Fall 2014

On some inferential problems with recurrent event models

Withanage Ajith Raveendra De Mel

ON SOME INFERENTIAL PROBLEMS WITH RECURRENT EVENT MODELS

by

WITHANAGE AJITH RAVEENDRA DE MEL

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS & STATISTICS

2014

Approved by:

AKIM ADEKPEDJOU, Advisor
V. A. Samaranayake
K. D. Zamba
Xuerong Wen
John Singler

DEDICATION

This dissertation is dedicated to my wife, Sanjeewanie, and my kids Thenuki and Senithi for their constant love and support.

ABSTRACT

Recurrent events (RE) occur in many disciplines, such as biomedical, engineering, actuarial science, sociology, economy to name a few. It is then important to develop dynamic models for their modeling and analysis. Of interest with data collected in a RE monitoring are inferential problems pertaining to the distribution function F of the time between occurrences, or that of the distribution function G of the monitoring window, and their functionals such as quantiles, mean. These problems include, but not limited to: estimating F parametrically or nonparametrically; goodness of fit tests on an hypothesized family of distributions; efficient of tests; regression-type models, or validation of models that arise in the modeling and analysis of RE. This dissertation work focuses on several inferential problems of significant importance with these types of data. The first one we dealt with is the problem of *informative monitoring*. Informative monitoring occurs when G contains information about F, and the information is accounted for in the inferential process through a Lehman-type model, $1 - G = (1 - F)^{\beta}$, so called generalized Koziol-Green model in the literature. We propose a class of inferential procedures for validating the model. The research work proceeds with the development of a flexible, *random cells* based chi-square goodness of fit test for an hypothesized family of distributions with unknown parameter. The cells are random in the sense that they are cut free, are function of the data, and are not predetermined in advance as is done in standard chi-square type tests. A *minimum chi-square estimator* is used to construct the test statistic whose power is assessed against a sequence of Pitman-like alternatives. The last problem we considered is that of an *efficiency*, *optimality*, and *comparison* of various statistical tests on RE that are derived in this work and existed in the literature. The efficiency and optimality are obtained by extending the theory of Bahadur and Wieand to RE. Asymptotic properties of the different estimators and or statistics are presented via empirical processes tools. Small sample results using intensive simulation study of the various procedures are presented, and these show good approximation of the truth. Real recurrent event data from the engineering and biomedical studies are utilized to illustrate the various methods.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Akim Adekpedjou, for suggesting the topic of this dissertation and for his patience and thoughtful discussion at any time during the research. Without you, this dissertation would not have been possible. I would like to give special thank to Dr. K. D. Zamba for his valuable inputs over the course of this research work. Thanks also go to my advisory committee, Dr. V. A. Samaranayake, Dr. K. D. Zamba, Dr. Xuerong Wen, and Dr. John Singler.

The financial assistance of the Department of Mathematics and Statistics during the course of my graduate studies is gratefully acknowledged.

I also wish to thank my family and my parents for their encouragement during the course of this work. Special thank goes to my wife, Sanjeewanie, my daughters, Thenuki and Senthi.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

In life testing, medical follow-up studies, and other fields, it is often impossible to observe the lifetimes of all experimental units in the study. What makes measuring durations difficult is time itself. In most cases, it is highly likely that all the events have not been observed by the time one wants to make inferences about lifetimes. For example, a medical professional will not wait fifty years for each individual in the study to pass away before closing the study. He or she is interested in the effectiveness of improving lifetimes after only a few years. The individuals in the study who have not died by the end of the study period are labeled as right-censored: all information we have on these individuals are their current lifetimes durations which is naturally less than their actual lifetimes. The simplest kind of censoring is that of single censoring which occurs when all observations are censored at the same time. There are two types of single censoring: Type I censoring and Type II censoring. In Type I censoring, the censoring time is predetermined. Type II censoring occurs if an experiment stops when a predetermined number of failures are observed; the remaining subjects are then right-censored. In many studies, observations are not censored at the same time, which is frequently referred to as arbitrary censored data. For instance, in a clinical trial, censoring occurs because of event from causes that are not related to what is being investigated in the study such as: self removal from study, drop out.

Under the random censorship model, we assume that $X_1, X_2, \ldots, X_n$ are independent nonnegative random variables with continuous distribution function $F(x) = P(X \leq x)$. The censoring variables $Y_1, Y_2, \ldots, Y_n$ are also nonnegative and are assumed to be a random sample, drawn independently of the X_i's from a population with continuous distribution function $G(y) = P(Y \leq y)$. The Y_is right-censored the X_is. The observable random variables are $Z_i = \min(X_i, Y_i)$ and $\delta_i = I\{Z_i = X_i\}$, δ_i indicates whether Z_i is an uncensored observation or not. In this model, the X_is represent times to an endpoint event (e.g., death, relapse, malfunctioning) and the Y_is represent censoring times. In the random censorship model, informative censoring occurs when the distribution function G is informative about the distribution function F.

In recurrent events, we consider an event process of $i = 1, 2, \ldots, n$ units wherein the jth event occurred at calendar time $S_{i,j}$. Suppose that for unit i the recurrent events are observed over a random interval $[0, \tau_i]$ where the τ_is are independent and identically distributed (i.i.d.) with an absolutely continuous distribution function $G(t) = P(\tau \le t)$. Let $T_{i,j} = S_{i,j} - S_{i,j-1}$ be the time between two occurrences of the event, so called gap time, or inter-event time and these are assumed to be i.i.d with absolutely continuous distribution function $F(t) = P(T_{i,j} \le t)$. For the ith unit, the $T_{i,j}$s could be viewed as the time elapsed between the $(j-1)$th and the jth occurrences. If K_i is the total number of occurrences for unit i, then the observable for n units is n i.i.d. copies $\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_n$ with

$$\mathcal{O}_i = (K_i, \tau_i, T_{i,1}, \ldots, T_{i,K_i}, \tau_i - S_{i,K_i}, (X_i(s) : s \le \tau_i)), \tag{1.1}$$

Note that, the observation time $\tau_i - S_{i,K_i}$ is redundant since it is determined by the other observable variables. It is retained, however, to emphasize the fact that this variable right-censors T_{i,K_i+1}. $\mathbf{X} = \{X_i(s) : s \ge 0\}$ is a possibly $m-$ dimensional time dependent covariates vector associated with the unit. In reliability or engineering studies, the components of this covariate vector could be related to environmental or operating condition characteristics; in biomedical studies, they could be blood pressure, treatment assigned, initial tumor size, etc. In biomedical studies, for example, the interoccurrence times for a unit may be associated or correlated, possibly because of observed individual biological variation Z or so-called frailties, thereby rendering the i.i.d. assumption of interoccurrence times restrictive in biomedical studies. One obvious generalization allowing association between interoccurrence times is a frailty model. We will not investigate frailty modelling in this thesis. A pictorial representation of recurrent event data for the ith unit is given in Figure 1.1, cf. Adekpedjou, A. [3].

Note that the components of the random vector $(T_{i,1}, T_{i,2}, \ldots, T_{i,K_i}, T_{i,K_i+1})$ are not anymore i.i.d. random variables from F, in particular, the distribution of T_{i,K_i} is not anymore F. Indeed, the interocurrence times vector satisfies the sum-quota constraint

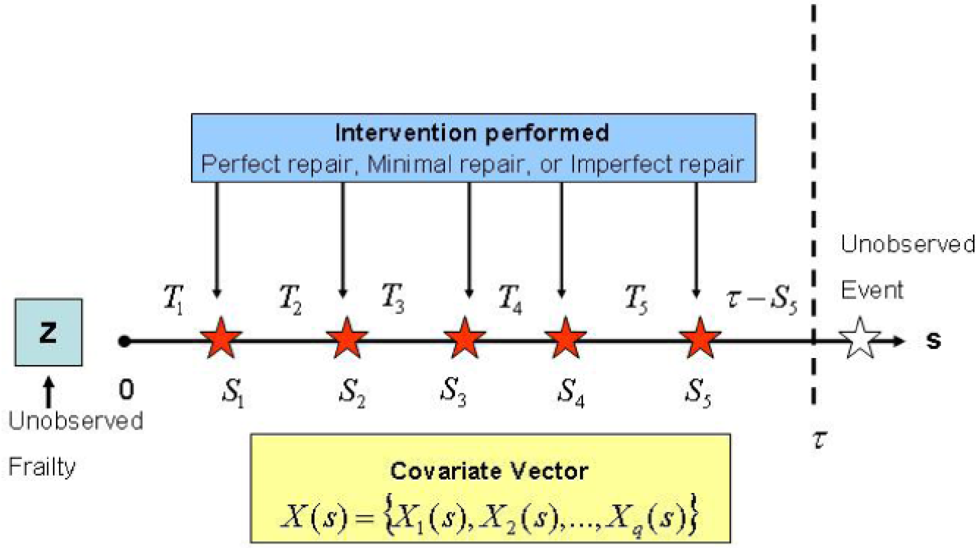

Figure 1.1. A pictorial representation of recurrent event data.

given by $\sum_{j=1}^{K_i} T_{i,j} \le \tau_i < \sum_{j=1}^{K_i+1} T_{i,j}$. Consequently, this recurrent event model has both an informative censoring mechanism as well as a dependent censoring structure. We will discuss informative censoring in detail in Section 4.

Recurrent event data frequently arise in a wide variety of settings including the biomedical, psychiatry, engineering, social sciences, and economics. Examples of such events in the health and biomedical sciences are drug abuse of teenagers or adults, recurrent hospitalization of patients with chronic diseases. In psychiatric studies the onset of depression and dementia are instances of recurrent events, in engineering and reliability settings, the break down of mechanical or electronic systems. In sociology, absenteeism rate of employees and the recurrence of war and conflict in geographical regions. In actuarial science, such as keeping track of a claim from a given insurer, are potential examples of these types of data.

Due to recurrent events high prevalence and importance in many diverse areas, it is essential to develop stochastic models and statistical methods appropriate for analyzing them. These analyses include, but not limited to: estimation of model parameters such as the survivor function $\bar{F}(t) = 1 - F(t)$, the cumulative hazard rate function $\Lambda(t) = \int_0^t \lambda(w)dw$, where $\lambda(w)$ is the hazard rate function of F given by $\lambda(w) =$

$f(w)/\bar{F}(w),\ f(w) = dF(w)/dw$. Other major inferential problems include goodness of fit tests pertaining to the distribution function F, along the lines of Kolmogrorov-Smirnov, Cramér-von Mises, or the Pearson's chi-square type tests.

There has been a sustain interest in the general problem of testing goodness of fit for a parametric family of distributions, especially the development of chi-square type tests, since the pioneering works of Pearson [27] and Pearson [28]. These interests and motivation come from the fact that in survival analysis, for instance, there may be physical reasons of having a parametric family for the underlying failure time distribution. In reliability studies, extreme values distributions such as Gumbel, Fréchet, or Weibull come as limit of distributions of parallel or series systems. In actuarial science, if the parametric distribution provide good fit to the data, they could be used for modeling large claims. In the area of failure time data analysis-under valid assumptions-parametrically driven estimates of relative hazard, survival time or their functionals such as mean or median, tend to have smaller standard errors than they would in non-parametric settings.

Going back to the definition of right censored data, if we have full knowledge of the data, then, the empirical distribution function given by $\hat{F}_n(x) = \sum I(X_i \leq x)/n$ can be used as an estimate for F. However, we have an incomplete data. Kaplan and Meier (1958) developed a nonparametric estimator of F called $\hat{F} = \hat{F}_{KM}$, based on censored data. The asymptotic properties of $\hat{F}_{KM}$ were established by Efron [26], Breslow and Crowley [17] and Gill [31], among others. If there is no censoring, $\hat{F}_{KM}$ reduces to the empirical distribution function $\hat{F}$, which is the basis of Pearson's procedure. However, the efficiency of KM estimator is lost if informative censoring is present. Koziol and Green [44] proposed an appealing and convenient model to assess informative censoring in the single event settings . In their model, they assume existence of some parameter $\beta \geq 0$ such that F and G are related via

$$\bar{G}(\cdot) = \bar{F}^{\beta}(\cdot),$$

and β was interpreted as the censoring parameter and we have $P(X_i \leq Y_i) = \frac{1}{\beta}$.

A similar result has been derived with recurrent event data. As pointed out in Peña et al. [60], recurrent event data have additional features that require attention in performing statistical inference. Two of these important features are (i) because of the sum-quota data accrual scheme, the number of observed event occurrences is informative about the inter-event distribution even if G is unrelated to F; and (ii) the variable that right-censors the last inter-event time is dependent on the pervious inter-event times. Thus, there is both an informative and dependent censoring in recurrent event data. If both an informative censoring mechanism as well as a dependent censoring structure of recurrent event data accrual are not properly accounted for, the efficiency properties may be lost. This was demonstrated analytically and through simulation studies in Peña et al. [60] and Adekpedjou et al. [5].

Adekpedjou et al. [5] generalized the Koziol-Green(KG) model (henceforth GKG) to recurrent event data. They postulate the existence of a parameter $\beta > 0$ such that $\bar{G} = \bar{F}^{\beta}$, where $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$ are the survival functions of the interoccurrence times and τ_is respectively. The GKG model is equivalent to $\Lambda_G = \beta\Lambda_F$, where Λ_G and Λ_F are the cumulative hazard rate functions of G and F respectively. In this case, the parameter β determines the length of the monitoring period relative to the interoccurrence times and a better interpretation is a monitoring parameter. More details on the GKG model can be found in Adekpedjou et al. [5].

We now provide the specific aims of this dissertation.

1.1. SPECIFIC AIMS

1. To propose procedures for assessing the validity of the generalized Koziol-Green model with recurrent events by comparing two competing estimators of the cumulative hazard rate processes. To that end, a chi-square and Kolmogorov-Smirnov type tests are developed in Section 4.

2. To derive a chi-square test based on random cells with recurrent event data for testing a composite hypothesis $F \in \mathcal{F}_{\boldsymbol{\theta}} = \{F(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}$, $\boldsymbol{\theta}$ unknown. Random cells will allow flexibility in obtaining the bounds of the cells instead of being pre-determined in advance. This is done in Section 5.

3. By considering Pitman-type alternatives, we will investigate the asymptotic power of the proposed chi-square test with random cells in Section 6.

4. The choice of the matrix used to scale the vector of differences between observed and expected frequencies in proposed chi-square procedure tests is an interesting and important problem. We investigate the optimality properties of three choices on the basic of asymptotic approximate Bahadur slopes and asymptotic Pitman efficiencies. This is done in Section 7.

2. PROBABILITY BACKGROUND AND NOTATION

2.1. SOME ORDER RELATIONS

We shall use some notation and results given by Mann and Wald [49] for stochastic limits. Let $\{a_n\}$ be a sequence of $k-$dimensional vectors, $\{V_n\}$ be a sequence of $k-$dimensional random vectors, $\{q_n\}$ a sequence of positive functions of n. The following notation will be used.

$$
\begin{aligned}
a_n &= o(q_n) \quad \text{if} \quad \lim_{n\to\infty} \|a_n/q_n\| = 0, \\
a_n &= O(q_n) \quad \text{if} \quad \|a_n/q_n\| < M, \quad \text{for all } n, \text{ where } M \text{ is some positive constant,} \\
V_n &= o_p(q_n) \quad \text{if} \quad V_n/q_n \quad \text{converges to zero in probability,} \\
V_n &= O_p(q_n) \quad \text{if} \quad V_n/q_n \quad \text{is bounded in probability.}
\end{aligned}
$$

One can show that all the ordinary operation rules regarding O and o are also applicable to O_p and o_p. For examples, if $V_n = O_p(n^{\frac{1}{2}})$ and $W_n = o_p(1)$, or $V_n = O_p(1)$ and $W_n = o_p(n^{\frac{1}{2}})$, then $V_n' W_n = o_p(n^{\frac{1}{2}})$. Next, some basic probability results are listed below:

1. If V is a random variable and independent of n, then $V = O_p(1)$,

2. If V_n converges in distribution, then $V_n + o_p(1)$ converges to the same distribution as V_n,

3. If V_n converges to V in probability $(V_n \xrightarrow{p} V)$ and g is a continuous function, then $g(V_n)$ converges to $g(V)$ in probability. In other words, $V_n - V = o_p(1)$,

4. If V_n converges to V in distribution $(V_n \xrightarrow{d} V)$ and g is a continuous function, then $g(V_n)$ converges to $g(V)$ in distribution.

2.2. PROBABILITY BACKGROUND

In this subsection, we describe some important results pertaining to the distribution of quadratic forms of $k-$dimensional random variables. Detail discussion and proofs of the following results can be found in Chen [20].

Lemma 2.1. *If $\{V_n\}$ is a sequence of $k-$dimensional random variables which converges in distribution to V, then $V_n = O_p(1)$.*

Lemma 2.2. *Let $\{V_n\}$ and V be defined as in Lemma (2.1) and $\{A_n\}$ be a sequence of $k \times k$ random matrices such that $A_n \xrightarrow{p} \boldsymbol{0}$, then $V_n' A_n V_n \xrightarrow{p} 0$.*

Proof: Applying Lemma (2.1) and Landau rules, $V_n' A_n V_n = O_p(1).o_p(1).O_p(1) = o_p(1)$, the lemma follows. $\square$

Lemma 2.3. *Let $\{V_n\}$ be a sequence of $k-$dimensional random vectors such that $V_n \xrightarrow{d} N_k(0, \Sigma)$, where Σ is a positive semi-definite symmetric constant matrix, and Σ^- be a generalized inverse of Σ. Then $V_n' \Sigma^- V_n \xrightarrow{d} \chi^2(s)$, the chi-square distribution with $s =$ rank(Σ) degrees of freedom.*

Proof: Let the random vector V be distributed according to $N_k(0, \Sigma)$, and define the function g from $\mathbb{R}^k$ to $\mathbb{R}$ by $g(V) = V' \Sigma^- V$. Since $g(V)$ is a polynomial and is a continuous function of V, which implies basic probability results that $g(V_n) \xrightarrow{d} g(V)$. A well known result is that $V \sim N_k(0, \Sigma)$, then $V' \Sigma^- V \sim \chi^2(s)$, where $s = \text{rank}(V)$. Hence $g(V_n) = V_n' \Sigma^- V_n \xrightarrow{d} \chi^2(s)$. Chen [20]. $\square$

Lemma 2.4. *Let $\{A_n\}$ be a sequence of $k \times k$ random matrices, and A be some $k \times k$ nonsingular constant matrix. If $A_n \xrightarrow{p} A$, then for any given $\epsilon > 0$, there exists N_ϵ such that for $n > N_\epsilon$, $P(A_n$is nonsingular$) > 1 - \epsilon$.*

Lemma 2.5. *Let $\{A_n\}$ and A be defined as in Lemma 2.4, then $A_n \xrightarrow{p} A$ implies that $A_n^{-1} \xrightarrow{p} A^{-1}$.*

Lemma 2.6. *Let $\{G_n(y)\}$ be a sequence of monotone random functions such that $G_n(y) \xrightarrow{p} G(y)$, where $G(y)$ is continuous over the closed interval $[0, a]$. Then for $\epsilon > 0$ and $\delta > 0$ there exits $N_{(\epsilon,\delta)}$ such that $n > N_{(\epsilon,\delta)}$,*

$$P\left(\sup_{0 \le y \le a} |G_n(y) - G(y)| \le \delta \right) > 1 - \epsilon.$$

Lemma 2.7. *Let* $\{Q_n(y)\}$ *and* $\{G_n(y)\}$ *be sequences of monotone random functions such that* $\{G_n(y)\}$ *are uniformly bounded,* $Q_n(y) \xrightarrow{p} Q(y)$ *and* $G_n(y) \xrightarrow{p} G(y)$*, where* $Q(y)$ *and* $G(y)$ *are bounded and continuous. Then, for arbitrary* $0 \leq a < b < \infty$,

$$\int_a^b Q_n(y)\, dG_n(y) \xrightarrow{p} \int_a^b Q_(y)\, dG_(y).$$

3. BACKGROUND ON RECURRENT EVENT MODELING

3.1. INTRODUCTION

The theory of counting processes via the intensity processes plays a vital role in the modeling and analysis of survival data, in particular recurrent events. In this section, we briefly review relevant stochastic processes that are used in the modeling of recurrent events. We also include some important results pertaining to stochastic process, martingale, and weak convergence. A detail discussion of this framework can be found in Andersen et al. [65], Fleming and Harrington [29], Chung and Williams [24].

Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space and $T = [0, \tau] \subset \mathbb{R}$ be an interval of time.

Definition 3.1. *A filtration* $\mathbf{F} = \{\mathfrak{F}_t : t \in T\}$ *on* $(\Omega, \mathfrak{F}, P)$ *is an increasing family of* $\sigma-$*algebras, that is* $\forall t \leq s$,

$$\mathfrak{F}_t \subseteqq \mathfrak{F}_s \subseteqq \mathfrak{F}.$$

Definition 3.2. $(\Omega, \mathfrak{F}, \mathbf{F}, P)$ *is called a filtered probability space or a stochastic basis.*

Note here that in the case of a stochastic process, $\mathfrak{F}_t$ could be taken to be all information generated by the process up to time t, and is called the natural history of the process. From now on, we denote by $\mathbf{F}$ the natural filtration associated with the probability space $(\Omega, \mathfrak{F}, P)$.

Definition 3.3. *A stochastic process* $\mathbf{X} = \{X(t) : \ t \geq 0\}$ *is called cadlag if its sample paths are right continuous with left hand limits for almost all* ω. *Furthermore, the set of all cadlag functions is called Skorohod space.*

Definition 3.4. *A stochastic process* $\mathbf{X}$ *is adapted to a filtration* $\mathbf{F}$ *if, for every* $t \geq 0$, $X(t)$ *is* $\mathfrak{F}_t-$*measureable.*

Definition 3.5. *A counting process is a stochastic process* $\{N(t) : \ t \geq 0\}$ *adapted to a filtration* $\mathbf{F}$ *with* $N(0) = 0$ *and* $N(t) < \infty$ *almost surely (a.s.), and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with jumps of size* $+1$.

Definition 3.6. *Let* $\mathbf{X}$ *be a right-continuous stochastic process with left-hand limits and* $\mathbf{F}$ *a filtration defined on a common probability space.* $\mathbf{X}$ *is called a martingale with respect to* $\mathbf{F}$ *if*

1. $\mathbf{X}$ *is adapted to* $\mathbf{F}$

2. $E(|X(t)|) < \infty$ *for all* $t < \infty$

3. $E(|X(t+s)|\mathfrak{F}_t|) = X(t-)$ *a.s. for all* $s \geq 0,\ t \geq 0$

$\mathbf{X}$ *is called a submarginale if item* 3 *in Definition 3.6 is replaced by*

$$E(|X(t+s)|\mathfrak{F}_t|) \geq X(t) \ \ a.s.$$

and a supermarginale when item 3 *is replaced by*

$$E(|X(t+s)|\mathfrak{F}_t|) \leq X(t) \ \ a.s.$$

Lemma 3.7. *Let* $\mathbf{F}$ *be a filtration and* $\mathbf{X}$ *a left-continuous real-valued adapted to* $\mathbf{F}$*. Then* $\mathbf{X}$ *is predictable.*

One appealing characterization of a predictable process is based on its sample path property.

3.2. BACKGROUND ON RECURRENT EVENT MODELING

We reconsider the recurrent event data as in (1.1), that's

$$\mathcal{O}_i = (K_i, \tau_i, T_{i,1}, \ldots, T_{i,K_i}, \tau_i - S_{i,K_i}, (X_i(s) : s \leq \tau_i)).$$

We now briefly review the relevant stochastic processes that are used in the modeling of recurrent event data. For more details on these processes, notations, and some derivations, we refer the reader to Peña, Strawderman, and Hollander [60]. We begin by defining the calendar time processes. For a calendar time s, let

$$N_i^\dagger = \{N_i^\dagger(s) :\ s \leq \tau_i\},\ \ Y_i^\dagger = \{Y_i^\dagger(s) :\ s \geq 0\} \quad \text{and} \quad N_i^\tau(s) = I\{\tau_i \leq s\} \tag{3.1}$$

where

$$N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq s \wedge \tau_i\}, \quad \text{and} \quad Y_i^\dagger(s) = I\{\tau_i \geq s\}, \tag{3.2}$$

for $i = 1, \ldots, n,\ \ s \geq 0$. For i^{th} subject, the $N_i^\dagger$ process determine the event occurrences up to time τ_i whereas the $Y_i^\dagger$ process determines if the unit is at-risk for a recurrent event. The at-risk process associated with $N_i^\tau(s)$ is $Y_i^\tau(s) = I\{\tau_i \geq s\}$, which is also equal to $Y_i^\dagger(s)$. The renewal function associated with F is $\rho_F(t) = \sum_{n=1}^\infty F^{*(n)}(t) I\{t \geq 0\}$, where $F^{*(n)}(\cdot)$ is the nth convolution of F, that is the distribution of $S_{i,n} = \sum_{k=1}^n T_{i,k}$.

Let $\mathbf{G} = \{\mathfrak{g}_s \ : \ s \geq 0\}$ be a filtration such that $\{(N_i^\dagger(s), Y_i^\dagger(s)) : s \geq 0\}$ is $\mathbf{G}-$adapted. Let $A_i^\dagger(s) = \int_0^s Y_i^\dagger(v)\lambda(v - S_{iN_i^\dagger(v-)})dv$ be the associated compensator of the process $N_i^\dagger(s)$. The process $M_i^\dagger(s) = N_i^\dagger(s) - A_i^\dagger(s),\ s \geq 0$ is a local square-integrable martingale with respect to the filtration $\mathbf{G}$. In particular, this result holds if one takes $\mathbf{G}$ to be the natural filtration generated by $\mathfrak{g}_s = \mathfrak{F}_0 \bigvee \sigma\{(N_i^\dagger(s), Y_i^\dagger(s)) \ : \ v \leq s,\ i = 1, \ldots, n\}$.

We now introduce appropriate processes that are indexed by both calendar time s, and gap time t. These are the basic processes considered in Peña, Strawderman, and Hollander [60] and also Sellke [68] and provide the crucial connection between the gap time formulation and that based on calendar time. Define the time elapsed since the last event, so called "backward recurrence time" by $R_i(s) = s - S_{iN_i^\dagger(s-)}$, where $s-$ is the time just before s. Following Sellke [68], we define the doubly-indexed process $Z_i(s,t)$ by

$$Z_i(s,t) = I\{R_i(s) \leq t\} \quad \text{for} \quad i = 1, \ldots, n.$$

Note that for fixed t, $Z_i(\cdot, t)$ is a $\mathbf{G}-$adpated, bounded, and has left-continuous paths, hence is a $\mathbf{G}-$predictable bounded process. The doubly indexed processes are:

$$\begin{aligned}
N_i(s,t) &= \int_0^s Z_i(v,t) N_i^\dagger(dv) \\
A_i(s,t) &= \int_0^s Z_i(v,t) A_i^\dagger(dv) \\
M_i(s,t) &= \int_0^s Z_i(v,t) M_i^\dagger(dv) = N_i(s,t) - A_i(s,t).
\end{aligned}$$

For the ith subject, $N_i(s,t)$ determines the number of observed events occurring over the calendar period $[0,s]$ whose inter-event times (i.e., gap times) are at most t. $A_i(s,t)$ is the natural compensator of $N_i(s,t)$ and therefore $M_i(s,t)$ is a square integrable martingale for fixed t with respect to the filtration $\mathbf{G}$. For estimation purpose, a multiplicative form for $A_i(s,t)$ is needed. This is given by

$$A_i(s,t) = \int_0^t Y_i(s,w)\lambda(w)dw \tag{3.3}$$

where $Y_i(s,t)$ is

$$Y_i(s,t) = \sum_{j=1}^{N_i^{\dagger}((s\wedge\tau_i)-)} I\{T_{ij} \geq t\} + I\{(s\wedge\tau_i) - S_{iN_i^{\dagger}((s\wedge\tau_i)-)} \geq t\} \text{ for } i = 1,\ldots,n.$$

The aggregate processes over n units are then

$$N(s,t) = \sum_{i=1}^{n} N_i(s,t), \quad A(s,t) = \sum_{i=1}^{n} A_i(s,t) \text{ and } M(s,t) = \sum_{i=1}^{n} M_i(s,t). \tag{3.4}$$

By (3.3), $A(s,t) = \sum_{i=1}^{n} A_i(s,t) = \int_0^t Y(s,w)\lambda(w)dw$ where $Y(s,t) = \sum_{i=1}^{n} Y_i(s,t)$. The following two results deal with the uniform convergence of $Y(s,t)/n$ and $Y^{\tau}(s)/n = \sum_{i=1}^{n} Y_i^{\tau}(s)/n$ as $n \to \infty$. We have

$$\sup_{v\in[0,s^*]} |n^{-1}Y^{\tau}(v) - \bar{G}(v)| \xrightarrow{p} 0 \text{ and } \sup_{(v,w)\in[0,s^*]\times[0,t^*]} |n^{-1}Y(v,w) - y(v,w;\beta)| \xrightarrow{p} 0, \tag{3.5}$$

where $s^* = \max_{i\leq i\leq n} \tau_i,\ t^* = \max T_{i,j}$ and the function $y(v,w;\beta)$ is given in Adekpedjou et al. [5].

3.3. ESTIMATORS OF F AND Λ

Peña et al. [60] developed, based on the data in (1.1), a NPMLE of the survival function $\bar{F}(s,t)$—which is a Kaplan-Meier type estimator, and given by

$$\hat{\bar{F}}(s^*,t) = \prod_{w\leq t}\left[1 - \frac{N(s^*,dw)}{Y(s^*,w)}\right], \tag{3.6}$$

where $\prod$ denotes product integration. Furthermore, they showed that, over an appropriate Skorohod space

$$\sqrt{n}[\hat{\bar{F}}(s^*,t) - \bar{F}(t)] \xrightarrow{d} W, \tag{3.7}$$

where W is a zero-mean Gaussian process with some variance-covariance matrix $\Sigma_1(s^*, t)$. A NPMLE of the cumulative hazard function of the gap times is given by

$$\hat{\Lambda}(s^*, t) = \int_0^t \frac{J(s^*, w)}{Y(s^*, w)} \, N(s^*, dw), \quad 0 \le t < \infty,$$

where $J(s^*, w) = I\{Y(s^*, t) > 0\}$.

4. A CLASS OF INFERENCES PROCEDURES FOR VALIDATING THE GENERALIZED KOZIOL-GREEN MODEL WITH RECURRENT EVENTS

4.1. INTRODUCTION

As we explained in Section 1, informative monitoring occurs in random censorship model when the distribution function of the end of the monitoring period random variable is informative about the distribution function of the failure times. One major concern for researchers is how to model informative monitoring. There have been several models suggested in the literature for dealing with the property. Link [48] proposed a model where the censoring variable is related to the frailty of the individual. Wang et al. [76] proposed various models where the occurrence of recurrent events is modeled by a subject specific non-stationary Poisson process via a latent variable. Siannis [71] considered a parametric model where the parameter represents the level of dependence between the failure and censoring process. In this section, we employ the generalized KG model (henceforth GKG) for recurrent events discussed in section one.

The KG model has been utilized in studying efficiency aspects under informative censoring in single event settings. Chen et al. [21] obtained exact properties of the Kaplan-Meier estimator under the KG model, and Cheng and Lin [22] derived an estimator of the survivor function utilizing the informative structure. With recurrent event settings, the reference is Adekpedjou, Peña, and Quiton [5] where the GKG model has been used for modeling informative monitoring with recurrent events thereby enabling the derivation of an estimator of the cumulative hazard function and assessing efficiency loss when it is ignored. In both settings, the conclusion that transpired is that ignoring informative censoring/monitoring in the estimation process can lead to loss in statistical efficiency and/or biased estimators. Although the model might seem only of technical relevance, in many applications, such as biomedical studies, it is of substantial importance. See for instance Koziol and Green [44] where the model was used to develop a Cramér-von Mises type statistic to check cancer deaths among oestrogen patients. Henze [36], Herbst [37] and Kirmani and Dauxois [46] proposed procedures for checking the assumption of the KG model in the single event settings. For a review of some of their proposed procedures, see Kay [41].

To the best of our knowledge, no formal procedures have been proposed in the literature for assessing the validity of the assumption of the GKG model with recurrent events. In this section, we develop procedures for checking the validity of the informative monitoring model based on the difference of two consistent estimators of the cumulative hazard rate $\Lambda(t)$. These procedures are based on the asymptotic properties of a scaled difference of these estimators. The asymptotic property of the properly scaled process is used to construct several inferential procedures for validating the aforementioned model.

4.2. ESTIMATORS OF THE CUMULATIVE HAZARD FUNCTION Λ

Estimators of the cumulative hazard function of the gap times that ignore informative monitoring denoted by $\tilde{\Lambda}$ and one that accounts for it, denoted by $\hat{\Lambda}$ have been proposed in Peña et al. [60] and Adekpedjou and Peña [6] respectively. The estimator $\hat{\Lambda}$ accounts for informative monitoring through the GKG model. With $s^\star$ satisfying the condition $s^\star > \max \tau_i$, the estimating equation for β (cf. Adekpedjou and Peña [6]) is given by

$$\int_0^{s^\star} \left\{ \frac{\beta Y^\tau(w)}{Y(s^\star, w) + \beta Y^\tau(w)} \right\} [N^\tau(dw) + N(s^\star, dw)] = N^\tau(s^\star).$$

Upon estimating β, the two estimators of $\Lambda(t)$ are given by

$$\tilde{\Lambda}(s^\star, t) = \int_0^t \frac{J(s^\star, w)}{Y(s^\star, w)} N(s^\star, dw) \tag{4.1}$$

and

$$\hat{\Lambda}(s^\star, t; \hat{\beta}) = \int_0^t J(s^\star, w; \hat{\beta}) \left\{ \frac{N(s^\star, dw) + N^\tau(dw)}{Y(s^\star, w) + \hat{\beta} Y^\tau(w)} \right\}, \tag{4.2}$$

where $J(s^\star, t) = I\{Y(s^\star, t) > 0\}$ and $J(s^\star, t; \hat{\beta}) = I\{Y(s^\star, t) + \hat{\beta} Y^\tau(t) > 0\}$. Asymptotic properties such as consistency and weak convergence of $\hat{\beta}$, $\hat{\Lambda}$, and $\tilde{\Lambda}$, properly standardized can be found in the aforementioned articles.

4.3. ASYMPTOTIC PROPERTIES OF THE SCALED PROCESS

We want to check whether the GKG model is adequate for modeling informative monitoring in a particular population. A graphical approach is based on the fact that $\hat{\Lambda}(s^\star, t; \hat{\beta})$ and $\tilde{\Lambda}(s^\star, t)$ both uniformly converge to the true cumulative hazard rate, $\Lambda(t)$. If the model is valid, then a plot containing both of these estimators should agree reasonably well. Similar graphical checking procedures have been utilized by Nelson [56] and Nelson [57]. Examples of applying this method in the single event settings are described in Aalen [1]. However, a graphical comparison alone will not be enough for the purpose of checking the validity of the GKG model. To construct a proper validating procedure, we will work with a scaled difference of the estimators in (4.1) and (4.2) and derive asymptotic properties of the resulting empirical process, thereby enabling the construction of several goodness of fit procedures. We begin by defining the scaled process $Z_n(s^\star, t; \hat{\beta})$ by

$$Z_n(s^\star, t; \hat{\beta}) = \sqrt{n}(\hat{\Lambda}(s^\star, t; \hat{\beta}) - \tilde{\Lambda}(s^\star, t)). \tag{4.3}$$

We wish to study the asymptotic properties of the empirical process $\{W_n(s, t; \hat{\beta}) : s > 0, t \in [0, t^\star]\}$ given by

$$W_n(s, t; \hat{\beta}) = \sqrt{n} \int_0^t d\{\hat{\Lambda}(s, w; \hat{\beta}) - \tilde{\Lambda}(s, w)\}. \tag{4.4}$$

Let $\mathbf{a}^t$ denote the transpose of a matrix or vector $\mathbf{a}$ and $Dg(\mathbf{a})$ the diagonal matrix formed from the vector $\mathbf{a}$. To make notation compact, let $\mathbb{Y}(s, t) = Y(s, t) + \hat{\beta} Y^\tau(t)$ and

$\mathbb{M}(s,t) = [M(s,t), M^{\tau}(t)]^t$. From (4.1) and (4.2), we have

$$\begin{aligned} W_n(s,t;\hat{\beta}) &= \sqrt{n}\int_0^t d\{\hat{\Lambda}(s,w;\hat{\beta}) - \tilde{\Lambda}(s,w)\} \\ &= \sqrt{n}\int_0^t \left\{\frac{(N(s,dw)+N^{\tau}(dw))}{\mathbb{Y}(s,w)} - \frac{N(s,dw)}{Y(s,w)}\right\} \\ &= \sqrt{n}\int_0^t \left\{\frac{\{M(s,dw)+M^{\tau}(dw)\}}{\mathbb{Y}(s,w)} - \frac{M(s,dw)}{Y(s,w)}\right\} \\ &= \sqrt{n}\int_0^t \left\{\frac{1}{\mathbb{Y}(s,w)} - \frac{1}{Y(s,w)}\right\} M(s,dw) + \sqrt{n}\int_0^t \frac{M^{\tau}(dw)}{\mathbb{Y}(s,w)} \\ &\equiv \sqrt{n}\int_0^t f(s,w;\hat{\beta})M(s,dw) + \sqrt{n}\int_0^t g(s,w;\hat{\beta})M^{\tau}(dw), \end{aligned} \tag{4.5}$$

where $f(s,t;\hat{\beta}) = \frac{1}{\mathbb{Y}(s,t)} - \frac{1}{Y(s,t)}$ and $g(s,t;\hat{\beta}) = \frac{1}{\mathbb{Y}(s,t)}$. To obtain the asymptotic properties of $W_n(s,t;\hat{\beta})$, we need those of the vector of the martingale transform processes $\mathbf{U}_n(s,t;\hat{\beta})$ given by

$$\mathbf{U}_n(s,t;\hat{\beta}) = \frac{1}{\sqrt{n}}\int_0^t \mathbf{H}(s,w;\hat{\beta})\mathbb{M}(s,dw), \tag{4.6}$$

where

$$\mathbf{H}(s,t;\hat{\beta}) = \begin{bmatrix} \frac{n}{\mathbb{Y}(s,t)} & \frac{n}{\mathbb{Y}(s,t)} \\ \frac{-n}{Y(s,t)} & 0 \end{bmatrix}.$$

Processes of form given in (4.6) are a generalization of the process given in equation (A.1) of Peña et al. [61]. With a view toward the asymptotic distribution of $\mathbf{U}_n(s,t;\hat{\beta})$, we require some regularity conditions.

Regularity Conditions:

1. The components $\{H_{ij}(s,t;\hat{\beta}) : 0 \le s \le s^{\star}; 0 \le t \le t^{\star}\}$ of $\mathbf{H}(s,t;\hat{\beta})$ are left-continuous and bounded in (s,t) for $t^{\star} > 0$.

2. There exist deterministic and bounded functions $h_{ij}(s,t;\beta)$ with $\mathbf{h}(s,t;\beta) = \{h(s,t;\beta)\}_{ij}$ such that for all i and j, as $n \to \infty$,

$$\sup_{w\in[0,t^{\star}]} |H_{ij}(s,w;\hat{\beta}) - h_{ij}(s,w;\beta)| \to 0.$$

Theorem 4.1. *Under the above regularity conditions, and as $n \to \infty$, the vector-valued process $\{\boldsymbol{U}_n(s^\star, t; \hat{\beta}) : t \in [0, t^\star]\}$ converges weakly on the product Skorohod's space $D[0, t^\star]^2$ to a 2×1 zero-mean Gaussian process $\{\boldsymbol{U}^\infty(s^\star, t; \beta) : t \in [0, t^\star]\}$ with covariance matrix function given by*

$$\boldsymbol{\Sigma}(s^\star, t; \beta) = \int_0^t \mathbf{h}(s, w; \beta) Dg(y(s^\star, w), \beta \bar{G}(w)) \mathbf{h}(s, w; \beta)^t \lambda(w) dw.$$

Proof: The regularity conditions along with the consistency of $\hat{\beta}_n$ and the uniform convergence given in (3.5) ensure that $\mathbf{H}(s^\star, t; \hat{\beta}_n)$ converges uniformly to $\mathbf{h}(s^\star, t; \beta)$. Let $\mathbf{H}(s, t; \hat{\beta}) = [\mathbf{H}_{\cdot 1}(s, t; \hat{\beta}), \mathbf{H}_{\cdot 2}(s, t; \hat{\beta})]$, where $\mathbf{H}_{\cdot 1}$ and $\mathbf{H}_{\cdot 2}$ represent the first and second column of $\mathbf{H}$ respectively with a similar definition for $\mathbf{h}$. We may then write $\mathbf{U}_n(s^\star, t; \hat{\beta}) = \mathbf{U}_n^1(s^\star, t; \hat{\beta}) + \mathbf{U}_n^2(s^\star, t; \hat{\beta})$ where

$$\mathbf{U}_n^1(s^\star, t; \hat{\beta}) = \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}_{\cdot 1}(s^\star, w; \hat{\beta}) M(s^\star, dw),$$

and

$$\mathbf{U}_n^2(s^\star, t; \hat{\beta}) = \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}_{\cdot 2}(s^\star, w; \hat{\beta}) M^\tau(dw).$$

From Peña et al. [61], $\mathbf{U}_n^1(s^\star, t; \hat{\beta})$ converges on $D[0, t^\star]$ to a zero-mean Gaussian process with covariance matrix function

$$\boldsymbol{\Sigma}_1(s^\star, t; \beta) = \int_0^t \mathbf{h}_{\cdot 1}(s^\star, w; \beta)^{\otimes 2} y(s^\star, w) \lambda(w) dw.$$

By the martingale central limit theorem, $\mathbf{U}_n^2(s^\star, t; \hat{\beta})$ converges on $D[0, t^\star]$ to a zero-mean Gaussian process with covariance matrix function

$$\boldsymbol{\Sigma}_2(s^\star, t; \beta) = \int_0^t \mathbf{h}_{\cdot 2}(s^\star, w; \beta)^{\otimes 2} \beta \bar{G}(w) \lambda(w) dw.$$

Since $\langle M_i^\dagger, M_j^\tau \rangle(dw) = 0$ for all i and j, it follows that the asymptotic covariance of $\mathbf{U}_n(s^\star, t; \hat{\beta})$ is $\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$ which is the expression given in the statement of the theorem. $\square$

The next corollary pertains to the asymptotic distribution of the statistics $W_n(s^\star, t; \hat\beta)$.

Corollary 4.2. *As $n \to \infty$, the process $\{W_n(s,t;\hat\beta) : t \in [0,t^\star]\}$ converges weakly to a zero-mean Gaussian process $\{W^\infty(s,t;\beta) : t \in [0,t^\star]\}$ with covariance function given by*

$$cov(W(s,t_1;\beta), W(s,t_2;\beta)) = \int_0^{t_1 \wedge t_2} \frac{\beta\lambda(w)\bar G(w)}{y(s,w)\boldsymbol{y}(s,w;\beta)} dw,$$

where $\boldsymbol{y}(s,t;\beta) = y(s,t) + \beta\bar G(t)$.

Proof: We can write the two terms in the right hand side of (4.5) in a vector form as

$$U_n(s,t;\hat\beta) = \frac{1}{\sqrt{n}} \int_0^t \begin{pmatrix} f(s,w;\hat\beta) & 0 \\ 0 & g(s,w;\hat\beta) \end{pmatrix} \mathbb{M}(s,dw).$$

Furthermore, let

$$\begin{aligned} \zeta_n(s,t;\hat\beta) &= \sqrt{n}\int_0^t f(s,w;\hat\beta) M(s,dw) \\ \text{and} \quad \xi_n(s,t;\hat\beta) &= \sqrt{n}\int_0^t g(s,w;\hat\beta) M^\tau(dw). \end{aligned}$$

Convergence to a zero-mean Gaussian process of $(\zeta_n(s,t;\hat\beta), \xi_n(s,t;\hat\beta))$ to $(\zeta^\infty(s,t;\beta), \xi^\infty(s,t;\beta))$ on $D[0,t^\star]^2$ is guaranteed by Theorem 4.1. In addition, the finite dimensional distributions of $(\zeta_n(s,t;\hat\beta), \xi_n(s,t;\hat\beta))$ converge to those of $(\zeta^\infty(s,t;\beta), \xi^\infty(s,t;\beta))$, that is for any $0 \le t_1 < t_2 < ... \le t_k \le t^\star$, we have:

$$\left[(\zeta_n(s,t_1;\hat\beta), \xi_n(s,t_1;\hat\beta)), ..., (\zeta_n(s,t_k;\hat\beta), \xi_n(s,t_k;\hat\beta))\right] \xrightarrow{d}$$
$$\left[(\zeta^\infty(s,t_1;\beta), \xi^\infty(s,t_1;\beta)), ..., (\zeta^\infty(s,t_k;\beta), \xi^\infty(s,t_k;\beta))\right].$$

Define the map $\phi : D[0,t^\star]^2 \to D[0,t^\star]$ by $\phi(\gamma_1(t_1), \gamma_2(t_2)) = \gamma_1(t_1) - \gamma_2(t_2)$. The mapping is measurable and continuous under the sup norm. It follows by the continuous mapping theorem that $\phi(\zeta_n(s,t;\hat\beta), \xi_n(s,t;\hat\beta)) \xrightarrow{d} \phi(\zeta^\infty(s,t), \xi^\infty(s,t))$. $\square$

4.4. CLASS OF INFERENCES PROCEDURES

Corollary 4.2 gave the limiting distribution of the goodness of fit process $W_n(s,t;\hat{\beta})$. The model should be rejected if $W_n(s,t;\hat{\beta})$ is significantly different from zero. Many different goodness of fit statistics can be constructed as functionals of $W_n(s,t;\hat{\beta})$. These include the Kolmogorov-Smirnov and chi-square tests.

4.4.1. Kolmogorov-Smirnov Test.

The Kolmogorov-Smirnov test is constructed as functionals of the process $W_n(s,t;\hat{\beta})$. The resulting test statistic is given by

$$Q_1^{(n)}(s,t;\hat{\beta}) = \sup_{w\in[0,t^\star]} |W_n(s,w;\hat{\beta})| \tag{4.7}$$

The test in (4.7) is a function of the gap time and its null distribution is usually mathematically intractable. Before proceeding with the test statistic, we need a consistent estimator of $\sigma^2(s,t;\beta) = \mathrm{Var}(W^\infty(s,t;\beta))$ given by

$$\hat{\sigma}^2(s,t;\hat{\beta}) = n\int_0^t \frac{\hat{\beta}Y^\tau(w)N(s,dw)}{Y(s,w)\{Y(s,w)+\hat{\beta}Y^\tau(w)\}^2}.$$

For $t > 0$, the process $W_n(s,t;\hat{\beta})$ converges weakly to a time transformed Brownian motion process, $\{B(\sigma^2(s,t;\beta)); t \geq 0\}$.

A test statistic can be constructed as

$$Q_3^{(n)}(s,t^\star;\hat{\beta}) = \frac{\sup_{w\in[0,t^\star]} \left|W_n(s,w;\hat{\beta})\right|}{\sqrt{\hat{\sigma}^2(s,t^\star;\hat{\beta})}} \xrightarrow{d} \frac{\sup_{w\in[0,t^\star]} |B(\sigma^2(s,w;\beta))|}{\sqrt{\sigma^2(s,t^\star;\beta)}} \stackrel{d}{=} \sup_{t\in[0,1]} |B(0,t)|. \tag{4.8}$$

Critical points can be obtained using well known facts about the distribution of the functional $\sup_{t\in[0,1]} |B(0,t)|$. The test rejects if $Q_3^{(n)}(s,t^\star;\hat{\beta}) > b_\alpha$, where b_α is the $1-\alpha$ upper quantile of the distribution of $\sup_{t\in[0,1]} |B(0,t)|$. A derivation of the distribution of the supremum of a Brownian motion process can be found in Billingsley [15] and a table is given in Walsh [75]. See also Shorack and Wellner ([69], page. 239).

4.4.2. Chi-Square Test. Let $0 = t_0 < t_1 < ... < t_k = t^\star$ be a subdivision of the gap time interval $[0, t^\star]$ into k cells $I_i = (t_{i-1}, t_i]$, $i = 1, ..., k$. Consider the processes

$$R_{n,i}(s, I_i; \hat{\beta}) = \sqrt{n} \int_{I_i} d\{\hat{\Lambda}(s, w; \hat{\beta}) - \tilde{\Lambda}(s, w))\}.$$

The vector $\mathbf{R}_n$ with elements $R_{n,i}$ converges in distribution to $\mathbf{R}$ which is k-variate with elements R_i. The convergence is to a k-variate normal distribution with covariance matrix equal to $\Sigma = Dg(\sigma_{11}^2(s, t; \beta), ..., \sigma_{kk}^2(s, t; \beta))$ where for $\mathbf{y}(s, t; \beta) = y(s, t) + \beta\bar{G}(t)$,

$$\sigma_{ii}^2(s, t; \beta = \int_{I_i} \frac{\beta\lambda(w)\bar{G}(w)dw}{y(s, w)\mathbf{y}(s, w; \beta)}, \quad \text{for} \quad i = 1, ..., k.$$

The chi-square test statistic is

$$X_n^2 = \mathbf{R}_n^t \hat{\Sigma}_n^- \mathbf{R}_n = (R_{n,1}, ..., R_{n,k})^t \hat{\Sigma}_n^- (R_{n,1}, ..., R_{n,k}). \tag{4.9}$$

Now, $X_n^2 \xrightarrow{d} X^2 = \mathbf{R}^t \Sigma^{-1} \mathbf{R} \sim \chi^2(k)$ by the continuous mapping theorem, where $k = \text{Rank}(\Sigma)$. The test rejects the GKG model if $X_n^2 > \chi_{k,\alpha}^2$, the $1 - \alpha$ upper quantile of a chi-square distribution.

4.5. SIMULATION STUDY AND APPLICATION

4.5.1. Simulation Study: Type I Error Analysis. A simulation study was conducted to investigate the finite null distribution of the proposed test statistics given in equations (4.8) and (4.9). A Weibull intensity of the form $\lambda(t; \theta_1, \theta_2) = \theta_1\theta_2(\theta_1 t)^{\theta_2 - 1}$ was considered. Therefore, under the GKG model, the censoring distribution will also be Weibull with a scale parameter of $[\theta_1\beta^{-\theta_2}]^{-1}$ and a shape parameter of θ_2. We performed 1000 simulations with sample sizes $n \in \{30, 50, 100, 200\}$ and parameter choices of $\theta_1 = 1$, $\theta_2 \in \{.8, 1, 2\}$, and $\beta \in \{.3, .5, .7\}$. For the chi-square test (see equation (4.9)) we divided the gap time into 5 cells based on the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles. Estimated type I error rates were calculated based on the appropriate asymptotic upper 5^{th} percentiles. The results of this simulation study are given in Table 4.1. The estimated type I error rates for both tests are conservative for small samples ($n \in \{30, 50\}$). The Kolomgrov-

Smirnov test statistic is more conservative than the chi-square test. Both tests reach the appropriate type I error rate for larger sample sizes ($n \in \{100, 200\}$).

Table 4.1. Table of simulated type I error rates for the chi-squared and Kolmogorov-Smirnov type tests. The respective estimated error rates are given in the chi-square Test and KS-Test columns.

n	θ_2	β	χ^2-Test	KS-Test	n	θ_2	β	χ^2-Test	KS-Test
30	0.8	0.3	0.032	0.017	200	1	0.5	0.053	0.053
50	0.8	0.3	0.049	0.026	300	1	0.5	0.051	0.052
100	0.8	0.3	0.036	0.034	500	1	0.5	0.055	0.054
200	0.8	0.3	0.053	0.049	30	1	0.7	0.022	0.022
300	0.8	0.3	0.050	0.051	50	1	0.7	0.028	0.033
500	0.8	0.3	0.051	0.050	100	1	0.7	0.037	0.044
30	0.8	0.5	0.017	0.013	200	1	0.7	0.048	0.051
50	0.8	0.5	0.037	0.033	300	1	0.7	0.048	0.049
100	0.8	0.5	0.035	0.044	500	1	0.7	0.050	0.051
200	0.8	0.5	0.048	0.051	30	2	0.3	0.017	0.012
300	0.8	0.5	0.049	0.051	50	2	0.3	0.032	0.027
500	0.8	0.5	0.051	0.050	100	2	0.3	0.036	0.047
30	0.8	0.7	0.013	0.014	200	2	0.3	0.048	0.048
50	0.8	0.7	0.029	0.022	300	2	0.3	0.047	0.047
100	0.8	0.7	0.042	0.035	500	2	0.3	0.051	0.051
200	0.8	0.7	0.049	0.051	30	2	0.5	0.030	0.018
300	0.8	0.7	0.055	0.053	50	2	0.5	0.032	0.028
500	0.8	0.7	0.051	0.052	100	2	0.5	0.039	0.039
30	1	0.3	0.034	0.020	200	2	0.5	0.049	0.048
50	1	0.3	0.030	0.022	300	2	0.5	0.048	0.053
100	1	0.3	0.045	0.033	500	2	0.5	0.047	0.052
200	1	0.3	0.051	0.047	30	2	0.7	0.030	0.028
300	1	0.3	0.047	0.052	50	2	0.7	0.035	0.033
500	1	0.3	0.048	0.052	100	2	0.7	0.049	0.043
30	1	0.5	0.017	0.017	200	2	0.7	0.051	0.053
50	1	0.5	0.034	0.030	300	2	0.7	0.048	0.048
100	1	0.5	0.030	0.038	500	2	0.7	0.050	0.051

4.5.2. Simulation Study: Power Analysis. We investigated the power of our proposed tests via a computer simulation study. The baseline intensity function has the Weibull form as given in Section 4.1 with the scale parameter $\theta_1 = 1$ and shape parameter $\theta_2 \in \{0.4, 0.6, 1.5, 2\}$. The censoring distribution was either unit exponential or a uniform distribution over $[0, \theta]$ where $\theta \in \{2, 4, 8\}$. Our sample sizes were $n \in \{30, 50, 100, 200\}$ and we performed 1000 simulation replications. Table 4.2 summarizes the results when the censoring distribution is unit exponential. Overall as sample size increases so does the power of both tests. The chi-square test is significantly more powerful than the Kolmogorov-Smirnov test especially for values of $\theta_2 \in \{0.4, 0.6\}$. Both tests have very low power when $n = 30$ and $\theta_2 = 0.4$ and this is also the case for the Kolmogorov-Smirnov test when $n = 30$ and $\theta_2 = 0.4$. In these situations the sample size is small and the Weibull intensity is decreasing producing relative few events per unit. Table 4.3 summarizes the results when the censoring distribution is uniform. The results are similar to that of the censoring distribution being unit exponential. As sample size increases so does the power and the chi-square test outperforms the Kolmogorov-Smirnov test. Some cases exist where power is very low for small sample sizes but overall the tests perform relatively well. Simulation results for both the type 1 error analysis and the power strongly suggest that the chi-square test is preferred over the Kolmogorov-Smirnov form.

Table 4.2. Simulated powers for Weibull family with $\tau \sim \exp(1)$

n	θ_2	Power-χ^2	Power-KS
30	0.4	0.156	0.019
50	0.4	0.695	0.118
100	0.4	0.997	0.604
200	0.4	1.000	0.911
30	0.6	0.035	0.005
50	0.6	0.114	0.022
100	0.6	0.550	0.128
200	0.6	0.953	0.404
30	1.5	0.182	0.145
50	1.5	0.267	0.255
100	1.5	0.538	0.411
200	1.5	0.864	0.572
30	2.0	0.382	0.337
50	2.0	0.603	0.548
100	2.0	0.923	0.772
200	2.0	0.998	0.928

4.5.3. Application. To illustrate the different tests, we analyze the bladder cancer data given by Byar [19] that is also utilized by Wei, Lin, and Weissfeld [77]. This data came from a study conducted by the Veterans Administration Cooperative Urological Research Group. Eighty five patients entered the study with bladder tumors that were removed through a surgical intervention. Multiple recurrences of the tumors were then observed for the patients with four being the maximum number of recurrences documented. Surgical intervention was applied each time there was a recurrence of the tumors. Using the GKG model, we obtained an estimate of 0.778 for β. The Kolmogorov-Smirnov test statistic is equal to 1.72. Comparing this value to the upper 5^{th} percentile of the supreme of a standardized Brownian motion results in there not being enough evidence to reject the null hypothesis. For the chi-square test we divided the gap time into 5 cells and obtained 38.72 for the test statistic. This results in a p-value of $2.7 \times 10^{-5} < .05$, implying that the null hypothesis should be rejected. The conservative nature of the Kolmogorov-Smirnov test helps to explain the different conclusions of these tests. We also simultaneously plotted the estimated cumulative hazard functions obtained under

Table 4.3. Simulated powers for Weibull family with $\tau \sim$ U(0,α)

n	χ^2	θ_2	α	n	χ^2	θ_2	α	n	KS	θ_2	α	n	KS	θ_2	α
30	0.28	0.4	2	30	0.07	1.5	4	30	0.03	0.4	2	30	0.03	1.5	4
50	0.87	0.4	2	50	0.11	1.5	4	50	0.20	0.4	2	50	0.04	1.5	4
100	1.00	0.4	2	100	0.16	1.5	4	100	0.71	0.4	2	100	0.06	1.5	4
200	1.00	0.4	2	200	0.32	1.5	4	200	0.98	0.4	2	200	0.08	1.5	4
30	0.05	0.6	2	30	0.18	2.0	4	30	0.01	0.6	2	30	0.07	2.0	4
50	0.34	0.6	2	50	0.32	2.0	4	50	0.05	0.6	2	50	0.16	2.0	4
100	0.91	0.6	2	100	0.61	2.0	4	100	0.46	0.6	2	100	0.27	2.0	4
200	1.00	0.6	2	200	0.90	2.0	4	200	0.93	0.6	2	200	0.36	2.0	4
30	0.05	1.5	2	30	0.28	0.4	8	30	0.03	1.5	2	30	0.14	0.4	8
50	0.04	1.5	2	50	0.85	0.4	8	50	0.03	1.5	2	50	0.44	0.4	8
100	0.06	1.5	2	100	1.00	0.4	8	100	0.02	1.5	2	100	0.92	0.4	8
200	0.16	1.5	2	200	1.00	0.4	8	200	0.04	1.5	2	200	1.00	0.4	8
30	0.09	2.0	2	30	0.04	0.6	8	30	0.10	2.0	2	30	0.06	0.6	8
50	0.18	2.0	2	50	0.23	0.6	8	50	0.14	2.0	2	50	0.22	0.6	8
100	0.31	2.0	2	100	0.86	0.6	8	100	0.15	2.0	2	100	0.66	0.6	8
200	0.60	2.0	2	200	1.00	0.6	8	200	0.18	2.0	2	200	0.95	0.6	8
30	0.30	0.4	4	30	0.11	1.5	8	30	0.07	0.4	4	30	0.01	1.5	8
50	0.87	0.4	4	50	0.20	1.5	8	50	0.33	0.4	4	50	0.05	1.5	8
100	1.00	0.4	4	100	0.26	1.5	8	100	0.84	0.4	4	100	0.10	1.5	8
200	1.00	0.4	4	200	0.47	1.5	8	200	0.99	0.4	4	200	0.15	1.5	8
30	0.05	0.6	4	30	0.23	2.0	8	30	0.04	0.6	4	30	0.01	2.0	8
50	0.32	0.6	4	50	0.43	2.0	8	50	0.13	0.6	4	50	0.11	2.0	8
100	0.91	0.6	4	100	0.69	2.0	8	100	0.61	0.6	4	100	0.25	2.0	8
200	1.00	0.6	4	200	0.93	2.0	8	200	0.97	0.6	4	200	0.42	2.0	8

both the model considered in Peña et al. [60] and the GKG model. The plot given in Figure 4.1 supports the conclusion of the chi-square test that the GKG model does not hold.

4.6. CONCLUSION

In this section we have developed a class of inference procedures for validating the GKG model based on scaled difference of two competing estimators of the cumulative hazard possessing nice asymptotic properties. A class of validating procedures was developed based on the asymptotic properties of the scaled process thereby enabling construction of goodness-of-fit type tests. A weight process $K_n(s,t;\hat{\beta})$ converging almost surely to some

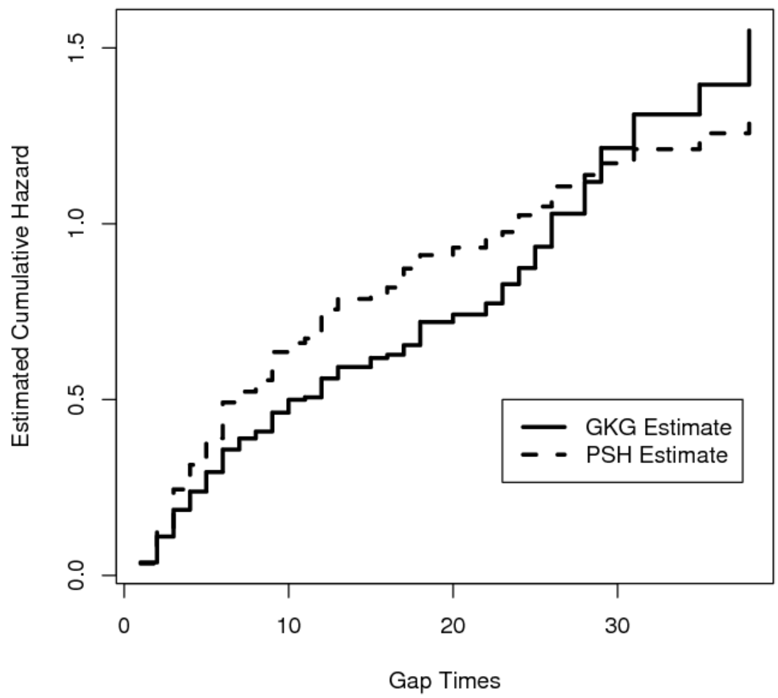

Figure 4.1. A plot of the estimated cumulative hazard functions for the bladder cancer data of Byar [19] utilizing the generalized Koziol-Green model and that considered by Peña et al. [60].

deterministic function, could be included in the process in (4.4). Such inclusion could lead to a more general type of process of the form

$$W_n(s,t;\hat{\beta}) = \sqrt{n}\int_0^t K_n(s,t;\hat{\beta})d\{\hat{\Lambda}(s,t;\hat{\beta}) - \hat{\Lambda}(s,w)\}.$$

However, because of technical difficulty with regard to the predictability of such a process the weight has been taken to be one in the current work. We think the weight can be approximated by a predictable process and will be investigated in future research. The two estimators used here to develop the class of tests can be extended to include covariates. Similar tests like those derived here that include covariates may then be obtained using the same technique. Another extension to the current work is to relax the assumption of the independence and allow the inter-event times to be dependent. Frailties can be used to account for correlation among the inter-event times.

5. A RANDOM CELLS-BASED CHI-SQUARE TEST WITH RECURRENT EVENT DATA

5.1. INTRODUCTION

With complete data, that is with no censoring, for testing "goodness of fit" the classical test is the chi-square goodness of fit test. Statistics of the chi-square types are defined in terms of cells which are fixed prior to taking observations. The distribution theory was originally developed for the multinomial distribution. Let $N_1^{(n)}, N_2^{(n)}, \ldots, N_k^{(n)}$ be observed frequencies in a multinomial distribution with number of trials n and outcomes probabilities $p_1(\boldsymbol{\theta}), p_2(\boldsymbol{\theta}), \ldots, p_k(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \mathbb{R}^q$, and $\hat{\boldsymbol{\theta}}_n$ be the maximum likelihood estimate (m.l.e.) of the parameter $\boldsymbol{\theta}$, then the statistic

$$X_n^2(\hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^{k} \frac{(N_i^{(n)} - np_i(\hat{\boldsymbol{\theta}}_n))^2}{np_i(\hat{\boldsymbol{\theta}}_n)} \tag{5.1}$$

is distributed in the limit as chi-square with $k - q - 1$ degrees of freedom.

We shall be testing the hypothesis that the distribution function (d.f.) of the sample $X_1, X_2, \ldots, X_n$ belongs to a prescribed family $\mathcal{F}_{\boldsymbol{\theta}} = \{F(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}$, where Θ is an open subset of $\mathbb{R}^q$. Dividing the range of values of the X_is into k cells, this problem can be reduced to the classical chi-square goodness of fit test. But the resulting test statistic in (5.1) can be applied directly only in the case when the cells are chosen by using the actual observations and $\hat{\boldsymbol{\theta}}_n$, the m.l.e. of $\boldsymbol{\theta}$, is determined using the class frequencies $(N_1^{(n)}, N_2^{(n)}, \ldots, N_k^{(n)})$. This m.l.e. is called a multinomial m.l.e (m.m.l.e.). Furthermore, the statistic in (5.1) is distributed in the limit as chi-square distribution with $k - 1$ degrees of freedom only if the parameter $\boldsymbol{\theta}$ is considered to be known and this remarkable result is due to Pearson [59].

Fisher [27] and Fisher [28] showed that if the parameter $\boldsymbol{\theta}$ is estimated by the value $\bar{\boldsymbol{\theta}}_n$ minimizing the statistic $X_n^2(\boldsymbol{\theta})$, then $X_n^2(\bar{\boldsymbol{\theta}}_n)$ is distributed in the limit as chi-square with $k - q - 1$ degrees of freedom. The estimator $\bar{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ obtained in this fashion is called the minimum chi-square estimator. It is important to note that Fisher's result is valid only if $\boldsymbol{\theta}$ is estimated by the minimum chi-square estimator or any estimator asymptotically equivalent to it.

If one uses the m.l.e $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ based on the original sample values $X_1, X_2, \ldots, X_n$, not on the frequencies $(N_1^{(n)}, N_2^{(n)}, \ldots, N_k^{(n)})$, then the statistic in (5.1) is no longer distributed in the limit as chi-square but is distributed in the limit as

$$\chi^2_{k-q-1} + \sum_{j=1}^{q} \lambda_j \, y_j^2$$

where the $y_1, y_2, \ldots, y_q$ are normal $(0, 1)$ random variables, mutually independent and independent of χ^2_{k-q-1}. The $\lambda_1, \lambda_2, \ldots, \lambda_q$ are scalers between 0 and 1 (cf. Chernoff and Lehman [23]).

Instead of fixing the cells in advance, one can allow the cell boundaries to be data dependent, that are cut free. Further, we require that the cells settle down as the sample size increases. Cells obtained following this fashion are called random cells. By doing so, the limiting distribution of the chi-square statistic will not depend on the unknown parameter $\boldsymbol{\theta}$. This approach of constructing chi-square statistic is more flexible, guarantees that the cell probabilities will not be small and is increasingly practicable. The cell frequencies are no longer multinomial and the limiting distribution of the vector of standardized frequencies is now obtained using empirical process techniques.

The problem of goodness of fit for censored data is to test the null hypothesis that F is a member of the family $\mathcal{F}_{\boldsymbol{\theta}} = \{F(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}$ of distribution functions indexed by a parameter $\boldsymbol{\theta}$ running over a space Θ. Since censored observations contain only partial information about the underlying distribution of the data, one cannot use the usual empirical distribution function $\hat{F}(x) = \sum I(X_i \leq x)/n$ to calculate the cell frequencies $N_i^{(n)}$ in (5.1). Suppose now that we have an estimator $\hat{F}$ of F with the property that, whether or not the parametric model holds, we have

$$\sqrt{n}\,(\hat{F} - F) \xrightarrow{d} W$$

for some Gaussian process W, on an appropriate Skorohod space. With random censorship model, the most commonly used nonparametric estimator for the survival function $\bar{F} = 1 - F$ is the KM-estimator (cf. Kaplan and Meier [40]).

In the single events, the seminal papers dealing with the problem of chi-square goodness of fit with fixed or data-dependent cells include those of Hjort [38], Kim [43], Li and Doss [47], Habib and Thomas [33], Akritas [9], Hollander and Peña [39], Moore and Spruill [52], Pollard [63], Mihalka and Moore [51], Ruymgaart [67], and Dahiya and Gurland [25] to name a few. Akritas proposed a chi-square test based on KM type estimator of censoring time, while Hollander and Peña proposed a Pearson-type of test based on KM estimator of failure times to test a simple null distribution. Asymptotic properties of their proposed statistics were obtained using asymptotic properties of the KM estimator given in Breslow and Crowely [17]. Li and Doss developed chi-square test based on random cells for right-censored and left truncated data in the single event settings. In fact, their test is applicable every time there exits an estimator $\hat{F}(t)$ of $F(t)$ satisfying the asymptotic property (5.1), where W is a zero-mean continuous Gaussian process whose variance-covariance matrix is non-singular. Other chi-square tests based on random cells have also been developed by Moore [53], Pollard [63], among others. A discussion on the use of random cells is given in Ruymgaart [67].

The situation where the event is recurrent has been dealt with, albeit not as thoroughly yet as in the single event. In recurrent event settings, the goodness of fit problem has been considered by Presnell, Hollander, and Sethuraman [64], Agustin and Peña [7], and Agustin and Peña [8], Stocker and Adekpedjou [73], Adekpedjou and Zamba [4]. Presnell et al. [7] proposed tests for the minimum repair assumption in the imperfect repair model. Agustin and Peña [8] proposed goodness of fit test for the Block, Borges, and Savits [16] model whereas Agustin and Peña [8] developed goodness of fit test for an extended Block et al. [16] model that include covariates. Stocker and Adekpedjou [73] developed a class of tests for the hazard rate function that include chi-square, Kolmogorov-Smirnov, Cramér-von Mises and obtained asymptotic properties of their tests using empirical process techniques and Khmaladze transformation (cf. Khmaladze [42]). Adekpedjou and Zamba [4] developed a chi-square goodness of fit for testing the hypothesis of completely known distribution with fixed cells based on a nonparametric maximum likelihood estimator of F.

Assume that the inter-event times in a recurrent event setting are i.i.d. with a common absolutely continuous distribution function $F(t;\boldsymbol{\theta})$. The parameter $\boldsymbol{\theta}$ ranges over an open set Θ in $\mathbb{R}^q$. The major goals of this section is to develop a chi-square goodness of fit for testing the null hypothesis that F belongs to some parametric family of distributions. The null hypothesis of interest is

$$H_0 : F(\cdot) \in \mathcal{F}_{\boldsymbol{\theta}} = \{F(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}. \tag{5.2}$$

All the tests considered in recurrent event settings are based on fixed cell boundaries. In the one we propose, the cells are random and data-driven and generalized the work of Li and Doss [47] to the situation where the event is recurrent. Furthermore, it encompasses a wide range of tests including-fixed null with random cells, fixed cells with composite hypothesis- and is different from those proposed in the literature of recurrent events. We use the NPMLE of the distribution function of the inter-event time in (3.6) to obtain the observed frequencies and we assume the asymptotic property (3.7) is in force. The expected frequencies are obtained using the estimator of $\boldsymbol{\theta}$ that minimizes a quadratic form obtained from the suitably standardized vector of " observed - expected " frequencies. The importance of using the minimum chi-square estimator in the construction of chi-square statistic is discussed in Harris and Kanji [35].

Let $T = [0, t^*]$, where $t^* = \max_{i,j} T_{i,j}$ is the largest gap-time. The set T could also be taken to be $[0, s^*]$. Contrary to chi-square test based on fixed cells, we consider a subdivision of $[0, t^*]$ given by $0 = t_0^n < t_1^n < \cdots < t_k^n = t^\star$ where the end-points are functional of the data, namely $t_j^n = t_j^n(\mathcal{O}_i; i = 1, ..., n)$. The random cells are given by $I_i^n = [t_{l-1}^n, t_l^n)$ for $l = 1, \ldots, k$, and we require them to settle down as sample size increases. That is $t_l^n \xrightarrow{p} t_l$ and $I_l^n \xrightarrow{p} I_l = [t_{l-1}, t_l)$ as $n \to \infty$, under $F(\cdot, \boldsymbol{\theta}_0)$, where $t_l \in [0, t^*]$ and $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Here the notation $\xrightarrow{p}$ means convergence in probability.

Set $\mathbf{t}^n = (t_1^n, \ldots, t_{k-1}^n)$ and $\mathbf{t} = (t_1, ..., t_{k-1})$. The number of $T_{i,j}$ falling in the lth random cell I_l^n $(l = 1, ..., k)$ by calendar time s, that is the observed cells frequencies using the

NPMLE is defined by

$$\hat{p}_l^n(s) = \int_{I_l^n} \hat{F}(s, dw) = \hat{F}(s, t_l^n) - \hat{F}(s, t_{l-1}^n). \tag{5.3}$$

The expected random cell frequencies, that is the expected number of $T_{i,j}$ falling in I_l^n is given by

$$p_l^n(\boldsymbol{\theta}) = \int_{I_l^n} F(dw, \boldsymbol{\theta}) = F(t_l^n, \boldsymbol{\theta}) - F(t_{l-1}^n, \boldsymbol{\theta}), \tag{5.4}$$

and these are expected, as $n \to \infty$, to stabilize to

$$p_l(\boldsymbol{\theta}) = \int_{I_l} F(dw, \boldsymbol{\theta}) = F(t_l, \boldsymbol{\theta}) - F(t_{l-1}, \boldsymbol{\theta}). \tag{5.5}$$

In the sequel, we introduce the corresponding vector of observed cells frequencies, expected random cells frequencies, and limiting values of expected random cells frequencies by

$$\hat{\mathbf{p}}^n(s) = [\hat{p}_l^n(s)]_{k\times 1}, \quad \mathbf{p}^n(\boldsymbol{\theta}) = [p_l^n(\boldsymbol{\theta})]_{k\times 1}, \quad \text{and} \quad \mathbf{p}(\boldsymbol{\theta}) = [p_l(\boldsymbol{\theta})]_{k\times 1}, \tag{5.6}$$

respectively. Let the lth element of a $k \times 1$-vector $\mathbf{U}_n(s, t; \boldsymbol{\theta})$ of "observed-expected" frequencies over the random cells I_l^n be defined by

$$U_n^l(s, t; \boldsymbol{\theta}) = \sqrt{n}[\hat{p}_l^n(s) - p_l^n(\boldsymbol{\theta})], \quad l = 1, ..., k. \tag{5.7}$$

In general, a chi-square statistic has the form $\mathbf{U}_n'(s, t; \boldsymbol{\theta}_n)\hat{\boldsymbol{\Sigma}}\mathbf{U}_n(s, t; \boldsymbol{\theta}_n)$, where $\mathbf{a}'$ denote the transpose of a vector $\mathbf{a}$, $\boldsymbol{\theta}_n$ is an estimator of $\boldsymbol{\theta}$ having some nice asymptotic properties, and $\hat{\boldsymbol{\Sigma}}$ is a $k \times k$ matrix that could possibly depends on $\boldsymbol{\theta}_n$. The matrix $\hat{\boldsymbol{\Sigma}}$ is-most of the time- an estimate of the Moore-Penrose generalized inverse of a consistent estimator of the in-probability limit of the variance-covariance matrix of the limiting distribution of $\mathbf{U}_n(s, t; \boldsymbol{\theta})$. If one lets $\hat{\boldsymbol{\Sigma}} = \mathbf{I}_{k\times k}$, this reduces to the classical chi-square test proposed by Fisher [28]. At any rate, the true limiting matrix $\boldsymbol{\Sigma}$ is in general assumed to satisfy some regularity conditions such as positive definite and non-singularity.

We now impose some assumptions that are crucial for the proof of our asymptotic results. These are the classical conditions imposed on the hypothesized distribution and the expected frequencies under H_0 for chi-square type tests -however slightly changed to accommodate recurrent events.

Assumption I. There exists a neighborhood $\mathcal{N}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ on which $F(t, \boldsymbol{\theta})$ is continuous and differentiable on $[0, t^*] \times \mathcal{N}(\boldsymbol{\theta}_0)$ and the derivative of all orders are continuous in t and $\boldsymbol{\theta}$.

Assumption II. The $k \times q$ matrix

$$\nabla_{\boldsymbol{\theta}'}\mathbf{p}(\boldsymbol{\theta}) = \begin{bmatrix} \nabla_{\boldsymbol{\theta}_1} p_1(\boldsymbol{\theta}) & \cdots & \nabla_{\boldsymbol{\theta}_q} p_1(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \nabla_{\boldsymbol{\theta}_1} p_k(\boldsymbol{\theta}) & \cdots & \nabla_{\boldsymbol{\theta}_q} p_k(\boldsymbol{\theta}) \end{bmatrix}_{k \times q}$$

where $\nabla_{\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} \equiv (\partial/\partial\theta_j, j = 1, 2, \ldots, q)^t$ and $\nabla_{\theta_j} p_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} p_i(\boldsymbol{\theta})$ is of rank q for all $\boldsymbol{\theta} \in \Theta$.

5.2. PRELIMINARY RESULTS

Our first result in this subsection pertains to the asymptotic distribution of $\mathbf{U}_n(s, t; \boldsymbol{\theta}_0)$ under H_0.

Theorem 5.1. *Under H_0, $\boldsymbol{U}_n(s, t; \boldsymbol{\theta}_0)$ converges in distribution to $N_k(0, \Sigma)$ where $\Sigma = J\Sigma_1(s, t; \boldsymbol{\theta}_0)J'$ and the matrix J is given by*

$$J = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}_{k \times (k-1)}$$

Furthermore, rank(J) = k - 1.

Proof: Define the product-limit type process by $W_n(s, t; \boldsymbol{\theta}_0) = \sqrt{n}[\hat{\bar{F}}(s, t) - \bar{F}(t, \boldsymbol{\theta}_0)]$, where as before $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. With $\boldsymbol{\xi}_n = [W_n(s, t_j^n; \boldsymbol{\theta}_0)]_{(k-1)\times 1}$, it is easily shown that $\mathbf{U}_n(s,t;\boldsymbol{\theta}_0) = -J\boldsymbol{\xi}_n$. Define $\boldsymbol{\xi}_n^{(1)} = [W_n(s,t_j;\boldsymbol{\theta}_0)]_{(k-1)\times 1}$ and $\boldsymbol{\xi}_n^{(2)} = [W_n(s,t_j^n;\boldsymbol{\theta}_0) - W_n(s,t_j;\boldsymbol{\theta}_0)]_{(k-1)\times 1}$. Then $\boldsymbol{\xi}_n = \boldsymbol{\xi}_n^{(1)} + \boldsymbol{\xi}_n^{(2)}$. Observe that $W_n(s,t;\boldsymbol{\theta}_0)$ is a type of process given in (3.7) and its weak convergence to say $W(s,t;\boldsymbol{\theta}_0)$ is obtained by applying (3.7) to parametric distribution. Furthermore, Peña et al. [61] proved convergence of finite dimensional distributions of $W(s,t;\boldsymbol{\theta}_0)$ to Gaussian distributions for any $t_1 < t_2 < \cdots < t_{k-1}$. The proof is complete if we can show that $\boldsymbol{\xi}_n^{(2)}$ converges in probability to a $(k-1)$- dimensional vector $\mathbf{0}_{(k-1)\times 1}$. By the representation theorem of Pollard (cf. Pollard [63]), there exists a new probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, new processes $\tilde{W}_n(s,t;\boldsymbol{\theta}_0)$ and $\tilde{W}(s,t;\boldsymbol{\theta}_0)$ such that $W_n(s,t;\boldsymbol{\theta}_0) \stackrel{d}{=} \tilde{W}_n(s,t;\boldsymbol{\theta}_0)$ [where $\stackrel{d}{=}$ means "equal in distribution"] and $W(s,t;\boldsymbol{\theta}_0) \stackrel{d}{=} \tilde{W}(s,t;\boldsymbol{\theta}_0)$. Moreover, $\tilde{W}_n(s,t;\boldsymbol{\theta}_0)$ converges weakly to $\tilde{W}(s,t;\boldsymbol{\theta}_0)$ and the new processes have the same finite distributions as the old ones on their respective probability spaces. Since $\tilde{W}(s,t;\boldsymbol{\theta}_0)$ has continuous sample paths, the distribution equalities imply

$$\sup_{0\le t\le t^\star} |\tilde{W}_n(s,t;\boldsymbol{\theta}_0) - \tilde{W}(s,t;\boldsymbol{\theta}_0)| \stackrel{as}{\to} 0$$

as $n \to \infty$. Next, for $l = 1, ..., k$, introduce $\tilde{t}_l^n = t_l^n(\tilde{\mathcal{O}}_i : i = 1, ..., n)$, the counterparts of t_l^n on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that $\tilde{t}_l^n \stackrel{d}{=} t_l^n$. In addition, $\tilde{t}_l^n \stackrel{p}{\to} t_l$ since $t_l^n \stackrel{p}{\to} t_l$. It then follows, for large n that (we drop the argument $\boldsymbol{\theta}_0$ for brevity)

$$\begin{aligned}
|\tilde{W}_n(s,\tilde{t}_l^n) - \tilde{W}_n(s,t_l)| &\le |\tilde{W}_n(s,\tilde{t}_l^n) - \tilde{W}(s,\tilde{t}_l^n)| + |\tilde{W}(s,\tilde{t}_l^n) - \tilde{W}(s,t_l)| \\
&\quad + |\tilde{W}(s,t_l) - \tilde{W}_n(s,t_l)| \\
&\le 2\sup_{0\le t\le t^\star} |\tilde{W}_n(s,t) - \tilde{W}(s,t)| + |\tilde{W}(s,\tilde{t}_l^n) - \tilde{W}(s,t_l)| \qquad (5.8) \\
&= o_p(1).
\end{aligned}$$

as $n \to \infty$, where the last inequality is obtained by using the continuous sample path property of $\tilde{W}(s,t;\boldsymbol{\theta}_0)$. Because $\tilde{W}(s,t;\boldsymbol{\theta}_0)$ has continuous sample paths, an application of the functional continuous mapping theorem implies that the right hand side of (5.8) is negligible. Therefore, $\boldsymbol{\xi}_n^{(2)} \stackrel{p}{\to} \mathbf{0}$ and the result in the first part of the theorem follows by applying Slustky theorem. $\square$

Let $\Xi(s,t;\boldsymbol{\theta})$ be the square root of the Moore-Penrose generalized inverse of $\Sigma(s,t;\boldsymbol{\theta})$. Then $\Xi(s,t;\boldsymbol{\theta})$ is a $k\times k$ symmetric matrix whose elements are function of $(\mathbf{t},\boldsymbol{\theta})$ for fixed s. The variance-covariance matrix Σ_1 of W in (3.7) is a non-singular matrix because the process W is of the form $\bar{F}V$, where V is some Gaussian martingale (cf. Fleming and Harrington [29])-and this is true even in the finite dimensional distributions case. We now impose some conditions on $\Xi(s,t;\boldsymbol{\theta})$.

Condition 1: $\Xi(s,t;\boldsymbol{\theta})$ is continuous at $(t,\boldsymbol{\theta})$ for a fixed s.

Condition 2: $\Xi^{-1}(s,t;\boldsymbol{\theta})$ exists and bounded on $([0,s^*]\times[0,s^*]\times\Theta)$.

Condition 3: $\nabla_\theta(\Xi^2(s,t;\boldsymbol{\theta}))$ exists at every $(t,\boldsymbol{\theta})\in(N(t)\times\Theta)$ and $\nabla_{\boldsymbol{\theta}}(\Xi^2(s,t;\boldsymbol{\theta}))$ is continuous at $(t,\boldsymbol{\theta}_0)$.

For brevity, let $\Xi(s,\mathbf{t}^n;\boldsymbol{\theta})\equiv\Xi_n(s,\boldsymbol{\theta})$ be the matrix obtained with $\mathbf{t}$ replaced by the random boundaries vector $\mathbf{t}^n$. Define

$$\mathbf{V}_n(s,t;\boldsymbol{\theta})=\Xi_n(s,\boldsymbol{\theta})\mathbf{U}_n(s,t;\boldsymbol{\theta}). \tag{5.9}$$

The limiting distribution of $\mathbf{V}_n(s,t;\boldsymbol{\theta}_0)$ under H_0 is straightforward from Theorem 5.1. Let $\{\bar{\boldsymbol{\theta}}_n(s^*,t^*):n=1,2,...\}$ be the sequence of $\boldsymbol{\theta}$-values minimizing the sequence of quadratic forms $\{\mathbf{V}'_n(s^*,t^*;\boldsymbol{\theta})\mathbf{V}_n(s^*,t^*;\boldsymbol{\theta}):n=1,2,...\}$. To proceed, we need an assumption pertaining to the sequence $\{\bar{\boldsymbol{\theta}}_n(s^*,t^*):n=1,2,...\}$. Observe that $\bar{\boldsymbol{\theta}}_n(s^*,t^*)$ is a process that varies as calendar time progresses. Note that in the above $\bar{\boldsymbol{\theta}}_n(s^*,t^*)$ is the modified minimum chi-square estimator of $\boldsymbol{\theta}$. When $t^*\to\infty$, the modified minimum chi-square reduces to the minimum chi-square estimator $\bar{\boldsymbol{\theta}}_n(s^*)$. In the most cases, closed form solutions for zeros of $\mathbf{V}'_n(s^*,t^*;\boldsymbol{\theta})\mathbf{V}_n(s^*,t^*;\boldsymbol{\theta})$ do not exist, numerical methods such as the Newton-Raphson algorithm or the Nelder-Mead (cf. Nelder and Mead [55]) will be needed to minimize the quadratic form. In what follows, for a $q\times 1$ vector $\mathbf{a}$, let $\|\mathbf{a}\|_q^2=\sum_{t=1}^q a_t^2$.

Assumption III $\|\bar{\boldsymbol{\theta}}_n(s^*,t^*)-\boldsymbol{\theta}_0\|_q^2=o_p(1)$ as $n\to\infty$.

Remark 5.1. Assumption III is satisfied by any minimum chi-square estimator or an estimator asymptotically equivalent to it. A minimum chi-square estimator is obtained by using the minimum Hellinger distance estimator, the Kullback-Lieber separator, or the Haldane discrepancy. All these methods of estimation provide estimators which are consistent and asymptotically efficient and the resulting estimators are classified as minimum chi-square estimators (cf. Harris and Kanji [35]).

From now on, we abbreviate $\bar{\boldsymbol{\theta}}_n(s^*, t^*)$ by $\bar{\boldsymbol{\theta}}_n$. The next two lemmas pertain to a Taylor-type expansion around $\bar{\boldsymbol{\theta}}_n$ for $\mathbf{p}^n(\bar{\boldsymbol{\theta}}_n)$ and $\nabla_{\boldsymbol{\theta}'}\mathbf{p}^n(\bar{\boldsymbol{\theta}}_n)$. These two lemmas prove to be crucial in most of our asymptotic proofs.

Lemma 5.2. *Under the assumptions and conditions enumerated above, we have:*
[i] $\boldsymbol{p}^n(\bar{\boldsymbol{\theta}}_n) = \boldsymbol{p}(\boldsymbol{\theta}_0) + o_p(1)$ *and [ii]* $\nabla_{\boldsymbol{\theta}'}\boldsymbol{p}^n(\bar{\boldsymbol{\theta}}_n) = \nabla_{\boldsymbol{\theta}'}\boldsymbol{p}(\boldsymbol{\theta}_0) + o_p(1)$.

Proof: [i] Observe that for each $l = 1, ..., k$, adding and subtracting $p_l^n(\boldsymbol{\theta}_0)$ in $p_l^n(\bar{\boldsymbol{\theta}}_n) - p_l(\boldsymbol{\theta}_0)$, we get $p_l^n(\bar{\boldsymbol{\theta}}_n) - p_l(\boldsymbol{\theta}_0) = p_l^n(\bar{\boldsymbol{\theta}}_n) - p_l^n(\boldsymbol{\theta}_0) + p_l^n(\boldsymbol{\theta}_0) - p_l(\boldsymbol{\theta}_0)$. Using (5.4) and (5.5), we obtain

$$\begin{aligned} p_l^n(\bar{\boldsymbol{\theta}}_n) - p_l(\boldsymbol{\theta}_0) &= [\bar{F}(t_{l-1}^n, \bar{\boldsymbol{\theta}}_n) - \bar{F}(t_l^n, \bar{\boldsymbol{\theta}}_n)] - [\bar{F}(t_{l-1}^n, \boldsymbol{\theta}_0) - \bar{F}(t_l^n, \boldsymbol{\theta}_0)] \\ &\quad + [\bar{F}(t_{l-1}^n, \boldsymbol{\theta}_0) - \bar{F}(t_l^n, \boldsymbol{\theta}_0)] - [\bar{F}(t_{l-1}, \boldsymbol{\theta}_0) - \bar{F}(t_l, \boldsymbol{\theta}_0)] \\ &= [\bar{F}(t_{l-1}^n, \bar{\boldsymbol{\theta}}_n) - \bar{F}(t_{l-1}^n, \bar{\boldsymbol{\theta}}_n)] - [\bar{F}(t_l^n, \boldsymbol{\theta}_0) - \bar{F}(t_l^n, \boldsymbol{\theta}_0)] \\ &\quad + [\bar{F}(t_{l-1}^n, \bar{\boldsymbol{\theta}}_0) - \bar{F}(t_{l-1}, \bar{\boldsymbol{\theta}}_0)] - [\bar{F}(t_l^n, \boldsymbol{\theta}_0) - \bar{F}(t_l, \bar{\boldsymbol{\theta}}_n)]. \quad (5.10) \end{aligned}$$

An application of the mean value theorem to each bracket in the right hand side of (5.10) yields

$$\begin{aligned} p_l^n(\bar{\boldsymbol{\theta}}_n) - p_l(\boldsymbol{\theta}_0) &= \left[\boldsymbol{\nabla}_{\boldsymbol{\theta}'}\bar{F}(t_{l-1}^n, \bar{\boldsymbol{\theta}}_n^*) - \nabla_{\boldsymbol{\theta}'}\bar{F}(t_l^n, \bar{\boldsymbol{\theta}}_n^{**})\right](\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &\quad + \nabla_t\bar{F}(t^*, \boldsymbol{\theta}_0)(t_{l-1}^n - t_{l-1}) - \nabla_t\bar{F}(t^{\star\star}, \boldsymbol{\theta}_0)(t_l^n - t_l), \end{aligned}$$

where $\bar{\boldsymbol{\theta}}_n^{**}$ and $\bar{\boldsymbol{\theta}}_n^*$ both lie in the line segment between $\bar{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, $t^\star \in (t_{l-1}^n, t_{l-1})$ and $t^{\star\star} \in (t_{l-1}^n, t_l)$. Additionally, $\boldsymbol{\theta}_n^* \xrightarrow{p} \boldsymbol{\theta}_0$, $\boldsymbol{\theta}_n^{\star\star} \xrightarrow{p} \boldsymbol{\theta}_0$, $t^\star \xrightarrow{p} t_{l-1}$ and $t^{\star\star} \xrightarrow{p} t_l$. An application of Assumption I and consistency of $\bar{\boldsymbol{\theta}}_n$ completes the proof of [i]. To prove part [ii], note that by Assumption I again, $\nabla_{\boldsymbol{\theta}}\mathbf{p}(t,\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ and t. Because $\bar{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$, the result follows by an application of the continuous mapping theorem. □

Remark 5.2. The above lemma is similar to assumption A4 of Pollard [63]. In the single event setting, Li and Doss [47] observed similar result for the vector $p^n(\bar{\boldsymbol{\theta}}_n)$ and its derivative, but did not provide a proof of them.

We now provide an intermediate result that will be used in the proof of the asymptotic distribution of the chi-square statistic. Introduce the $k \times q$ matrix $\mathbf{B}(s,t;\boldsymbol{\theta}_0)$ with (i,j)th entry equals to

$$\Xi(s,t;\boldsymbol{\theta}_0)_{i,j}\frac{\partial p_i(t,\boldsymbol{\theta}_0)}{\partial \theta_j} \tag{5.11}$$

for $i = 1,...,k$ and $j = 1,...,q$, so that $\mathbf{B}(s,t;\boldsymbol{\theta}_0) = \Xi(s,t;\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}_0'}\mathbf{p}(t,\boldsymbol{\theta}_0)$. The next proposition deals with the negligibility of $\mathbf{B}'(s,t;\boldsymbol{\theta}_0)\mathbf{V}_n(s,t;\bar{\boldsymbol{\theta}}_n)$.

Proposition 5.3. *Under the regularity conditions and assumptions, we have:*

$$\boldsymbol{B}'(s,t;\boldsymbol{\theta}_0)V_n(s,t;\bar{\boldsymbol{\theta}}_n) = o_p(1).$$

Proof: Since $\bar{\boldsymbol{\theta}}_n$ is the value of $\boldsymbol{\theta}$ minimizing the quadratic form $\mathbf{V}_n'(s,t;\boldsymbol{\theta})\mathbf{V}_n(s,t;\boldsymbol{\theta})$, it follows that for $j = 1,...,q$

$$\nabla_{\theta_j}[\mathbf{V}_n'(s,t;\bar{\boldsymbol{\theta}}_n)\cdot\mathbf{V}_n(s,t;\bar{\boldsymbol{\theta}}_n)] = 0, \tag{5.12}$$

where the symbol $\cdot$ represents the dot operator. Using the definition of $\mathbf{V}_n(s,t;\boldsymbol{\theta})$, (5.12) is equivalent to

$$\nabla_{\theta_j}[\mathbf{U}_n'(s,\bar{\boldsymbol{\theta}}_n)\Xi^2(s,\bar{\boldsymbol{\theta}}_n)\mathbf{U}_n(s,\bar{\boldsymbol{\theta}}_n)] = 0. \tag{5.13}$$

Expending (5.13) and differentiating $\mathbf{U}_n'(s,\bar{\boldsymbol{\theta}}_n)$ with respect to θ_j, we obtain

$$-2\sqrt{n}\nabla_{\theta_j}[\mathbf{p}'^n(\bar{\boldsymbol{\theta}}_n)]\Xi^2(s,\bar{\boldsymbol{\theta}}_n)\mathbf{V}_n(s,\bar{\boldsymbol{\theta}}_n) + \mathbf{U}_n'(s,\bar{\boldsymbol{\theta}}_n)\nabla_{\theta_j}(\Xi^2(s,\bar{\boldsymbol{\theta}}_n))\mathbf{U}_n(s,\bar{\boldsymbol{\theta}}_n) = 0.$$

By Condition II, $\Xi^{-1}(s, \bar{\boldsymbol{\theta}}_n)$ exists and is bounded, therefore, $\mathbf{U}_n(s, \bar{\boldsymbol{\theta}}_n) = \Xi^{-1}(s, \bar{\boldsymbol{\theta}}_n)\mathbf{V}_n(s, \bar{\boldsymbol{\theta}}_n)$. Furthermore, the asymptotic distribution of $\mathbf{U}_n(s, \boldsymbol{\theta}_0)$ yields

$$\|\mathbf{U}_n(s, \boldsymbol{\theta}_0)\| = O_p(1). \tag{5.14}$$

The statement of the proposition follows by applying the Landau rules of multiplication of little o and big O by bounded elements. $\square$

Theorem 5.4. *Under the null hypothesis H_0 and Assumptions I and II above, we have*

$$\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = [\boldsymbol{B}'(s^*, t^*; \boldsymbol{\theta}_0)\boldsymbol{B}(s^*, t^*; \boldsymbol{\theta}_0)]^{-1}\boldsymbol{B}'(s^*, t^*; \boldsymbol{\theta}_0)\, \boldsymbol{V}_n(s^*, t^*; \boldsymbol{\theta}_0) + o_p(1)$$

Proof: Start out by adding and subtracting $\mathbf{p}(\boldsymbol{\theta}_0)$ in (5.9) to obtain (we drop the gap time argument in $\mathbf{V}$ and $\mathbf{U}$ for simplicity)

$$\begin{aligned}
\mathbf{V}(s, \bar{\boldsymbol{\theta}}_n) &= \Xi(s, \bar{\boldsymbol{\theta}}_n)\sqrt{n}[\hat{\mathbf{p}}^n(s) - \mathbf{p}^n(\boldsymbol{\theta}_0)] - \Xi(s, \bar{\boldsymbol{\theta}}_n)\sqrt{n}[\hat{\mathbf{p}}^n(\bar{\boldsymbol{\theta}}_n) - \mathbf{p}^n(\boldsymbol{\theta}_0)] \\
&= \Xi(s, \bar{\boldsymbol{\theta}}_n)\mathbf{U}(s, \boldsymbol{\theta}_0) - \Xi(s, \bar{\boldsymbol{\theta}}_n)\sqrt{n}[\mathbf{p}^n(\bar{\boldsymbol{\theta}}_n) - \mathbf{p}^n(\boldsymbol{\theta}_0)].
\end{aligned}$$

Using (5.3), Assumption I, and Lemma 5.2, we obtain

$$\mathbf{V}_n(s, \bar{\boldsymbol{\theta}}_n) = \Xi(s, \bar{\boldsymbol{\theta}}_n)\mathbf{U}_n(s, \boldsymbol{\theta}_0) - \Xi(s, \bar{\boldsymbol{\theta}}_n)[\nabla_{\boldsymbol{\theta}'}\mathbf{p}(\boldsymbol{\theta}_0) + o_p(1)]\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \tag{5.15}$$

An application of Condition I, Lemma 5.2 to $\Xi(s; \bar{\boldsymbol{\theta}}_n)$, and by applying the Landau rules, we have

$$\begin{aligned}
\mathbf{V}(s, \bar{\boldsymbol{\theta}}_n) &= [\Xi(s, \boldsymbol{\theta}_0) + o_p(1)]\mathbf{U}_n(s, \boldsymbol{\theta}_0) - [\Xi(s, \boldsymbol{\theta}_0) + o_p(1)][\nabla_{\boldsymbol{\theta}'}\mathbf{p}(\boldsymbol{\theta}_0) + o_p(1)]\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&= \mathbf{V}_n(s, \boldsymbol{\theta}_0) + o_p(1)\mathbf{U}_n(s, \boldsymbol{\theta}_0) - [\Xi(s, \boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}'}\mathbf{p}(\boldsymbol{\theta}_0) + o_p(1)]\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&= \mathbf{V}_n(s, \boldsymbol{\theta}_0) - [\mathbf{B}(s, \boldsymbol{\theta}_0) + o_p(1)]\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1).
\end{aligned} \tag{5.16}$$

Multiplying (5.16) by $\mathbf{B}'(s, t; \boldsymbol{\theta}_0)$ and using the fact that $\mathbf{B}'(s, t; \boldsymbol{\theta}_0)\mathbf{V}_n(s, \bar{\boldsymbol{\theta}}_n)$ is negligible by Proposition 5.3 gives the desired result. $\square$

5.3. THE TEST STATISTIC AND LARGE SAMPLE PROPERTIES

With $\bar{\boldsymbol{\theta}}_n$ being the minimum chi-square estimator, let $\mathbf{V}_n(s, \bar{\boldsymbol{\theta}}_n)$ be the value of $\mathbf{V}_n(s, \boldsymbol{\theta})$ at $\bar{\boldsymbol{\theta}}_n$. Furthermore, let $\mathbf{A}(s, t; \boldsymbol{\theta})$ be the $k \times k$ matrix defined by

$$\mathbf{A}(s, t; \boldsymbol{\theta}) = I_k - \mathbf{B}(s, t; \boldsymbol{\theta})[\mathbf{B}'(s, t; \boldsymbol{\theta})\mathbf{B}(s, t; \boldsymbol{\theta})]^{-1}\mathbf{B}'(s, t; \boldsymbol{\theta}).$$

Theorem 5.5. *Under the regularity conditions and assumptions stated above, and under H_0, we have*

$$\boldsymbol{V}_n(s^*, t^*; \bar{\boldsymbol{\theta}}_n) \xrightarrow{d} N_k(\mathbf{0}, \Gamma(s^*, t^*; \boldsymbol{\theta}_0)),$$

where $\Gamma(s^*, t^*; \boldsymbol{\theta}_0) = \boldsymbol{A}'(s^*, t^*; \boldsymbol{\theta}_0)\Xi(s^*, t^*; \boldsymbol{\theta}_0)\Sigma\Xi(s^*, t^*; \boldsymbol{\theta}_0)\boldsymbol{A}(s^*, t^*; \boldsymbol{\theta}_0)$.

Proof: From Theorem 5.1 and the definition of $\mathbf{V}_n(s^*, t^*; \boldsymbol{\theta})$, the asymptotic distribution of $\mathbf{V}_n(s^*, t^*; \boldsymbol{\theta}_0)$ under H_0 is given by

$$\mathbf{V}(s^*, t^*; \boldsymbol{\theta}_0) \to N(\mathbf{0}, \Omega(s^*, t^*; \boldsymbol{\theta}_0)) \tag{5.17}$$

where $\Omega(s^*, t^*; \boldsymbol{\theta}_0) = \Xi(s^*, t^*; \boldsymbol{\theta}_0)\Sigma\Xi'(s^*, t^*; \boldsymbol{\theta}_0)$. Taylor expending of $\mathbf{V}_n(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$ around $\boldsymbol{\theta}_0$ and an application of Theorem 5.4 to $\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ successively yields

$$\begin{aligned}
\mathbf{V}_n(s^*, \bar{\boldsymbol{\theta}}_n) &= \mathbf{V}_n(s^*, \boldsymbol{\theta}_0) - [\mathbf{B} + o_p(1)]\sqrt{n}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1) \\
&= \mathbf{V}_n(s^*, \boldsymbol{\theta}_0) - [\mathbf{B} + o_p(1)][(\mathbf{B}\mathbf{B}')^{-1}\mathbf{B}'\mathbf{V}_n(s^*, \boldsymbol{\theta}_0)] + o_p(1) \\
&= [I - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}']\mathbf{V}_n(s^*, \boldsymbol{\theta}_0) + o_p(1).
\end{aligned}$$

Thus the limiting distribution of $\mathbf{V}_n(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$ follows upon applying standard results of multivariate normal distributions together with the limiting distribution of $\mathbf{V}_n(s^*, \boldsymbol{\theta}_0)$ under H_0. $\square$

To obtain a statistic with limiting chi-square distribution, it suffices to find a uniformly consistent estimator $\hat{\Gamma}(s, t; \boldsymbol{\theta}_0)$ of $\Gamma(s, t; \boldsymbol{\theta}_0)$. We now provide a discussion on the choice of a consistent estimator of $\Gamma(s, t; \boldsymbol{\theta}_0)$. We first present a consistent estimator of $\Sigma_1(s, t; \boldsymbol{\theta}_0)$. The limiting variance-covariance matrix $\Sigma_1(t_1, t_2, ..., t_{k-1})$ of the finite dimensional distributions of W for any $0 < t_1 < t_2 < \cdots < t_{k-1}$ under H_0 is given by

$$\Sigma_1(s,t;\boldsymbol{\theta}_0) = \begin{bmatrix} \Sigma_1(s,t_1;\boldsymbol{\theta}_0) & \Sigma_1(s,t_1;\boldsymbol{\theta}_0) & \cdots & \Sigma_1(s,t_1;\boldsymbol{\theta}_0) \\ \Sigma_1(s,t_1;\boldsymbol{\theta}_0) & \Sigma_1(s,t_2;\boldsymbol{\theta}_0) & \cdots & \Sigma_1(s,t_2;\boldsymbol{\theta}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_1(s,t_1;\boldsymbol{\theta}_0) & \Sigma_1(s,t_2;\boldsymbol{\theta}_0) & \cdots & \Sigma_1(s,t_{k-1};\boldsymbol{\theta}_0) \end{bmatrix};$$

where each $\Sigma_1(s,t_j;\boldsymbol{\theta}_0)$, $j = 1,...,k-1$ is given by (cf. Peña et al. [60])

$$\Sigma_1(s,t_j;\boldsymbol{\theta}_0) = -\bar{F}^2(t_j,\boldsymbol{\theta}_0)\int_0^{t_j} \frac{\bar{F}(dw,\boldsymbol{\theta}_0)}{\bar{F}(w,\boldsymbol{\theta}_0)y(s,w;\boldsymbol{\theta}_0)}. \tag{5.18}$$

With $Y(s,t) = \sum_{i=1}^{n} Y_i(s,t)$ being the generalized at-risk process, it has been shown in the aforementioned paper that a uniformly consistent estimator of $y(s,t)$ is $\bar{Y}(s,t)$. Therefore, a natural estimator of $\Sigma_1(s,t_j;\boldsymbol{\theta}_0)$ under H_0 for each j is

$$\hat{\Sigma}_1(s,t_j;\bar{\boldsymbol{\theta}}_n) = -\bar{F}^2(t,\bar{\boldsymbol{\theta}}_n)\int_0^{t_j} \frac{\bar{F}(dw,\bar{\boldsymbol{\theta}}_n)}{\bar{F}(w,\bar{\boldsymbol{\theta}}_n)\bar{Y}(s,w)}. \tag{5.19}$$

Lemma 5.6. *Under H_0, for $j = 1,...,k-1$, we have*

$$\sup_{t_j\in[0,t^\star]} \left|\hat{\Sigma}_1(s,t_j;\bar{\boldsymbol{\theta}}_n) - \Sigma_1(s,t_j;\boldsymbol{\theta}_0)\right| \stackrel{as}{\to} 0$$

as $n \to \infty$.

Proof: This just follows from the continuity of $\bar{F}(t,\boldsymbol{\theta})$ and the consistency of the minimum chi-square estimator. □

Remark 5.3. It is desirable to obtain closed form expression for the components of $\Sigma_1(s,t_j;\boldsymbol{\theta}_0)$. However, this may not be possible if the tested parametric family is different from the exponential family, or gamma family because the expression of $y(s,t;\boldsymbol{\theta})$ involves computing the renewal function $\rho(t)$. Closed form expression for the renewal function exists for exponential and Gamma distributions, consequently one may not be able to obtain $\Sigma_1(s,t_j;\boldsymbol{\theta}_0)$. In those situations, Lemma 5.6 can be used. Suppose for instance

$F \in \mathcal{F}_\theta = \{F(t, \boldsymbol{\theta}) = 1 - \exp(-\theta t) : \theta > 0\}$ and the τ_is are assumed to follow exponential distribution with parameter μ. Under the two models, we have, as $s \to \infty \quad y(\infty, t; \boldsymbol{\theta}) = (1 + \theta_0 \mu^{-1})\exp[-(\theta_0 + \mu)t]$.

Hence, an expression of $\Sigma_1(\infty, t_j; \boldsymbol{\theta}_0)$ is given by

$$\Sigma_1(\infty, t_j; \boldsymbol{\theta}_0) = \frac{\theta_0 \mu \exp(-2\theta_0 t_j)\{\exp[(\theta_0 + \mu)t_j] - 1\}}{(\theta_0 + \mu)^2}.$$

A consistent estimator of $\Gamma(s, t; \boldsymbol{\theta}_0)$ is then obtained by just replacing $\boldsymbol{\theta}_0$ by the minimum chi-square estimator everywhere in $\Gamma(s, t; \boldsymbol{\theta}_0)$ and is given by $\hat{\Gamma}(s, t; \bar{\boldsymbol{\theta}}_n)$. Note also that it is important that the two matrices (true and estimate) have the same structure. This is critical for the Moore-Penrose generalized inverse of the aforementioned matrices to converge in probability for large n.

The next result pertains to the rank of the matrices $\Gamma(s, t; \boldsymbol{\theta}_0)$ and $\hat{\Gamma}(s, t; \bar{\boldsymbol{\theta}}_n)$. This will be used later to obtain the number of degrees of freedom of our chi-square statistic.

Theorem 5.7. *Under the regularity condition and assumptions, we have*
[i] rank$(\Gamma(s, t; \boldsymbol{\theta}_0)) = k - q - 1$
[ii] P(rank$(\hat{\Gamma}(s, t; \bar{\boldsymbol{\theta}}_n)) = k - q - 1) \to 1$ *as* $n \to \infty$.

Proof: [*i*] Abbreviate $\Sigma_1(s, t; \boldsymbol{\theta}_0)$, $\Xi(s, t; \boldsymbol{\theta}_0)$, $\mathbf{A}(s, t; \boldsymbol{\theta}_0)$ and $\Gamma(s, t; \boldsymbol{\theta}_0)$ by Σ_1, Ξ, $\mathbf{A}$ and Γ respectively. We begin by noting that because Σ_1 is a positive definite matrix, there exists a matrix T via Cholesky decomposition such that $\Sigma_1 = TT'$. Therefore,

$$\text{rank}(\Gamma) = \text{rank}[\mathbf{A}\Xi J T T' J' \Xi \mathbf{A}] = \text{rank}[(\mathbf{A}\Xi J T)(\mathbf{A}\Xi J T)'],$$

and rank(Γ) reduces to rank$[\mathbf{A}\Xi J T]$ (cf. Theorem 1.2.11 of Graybill [32]). The latter can be further reduced to rank$[\mathbf{A}\Xi J]$ by a second application of the same theorem since T is nonsingular. The proof of [*i*] will be completed if we can show that rank$[\mathbf{A}\Xi J] = k - q - 1$. Using results from linear models theory, that amounts to showing that

$$\text{rank}(\mathbf{A}\Xi J) = \dim \mathcal{C}(\mathbf{A}\Xi J) = k - \dim \mathcal{C}^{\perp}(\mathbf{A}\Xi J),$$

where $\mathcal{C}(\mathbf{A}\Xi J)$ is the column space spanned by the columns of the matrix $\mathbf{A}\Xi J$ and $\mathcal{C}^{\perp}(\mathbf{A}\Xi J)$ is the space of all vectors orthogonal to $\mathcal{C}(\mathbf{A}\Xi J)$. Let $\mathbf{B}(s, \boldsymbol{\theta})$ be the matrix with (i, j)th defined in (5.11). It is clear that $\text{rank}(\mathbf{B}) = q$ and $\dim \mathcal{C}(\mathbf{B}) = q$. Let $\mathbf{e} = \Xi^{-1}\mathbf{1}$, where $\mathbf{1}$ is a $k \times 1$ vector with $(\mathbf{1})_i = 1$, for $i = 1, ..., k$. Then it is straightforward to see that using matrix multiplication

$$\mathbf{B}'\mathbf{e} = \nabla_{\boldsymbol{\theta}'}[\mathbf{1}' \cdot \mathbf{p}(\boldsymbol{\theta})]' = 0. \tag{5.20}$$

From (5.20), it follows that $\mathbf{e}$ is orthogonal to $\mathcal{C}(\mathbf{B})$ and consequently $\dim \mathcal{C}[\mathbf{B}, \mathbf{e}] = q+1$. To complete the proof, we need to show that the space of all vectors orthogonal to $\mathcal{C}[\mathbf{B}, \mathbf{e}]$ is $\mathcal{C}[\mathbf{A}\Xi J]$. Observe that

$$\begin{aligned} (\mathbf{A}\Xi J)'\mathbf{B} &= J'\Xi\mathbf{A}\mathbf{B} = J'\Xi\mathbf{0} = \mathbf{0}, \quad \text{and} \\ (A\Xi J)'\mathbf{e} &= J'\mathbf{1} = \mathbf{0}. \end{aligned}$$

Therefore $\mathcal{C}(\mathbf{A}\Xi J)$ is orthogonal to $\mathcal{C}[\mathbf{B}, \mathbf{e}]$, hence $\mathcal{C}^{\perp}(\mathbf{A}\Xi)J) \supseteq \mathcal{C}[\mathbf{B}, \mathbf{e}]$. In a similar way, we can prove the inverse inclusion, and the result follows.
$[ii]$ This part follows from part $[i]$ and standard results on rank of uniformly consistent estimator of matrices. $\square$

We are now ready to construct our test statistic. Let $\Gamma^{-}(s^*, t^*; \boldsymbol{\theta}_0)$ and $\hat{\Gamma}^{-}(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$ denote the Moore-Penrose generalized inverse of $\Gamma(s^*, t^*; \boldsymbol{\theta}_0)$ and $\hat{\Gamma}^{-}(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$ respectively. From $[i]$ and $[ii]$ of Theorem 5.7, it follows that

$$\hat{\Gamma}^{-}(s^*, t^*; \bar{\boldsymbol{\theta}}_n) \xrightarrow{p} \Gamma^{-}(s^*, t^*; \boldsymbol{\theta}_0). \tag{5.21}$$

The result in (5.21) is key to obtaining the asymptotic distribution of the chi-square statistic, given in the next theorem.

Theorem 5.8. *Define* $\bar{Q}(s^*, t^*) = \boldsymbol{V}'(s^*, t^*; \bar{\boldsymbol{\theta}}_n)\hat{\Gamma}^-(s^*, t^*; \bar{\boldsymbol{\theta}}_n)\,\boldsymbol{V}(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$. *Then, under* H_0

$$\bar{Q}(s^*, t^*) \xrightarrow{d} \chi^2(k - q - 1),$$

and the test reject the hypothesized family of distributions at level α *if* $\bar{Q}(s^*, t^*) \geq \chi^2(k - q - 1, \alpha)$*, where* $\chi^2(k - q - 1, \alpha)$ *is the upper* α *point of* $\chi^2(k - q - 1)$*.*

Proof: Result in (5.21) and Theorem 5.4 together with results on multivariate normal distribution (cf. Moore [54]) provide the result. □

5.4. SIMULATION DESIGN

We perform a Monte Carlo simulation study using the **R** software. The goal of the simulation study is to assess estimation using minimum chi-square methods, and to gauge the performance of our proposed random cells test with respect to nominal and achieved significance levels. For the sake of brevity, we design our simulations using 3 random partitions (I_l) of the monitoring period; although more than three partitions can be considered. In order to carry out the simulation, recurrent event data must be generated during a study monitoring period $[0, \tau]$. To do so, we use a structure that reconciles the interocurrence time survivor function $\bar{F}$ and the length of the monitoring period τ-by implication, the censoring distribution $\bar{G}$. A well-known structure that reconciles $\bar{F}$ and $\bar{G}$ in the presence of recurrent event data is the generalized Koziol-Green (GKG) model (Koziol & Green [44]). The GKG model for a recurrent event settings postulates the existence of a monitoring parameter $\beta > 0$ such that $\bar{G}(t) = \bar{F}(t)^{\beta}$. The parameter β controls the events intensity over the monitoring period and is reasonably constrained to (0, 1] for practical relevance-constraint that leads to more observed recurrences. We set the value of β to 0.3 and consider estimating and testing within two parametric models: the Exponential and the Weibull parametric lifetime models. To find the minimum chi-square estimator, the quadratic form $\mathbf{V}'(s^*, t^*; \boldsymbol{\theta})\mathbf{V}(s^*, t^*; \boldsymbol{\theta})$ was used. It is to be noted that any other estimator that is asymptotically equivalent to the minimum chi-square estimator (MCSE) provides the same asymptotic result as the MCSE. In the case of recurrent events, among the class of MCSE, the minimum Hellinger distance estimator (cf. Beran [13]) has proved to have consistently provided unbiased estimators of the parametric family and has been shown to belong to the class of MCSE (cf. Berkson [14]). Moreover, the minimum Hellinger distance estimator has also facilitated convergence issues related to the quadratic form when generalized inverses are used. Its consistency and asymptotic efficiency were discussed in Harris and Kanji [35].

Model I: The inter event time survivor function follows an exponential distribution with parameter θ and $\bar{F}(t,\theta) = \exp(-\theta t)$.

Model II: True inter event time survivor function follows a Weibull distribution; $\bar{F}(t,\theta) = \exp(-t^{\theta})$, where the scale parameter is taken to be 1.

Estimation

For models I and II, we vary θ in {5.00, 3.00, 2.00, 1.75, 1.50, 1.00, .75, .50, .25}. We view these values as 'true' parameters for the sake of simulation and to gauge how well they are recuperated by our estimation methods. Under model I, the parametric values range from convex-shape densities with lighter right tail than $\bar{F}(t,\theta = 1)$ (i.e all $\theta > 1$), to convex-shape densities with heavier tail than $\bar{F}(t,\theta = 1)$ (i.e for all $\theta < 1$).
Under model II, the parameter values correspond to density functions with lighter tails than $\bar{F}(t,\theta = 1)$, but starting out from concave-shape density and progressively moving towards highly skewed and convex-shape densities. We consider small sample as well as large sample estimation (n in { 20, 30, 50, 75, 100, 200}). For each combination of (θ, n), we run 1000 replications to estimate the parameter and carry out the test. Figures 5.1-5.4 provide graphical displays of the distribution of the parameter estimates for various values of θ as sample size varies. From the figures, it is trivial and anticipated that as n gets larger, the parametric family is better estimated with minimum bias. It is also noteworthy that larger values of the parameter require larger sample sizes to benefit from more accurate estimation. Tables 5.1 and 5.2 display the value of a 'true' parameter (i.e. θ as set by the simulated data), the average value of the parameter across 1000 replications ($\bar{\theta}$), the median value of the parameter ($\dot{\theta}$) and the standard deviation around the

estimation of $\bar{\theta}$. The recurrent event data are generated using these parameters as true parametric settings. $\bar{\theta}$ and $\dot{\theta}$ are the estimated values from the simulation after we apply the minimum chi-square framework. We found out that for the Weibull model, the more concave shape densities (i.e $\theta \gg 1$) are prone to bigger estimation errors than the convex counterparts. Similar conclusions can be drawn about the exponential model where, relative to $\bar{F}(t, \theta = 1)$, heavier tail densities are better estimated than lighter tail densities. In general though, the statistics from Tables 5.1 and 5.2 along with Figures 5.1-5.4 have given us enough confidence to build a testing mechanism around the minimum chi-square estimation approach.

Testing:

For testing the parametric family against the NPMLE, the estimated values of $\bar{\theta}$, say $\bar{\boldsymbol{\theta}}_n$ were plugged into the quadratic form of Theorem 5.8 to obtain test statistics that are compared to the chi-square distribution with one degree of freedom. The test has been set to reach a nominal significance level of 0.05. The achieved significance is represented by the proportion of these quadratic forms that cross the upper .95 quantile of the limiting chi-square distribution (i.e. 3.8415). Table 5.3 displays such results for selected values of θ. We use two different tests based on two different estimations of $\Gamma(s, t, \bar{\theta}_n)$. The first estimator is based on the parametric cumulative hazard functions as outlined in equation (5.19). The resulting test is labeled $\textbf{Test}_1$. The second estimator substitutes the parametric estimation of equation (5.19) by its non-parametric equivalent. The second test is labeled $\mathfrak{Test}_2$ on Table 5.3. For the exponential model, the tests are anti-conservative in small samples and tend to be conservative as sample size increases. For Weibull parametric model, the same pattern is apparent; but the test built around a non-parametric estimation of the integrated hazard is very conservative when the parameter is small. The Weibull family with decreasing hazard reacts extremely conservatively to the test in large samples than the Weibull family with increasing hazard. Anti-conservativeness remains an issue in small samples; though more pronounced when the hazard increases in time.

By and large, reliability growth suffers from small and large samples conservativeness while reliability deterioration suffers from small samples anti-conservativeness and a large samples conservativeness of the test.

Table 5.1. Exponential Parametric Family: $\bar{F}(t,\theta) = \exp(-\theta t)$

n	θ	$\bar{\theta}$	SD	$\dot{\theta}$	n	θ	$\bar{\theta}$	SD	$\dot{\theta}$
20	0.25	0.26	0.03	0.26	75	0.25	0.25	0.01	0.25
	0.50	0.53	0.06	0.53		0.50	0.50	0.03	0.50
	0.75	0.79	0.09	0.79		0.75	0.75	0.04	0.75
	1.00	1.05	0.12	1.05		1.00	1.00	0.06	0.99
	1.25	1.32	0.15	1.32		1.25	1.25	0.07	1.24
	1.50	1.58	0.18	1.58		1.50	1.50	0.08	1.49
	1.75	1.84	0.21	1.84		1.75	1.75	0.01	1.74
	2.00	2.10	0.24	2.11		2.00	2.00	0.11	1.99
	3.00	3.16	0.36	3.16		3.00	3.00	0.17	2.98
	5.00	5.26	0.59	5.27		5.00	5.00	0.28	4.97
30	0.25	0.26	0.02	0.26	100	0.25	0.25	0.01	0.25
	0.50	0.52	0.05	0.52		0.50	0.51	0.02	0.51
	0.75	0.78	0.07	0.77		0.75	0.76	0.04	0.76
	1.00	1.04	0.10	1.03		1.00	1.01	0.05	1.01
	1.25	1.30	0.12	1.29		1.25	1.26	0.06	1.27
	1.50	1.56	0.15	1.55		1.50	1.52	0.07	1.52
	1.75	1.82	0.17	1.80		1.75	1.77	0.09	1.77
	2.00	2.07	0.20	2.06		2.00	2.02	0.10	2.03
	3.00	3.11	0.30	3.09		3.00	3.03	0.15	3.04
	5.00	5.19	0.50	5.15		5.00	5.05	0.25	5.06
50	0.25	0.25	0.02	0.25	200	0.25	0.25	0.01	0.25
	0.50	0.51	0.04	0.50		0.50	0.50	0.02	0.50
	0.75	0.76	0.06	0.75		0.75	0.75	0.03	0.75
	1.00	1.01	0.08	1.00		1.00	1.00	0.04	1.00
	1.25	1.26	0.09	1.25		1.25	1.25	0.05	1.25
	1.50	1.52	0.11	1.50		1.50	1.50	0.06	1.50
	1.75	1.77	0.13	1.76		1.75	1.75	0.07	1.74
	2.00	2.02	0.15	2.01		2.00	2.00	0.08	1.99
	3.00	3.03	0.23	3.01		3.00	3.00	0.12	2.99
	5.00	5.05	0.38	5.02		5.00	5.00	0.19	4.98

Table 5.2. Weibull Parametric Family: $\bar{F}(t,\theta) = \exp(-t^{\theta})$

n	θ	$\bar{\theta}$	SD	$\dot{\theta}$	n	θ	$\bar{\theta}$	SD	$\dot{\theta}$
20	0.25	0.25	0.01	0.25	75	0.25	0.25	0.01	0.25
	0.50	0.51	0.03	0.51		0.50	0.50	0.01	0.50
	0.75	0.76	0.07	0.76		0.75	0.76	0.04	0.76
	1.00	1.04	0.13	1.02		1.00	1.01	0.06	1.01
	1.25	1.30	0.17	1.27		1.25	1.28	0.10	1.27
	1.50	1.54	0.24	1.50		1.50	1.54	0.13	1.54
	1.75	1.94	0.39	1.88		1.75	1.80	0.15	1.79
	2.00	2.17	0.41	2.15		2.00	2.07	0.17	2.06
	3.00	3.37	0.90	3.18		3.00	3.11	0.34	3.11
	5.00	5.57	1.43	5.29		5.00	5.33	0.75	5.29
30	0.25	0.25	0.01	0.25	100	0.25	0.25	0.01	0.25
	0.50	0.50	0.03	0.50		0.50	0.50	0.01	0.50
	0.75	0.76	0.07	0.76		0.75	0.75	0.03	0.74
	1.00	1.02	0.09	1.02		1.00	1.01	0.05	1.01
	1.25	1.28	0.15	1.26		1.25	1.26	0.07	1.27
	1.50	1.53	0.20	1.52		1.50	1.53	0.11	1.54
	1.75	1.87	0.26	1.81		1.75	1.79	0.14	1.79
	2.00	2.10	0.33	2.08		2.00	2.07	0.16	2.04
	3.00	3.15	0.58	3.11		3.00	3.10	0.26	3.09
	5.00	5.42	1.13	5.40		5.00	5.21	0.50	5.18
50	0.25	0.25	0.01	0.25	200	0.25	0.25	0.01	0.25
	0.50	0.50	0.02	0.50		0.50	0.50	0.01	0.50
	0.75	0.76	0.04	0.76		0.75	0.75	0.02	0.75
	1.00	1.02	0.08	1.02		1.00	1.00	0.04	1.00
	1.25	1.28	0.10	1.28		1.25	1.25	0.05	1.25
	1.50	1.55	0.14	1.54		1.50	1.51	0.08	1.51
	1.75	1.81	0.18	1.79		1.75	1.76	0.09	1.76
	2.00	2.08	0.22	2.08		2.00	2.02	0.13	2.02
	3.00	3.18	0.43	3.15		3.00	3.04	0.18	3.05
	5.00	5.29	0.90	5.14		5.00	5.07	0.34	5.07

Table 5.3. Observed Significance

		$e^{-\theta t}$		$e^{-t^{\theta}}$	
θ	n	$\mathfrak{Test}_1$	$\mathfrak{Test}_2$	$\mathfrak{Test}_1$	$\mathfrak{Test}_2$
	20	0.090	0.100	0.030	0.001
	30	0.120	0.100	0.010	0.010
0.5	50	0.070	0.070	0.001	0.001
	75	0.050	0.050	0.001	0.001
	100	0.060	0.050	0.001	0.001
	200	0.030	0.030	0.001	0.001
	20	0.090	0.100	0.110	0.100
	30	0.120	0.100	0.120	0.100
1.0	50	0.070	0.070	0.070	0.070
	75	0.050	0.050	0.050	0.050
	100	0.060	0.050	0.060	0.050
	200	0.030	0.030	0.030	0.040
	20	0.090	0.100	0.190	0.100
	30	0.120	0.100	0.140	0.110
3.0	50	0.070	0.070	0.070	0.060
	75	0.050	0.050	0.040	0.020
	100	0.060	0.050	0.020	0.010
	200	0.030	0.030	0.020	0.001

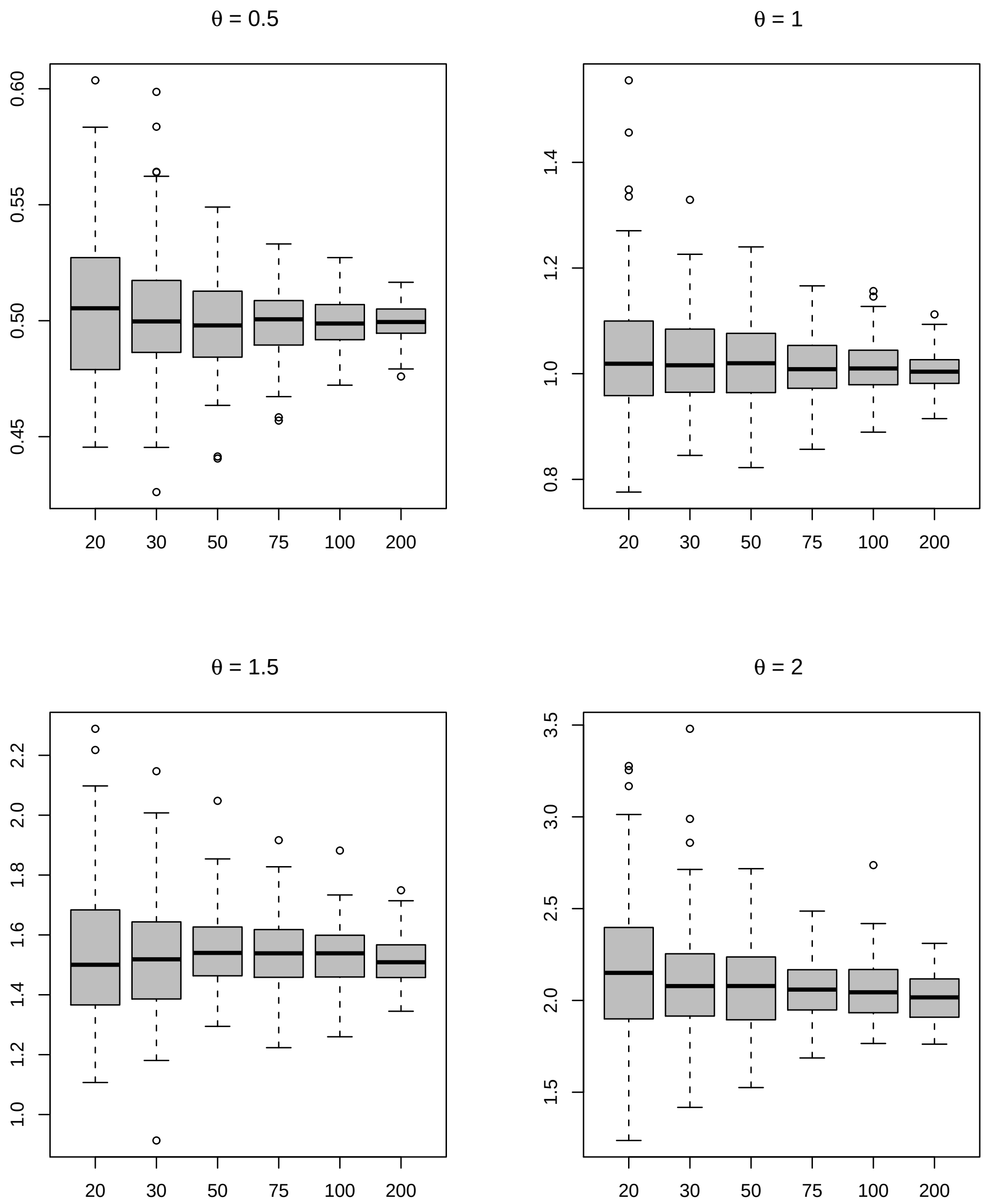

Figure 5.1. Minimum χ^2 parameter estimation as function of n under the Weibull model.

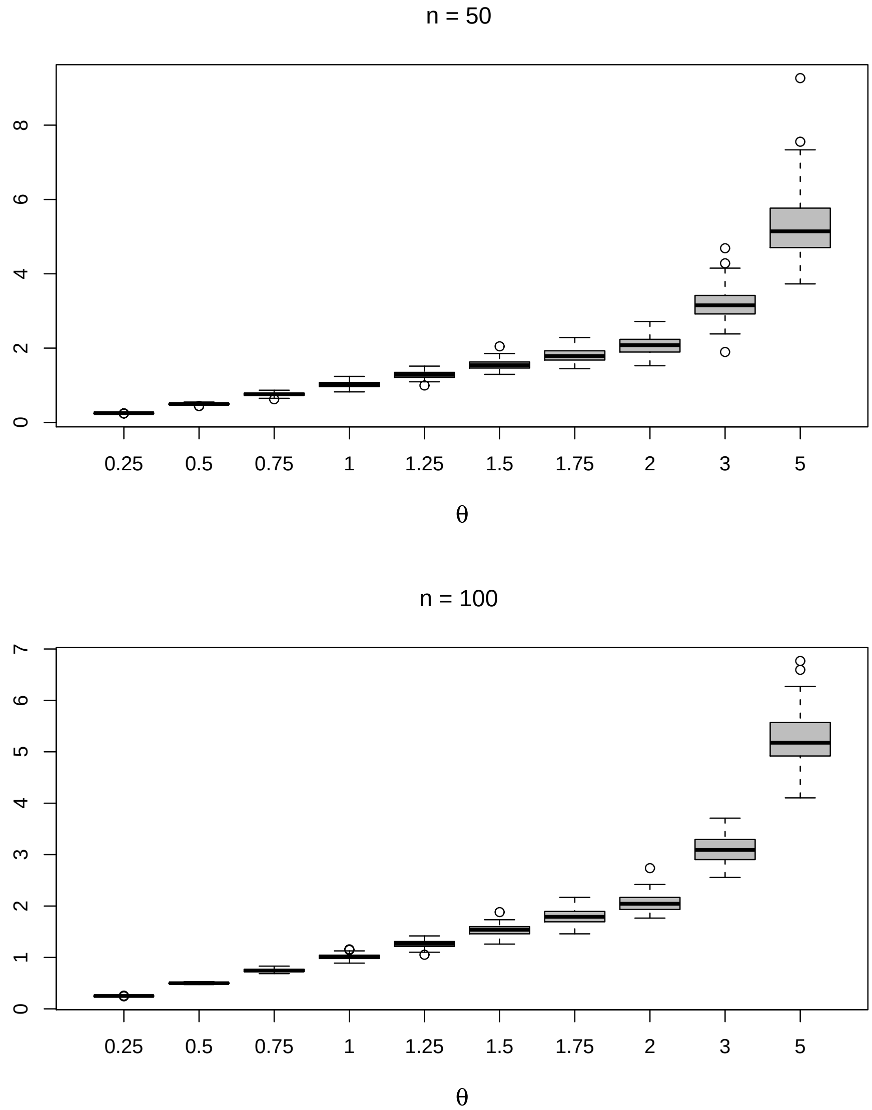

Figure 5.2. Minimum χ^2 parameter estimation as function of θ under the Weibull model.

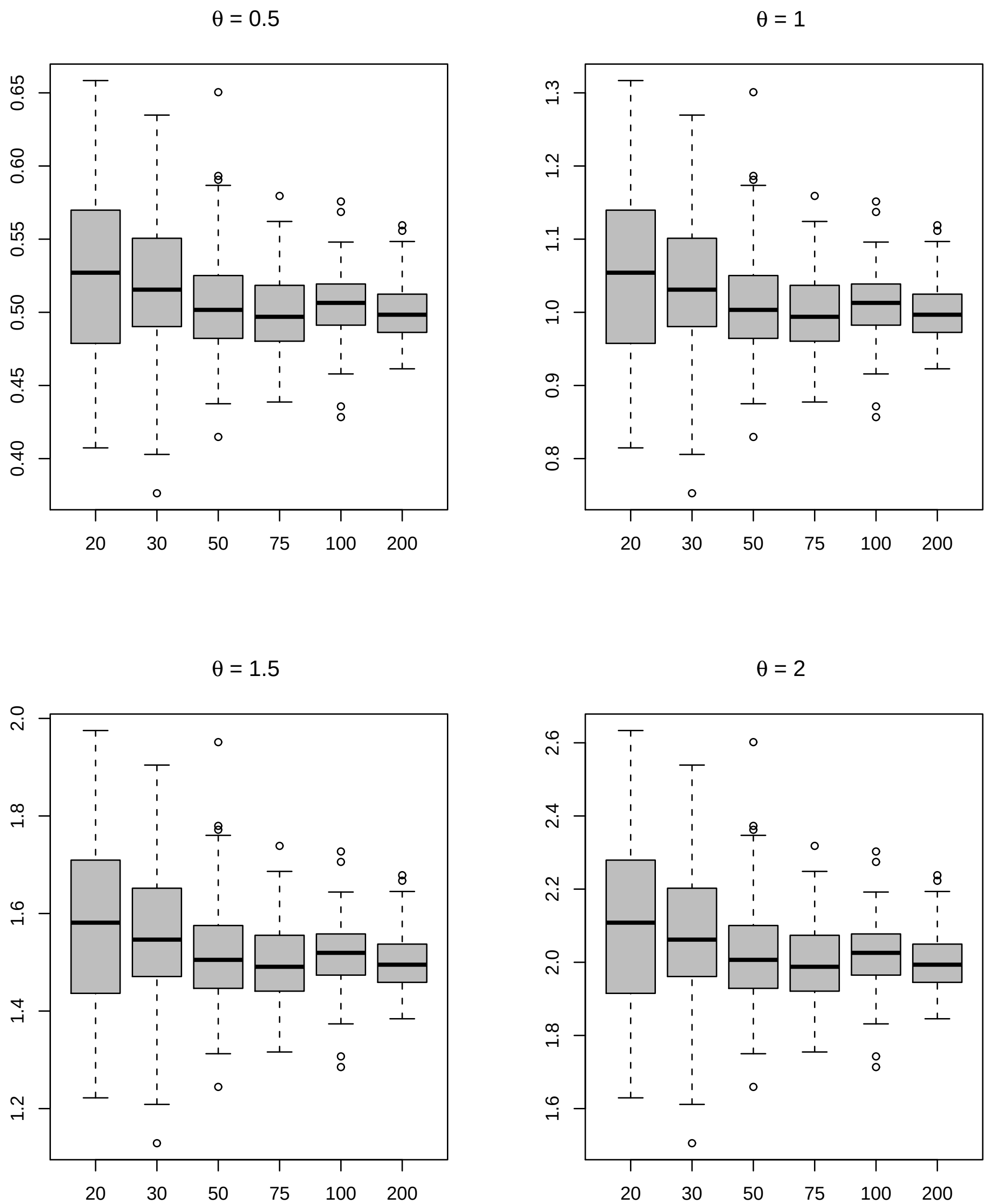

Figure 5.3. Minimum χ^2 parameter estimation as function of n under Exponential model.

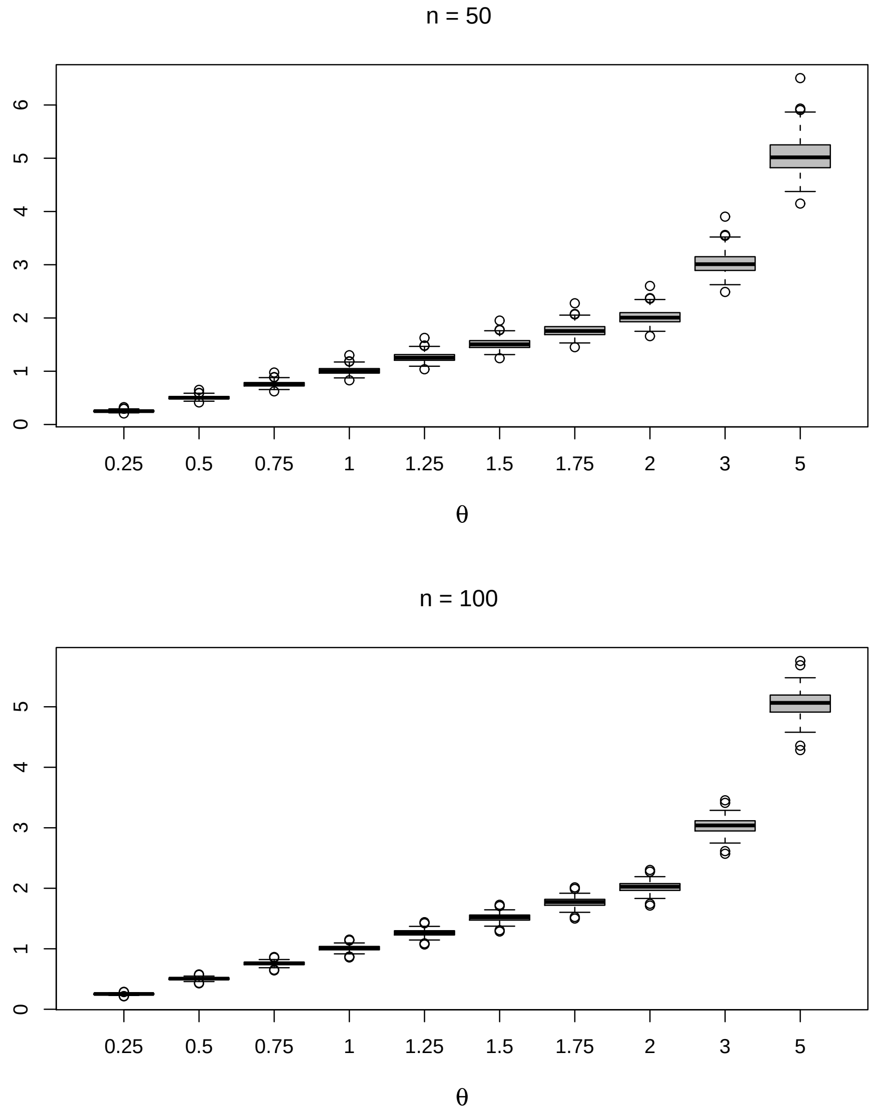

Figure 5.4. Minimum χ^2 parameter estimation as function of θ under Exponential model.

5.5. APPLICATION

For our proposed test to be acceptable, its performance in real-life data setting is desirable. The data we chose as illustrative example has surfaced in more than one publication. The primary objective of the data collection was to obtain information regarding the distribution of failure intervals for the air conditioning system of a fleet of Boeing 720 jet airplanes. Successive failure times of the air conditioning (AC) system were recorded for each member of the fleet. After roughly 2000 hours of service, the planes received a major overhaul. The failure time interval containing the major overhaul is not recorded in the data since the length of that failure time may have been affected by the major overhaul (See Proschan [66]).

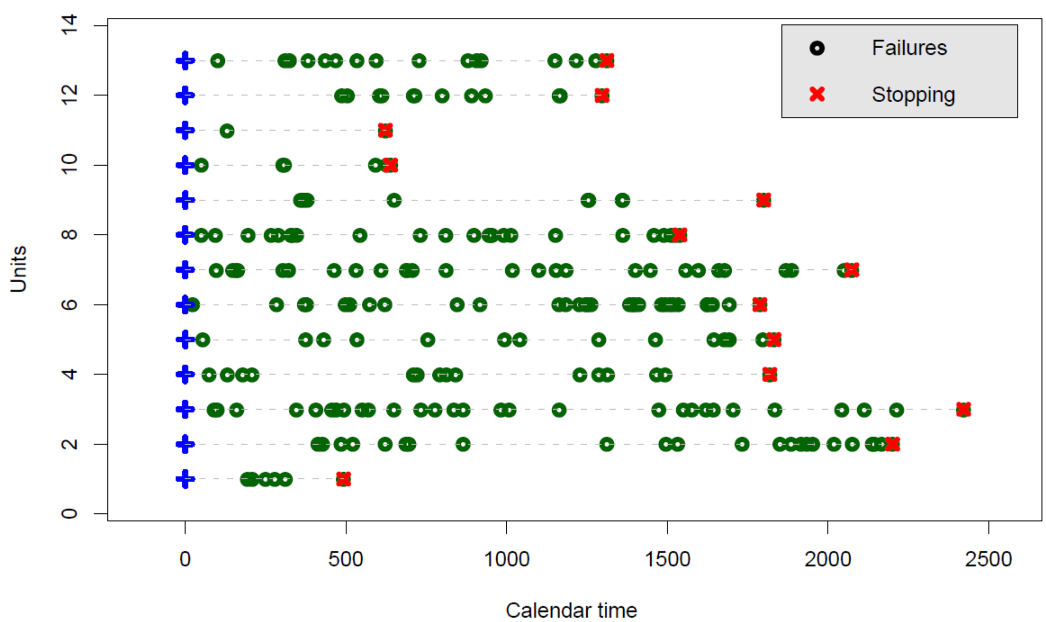

Figure 5.5. Air Conditioning Data, successive failures of AC machines for 13 jets.

In line with the goal of the data collection, once the distributional properties of the inter event times are known, reliability predictions, maintenance scheduling and decisions can be made regarding the entire fleet of Boeing 720 jets. Proschan [66] has demonstrated evidence against an exponential fit to the pooled data and has argued for a decreasing failure rate. Park and Kim [58] and Gaudoin et al. [30] suggested that these data may have come from a non homogenous Poisson process with a power law (Weibull) intensity process. Our goal is to handle these data along the line of recurrent event analysis, identify the NPMLE estimate of the survivor function, use the minimum chi-square parametric estimation to find the parameter of the Weibull parametric family with little to no evidence against the NPMLE, and suggest this parameter setting for maintenance scheduling and prediction purposes. We apply the minimum chi-square method to the AC data to estimate and test the parametric family most likely to represent these data.

We fit a two-parameter Weibull lifetime model (i.e. $\bar{F}(t) = e^{-\gamma t^{\theta}}$; shape parameter θ and scale parameter γ) to these data. To be consistent with the current study, we denote the parameters by $\boldsymbol{\theta} = (\theta, \gamma) \equiv (\theta_1, \theta_2)$. The minimum chi-square approach based on the Hellinger distance estimation has estimated the shape parameter as $\hat{\theta} = 0.868$. The scale parameter was estimated to be $\hat{\gamma} = 0.01247$. Figure 5.6. presents a pictorial representation of an overlay of the non-parametric estimator of the inter event time survivor function and the estimated parametric family as suggested by the minimum chi-square estimation approach. With the scale parameter set to its value of 0.01247 and the shape parameter value of 0.868 taken as a 'true' value, we mis-specify the shape parameter to assess the robustness of the minimum chi-square approach to parameter mis-specification. $\mathfrak{Test}_1$ and $\mathfrak{Test}_2$ are again used to assess this robustness. Table 5.3 presents the results of our tests under parameter mis-specification. On the table, R stands for a decision to reject the parametric family while A stands for a decision not to reject it. The message contains in the table is that $\mathfrak{Test}_1$ displays a symmetrically oriented robustness while $\mathfrak{Test}_2$ displays a one-sided robustness.

Table 5.4. Testing Under Mis-speciffied parameters

θ	$\mathfrak{Test}_1$	$\mathfrak{Decision}$	$\mathfrak{Test}_2$	$\mathfrak{Decision}$
0.06	50.357	R	0.860	A
0.16	32.534	R	0.858	A
0.26	20.830	R	0.853	A
0.36	13.191	R	0.844	A
0.46	08.228	R	0.828	A
0.56	05.006	R	0.797	A
0.66	02.904	A	0.735	A
0.76	01.510	A	0.608	A
0.86	00.581	A	0.367	A
0.96	00.086	A	0.080	A
1.06	00.002	A	0.003	A
1.16	00.044	A	0.082	A
1.26	00.112	A	0.297	A
1.36	00.426	A	1.603	A
1.46	03.583	A	1.91+01	R
1.56	9.2+01	R	7.50+02	R
1.66	1.5+04	R	1.68+05	R
1.76	3.2+07	R	5.04+08	R
1.86	3.0+12	R	6.88+13	R
1.96	8.3+19	R	2.65+21	R

5.6. CONCLUSION

We develop a chi square goodness of fit test for testing whether the distribution of the time between failures for a recurrent event process belongs to some parametric family of distributions. The chi square test developed is adaptive in the sense that its cells boundaries are data dependent. The test developed is more flexible and guarantees that cells probabilities will not be small as would have been the case if fixed cells probabilities were chosen. A typical chi square test based on fixed cell probabilities is known not to perform well when some of the cell partitions have little to no observation. The data-driven cell partition is a way to address this shortcoming. Our test statistic is shown to be asymptotically chi square and asymptotic results were demonstrated using the theoretical

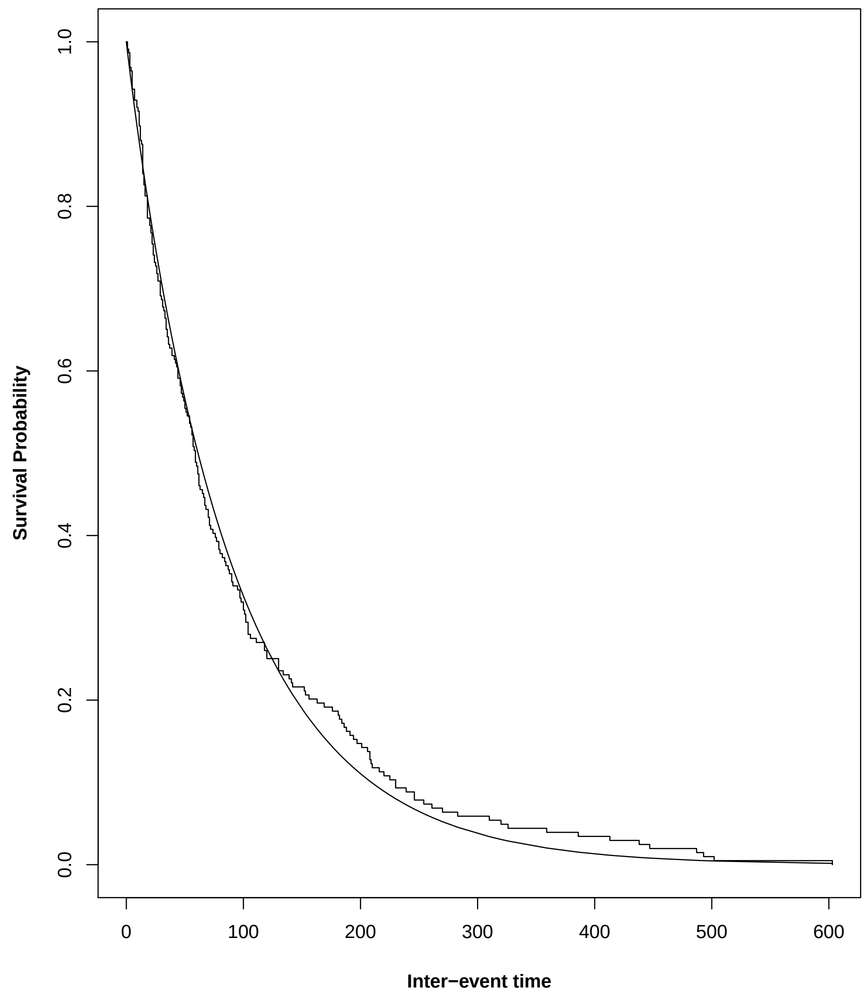

Figure 5.6. Overlay of non parametric estimate of the Boeing 720 air conditioning survivor function (step function) and the Weibull parametric estimate $e^{-\gamma t^{\theta}}$; $\gamma = 0.0124$; $\theta = 0.868$ (solid continuous line). The minimum chi-square tests did not have enough evidence to reject the parametric survivor function.

tools of empirical processes. There are many ways of estimating the unknown parameter of the underlying parametric distribution. The discussion in Berkson [14] can be applied in this setting. We found the minimum chi square estimator appealing because of its nice asymptotic properties and its equivalence to other estimators of $\boldsymbol{\theta}$ in large sample. We have assumed that the random cells converge in probability to deterministic values. This assumption could be replaced by an assumption of tightness on cells boundaries. Additionally, asymptotic independence of $\mathbf{U}_n(s, t; \bar{\boldsymbol{\theta}}_n)$ with the random cells will hold if the random cells converge to some random element on $D[0, t^*]$−which consequently ensures the required independence.

We have not addressed the choice of the optimal number of cells (k) in current the development. The optimal choice of k would undoubtedly reflects on the performance of the test and on its achieved significance. There has been discussion in the literature about how many class intervals should be chosen and how these intervals should be weighted for optimal performance of the resulting chi square test. For example, Mann and Wald [50] gave a mathematical formula relating the number of class to the significance level of the test such that these class intervals are chosen with equal probabilities. However, their procedure holds only for extremely large sample studies and its limitations have been studied by many researchers. Williams [79] discussed the advantages and limitations of Mann and Wald [50] approach and concluded that their approach required sample size can be cut to half without greatly affecting the power of the test. Hamdam [34], in the restricted case of normal distribution, approached the problem on the ground of maximizing the power of the test and concluded that a number of classes between 10 and 20 is adequate to ensure a test with reasonable power. To our knowledge, there has not been 'one size fits all' selection criterion that optimizes k in the setting of recurrent events as many parameters are in play. For example, in the case of recurrent event data under the generalized Koziol-Green model, the number of subject n, the monitoring parameter β which controls the accumulation of events during the study period, the type of variance estimator used, the significance level α, and the dimension of the parameter space (q) all have a final say on the choice of k. However, the ideal solution to the number of partitions would be to find a mathematical formula that optimizes k as a function of $(\alpha, \beta, n, \sigma, \tau, q)$. The issue around the choice of optimal k, in the presence of recurrent event data, needs more rigorous treatment and more space than can be devoted to it here. On a short note though, a heuristic approach is to choose $k \geq 2 + q$; so the test asymptotic chi square approximation can be carried.

Covariates were not included in the development of our test. Covariates are important in the process of having better knowledge of time to event, thereby goodness of fit for a specific parametric family of distributions. It would be interesting to develop goodness of fit for composite hypothesis when covariates are taken into consideration. Smooth goodness of fits along the lines of those developed in Peña [62] can be extended to recurrent events using hazard based regression models of the form $\lambda(t) = \lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{X}(t))$, where $\lambda_0(t)$ is a baseline hazard, $\boldsymbol{\beta}$ is a $q-$dimensional vector of regression coefficient and $\mathbf{X}$ is a $q-$dimensional vector of covariates which are possibly time dependent.

6. POWER OF THE DATA-DEPENDENT CHI-SQUARE TEST FOR RECURRENT EVENT DATA

6.1. INTRODUCTION

In Section 5, we considered the problem of goodness of fit test for testing the null hypothesis

$$H_0 : F(\cdot) \in \mathcal{F}_{\boldsymbol{\theta}} = \{F(\cdot;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}.$$

In this section, we study the local asymptotic power of our test. To that end, we embed the family $\mathcal{F}_{\boldsymbol{\theta}}$ in a bigger family where $\boldsymbol{\eta}$ is a nuisance parameter, that is

$$H_0 : F(\cdot) \in \mathcal{F}_{(\boldsymbol{\theta},\boldsymbol{\eta}_0)} = \{F(\cdot;\boldsymbol{\theta},\boldsymbol{\eta}_0) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q,\ \boldsymbol{\eta}_0 \in \mathbb{R}^s\},$$

versus the alternative hypothesis

$$H_{1n} : F(\cdot) \in \mathcal{F}_{(\boldsymbol{\theta},\boldsymbol{\eta}_n)} = \left\{F(\cdot;\boldsymbol{\theta},\boldsymbol{\eta}_n) : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q,\ \boldsymbol{\eta}_n = \boldsymbol{\eta}_0 + \frac{\boldsymbol{\delta}}{\sqrt{n}}\right\},$$

where $\boldsymbol{\eta}_0$ is a fixed known parameter vector in $\mathbb{R}^s$ and $\boldsymbol{\delta}$ is an arbitrary vector in $\mathbb{R}^s$. We assume $0 < s \leq q$. By doing so, we consider $(q+s)-$dimensional parametric family $\mathcal{F}_{(\boldsymbol{\theta},\boldsymbol{\eta}_n)}$ and assume that when $\boldsymbol{\eta}_n = \boldsymbol{\eta}_0$ the bigger parametric family corresponds to the simpler $q-$dimensional family $\mathcal{F}_{\boldsymbol{\theta}}$. We write $F(t;\boldsymbol{\theta},\boldsymbol{\eta}_0) = F(t;\boldsymbol{\theta})$, when $\boldsymbol{\eta}_n = \boldsymbol{\eta}_0$. Under this model, the composite hypothesis becomes, a simple hypothesis as

$$H_0 : \boldsymbol{\eta}_n = \boldsymbol{\eta}_0 \quad \text{versus} \quad H_{1n} : \boldsymbol{\eta}_n \neq \boldsymbol{\eta}_0.$$

The local asymptotic power assessment via the sequence H_{1n} is twofold: First, we can easily investigate the power of our chi-square test under fixed alternative with respect to the parameter $\boldsymbol{\eta}$ by treating the original null hypothesis as composite in $\boldsymbol{\theta}$. Next, we can directly apply the concepts of the approximate Bahadur asymptotic relative efficiency and the limiting asymptotic Pitman relative efficiency to compare test statistics for simple hypothesis testing as we explain later in this section. Other important advantage of this bigger parametric family with $s-$dimensional parameter $\boldsymbol{\eta}$ is that it allows us to cover wider and common family of alternatives.

To develop the asymptotic power we need to establish the following result pertaining to the distribution function $F(t;\boldsymbol{\theta},\boldsymbol{\eta}_n)$. Taylor expansion of $F(t;\boldsymbol{\theta},\boldsymbol{\eta}_n)$ give

$$F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_n) \doteq F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) + \boldsymbol{\nabla}_{\boldsymbol{\eta}'} F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)(\boldsymbol{\eta}_n - \boldsymbol{\eta}_0). \tag{6.1}$$

Let $\phi(t;\boldsymbol{\theta}_0) = \boldsymbol{\nabla}_{\boldsymbol{\eta}'} \log F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)$, then

$$\phi(t;\boldsymbol{\theta}_0) = \frac{1}{F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)} \boldsymbol{\nabla}_{\boldsymbol{\eta}'} F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) \boldsymbol{\nabla}_{\boldsymbol{\eta}'} F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0) = F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\phi(t;\boldsymbol{\theta}_0) \tag{6.2}$$

From (6.1) and (6.2), we obtain

$$F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_n) = F(t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\Big(1 + \phi(t;\boldsymbol{\theta}_0).\frac{\boldsymbol{\delta}}{\sqrt{n}}\Big). \tag{6.3}$$

6.2. PRELIMINARY RESULTS

Our first result in this subsection is on the asymptotic distribution of $\mathbf{U}_n(s,t;\boldsymbol{\theta}_0)$ under H_{1n}.

Lemma 6.1. *Under H_{1n}, $\boldsymbol{U}_n(s,t;\boldsymbol{\theta}_0)$ converges in distribution to $N_k(\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\theta_0)\cdot\boldsymbol{\delta},\ \Sigma)$.*

Proof: Start with

$$\begin{aligned}
\mathbf{U}_n(s,t;\boldsymbol{\theta}_0) &= \sqrt{n}(\mathbf{p}^n(s) - \mathbf{p}^n(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)) \\
&= \sqrt{n}(\mathbf{p}^n(s) - \mathbf{p}^n(\boldsymbol{\theta}_0,\boldsymbol{\eta}_n)) + \sqrt{n}(\mathbf{p}^n(\boldsymbol{\theta}_0,\boldsymbol{\eta}_n) - \mathbf{p}^n(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)) \\
&= \mathbf{U}_n(s,t;\boldsymbol{\theta}_0,\boldsymbol{\eta}_n) + \sqrt{n}(\mathbf{p}^n(\boldsymbol{\theta}_0,\boldsymbol{\eta}_n) - \mathbf{p}^n(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)) \qquad (6.4)
\end{aligned}$$

Now consider the lth element of $(\mathbf{p}^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_n) - \mathbf{p}^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))$,

$$\begin{aligned}
(\mathbf{p}_l^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_n) - \mathbf{p}_l^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)) &= \big[(F(t_l^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_n) - F(t_{l-1}^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_n)) \\
&\quad -(F(t_l^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0) - F(t_{l-1}^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0))\big] \\
&= \big[(F(t_l^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_n) - F(t_l^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)) \\
&\quad -(F(t_{l-1}^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_n) - F(t_{l-1}^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0))\big]
\end{aligned}$$

Using (6.3), we have

$$\begin{aligned}
(\mathbf{p}_l^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_n) - \mathbf{p}_l^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)) &= \big[F(t_l^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\phi(t_l^n, \boldsymbol{\theta}_0) \cdot \frac{\boldsymbol{\delta}}{\sqrt{n}} \\
&\quad -F(t_{l-1}^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\phi(t_{l-1}^n, \boldsymbol{\theta}_0) \cdot \frac{\boldsymbol{\delta}}{\sqrt{n}}\big] \\
&= \boldsymbol{\nabla}_{\boldsymbol{\eta}'}\big[F(t_l^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0) - F(t_{l-1}^n; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\big] \cdot \frac{\boldsymbol{\delta}}{\sqrt{n}} \\
&= \boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}_l^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \cdot \frac{\boldsymbol{\delta}}{\sqrt{n}}.
\end{aligned}$$

An application of (6.4) leads to $\mathbf{U}_n(s, t; \boldsymbol{\theta}_0) = \mathbf{U}_n(s, t; \boldsymbol{\theta}_0, \boldsymbol{\eta}_n) + \boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}^n(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \cdot \boldsymbol{\delta}$. Under the sequence of alternative hypotheses H_{1n}, $\mathbf{U}_n(s, t; \boldsymbol{\theta}_0, \boldsymbol{\eta}_n)$ converges in distribution to $N_k(\mathbf{0},\ \Sigma)$. Therefore, under H_{1n}, $\mathbf{U}_n(s, t; \boldsymbol{\theta}_0)$ converges to $N_k(\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \cdot \boldsymbol{\delta},\ \Sigma)$. □

The limiting distribution of $\mathbf{V}_n(s, t; \boldsymbol{\theta}_0)$ is straightforward. From Lemma 6.1, under H_{1n}, $\mathbf{V}_n(s, t; \boldsymbol{\theta}_0) \xrightarrow{d} N_k(\Xi(s, t; \boldsymbol{\theta}_0)\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \cdot \boldsymbol{\delta},\ \Xi(s, t; \boldsymbol{\theta}_0)\Sigma\Xi(s, t; \boldsymbol{\theta}_0))$.

Lemma 6.2. *Under H_{1n}, $\mathbf{V}_n(s, t; \bar{\boldsymbol{\theta}}_n)$ is distributed as $N_k(\boldsymbol{\mu},\ \Gamma(s, t; \boldsymbol{\theta}_0))$, where $\boldsymbol{\mu} = A(s, t; \boldsymbol{\theta}_0)\Xi(s, t; \boldsymbol{\theta}_0)\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \cdot \boldsymbol{\delta}$.*

Proof: Consider the equality $\mathbf{V}_n(s, t; \bar{\boldsymbol{\theta}}_n) = A(s, t; \boldsymbol{\theta}_0)\mathbf{V}_n(s, t; \boldsymbol{\theta}_0) + o_p(1)$ from Section 5. From that equality, the limiting distribution of $\mathbf{V}_n(s, t; \bar{\boldsymbol{\theta}}_n)$ under H_{1n} follows upon applying standard results of multivariate normal distributions together with the limiting distribution of $\mathbf{V}_n(s, t; \boldsymbol{\theta}_0)$ under H_{1n}. □

We are ready to obtain the asymptotic power of the chi-square test proposed in Section 5. In what follows, $\chi^2(r)$ will denote a central chi-square distribution with r degrees of freedom and its associated $100(1-\alpha)\%$ percentile will be denoted by $\chi^2(r,\alpha)$, and $\chi^2(r,\delta)$ will denote a non-central chi-square distribution with r degrees of freedom and noncentrality parameter δ.

Theorem 6.3. *If $\boldsymbol{\mu}$ is in the range space of $\Gamma(s^*,t^*;\boldsymbol{\theta}_0)$, then under H_{1n},*

$$\bar{Q}(s^*,t^*) \xrightarrow{d} \chi^2(k-q-1,\ \boldsymbol{\mu}'\Gamma^-(s^*,t^*;\boldsymbol{\theta}_0)\boldsymbol{\mu}).$$

Hence the local asymptotic power of $\bar{Q}(s^*,t^*)$ is

$$P\big(\chi^2(k-q-1,\ \boldsymbol{\mu}'\Gamma^-(s^*,t^*;\boldsymbol{\theta}_0)\boldsymbol{\mu}\big) \geq \chi^2(k-q-1,\alpha)).$$

Proof: Observe that

$$\begin{aligned} B'A(s^*,t^*;\boldsymbol{\theta}_0) &= B'(I_k - B(B'B)^{-1}B') \\ &= B' - B'B(B'B)^{-1}B' \\ &= B' - B' = \mathbf{0}. \end{aligned}$$

Clearly, we also have $B'\boldsymbol{\mu} = B'A(s^*,t^*;\boldsymbol{\theta}_0)\Xi(s,t;\boldsymbol{\theta}_0)\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\boldsymbol{\delta} = \mathbf{0}$. Therefore $\mathcal{C}(B)$ is orthogonal to $\mathcal{C}(\Gamma(s^*,t^*;\boldsymbol{\theta}_0))$, where $\mathcal{C}(B)$ denotes the column space of matrix B. Denote the orthogonal complement of $\mathcal{C}(B)$ by $\mathcal{C}^\perp(B)$. Next, define $\boldsymbol{e} = \Xi(s^*,t^*;\boldsymbol{\theta}_0)^{-1}\mathbf{1}$, where $\mathbf{1}$ is a $k\times 1$ vector with $(\mathbf{1})_i = 1$ for all $i = 1,\ldots,k$. Recall that $\operatorname{rank}(\Gamma(s^*,t^*;\boldsymbol{\theta}_0)) = \dim \mathcal{C}(\Gamma(s^*,t^*;\boldsymbol{\theta}_0)) = k-q-1$. Since $B'\boldsymbol{e} = 0$, one can show that $\mathcal{C}^\perp(\Gamma(s^*,t^*;\boldsymbol{\theta}_0)) = \mathcal{C}[B,\boldsymbol{e}]$ (See proof of Theorem 5.4). Furthermore, $\dim \mathcal{C}[B,\boldsymbol{e}] = q+1$ since $\operatorname{rank}(B) = q$. Then,

$$\begin{aligned} \boldsymbol{\mu}'\boldsymbol{e} &= (A(s^*,t^*;\boldsymbol{\theta}_0)\Xi(s^*,t^*;\boldsymbol{\theta}_0)\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\boldsymbol{\delta})'\boldsymbol{e} \\ &= \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\Xi(s^*,t^*;\boldsymbol{\theta}_0)A(s^*,t^*;\boldsymbol{\theta}_0)\boldsymbol{e} \\ &= \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\Xi(s^*,t^*;\boldsymbol{\theta}_0)((I_k - B(B'B)^{-1}B'))\boldsymbol{e} \\ &= \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\Xi(s^*,t^*;\boldsymbol{\theta}_0)\boldsymbol{e} \end{aligned}$$

where the last equality is obtained using the fact that $B'\boldsymbol{e} = 0$. Therefore

$$\begin{aligned}
\boldsymbol{\mu}'\boldsymbol{e} &= \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\Xi(s^*, t^*; \boldsymbol{\theta}_0)\Xi(s, t; \boldsymbol{\theta}_0)^{-1}\mathbf{1} \\
&= \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0)\mathbf{1} = \boldsymbol{\delta}'(\mathbf{1}'\boldsymbol{\nabla}_{\boldsymbol{\eta}'}\mathbf{p}(\boldsymbol{\theta}_0))' \\
&= \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}(\mathbf{1}'\mathbf{p}(\boldsymbol{\theta}_0)) = \boldsymbol{\delta}'\boldsymbol{\nabla}'_{\boldsymbol{\eta}'}(1) = 0.
\end{aligned}$$

Since $B'\boldsymbol{\mu} = 0$ and $\boldsymbol{\mu}'\boldsymbol{e} = 0$, we can conclude that $\boldsymbol{\mu}$ is in $\mathcal{C}(\Gamma(s^*, t^*; \boldsymbol{\theta}_0))$. Hence the result in the theorem follows from Lemma 6.2 and results about multivariate normal distributions. $\square$

7. COMPARISON OF TEST STATISTICS FOR RECURRENT EVENT DATA

7.1. INTRODUCTION

In Section 5 and Section 6, we considered the problem of goodness of fit test for testing the null hypothesis

$$H_0 : \boldsymbol{\eta}_n = \boldsymbol{\eta}_0 \quad \text{versus} \quad H_{1n} : \boldsymbol{\eta}_n \neq \boldsymbol{\eta}_0.$$

An important and interesting problem with data-dependent chi-square test is the optimal choice for the matrix $\Xi(s^*, t^*; \boldsymbol{\theta}_0)$ that appears in $\mathbf{V}_n(s, t; \boldsymbol{\theta})$. One may conjecture that choosing $\Xi(s^*, t^*; \boldsymbol{\theta}_0)$ to be the square root of a generalized inverse of Σ will lead to some optimality properties of the test statistic. In this section, we investigate some optimality properties of test statistics obtained using different choices of $\Xi(s^*, t^*; \boldsymbol{\theta}_0)$ via the approximate Bahadur efficiencies [10], [11], [12] and the limiting asymptotic relative Pitman efficiencies [78].

7.2. SOME POSSIBLE CLASS OF TEST STATISTICS

We describe two possible classes of test statistics that can be obtained using two choices for $\Xi(s, t; \boldsymbol{\theta})$.

Example 7.1. *Let $\Xi(s^*, t^*; \boldsymbol{\theta}_0) = I_k$ be the identity matrix. Then $\Xi(s^*, t^*; \boldsymbol{\theta}_0)$ satisfies Conditions 1 and 2 in Section 5. The test statistic $\bar{Q}_1(s^*, t^*)$ is formed with $\boldsymbol{V}_n(s^*, t^*; \boldsymbol{\theta}) = \boldsymbol{U}_n(s^*, t^*; \boldsymbol{\theta})$. The resulting test statistic is given by*

$$\bar{Q}_1(s^*, t^*) = \boldsymbol{V}'(s^*, t^*; \bar{\boldsymbol{\theta}}_n) \hat{\Gamma}_1^{-}(s^*, t^*; \bar{\boldsymbol{\theta}}_n) \boldsymbol{V}(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$$

where $\Gamma_1(s^*, t^*; \boldsymbol{\theta}_0) = \boldsymbol{A}(s^*, t^*; \boldsymbol{\theta}_0) \Sigma \boldsymbol{A}(s^*, t^*; \boldsymbol{\theta}_0)$. $\square$

Example 7.2. *Let* $\Xi(s^*,t^*;\boldsymbol{\theta}_0) = \Xi_1(s^*,t^*;\boldsymbol{\theta}_0) = diag(p_1^{-1/2}(\boldsymbol{\theta}),\cdots,p_k^{-1/2}(\boldsymbol{\theta}))$. *Then* $\Xi_1(s^*,t^*;\boldsymbol{\theta}_0)$ *satisfies Conditions 1 and 2. The test statistic* $\bar{Q}_2(s^*,t^*)$ *is formed with*

$$\begin{aligned} \boldsymbol{V}_n(s^*,t^*;\boldsymbol{\theta}) &= \Xi_1(s^*,t^*;\boldsymbol{\theta})\,\boldsymbol{U}_n(s^*,t^*;\boldsymbol{\theta}) \\ &= \sqrt{n}\,\Xi_1(s^*,t^*;\boldsymbol{\theta})(\hat{\boldsymbol{p}}(s^*) - \boldsymbol{p}^{(n)}(\boldsymbol{\theta})) \\ &= \Big(\frac{n\,\hat{p}_1(s^*) - n\,p_1^{(n)}(\boldsymbol{\theta})}{(n\,p_1^{(n)}(\boldsymbol{\theta}))^{1/2}},\cdots,\frac{n\,\hat{p}_k(s^*) - n\,p_k^{(n)}(\boldsymbol{\theta})}{(n\,p_k^{(n)}(\boldsymbol{\theta}))^{1/2}}\Big) \end{aligned}$$

which is the vector used in the classical i.i.d Pearson-Fisher setting. The resulting test statistic is given by

$$\bar{Q}_2(s^*,t^*) = \boldsymbol{V}'(s^*,t^*;\bar{\boldsymbol{\theta}}_n)\hat{\Gamma}_2^-(s^*,t^*;\bar{\boldsymbol{\theta}}_n)\,\boldsymbol{V}(s^*,t^*;\bar{\boldsymbol{\theta}}_n)$$

where

$$\Gamma_2(s^*,t^*;\boldsymbol{\theta}_0) = \boldsymbol{A}(s^*,t^*;\boldsymbol{\theta}_0)\Xi_1(s^*,t^*;\boldsymbol{\theta}_0)\Sigma\Xi_1(s^*,t^*;\boldsymbol{\theta}_0)\boldsymbol{A}(s^*,t^*;\boldsymbol{\theta}_0). \quad \square$$

The main advantage of the above examples is the simplicity. We need only minimize a simple quadratic form to obtain an estimator for $\boldsymbol{\theta}$.

As explained earlier, these three test statistics differ only by the matrix $\Xi(s^*,t^*;\boldsymbol{\theta}_0)$. In next section, we shall compare the asymptotic performance of test statistics $\bar{Q}(s^*,t^*)$, $\bar{Q}_1(s^*,t^*)$ and $\bar{Q}_2(s^*,t^*)$ by using the approximate Bahadur efficiencies and the limiting asymptotic relative Pitman efficiencies. Hence we investigate the optimal choice of this matrix with respect to the relative efficiencies.

7.3. COMPARISON OF TEST STATISTICS

When two or more test statistics are available to test the same hypothesis, we are faced with the problem of deciding which one to use. To make proper decision, one can use asymptotic relative efficiency of two sequences of tests. There are various concepts of asymptotic relative efficiencies for comparing the performance of two sequences of statistical tests for a given hypothesis testing problem. In this section, we describe the concept of the approximate Bahadur asymptotic relative efficiency, the limiting asymptotic Pitman relative efficiency, and the result of Wieand [78] that specifies condition under which the limit (as the alternative approaches the null hypothesis) of the Bahadur efficiency coincides with the limit (as the level α tends to zero) of the Pitman efficiency. First, we begin with the approximate Bahadur asymptotic relative efficiency.

7.3.1. Bahadur Efficiencies. The concept of the approximate Bahadur asymptotic relative efficiency (ARBE) was first introduced by Bahadur [10], [11], [12] and further analyzed by various authors. We briefly review the concept following the expositions given in Bahadur [10]. Suppose that there is a set of probability measures $\{P_\eta : \eta \in \Omega\}$ defined on a sample space $(X, \mathcal{B})$. Let H_0 be the hypothesis that $\eta \in \Omega_0$ where Ω_0 is a subset of Ω. Let $\{T_n\}$ be a sequence of real valued statistics, such that large values of T_n are significant, based on a random sample of size n, defined on $(X, \mathcal{B})$. Bahadur [10] defined $\{T_n\}$ to be a standard sequence if the following three conditions are met:

B1. There exists a continuous probability distribution function F such that, for each $\eta \in \Omega_0$ and $x \in \mathbb{R}$, $\lim_{n\to\infty} P_\eta(T_n < x) = F(x)$.

B2. There exists a constant a, $0 < a < \infty$, such that

$$\log\big(1 - F(x)\big) = -\tfrac{ax^2}{2}\big(1 + o(1)\big)$$

where $o(1) \to 0$ as $n \to \infty$.

B3. There exists a real valued function $b(\eta)$ on $\Omega - \Omega_0$ with $0 < b(\eta) < \infty$ such that, for each $\eta \in \Omega - \Omega_0$

$$\lim_{n\to\infty}\ P_\eta\left(\left|\tfrac{T_n}{\sqrt{n}} - b(\eta)\right| > x\right) = 0$$

for every $x > 0$.

Now suppose $\{T_n\}$ is a standard sequence. Then T_n has the asymptotic distribution F if H_0 is true, but otherwise $T_n \to \infty$ in probability. Consequently, large values of T_n are significant for testing H_0. Accordingly, for any given x, we define $1 - F(T_n(x))$ to be the level attained by T_n in the given case $n = 1, 2, \ldots$. In general, $1 - F(T_n(x))$ is only an approximate level, i.e., for given n and α, it does not equal the probability of T_n being as large as, or larger than $T_n(x)$ when H_0 is true.

In order to compare two tests based on standard sequences, Bahadur normalizes the tests so that they both have the same limiting distribution under H_0. For any standard sequence, he defines $K_n = -2\log(1 - F(T_n))$. The asymptotic distribution of K_n is given by

$$\begin{aligned} P(K_n \leq v) &= P\big(-\log(1 - F(T_n)) \leq v\big) \\ &= P\big(F(T_n) \leq 1 - e^{-v/2}\big) \\ &= 1 - e^{-v/2} \quad \text{for every } \ v > 0. \end{aligned}$$

Therefore, K_n is asymptotically distributed as χ^2_2 distribution with 2 degrees of freedom. He also showed that K_n/n converges to $ab^2(\eta)$ in probability for $\eta \in \Omega - \Omega_0$. The asymptotic or approximate slope of the sequence $\{T_n\}$ is $c(\eta) = ab^2(\eta)$ and the approximate efficiency of two standard sequences $\{T_n^{(1)}\}$ to $\{T_n^{(2)}\}$ is defined to be $e^B_{12}(\eta) = c_1(\eta)/c_2(\eta)$. Bahadur argued that the test based on $T_n^{(1)}$ is asymptotically less efficient than that based on $T_n^{(2)}$ if $c_1(\eta)/c_2(\eta) < 1$ and asymptotically more efficient if $c_1(\eta)/c_2(\eta) > 1$. The ratio is thus called the approximate Bahadur asymptotic relative efficiency (asymptotic Bahdur ARE) of the sequence $T_n^{(1)}$ relative to the sequence $T_n^{(2)}$. The theory of approximate slopes and related concepts is discussed more extensively in Bahadur [10], [11], [12].

We apply the method described in Bahdur [10] for the determination of the approximate slopes of the goodness of fit statistics by showing that the sequence of the square root of the test statistics are standard sequences.

7.3.2. Comparison of Test Statistics. Let us consider the model introduced in Section 6.1. Recall that our goodness of fit testing problem is equivalent to testing $H_0 : \boldsymbol{\eta} = \boldsymbol{\eta}_0$. $\bar{Q}(s^*, t^*) = \mathbf{V}'_n(s^*, t^*; \bar{\boldsymbol{\theta}}_n)\hat{\Gamma}^-(s^*, t^*; \bar{\boldsymbol{\theta}}_n)\mathbf{V}_n(s^*, t^*; \bar{\boldsymbol{\theta}}_n)$ was the statistic used back there. Let $T_{1n} = \sqrt{\bar{Q}(s^*, t^*)}$. Since the square root function is one-to-one and continuous function for nonnegative real numbers, one can easily show using transformed methods that the square root of χ^2 random variable with $k - q - 1$ degrees of freedom is distributed as a chi-distribution with the same degrees of freedom. Therefore T_{1n} converges in distribution to χ_{k-q-1}. Hence, there exists a continuous function $F(x)$ such that $\lim_{n\to\infty} P_{\boldsymbol{\eta}}(T_{1n} < x) = F(x)$ for all $x > 0$ and $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. Therefore, T_{1n} satisfies condition $B1$. The distribution function of χ^2_{k-q-1} statistic satisfies $B2$ with $a = 1$. To observe this, for each $x \in \mathbb{R}$ and $\eta = \eta_0$, notice that

$$P_{\boldsymbol{\eta}_0}(T_{1n} \le x) = P_{\boldsymbol{\eta}_0}(\bar{Q}(s^*, t) \le x^2) = F(x) \text{ from condition } B1.$$

To obtain condition $B2$, consider

$$\begin{aligned} 1 - F(x) &= \int_{x^2}^{\infty} \left(2^{r/2}\,\Gamma(r/2)\right)^{-1} e^{-z/2}\, z^{\frac{r-2}{2}}\, dz \text{ where } r = k - q - 1 \\ &= \left(2^{r/2}\,\Gamma(r/2)\right)^{-1} \int_{x^2}^{\infty} e^{-z/2}\, z^{\frac{r-2}{2}}\, dz. \end{aligned}$$

Using integration by parts, we write

$$1 - F(x) = \left(2^{r/2}\,\Gamma(r/2)\right)^{-1} \left[2\, e^{-x^2/2}\, x^{r-2} + 2 \int_{x^2}^{\infty} e^{-z/2} \left(\frac{r-2}{2}\right) z^{(\frac{r-2}{2})-1}\, dz\right].$$

Let $w = z/2$. Then the last equation reduces to

$$1 - F(x) = \left(2^{r/2}\,\Gamma(r/2)\right)^{-1} 2 \left[e^{-x^2/2}\, x^{r-2} + 2^{(\frac{r-2}{2})}\,(r-2) \int_{x^2/2}^{\infty} e^{-w}\, w^{(\frac{r-2}{2})-1}\, dw\right].$$

Using the following relationship

$$\int_{x^2/2}^{\infty} e^{-w}\, w^{\left(\frac{r-2}{2}\right)-1}\, dw = O\Big(e^{-x^2/2} \left(\frac{x^2}{2}\right)^{\left(\frac{r-2}{2}\right)-1}\Big) \text{ as } x \to \infty,$$

we get

$$1 - F(x) = \left(2^{r/2}\,\Gamma(r/2)\right)^{-1} 2 \left[e^{-x^2/2}\, x^{r-2} + 2^{(\frac{r-2}{2})}\,(r-2)\, O\Big(e^{-x^2/2} \left(\frac{x^2}{2}\right)^{\left(\frac{r-2}{2}\right)-1}\Big)\right]$$

Algebraic manipulations reduce the previous step to

$$\begin{aligned} 1 - F(x) &= e^{-x^2/2}\, x^{r-2} \Big[2 \big(2^{r/2}\, \Gamma(r/2) \big)^{-1} + O(x^{-2}) \Big] \\ &= e^{-x^2/2}\, x^{r-2}\, 2\, \big(2^{r/2}\, \Gamma(r/2) \big)^{-1} \big(1 + o(1) \big) \ \text{ as } \ x \to \infty. \end{aligned}$$

Hence

$$\log \big(1 - F(x)\big) = -\frac{x^2}{2} + (r-2)\, \log x + \log C_1 + o(1) \ \text{ as } \ x \to \infty,$$

where $C_1 = 2\, \big(2^{r/2}\, \Gamma(r/2)^{-1}\big)$ and we have

$$\log(1 - F(x)) = -\frac{x^2}{2} \left[1 - \frac{2\, (r-2)\, \log x}{x^2} + o(1) \right] \ \text{ as } \ x \to \infty.$$

So the last equation finally reduces to the desired form

$$\log\big(1 - F(x)\big) = -\frac{x^2}{2}\, \big[1 + o(1)\big] \ \text{ as } \ x \to \infty.$$

So condition $B2$ is satisfied by T_{1n} with $a = 1$. Since $\mathbf{V}_n(s^*, t^*; \bar{\boldsymbol{\theta}}_n) \overset{as}{\to} \Xi(s^*, t^*; \boldsymbol{\theta}_0)(\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))$ and $\hat{\Gamma}^{-}(s^*, t^*; \boldsymbol{\theta}_n) \overset{as}{\to} \Gamma^{-}(s^*, t^*; \boldsymbol{\theta}_0)$ as $n \to \infty$, T_{1n} also satisfied condition $B3$ with

$$\begin{aligned} b(\boldsymbol{\theta}_0, \boldsymbol{\eta}) &= [(\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))' \Xi(s^*, t^*; \boldsymbol{\theta}_0) \Gamma^{-}(s^*, t^*; \boldsymbol{\theta}_0) \Xi(s^*, t^*; \boldsymbol{\theta}_0) \cdot \\ &\quad (\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))]^{\frac{1}{2}} \end{aligned}$$

for $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$. Therefore $\{T_{1n}\}$ is a standard sequence. So the approximate Bahadur slope $b^2(\boldsymbol{\theta}_0, \boldsymbol{\eta})$ of $\bar{Q}(s^*, t^*)$ is

$$\begin{aligned} b^2(\boldsymbol{\theta}_0, \boldsymbol{\eta}) &= [(\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))' \Xi(s^*, t^*; \boldsymbol{\theta}_0) \Gamma^{-}(s^*, t^*; \boldsymbol{\theta}_0) \Xi(s^*, t^*; \boldsymbol{\theta}_0) \cdot \\ &\quad (\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))]. \end{aligned}$$

Likewise, the approximate Bahadur slopes of $\bar{Q}_1(s^*, t^*)$ and $\bar{Q}_2(s^*, t^*)$ are given by

$$b_1^2(\boldsymbol{\theta}_0, \boldsymbol{\eta}) = [(\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))'\Gamma_1^-(s^*, t^*; \boldsymbol{\theta}_0)(\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))]$$

and

$$\begin{aligned} b_2^2(\boldsymbol{\theta}_0, \boldsymbol{\eta}) &= [(\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))'\Xi_1(s^*, t^*; \boldsymbol{\theta}_0)\Gamma_2^-(s^*, t^*; \boldsymbol{\theta}_0)\Xi_1(s^*, t^*; \boldsymbol{\theta}_0) \cdot \\ &\quad (\mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) - \mathbf{p}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0))] \end{aligned}$$

respectively. We can now obtain the desired the approximate Bahadur efficiencies.

Theorem 7.3. *For $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$, the approximate Bahadur efficiency of $\bar{Q}(s^*, t^*)$ to $\bar{Q}_1(s^*, t^*)$ is*

$$e^B(\bar{Q}, \bar{Q}_1) = \frac{b^2(\boldsymbol{\theta}_0, \boldsymbol{\eta})}{b_1^2(\boldsymbol{\theta}_0, \boldsymbol{\eta})}.$$

The approximate Bahadur efficiency of $\bar{Q}(s^*, t^*)$ to $\bar{Q}_2(s^*, t^*)$ is

$$e^B(\bar{Q}, \bar{Q}_2) = \frac{b^2(\boldsymbol{\theta}_0, \boldsymbol{\eta})}{b_2^2(\boldsymbol{\theta}_0, \boldsymbol{\eta})}.$$

Thus, if two sequences of test statistics $\{\bar{Q}(s^*, t^*)\}$ and $\{\bar{Q}'(s^*, t^*)\}$ are such that conditions B1, B2, and B3 hold, then their Bahadur ARE $e^B(\bar{Q}, \bar{Q}')$ can be calculated, acoording to Theorem 6.6. If $e^B(\bar{Q}, \bar{Q}') > 1$ for some $\boldsymbol{\eta}$ then we should prefer the sequence $\{\bar{Q}(s^*, t^*)\}$ to $\{\bar{Q}'(s^*, t^*)\}$.

7.3.3. Connection to Pitman Efficiency. Since we are using the approximate Bahadur slope for comparison of the tests, it is natural to see whether the limiting asymptotic relative Pitman efficiency (ARPE) can be computed from the approximate Bahadur slopes. For the definition of Pitman efficiency we will use that of Wieand [78]. Let $T_n^{(1)}$ and $T_n^{(2)}$ be two sequences of statistics used to form tests of size α for testing $H_0 : \eta = \eta_0$ versus $H_1 : \eta = \eta_j$ where $\eta_j \neq \eta_0$.

For $0 < \beta < 1$ and sequences $\eta_j \to \eta_0$, $\beta_j^{(1)} \to \beta$, and $\beta_j^{(2)} \to \beta$, define $N(i, j)$ to be the smallest integer so that for every $N \geq N(i,j)$

$$P_{\eta_j}\left(T_N^{(i)} > t_N^{(i)}\right) \geq \beta_j^{(i)}, \quad i = 1, 2, \;\; j = 1, 2, \ldots$$

where $t_N^{(i)}$ is determined by $P_{\eta_0}\left(T_N^{(i)} > t_N^{(i)}\right) = \alpha, \quad i = 1, 2,$ and $j = 1, 2, \ldots$. We define the Pitman efficiency of sequences $T_n^{(1)}$ with respect to $T_n^{(2)}$ by

$$e_{12}(\alpha, \beta) = \lim_{n \to \infty} \frac{N(2,j)}{N(1,j)}$$

provided that the limit exists and is independent of the choice of the sequences η_j and $\beta_j^{(i)}$. If this is not the case, let

$$e_{12}^{-}(\alpha, \beta) = \inf_{\Pi} \liminf_{j \to \infty} \frac{N(2,j)}{N(1,j)} \text{ and } e_{12}^{+}(\alpha, \beta) = \sup_{\Pi} \limsup_{n \to \infty} \frac{N(2,j)}{N(1,j)}.$$

Here $\sup_{\Pi}$ ($\inf_{\Pi}$) represents the sup (inf) over all sequences $\{\eta_j\}$, $\{\beta_j^{(1)}\}$, and $\{\beta_j^{(1)}\}$, $i = 1, 2$ respectively and Π stands for any of above sequences.

Bahadur emphasizes that the most important property of a slope is its value in the immediate vicinity of the null hypothesis. Because the approximate slope and the exact slope of a test sequence typically coincide in a neighborhood of the null parameter, the main conclusions relevant to power considerations available from exact slope (cf, Bahadur [11]) also apply to approximate slope. Bahadur showed that for one-sided testing problems, the limiting approximate Bahadur efficiency of two asymptotically normal test sequences as the alternative parameter converges to the null value coincides with this Pitman efficiency as the alpha (α) level approximates zero. Wieand [78] has generalized Bahadur's remark to include test sequences with asymptotic distribution other than normal, and those used in two-sided testing problems. Further, it can be shown that Wieand's condition holds for the version of the goodness of fit statistic with estimated parameters given that it is satisfied by simple statistic, since limiting distributions do not depend on values of the parameters. Furthermore rates of convergence remain unaltered with asymptotically efficient estimators (See Koziol [45]). Wieand's condition is given below.

Wieand's (1976) Condition. Suppose for a standard sequence $\{T_n\}$ there exists an $\eta^* > 0$ such that for every $\epsilon > 0$ and $\delta \in (0,1)$, there exists a C such that for all $\eta \in I(\eta_0, \eta^*) = (\eta_0 - \eta^*, \eta_0 + \eta^*) \setminus \eta_0$ and $N > (C/b^2(\eta))$ we have

$$P_\eta\{|T_N/N^{\frac{1}{2}} - b(\eta)| < \epsilon b(\eta)\} > 1 - \delta.$$

Then T_N is said to satisfy Wieand's condition.

Remark 6.1

Note that in the above condition, C may depend on η^* but is otherwise independent of η. Wieand condition is somewhat stronger than Bahadur's condition $B3$, since it requires the convergence of $T_n/\sqrt{n}$ in probability at a specified rate. As a consequence, if Wieand's (1976) condition is verified for a standard sequence $\{T_n\}$ and given function $b(\eta)$, Bahadur's condition $B3$ is also satisfied.

Under conditions $B1$, $B2$ and Wieand (1976) condition with $b(\eta) \to 0$ as $\eta \to \eta_0$, Wieand showed that the limiting approximate Bahadur efficiency as $\eta \to \eta_0$ equals to the limiting asymptotic relative Pitman efficiency as the level of the test (α) approaches 0. [See Wieand [78]]. Oftentimes, the verification of Wieand (1976) condition is not straightforward, because in order to establish it, it is necessary to study the behavior of the test statistics under H_{1n}, and the knowledge of this behavior is often limited. The following lemma from Wieand [78] sometimes facilitates the verification of Wieand (1976) condition.

Lemma 7.4. *Suppose there is a family of sequences of statistics $U_n(\eta)$, which satisfies $P_\eta(U_n(\eta) < x) = Q(x)$ for every real number x where $Q(x)$ is a continuous distribution function and where the rate of convergence is independent of η in some neighborhood $N(\eta_0, \eta')$ of η_0. Then given any $\epsilon_1 > 0$ and $\delta_1 \in (0,1)$, there is a C' such that if $\eta \in N(\eta_0, \eta')$, $b(\eta) < 1$ and $N > C'/b^2(\eta)$, then $P_\eta(|U_n(\eta)/\sqrt{n}| < \epsilon_1 b(\eta)) > 1 - \delta_1$.*

BIBLIOGRAPHY

[1] Aalen, O. (1982), “Practical applications of the nonparametric statistical theory for counting processes,” *Statistical Research Report. Institute of Mathematics,*, University of Oslo.

[2] Aalen, O. (1978), ‘”Nonparametric inference for a family of counting processes,” *The Annals of Statistics,*, 6, 701-726.

[3] Adekpedjou, A. (2007), ”Estimation of gap-time distribution with recurrent event data under informative monitoring period” *PhD dissertation, Dept. Statistics, University of South Carolina-Columbia.*

[4] Adekpedjou, A. and Zamba, K. D. (2012), ”A chi-squared goodness of test for recurrent event data”, *Journal of Statistical Theory and Applications*, 11, 97- 119.

[5] Adekpedjou, A. and Peña, E. A. and Quiton, J, (2010), “Estimation and efficiency with recurrent event data under informative monitoring,” *Journal of Statistical Planning and Inference*, 140, 597-615.

[6] Adekpedjou, A. and Peña, E. A. (2012), “Semiparametric estimation with recurrent event data under informative monitoring,” *Journal of Nonparametric Statistics*, 24, 733-752.

[7] Agustin, Ma. Zenia N. and Peña, Edsel A. (2001), ”Goodness-of-fit of the distribution of time-to-first-occurrence in recurrent event models”, *Lifetime Data Annals*, 7, 289-306.

[8] Agustin, Ma. Zenia N. and Peña, Edsel A. (2005), ”A basic approach to goodness-of-fit testing in recurrent event models”, *Journal of Statistical Planning and Inference*, 133, 285-303.

[9] Akritas, Michael G. (1988), ”Pearson-type goodness-of-fit tests: the univariate case”, *Journal of the American Statistical Association*, 83, 222-230.

[10] Bahadur, R. R. (1960), ”Stochastic comparisons of tests,” *The Annals of Mathematical Statistics,*, 31, 276-295.

[11] Bahadur, R. R. (1967), ”Rates of convergence of estimates and test statistics,” *The Annals of Mathematical Statistics,*, 38, 303-324.

[12] Bahadur, R. R. (1971), ”Some limits theorems in statistics,” *SIMA, Philadelphia.*

[13] Beran, R. (1977), ”Minimum Hellinger distance estimates for parametric models”, *The Annals of Statistics*, 5, 462-463.

[14] Berkson, J. (1980), ”Minimum chi-square, not maximum likelihood!”, *The Annals of Statistics*, 8, 457-487.

[15] Billingsley, P. (1999), "Convergence of Probability Measures, second edition", *John Wiley & Sons Inc., New York*.

[16] Block, H.W., Borges, W.S., and Savits (1985), "Age-dependent minimal repair", *Journal of Applied, Probability*, 22, 370-385.

[17] Breslow, N. and Crowley, J. (1974), "A large sample study of the life table and product limit estimates under random censorship", *The Annals of Statistics*, 2, 437-453.

[18] Brémaud, P. (1981), "Point processes and queues", *Springer-Verlag, New York*, 84, 1065-1073.

[19] Byar, D. P. (1980), "The veterans administration study at cehmoprophylaxis for recurrent stage i bladder tumors comparisons of placebo, pyridoxice, and topical thiotepa in: Pavones-Macaluso. M. Smith, P. H. (Eds)," *Bladder Tumors and Other Topics in Urological Oncology Plenum, New York*, pp. 363-370.

[20] Chen, J, (1975), "Goodness of fit tests under random censorship" *PhD dissertation, Dept. Statistics, Oregon State University*.

[21] Chen, Y. Y, and Hollander, M. and Langberg, N. A. (1982), "Small-sample results for the Kaplan-Meier estimator," *Journal of the American Statistical Association*, 77, 141-144.

[22] Cheng, P., and Lin, G. D. (1987), "Maximum likelihood estimator of a survival function under the Koziol-Green proportional hazards model," *Statistics and Probability Letters*, 5, 75-80.

[23] Chernoff, H. and Lehmann, E.L.(1954), "The use of maximum lkelihood estimates in Chi-square tests for goodness of fit", *The Annals of Statistics*, 25, 579-586.

[24] Chung, K.L. and Wiliams, R. J. (1990) "Introduction to stochastic integration," *Probability and its applications,*, Boston, MA, second edition.

[25] Dahiya, Ram C. and Gurland, John (1972), "Pearson chi-squared test of fit with random intervals", *Biometrika*, 59, 147-153.

[26] Efron, B. (1967), "The two sample problem with censored data," *in Proceedings of the Fifth Berkely Symposium on Mathematical Statistics and Probability, IV, Berkeley, CA, University of California Press*, 831-853.

[27] Fisher, R.A. (1922), "On the interpretation of chi-square from contingency tables, and the calculation of P", *Journal of the Royal Statistical Society*, 85, 87-94.

[28] Fisher, R.A. (1924), "The condition under which chi-square measures the discrepancey between observation and hypothesis", *Journal of the Royal Statistical Society*, 87, 442-450.

[29] Fleming, Thomas R. and Harrington, David P. (1991), "Counting processes and survival analysis", *John Wiley & Sons Inc., New York*.

[30] Gaudoin, O. and Yang, B. and Xie, M. (2006), "Confidence intervals for the scale parameter of the power-law process", *Communications in Statistics-Theory and Methods*, 35, 1525-1538.

[31] Gill, R. (1983), "Large sample behaviour of the product-limit estimator on the hwole line", *The Annals of Statistics*, 11, 49-58.

[32] Graybill, Franklin A. (1976), "Theory and application of the linear model", *Duxbury Press, North Scituate, Mass.*

[33] Habib, M. G. and Thomas, D. R. (1986), "Chi-square goodness-of-fit tests for randomly censored data", *Ann. Statist.*, 14, 759- 765.

[34] Hamdan, M. A. (1963), "The number and width of classes in chi square test", *Journal of the American Statistical Association*, 58(303), 678- 689.

[35] Harris, R. R., and Kanji, G. K (1983), "On the Use of Minimum Chi-Square Estimation", *Journal of the Royal Statistical Society*, 32, 379-394.

[36] Henze, N. (1983), "A quick omnibus test for the proportional hazards model of random censorship," *Statistics*, 24, 253-263.

[37] Herbst, T. (1992), "Test of fit with the Koziol-Green model for random censorship," *Statistics and Risk Modeling*, 10, 163-171.

[38] Hjort, Nils Lid (1900), "Goodness of fit tests in models for life history data based on cumulative hazard rates", *The Annals of Statistics*, 18, 1221- 1258.

[39] Hollander, Myles and Peña, Edsel A. (1992), "A chi-squared goodness-of-fit test for randomly censored data", *Journal of the American Statistical Association*, 87, 458-463.

[40] Kaplan, E.L. and Meier, P.(1958), "Nonparametric estimation from incomplete observations", *Journal of the American Statistical Association*, 53, 457-481.

[41] Kay, R. (1984), "Goodness of fit methods for the proportional hazards regression model: a review," *Rev. Epidemiol. Sante Publique*, 32, 185-198.

[42] Khmaladze, E.V. (1988), "An innovation approach to goodness-of-fit tests in $\mathbb{R}$", *The Annals of Statistics*, 16, 1503-1516.

[43] Kim, Joo Han (1993), "Chi-square goodness-of-fit tests for randomly censored data", *The Annals of Statistics*, 21, 1621- 1639.

[44] Koziol, J.A., and Green, S.B. (1976), "A Cramér-von Mises statistic for randomly censored data", *Biometrika*, 63, 465-474.

[45] Koziol, J. A. (1986), "Relative efficiencies of goodness of fit procedures for assessing univariate normality", *Ann, Inst. Statis. Math.*, 38, Part A, 485-493.

[46] Krimani, S. and Dauxois, J. Y. (2004), "Testing the Koziol-Green model against monotone conditional odds for censoring," *Statistics and Probability Letters*, 66, 327-334.

[47] Li, Gang and Doss, Hani (1993), "Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data", *The Annals of Statistics*, 21, 772-797.

[48] Link, A. (1989), "A model for informative censoring," *Journal of the American Statistical Association*, 84, 749-752.

[49] Mann, H. B. and Wald, A. (1943), "On Stochastic limit and order realtionships", *The Annals of Statistics*, 14, 217- 226.

[50] Mann, H. B. and Wald, A. (1942), "On the choice of the number of class intervals in the application of the chi square test", *The Annals of Statistics*, 13(3), 306- 317.

[51] Mihalko, Daniel P. and Moore, David S. (1980), "Chi-square tests of fit for type II censored data", *The Annals of Statistics*, 3, 625-644.

[52] Moore, David S. and Spruill, M. C. (1975), "Unified large-sample theory of general chi-squared statistics for tests of fit", *The Annals of Statistics*, 3, 599-616.

[53] Moore, D. S. (1971), "A chi-square statistic with random cell boundaries", *The Annals of Statistics*, 42, 147-156.

[54] Moore, D. (1976), "Chi-square tests in studies in satistics, ed. R,V. Hogg", *Washington DC: Mathematical Association of America.*

[55] Nelder, J. A. and Mead, R (1965), "A simplex method for function minimization", *The Computer Journal*, 7, 308-313.

[56] Nelson, W. (1969), "Hazard plotting for incomplete failure data," *Technometrics*, 14, 27-52.

[57] Nelson, W. (1972), "Theory and application of hazard plotting for censored failure data," *Technometrics*, 14, 945-966.

[58] Park, W.J. and Kim, Y.G. (1992), "Goodness of fit tests for the power law process", *Reliability, IEEE transactions*, 41, 107-111.

[59] Pearson, K. (1900), "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably suported to have areisen from random sampling," *Phil. Mag.,*, 1, 157.

[60] Peña, E. A. and Strawderman, R. L. and Hollander, M.(2001), "Nonparametric estimation with recurrent event data", *Journal of the American Statistical Association*, 96, 1299- 1315.

[61] Peña, E. A. and Strawderman, R. L. and Hollander, M. (2000), "A weak convergence result relevant in recurrent and renewal models. In Recent advances in reliability theory (Bordeaux, 2000)", *Stat. Ind. Technol,*, 493- 514.

[62] Peña, E. A. (1998), "Smooth goodness of fit tests for composite hypothesis in hazard based models", *The Annals of Statistics*, 26(5), 1935- 1971.

[63] Pollard, D. (1979), "General Chi-square Goodness-of-fit Tests with Data-dependent Cells", *Z. Wahrscheinlichkeitheorie veiw. Gebiete*, 50, 317-331.

[64] Presnell, Brett and Hollander, Myles and Sethuraman, Jayaram (1994), "Testing the minimal repair assumption in an imperfect repair model", *Journal of the American Statistical Association*, 89, 289- 297.

[65] Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding (1993), "Statistical models based on counting processes", *Springer-Verlag, New York*.

[66] Proschan, F. (1963), "Theoretical explanation of observing decreasing failure rate", *Technomatrics*, 5, 375-383.

[67] Ruymgaart, F. H. (1975), "A note on chi-square statistics with random cell boundaries", *The Annals of Statistics*, 4, 965-968.

[68] Sellke, T. (1998), "Weak convergence of the Aalen estimator for a censored renewal process. In: Statistical Decision Theory and Related Topics," *Springer, New York*, 2, 183-194.

[69] Shorack, G. and Welliner, J. (1986), "Empirical processes with applications to statistics", *John Wiley & Sons Inc., New York*.

[70] Shorack, G., R. and Wellner, J. A. (1986), "Empirical processes with applications to statistics", *John Wiley & Sons Inc., New York*, 84, 1065-1073.

[71] Siannis, F. (2004), "Applications of a parametric model for informative censoring", *Biometrika*, 60, 704-714.

[72] Spiekerman, C. F. and Lin, D. Y.(1998), "Marginal regression models for multivariate failure time data", *Journal of the American Statistical Association*, 93, 1164-1175.

[73] Stocker, IV, Russell S. and Adekpedjou, Akim (2012), "Optimal goodness-of-fit tests for recurrent event data", *Lifetime Data Annals*, 17, 409- 432.

[74] Vander, Vaart and Gill, R. (1993), "Non and semiparametric maximum likelihood estimators and the Von Mises method-ii", *Scandinavian Journal of Statistics*, 20, 271-288.

[75] Walsh, J. (1962), "Percentiles Handbook at Nonparametric Statistics: Investigation of Randomness, Moments and Distributions. In Empricial Processes with Applications in Statistics," 1, Van Nostrand, Princeton, New Jersey.

[76] Wang, M. C. and Qin, J. and Chiang, C. (2001), "Analyzing recurrent event data with informative censoring," *Journal of the American Statistical Association*, 96, 1057-1065.

[77] Wei, L. J. and Lin, D. Y. and Weissfeld, L. (1989), "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions", *Journal of the American Statistical Association*, 84, 1065-1073.

[78] Wieand, H. S. (1976), "A condition under which the Pitman and Bahadur approaches to efficiency coinside", *The Annals od Statistics*, 5,

[79] Williams, A. C. (1950), "On the choice of the number of clases for the chi square test of goodness of fit", *Journal of the American Statistical Association*, 45(249), 77-86.

VITA

Withanage Ajith Raveendra De Mel attended Dharmasoka College, Ambalangoda, Sri Lanka. In September, 1996, he received his bachelor's degree in Mathematics from University of Kelaniya, Sri Lanka. In May, 2009, De Mel completed his MS in Statistics at Sam Houston State University, Texas, USA. In August, 2009, he joined Missouri University of Science and Technology to pursue his graduate studies and in May, 2012, he completed his second MS in Applied Mathematics with Statistics emphasis. At the same time he entered the PhD program, in the same discipline and in 2014, he completed his PhD in Applied Mathematics with Statistics emphasis.