostly been limited to describing hazard (or failure) rates, though many other important data-driven rates of occurrence can be described as a function of (i) time, and (ii) covariates. As such, the idea of modeling the time-dependent, state-of-the-system-dependent evolution of occurrence rates is extended in this work to modeling the rate of recovery of infrastructure systems following a disruptive event. As parameter μ is often used in maintenance context to describe repair, such notation is adopted in Eq. (2). The usefulness of $\mu(t,\mathbf{x}(t))$ lies in modeling how recovery rate changes over time (e.g., recovery rate would likely decrease over time after the initial disruption), perhaps more so when accounting for $\mathbf{x}(t)$. Recalling that resilience is a function of *robustness* and *rapidity*, when paired with robustness, this innovation in the PHM will help derive the concept of rapidity from data sources. Eq. (2) accounts for both time independent and time-varying covariates, $\mathbf{x}$ and $\mathbf{x}(t)$, respectively.

$$\mu(t,\mathbf{x}(t)) = \mu_0(t)\exp(\beta^{\mathrm{T}}\mathbf{x}+\beta^{\mathrm{T}}\mathbf{x}(t)) \tag{2}$$

Further, the likelihood $V(t,\mathbf{x}(t))$ of full recovery before time t and under condition $\mathbf{x}(t)$, shown in Eq. (3) is extended from the reliability literature: $V(t,\mathbf{x}(t))$ would be the equivalent of the cumulative distribution function for failure, or the probability that the event occurs prior to time t when $\mathbf{x}(t)$ is exhibited at t. This measure would provide a decision maker with an idea of how likely it is that recovery will occur by a given point in time and for a given set of covariates (potentially time-varying covariates representing state variables at time t).

$$V(t,\mathbf{x}(t)) = 1-\exp\left[-\int_0^t \mu(y,\mathbf{x}(y))dy\right] \tag{3}$$

A qualitative illustration of Eqs. (2) and (3) could include the response to and recovery of disrupted highway segments by dispatching emergency assistance vehicles. Time-invariant covariates could include the particular highway segment, number of vehicles involved, response category issued, and the location from which emergency assistance was transmitted. Time-varying covariates could include the number of responders expediting the cleanup. Periodic review of the accident could result in a 0/1 outcome of "cleanup still underway" or "cleanup complete and traffic resumed." $\mu(t,\mathbf{x}(t))$ would provide decision makers with an idea of how the recovery rate progresses over time with certain covariate values. Similarly, $V(t,\mathbf{x}(t))$ provides the likelihood that recovery will occur prior to time t under certain conditions models with the covariates.