The Annals of Statistics
1998, Vol. 26, No. 5, 1935–1971

SMOOTH GOODNESS-OF-FIT TESTS FOR COMPOSITE HYPOTHESIS IN HAZARD BASED MODELS[1]

By Edsel A. Peña

Bowling Green State University

Consider a counting process $\{N(t),\ t \in \mathscr{T}\}$ with compensator process $\{A(t),\ t \in \mathscr{T}\}$, where $A(t) = \int_0^t Y(s)\lambda_0(s)\,ds$, $\{Y(t),\ t \in \mathscr{T}\}$ is an observable predictable process, and $\lambda_0(\cdot)$ is an unknown hazard rate function. A general procedure for extending Neyman's smooth goodness-of-fit test for the composite null hypothesis H_0: $\lambda_0(\cdot) \in \mathscr{C} = \{\lambda_0(\cdot;\eta)\colon\ \eta \in \Gamma \subseteq \Re^q\}$ is proposed and developed. The extension is obtained by embedding $\mathscr{C}$ in the class $\mathscr{A}_k$ whose members are of the form $\lambda_0(\cdot;\eta)\exp\{\theta^t\psi(\cdot;\eta)\}$, $(\eta, \theta) \in \Gamma \times \Re^k$, where $\psi(\cdot;\eta) = (\psi_1(\cdot;\eta), \ldots, \psi_k(\cdot;\eta))^t$ is a vector of observable random processes satisfying certain regularity conditions. The tests are based on quadratic forms of the statistic $\int_0^\tau \psi(s;\hat{\eta})\,dM(s;\hat{\eta})$, where $M(t;\eta) = N(t) - \int_0^t Y(s)\lambda_0(s;\eta)\,ds$ and $\hat{\eta}$ is a restricted maximum likelihood estimator of η. Asymptotic properties of the test statistics are obtained under a sequence of local alternatives, and the asymptotic local powers of the tests are examined. The effect of estimating η by $\hat{\eta}$ is ascertained, and the problem of choosing the ψ-process is discussed. The procedure is illustrated by developing tests for testing that $\lambda_0(\cdot)$ belongs to (i) the class of constant hazard rates and (ii) the class of Weibull hazard rates, with particular emphasis on the random censorship model. Simulation results concerning the achieved levels and powers of the tests are presented, and the procedures are applied to three data sets that have been considered in the literature.

1. Introduction and setting. The problem of goodness-of-fit testing is one of the central themes of statistical theory and practice. In this paper we revisit the goodness-of-fit problem of testing a composite hypothesis by describing and implementing a new formulation for extending Neyman's (1937) smooth goodness-of-fit test to models which are specified through hazard or intensity functions. The new formulation is applicable even when the available data is incomplete because of censoring or truncation, hence the proposed goodness-of-fit procedures are applicable to survival models, dynamic reliability models, econometric models and other models which utilize the counting process framework. A related paper dealing with the Cox (1972) proportional hazards model is Peña (1998). Smooth goodness-of-fit tests [see Neyman (1937) or Rayner and Best (1989) for the classical formulation] have

Received August 1996; revised March 1998.

[1]Supported by Bowling Green State University Faculty Research Committee Basic Research Grant and DHHS/NIH/NIGMS Grant 1R01GM56182 to BGSU.

AMS 1991 *subject classifications*. Primary 62F03; secondary 62N05.

Key words and phrases. Censoring, counting process, generalized residual, local asymptotic relative efficiency, martingale central limit theorem, multiplicative intensity model, nuisance parameters and adaptiveness, Neyman's test, score test, test for exponentiality, test for Weibull.

the appealing property of having good power over a wide range of alternatives compared to other goodness-of-fit tests, to the extent that Rayner and Best (1990) implored practitioners to use a smooth test rather than other methods. Interest in this class of tests have recently increased as evidenced by the recent papers of Bickel and Ritov (1992), Eubank and Hart (1992), Fan (1996), Inglot, Kallenberg and Ledwina (1994), Kallenberg and Ledwina (1995), and Ledwina (1994).

In this paper we consider a univariate counting process $\mathbf{N} = \{N(t): t \in \mathscr{T} \equiv [0, \tau]\}$ defined on some filtered probability space $(\Omega, \mathscr{F}, \mathbf{P})$ with filtration $\mathbf{F} = \{\mathscr{F}_t: t \in \mathscr{T}\}$, together with an observable predictable process $\mathbf{Y} = \{Y(t): t \in \mathscr{T}\}$. The filtration $\mathbf{F}$ is usually the natural filtration $\mathbf{F}^N = \{\mathscr{F}_t^N: t \in \mathscr{T}\}$, where $\mathscr{F}_t^N$ is the σ-field generated by $\{(N(s), Y(s)): s \le t\}$ and $\mathscr{F}_0$, the latter containing all information at time zero. The statistical model of interest postulates that the compensator process of $\mathbf{N}$ is $\mathbf{A} = \{A(t): t \in \mathscr{T}\}$, where $A(t) = \int_0^t Y(w)\lambda(w)\,dw$ and $\lambda(\cdot)$ is some unknown hazard rate function. This is a special case of Aalen's (1978) multiplicative intensity model [see also Andersen, Borgan, Gill and Keiding (1993) and Andersen and Gill (1982)], and it subsumes many incomplete data models in survival analysis and reliability such as the random censorship model, Type II censorship model and left-truncation models. The goodness-of-fit problem is to test the null hypothesis H_0 which states that $\lambda(\cdot)$ belongs to a parametric class $\mathscr{C} = \{\lambda_0(\cdot; \eta): \eta \in \Gamma \subseteq \Re^q\}$ of hazard rate functions (e.g., the Weibull class), versus the alternative hypothesis H_1 which states that $\lambda(\cdot)$ does not belong to $\mathscr{C}$. The goodness-of-fit test is to be based on a realization of $(\mathbf{N}, \mathbf{Y}) = \{(N(t), Y(t)): t \in \mathscr{T}\}$.

If $N(t) = \sum_{i=1}^n I\{T_i \le t\}$ and $Y(t) = n - N(t-)$, where $I\{\cdot\}$ denotes the indicator function and $T_1, \dots, T_n$ are independent and identically distributed (i.i.d.) random variables with common distribution function $F(\cdot)$, then this reduces to the classical problem of testing that $F(\cdot)$ belongs to the parametric class of distribution functions given by $\{F_0(\cdot; \eta): \eta \in \Gamma\}$ where $F_0(\cdot; \eta) = 1 - \exp\{-\int_{-\infty}^{\cdot} \lambda_0(w; \eta)\,dw\}$. This problem has been the subject of many important papers, notably those of Chernoff and Lehmann (1954), Khmaladze (1981, 1993), Rao and Robson (1974) and many others. We refer the reader to Stephens (1992) for a historical review of this problem in the classical setting.

However, whereas in this classical setting the only nuisance parameter is η, in survival and reliability models where incomplete data is prevalent, other nuisance parameters associated with the mechanisms causing the incompleteness are also present in the model. Furthermore, when dealing with failure time data, the counting process framework affords more generality in terms of models and results [cf. Aalen (1975), Andersen, Borgan, Gill and Keiding (1993), Andersen and Gill (1982) and Fleming and Harrington (1991)]. Several papers dealing with the goodness-of-fit problem in the presence of censored or truncated data have appeared in the literature. Among those which consider the composite null hypothesis case are Akritas (1988), Gray and Pierce (1985), Habib and Thomas (1986), Hjort (1990), Khmaladze (1981, 1993), Kim (1993), and Li and Doss (1993).

The class of goodness-of-fit tests described in this paper extends Neyman's (1937) idea, but in a formulation more suitable and natural in the context of models specified through hazard functions. Neyman's smooth test has been extended to right-censored data by Gray and Pierce (1985), but as in Neyman (1937) and Thomas and Pierce (1979), their formulation utilized density functions. In contrast, ours will be anchored on the hazard rate functions $\lambda(\cdot)$. Although there is a bijection between density and hazard rate functions, we will see that the hazard-based formulation leads to more generality. The extension is obtained by embedding the class $\mathscr{C}$ in a larger family parametrized by a smoothing parameter k and (θ, η), and defined via

$$(1.1) \qquad \mathscr{A}_k = \left\{\lambda_k(\cdot\,; \theta, \eta) = \lambda_0(\cdot\,; \eta)\exp\{\theta^{t}\psi(\cdot\,; \eta)\}: \theta \in \Re^k\right\},$$

where ^{t} denotes vector–matrix transpose, $\theta = (\theta_1, \ldots, \theta_k)^{t}$, and $\psi(\cdot; \eta) = (\psi_1(\cdot; \eta), \ldots, \psi_k(\cdot; \eta))^{t}$ is a vector of (possibly random) processes, for which a precise set of properties will be given later. Allowing $\psi(\cdot; \eta)$ to be random leads to more generality and enables the extension to be applicable in general dynamic models in reliability and in other areas. Note that $\mathscr{A}_k$ reduces to $\mathscr{C}$ when $\theta = 0$. With this embedding of $\mathscr{C}$ into $\mathscr{A}_k$, we propose to develop the goodness-of-fit procedure as a test for the hypotheses

$$(1.2) \qquad H_0: \theta = 0,\ \eta \in \Gamma \quad \text{versus } H_1: \theta \neq 0,\ \eta \in \Gamma.$$

It should be pointed out that aside from η, there are other nuisance parameters, associated with the incompleteness mechanism and possibly infinite-dimensional, which are also left unspecified. One of the benefits of our formulation and our adoption of the stochastic process framework is that of being able to bypass direct estimation of these other nuisance parameters.

We outline the contents of this paper. Section 2 presents the development of the smooth goodness-of-fit test. The asymptotic properties of the test statistics are presented in Section 3, but their proofs are deferred to an Appendix. In particular, we present the limiting distribution of the test statistics under a sequence of local alternatives. The effects of estimating the nuisance parameter η are discussed in Section 4. Particular attention is given to the issue of when these effects disappear, an adaptiveness property akin to that in semiparametric inference. The choice of the process ψ is addressed in Section 5, and some specific choices are illustrated in Section 8. The local asymptotic powers of the tests are examined in Section 6, and the issue of whether it is necessary to make the ψ-process orthogonal to the gradient of the logarithm of $\lambda_0(\cdot; \eta)$ in order to achieve optimal asymptotic local power is resolved. In Section 7 the loss in efficiency under two types of model misspecification are examined, with the relative efficiency of tests measured using the local asymptotic relative efficiency of Woolson and Sen (1974). Section 9 presents results of simulation studies performed to ascertain the achieved levels and powers of the tests for small to moderate sample sizes. In Section 10 we present the results of applying specific smooth goodness-of-fit tests to three data sets considered in the literature. A summary and some concluding remarks are contained in Section 11.

2. Development of the smooth test. We assume the following basic conditions on the $\psi(\cdot;\eta)$ process and the class $\mathscr{C}$. If H_0 is true, we let η_0 denote the true value of η. Furthermore, $\psi(\cdot;\eta)$ is a locally bounded and predictable process, and $(\partial/\partial\eta)\lambda_0(\cdot;\eta)$ exists with $\lambda_0(t;\eta) > 0$ for each $(t,\eta) \in \mathscr{T} \times \Gamma$. Under our model, the partial log-likelihood process for (θ,η) [it will be the full log-likelihood under additional assumptions such as noninformative censoring] is given by [Andersen, Borgan, Gill and Keiding (1993)] $l(t;\theta,\eta) = \int_0^t \log[Y(s)\lambda_k(s;\theta,\eta)]\,dN(s) - \int_0^t Y(s)\lambda_k(s;\theta,\eta)\,ds$. Let $M(t;\theta,\eta) = N(t) - \int_0^t Y(s)\lambda_k(s;\theta,\eta)\,ds$, $t \in \mathscr{T}$. Then, under the condition that differentiation with respect to η and the integration operation can be interchanged, the score process is

$$U(t;\theta,\eta) = \begin{bmatrix} U_1(t;\theta,\eta) \\ U_2(t;\theta,\eta) \end{bmatrix} = \int_0^t \begin{bmatrix} \psi(s;\eta) \\ \dfrac{\partial}{\partial\eta}\log \lambda_k(s;\theta,\eta) \end{bmatrix} dM(s;\theta,\eta),$$

which, when evaluated at $(\theta,\eta) = (0,\eta_0)$, becomes

$$U(t;\eta_0) = \begin{bmatrix} U_1(t;\eta_0) \\ U_2(t;\eta_0) \end{bmatrix} = \int_0^t H(s;\eta_0)\,dM(s;\eta_0), \tag{2.1}$$

where $H(s;\eta) = [\psi(s;\eta)^{\mathrm{t}}, \rho(s;\eta)^{\mathrm{t}}]^{\mathrm{t}}$, $\rho(s;\eta) = (\partial/\partial\eta)\log\lambda_0(s;\eta)$ and $M(s;\eta) = M(s;0,\eta)$. Under H_0, $\{M(t;\eta_0): t \in \mathscr{T}\}$ is a square-integrable local martingale, and its predictable variation process is $\langle M(\cdot;\eta_0)\rangle(t) = A(t;\eta_0) \equiv \int_0^t Y(s)\lambda_0(s;\eta_0)\,ds$. Since $\{H(t;\eta_0): t \in \mathscr{T}\}$ is a locally bounded predictable process, then under H_0, $\{U(t;\eta_0): t \in \mathscr{T}\}$ is a square-integrable local martingale with predictable quadratic variation process

$$\langle U(\cdot;\eta_0)\rangle(t) = \int_0^t H(s;\eta_0)^{\otimes 2} Y(s)\lambda_0(s;\eta_0)\,ds, \tag{2.2}$$

where for a vector v, $v^{\otimes 0} = 1$, $v^{\otimes 1} = v$ and $v^{\otimes 2} = vv^{\mathrm{t}}$. If, under H_0, the true value η_0 is known, a test of H_0 can be based on the statistic $U_1(\tau;\eta_0) = \int_0^\tau \psi(s;\eta_0)\,dM(s;\eta_0)$. However, since η_0 is unknown, an asymptotically optimal test [cf. Bhat and Nagnur (1965), Choi, Hall and Schick (1996) and Neyman (1959)] can be obtained from the efficient score vector

$$\tilde{U}_1(\tau;\eta_0) = U_1(\tau;\eta_0) - \Sigma_{12}(\tau;\eta_0)\Sigma_{22}(\tau;\eta_0)^{-1}U_2(\tau;\eta_0),$$

upon replacing η_0 by a suitable estimator, and where

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

is a (possibly limiting) covariance matrix of U. We propose to replace η_0 by its restricted maximum likelihood estimator (RMLE) $\hat{\eta}$ obtained under the restriction $\theta = 0$. This RMLE satisfies $U_2(\tau;\hat{\eta}) = \int_0^\tau \rho(s;\hat{\eta})\,dM(s;\hat{\eta}) = 0$ and is the maximum likelihood estimator (MLE) of η under the condition that $\lambda(\cdot)$ belongs to $\mathscr{C}$. Asymptotic properties of this estimator have been obtained in

Borgan (1984) [see also Andersen, Borgan, Gill and Keiding (1993)]. Upon substituting $\hat{\eta}$ for η_0, the estimated efficient score vector becomes

$$U_1(\tau;\hat{\eta}) = \tilde{U}_1(\tau;\hat{\eta}) = \int_0^\tau \psi(s;\hat{\eta})\, dM(s;\hat{\eta}). \tag{2.3}$$

The process $\{M(t;\hat{\eta})\colon t \in \mathscr{T}\}$ is called a "martingale" residual process [cf. Barlow and Prentice (1988), Therneau, Grambsch and Fleming (1990)], although since $\hat{\eta}$ depends on information in $\mathscr{F}_\tau$, it is not a bonafide martingale. One may therefore view $U_1(\tau;\hat{\eta})$ as a weighted martingale residual, with weights determined by the $\psi(\cdot;\eta)$ process introduced in the class $\mathscr{A}_k$. We will discuss the choice of this process in later sections. To obtain the exact form of the test, we need the sampling distribution of $U_1(\tau;\hat{\eta})$ and its covariance matrix $\Xi(\tau;\eta_0)$ under H_0. With A^- denoting a generalized inverse of a matrix A, the proposed smooth goodness-of-fit test for H_0 of order k associated with $\psi(\cdot;\eta)$ takes the form

$$\text{Reject } H_0 \text{ if } S_k(\tau;\hat{\eta}) \equiv U_1(\tau;\hat{\eta})^{\mathrm{t}}\hat{\Xi}(\tau;\hat{\eta})^{-}\, U_1(\tau;\hat{\eta}) \geq c_\alpha, \tag{2.4}$$

where c_α is such that $\mathbf{P}\{S_k(\tau;\hat{\eta}) \geq c_\alpha | H_0\}$ equals α, or converges to α in some suitable asymptotic sense.

In general, the exact sampling distribution of $U_1(\tau;\hat{\eta})$, and hence that of $S_k(\tau;\hat{\eta})$, is difficult to obtain analytically. Computer-intensive methods may provide a viable way of approximately assessing the significance of an observed value of $S_k(\tau;\hat{\eta})$. In this paper we focus on the asymptotic properties of $\{U_1(\tau;\hat{\eta})\colon t \in \mathscr{T}\}$ under a sequence of local alternatives (in θ). This will enable us to assess the effects of substituting $\hat{\eta}$ for η_0, and to see how the asymptotic properties of $U_1(\cdot;\hat{\eta})$, and also of $S_k(\tau;\hat{\eta})$, are affected by changing the $\psi(\cdot;\eta)$ process. Knowledge of such effects may provide guidance in choosing the order k and the $\psi(\cdot;\eta)$ process to achieve certain desirable properties of the test.

3. Asymptotics. For our asymptotic analysis we consider a sequence of processes $\{(N^{(n)}(t), Y^{(n)}(t))\colon t \in \mathscr{T}\}$, $(n = 1, 2, \dots)$, with the nth member of the sequence defined on a probability space $(\Omega^{(n)}, \mathscr{F}^{(n)}, \mathbf{P}^{(n)})$ with filtration $\mathbf{F}^{(n)} = \{\mathscr{F}_t^{(n)}\colon t \in \mathscr{T}\}$. The sequence of compensator processes satisfies

$$A^{(n)}(t;\theta,\eta) = \int_0^t Y^{(n)}(s)\lambda_k^{(n)}(s;\theta,\eta)\, ds, \qquad n = 1, 2, \dots,$$

where

$$\lambda_k^{(n)}(s;\theta,\eta) = \lambda_0(s;\eta)\exp\{\theta^{\mathrm{t}}\psi^{(n)}(s;\eta)\},$$

and with $\psi^{(n)}(\cdot;\eta) = (\psi_1^{(n)}(\cdot;\eta), \dots, \psi_k^{(n)}(\cdot;\eta))^{\mathrm{t}}$ also depending on n. Note that k, as well as $\lambda_0(\cdot;\cdot)$, do not depend on n. The normalizing sequence of constants $\{a_n\colon n = 1, 2, \dots\}$ is an increasing sequence of positive real numbers. The sequence of hypotheses is

$$H_{0n}\colon \theta^{(n)} = 0,\ \eta \in \Gamma \quad \text{versus } H_{1n}\colon \theta^{(n)} = \gamma(1 + o(1))/a_n,\ \eta \in \Gamma,$$

where we assume that the true value of η, under H_{0n}, is η_0 and is independent of n, and γ is some $q \times 1$ vector of real numbers. The restricted MLE of η for the nth model is $\hat{\eta}^{(n)}$ which solves the equation $U_2^{(n)}(\tau;\eta) = 0$, where $U_2^{(n)}(t;\eta) = \int_0^t \rho^{(n)}(s;\eta)\, dM^{(n)}(s;\eta)$, $\rho^{(n)}(s;\eta) = (\partial/\partial n)\log \lambda_0(s;\eta)$, and $M^{(n)}(s;\eta) = N^{(n)}(s) - \int_0^t Y^{(n)}(s)\lambda_0(s;\eta)\, ds$. The score test statistic is therefore

$$U_1^{(n)}\big(t;\hat{\eta}^{(n)}\big) = \int_0^t \psi^{(n)}\big(s;\hat{\eta}^{(n)}\big)\, dM^{(n)}\big(s;\hat{\eta}^{(n)}\big), \qquad t \in \mathscr{T}.$$

Having just indicated which quantities or processes depend on n, for brevity we shall henceforth suppress writing the superscript $^{(n)}$, except in cases where confusion may arise.

We present the major asymptotic result needed for developing the smooth goodness-of-fit test. The proof of the theorem is deferred to the Appendix, together with intermediate results needed in its proof. The regularity conditions for the asymptotic results are also enumerated in the Appendix. Some of these conditions are needed to obtain results pertaining to the restricted MLE $\hat{\eta}$, and are similar to those in Borgan (1984), while others pertain to the behavior of the $\psi(\cdot;\eta)$ processes. The matrix function

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

which appears in the statement of Theorem 3.1, is defined in condition (V) in the Appendix. For ease of reference, the definition of its submatrices are given by $\boldsymbol{\Sigma}_{11}(t;\eta) = \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2} y(s)\lambda_0(s;\eta)\, ds$, $\boldsymbol{\Sigma}_{12}(t;\eta) = \int_0^t \psi^{(0)}(s;\eta)\rho(s;\eta)^{\mathrm{t}} y(s)\lambda_0(s;\eta)\, ds$, $\boldsymbol{\Sigma}_{22}(t;\eta) = \int_0^t \rho(s;\eta)^{\otimes 2} y(s)\lambda_0(s;\eta)\, ds$, and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^{\mathrm{t}}$, where $\psi^{(0)}(\cdot;\eta)$ is the limiting function of $\psi^{(n)}(\cdot;\eta)$, and $y(\cdot)$ is a limiting function of the standardized at-risk process $Y^{(n)}(\cdot)$. For precise descriptions of these functions, see condition (IV) in the Appendix.

THEOREM 3.1. *If conditions* (I)–(VIII) *are satisfied, then under* H_{1n}, $a_n^{-1}U_1(\cdot;\hat{\eta})$ *converges weakly, on Skorohod's space* $\mathscr{D}[0,\tau]^k$, *to a Gaussian process* $\tilde{Z}_1(\cdot;\eta_0)$ *with mean function*

$$\mathbf{E}\big\{\tilde{Z}_1(\cdot;\eta_0)\big\} \equiv \mu_1(\cdot;\eta_0) = \big\{\boldsymbol{\Sigma}_{11}(\cdot;\eta_0) - \boldsymbol{\Sigma}_{12}(\cdot;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)\big\}\gamma$$

and covariance matrix function

$$\operatorname{Cov}\big\{\tilde{Z}_1(t_1;\eta_0), \tilde{Z}_1(t_2;\eta_0)\big\}$$
$$\equiv \tilde{\boldsymbol{\Sigma}}_1(t_1,t_2;\eta_0) = \boldsymbol{\Sigma}_{11}(t_1 \wedge t_2;\eta_0) - \boldsymbol{\Sigma}_{12}(t_1;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(t_2;\eta_0).$$

With $\boldsymbol{\Sigma}_{11.2}(t;\eta_0) = \tilde{\boldsymbol{\Sigma}}_1(t,t;\eta_0)$, *the quantity*

$$S_k(\tau;\eta_0) = a_n^{-2}U_1(\tau;\hat{\eta})^{\mathrm{t}}\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0)^{-}\, U_1(\tau;\hat{\eta})$$

converges in distribution, under H_{1n}, *to a noncentral chi-squared distribution with degrees-of-freedom* $k^* = \operatorname{rank}(\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0))$ *and noncentrality parameter* $\gamma^{\mathrm{t}}\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0)\gamma$.

The quantity $S_k(\cdot;\eta_0)$ depends on the unknown parameter η_0 through $\mathbf{\Sigma}_{11.2}(\tau;\eta_0)$, hence is not a test statistic. To construct a test for H_0 we need to estimate $\mathbf{\Sigma}_{11.2}(\tau;\eta_0)$, or equivalently $\mathbf{\Sigma}_{11.2}(\tau;\eta_0)^-$, consistently. Two possible estimators of $\mathbf{\Sigma}(\cdot;\eta_0)$ are

$$\hat{\mathbf{\Sigma}}(\cdot;\hat{\eta}) = \frac{1}{a_n^2}\int_0^{\cdot}\begin{bmatrix}\psi(s;\hat{\eta})\\ \rho(s;\hat{\eta})\end{bmatrix}^{\otimes 2} Y(s)\lambda_0(s;\hat{\eta})\,ds;$$

$$\check{\mathbf{\Sigma}}(\cdot;\eta) = \frac{1}{a_n^2}\int_0^{\cdot}\begin{bmatrix}\psi(s;\hat{\eta})\\ \rho(s;\hat{\eta})\end{bmatrix}^{\otimes 2} dN(s).$$

These estimators, which are based on the predictable quadratic variation process and the optional variation process, respectively, are uniformly consistent estimators of $\mathbf{\Sigma}(\cdot;\eta_0)$ under H_{1n}. Our numerical studies indicate that for finite sample sizes it is advantageous to utilize a convex combination of the two estimators. In particular, we propose to use the estimator

$$\bar{\Sigma}(\cdot;\hat{\eta}) = \tfrac{1}{2}\left[\hat{\Sigma}(\cdot;\hat{\eta}) + \check{\Sigma}(\cdot;\hat{\eta})\right]. \tag{3.1}$$

As estimator of

$$\tilde{\mathbf{\Sigma}}_1(t_1,t_2;\eta_0) = \mathbf{\Sigma}_{11}(t_1\wedge t_2;\eta_0) - \mathbf{\Sigma}_{12}(t_1;\eta_0)\mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}\mathbf{\Sigma}_{21}(t_2;\eta_0),$$

we use $\bar{\mathbf{\Sigma}}_1(t_1,t_2;\hat{\eta})$ obtained by replacing the submatrices consisting of $\tilde{\mathbf{\Sigma}}_1$ by the corresponding submatrices in $\bar{\mathbf{\Sigma}}$. Note that $\mathbf{\Sigma}_{11.2}(\cdot;\eta_0)$ may also depend on nuisance parameters associated with the incompleteness mechanism. However, our estimator of this covariance matrix allows us to bypass the direct estimation of these nuisance parameters, an important consideration because the incompleteness mechanism need not be restrictive.

Our proposed quadratic test statistic now becomes

$$\bar{S}_k(\tau;\hat{\eta}) = a_n^{-2}U_1(\tau;\hat{\eta})^{\mathrm{t}}\bar{\mathbf{\Sigma}}_{11.2}(\tau;\hat{\eta})^{-}U_1(\tau;\hat{\eta}), \tag{3.2}$$

which has an asymptotic chi-squared distribution with k^* degrees-of-freedom under H_{0n}. The asymptotic α-level smooth test then rejects H_0 whenever $\bar{S}_k(\tau;\hat{\eta}) \geq \chi^2_{\hat{k}^*,\alpha}$, where $\hat{k}^*$ is the rank of the covariance matrix $\bar{\mathbf{\Sigma}}_{11.2}(\tau;\hat{\eta})$.

4. Effects of estimating the nuisance parameter.

From the intermediate result in Proposition B.1, we find that if η_0 is known, then the statistic $a_n^{-1}U_1(\tau;\eta_0)$ converges in distribution to a normal random vector $Z_1(\tau;\eta_0)$ with mean vector and covariance matrix

$$\mathbf{E}\{Z_1(\tau;\eta_0)\} = \mu_1(\tau;\eta_0) = \mathbf{\Sigma}_{11}(\tau;\eta_0)\gamma \quad\text{and}\quad \mathrm{Cov}\{Z_1(\tau;\eta_0)\} = \mathbf{\Sigma}_{11}(\tau;\eta_0).$$

In contrast, from Theorem 3.1, if η_0 is unknown and is estimated by the restricted MLE $\hat{\eta}$, then the statistic $a_n^{-1}U_1(\tau;\hat{\eta})$ converges in distribution to a normal random vector $\tilde{Z}_1(\tau;\eta_0)$ with mean vector and covariance matrix

$$\mathbf{E}\{\tilde{Z}_1(\tau;\eta_0)\} = \tilde{\mu}_1(\tau;\eta_0) = \mathbf{\Sigma}_{11.2}(\tau;\eta_0)\gamma$$

and

$$\mathrm{Cov}\{\tilde{Z}_1(\tau;\eta_0)\} = \mathbf{\Sigma}_{11.2}(\tau;\eta_0),$$

where $\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0) = \boldsymbol{\Sigma}_{11}(\tau;\eta_0) - \boldsymbol{\Sigma}_{12}(\tau;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)$. Thus, the asymptotic effect of replacing η_0 by $\hat{\eta}$ in $a_n^{-1}U_1(\tau;\eta_0)$ is manifested by a change in the mean vectors, with the magnitude of this change given by

$$\Delta\mu_1(\tau;\eta_0) = \mu_1(\tau;\eta_0) - \tilde{\mu}_1(\tau;\eta_0) = \boldsymbol{\Sigma}_{12}(\tau;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)\gamma,$$

and a decrease in covariance matrices given by $\boldsymbol{\Sigma}_{12}(\tau;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)$. Notice that, in an asymptotic sense, the estimation of η_0 by $\hat{\eta}$ has *no* effect on the distribution of $a_n^{-1}U_1(\tau;\hat{\eta})$ if and only if $\boldsymbol{\Sigma}_{12}(\tau;\eta_0) = 0$. Recalling that

$$\begin{aligned}\langle a_n^{-1}U_1(\cdot;\eta_0), a_n^{-1}U_2(\cdot;\eta_0)\rangle(\tau) &\\ \to \int_0^\tau \psi^{(0)}(s;\eta_0)\rho(s;\eta_0)^{\mathrm{t}} y(s)\lambda_0(s;\eta_0)\,ds \quad &\text{in probability}\\ = \boldsymbol{\Sigma}_{12}(\tau;\eta_0), &\end{aligned}$$

if $\psi^{(0)}(\cdot;\eta_0)$ and $\rho(\cdot;\eta_0)$ are orthogonal with respect to the inner product

$$\langle f, g\rangle = \int_0^\tau f(s)g(s)y(s)\lambda_0(s;\eta_0)\,ds$$

on the Hilbert space $\mathscr{H} = \{h: [0,\tau] \to \Re: \int_0^\tau h(s)^2 y(s)\lambda_0(s;\eta_0)\,ds < \infty\}$, then, in an asymptotic sense, not knowing η_0 does not matter since it could be replaced by the restricted MLE $\hat{\eta}$ without altering the limiting distribution of the test statistic. This is analogous to the notion of adaptive estimation in the presence of nuisance parameters [cf. Bickel, Klaasen, Ritov and Wellner (1993)], as well as to the notion of orthogonal parameters [cf. Cox and Reid (1987)]. On the other hand, if $\psi^{(0)}(\cdot;\eta_0)$ and $\rho(\cdot;\eta_0)$ are not orthogonal, then substituting $\hat{\eta}$ for η_0 in $a_n^{-1}U_1(\tau;\eta_0)$ could have an effect on the limiting variance. This result indicates that one should be cognizant that when η_0 is replaced by $\hat{\eta}$, though a consistent estimator of η_0, there could be a substantial change in the distributional properties of the test statistic, with the magnitude of this change determined by the interplay among the class $\mathscr{C}$, the choice of the ψ-process and the data structure [represented by $y(\cdot)$]. Thus, if in testing H_0: $\lambda(\cdot) \in \mathscr{C}$, one is to use the test statistics $\bar{S}_k^u(\tau;\hat{\eta}) = a_n^{-2}U_1(\tau;\hat{\eta})^{\mathrm{t}}\bar{\boldsymbol{\Sigma}}_{11}(\tau;\hat{\eta})^{-}U_1(\tau;\hat{\eta})$ instead of $\bar{S}_k(\tau;\hat{\eta}) = a_n^{-2}U_1(\tau;\hat{\eta})^{\mathrm{t}}\bar{\boldsymbol{\Sigma}}_{11.2}(\tau;\hat{\eta})^{-}U_1(\tau;\hat{\eta})$, then misleading conclusions may arise if the chosen ψ-process and the $\rho(\cdot)$ associated with $\mathscr{C}$ are highly related in the space $\mathscr{H}$. In classical settings this aspect has also been addressed by Chernoff and Lehman (1954), Durbin (1973, 1975), Loynes (1980), Pierce (1982), Randles (1982, 1984) and Stephens (1976). These results also have a bearing on the appropriate use of hazard-based generalized residuals in model validation and diagnostics; for instance, see Aban and Peña (1998), Baltazar–Aban and Peña (1995), Lagakos (1981) and Peña (1995, 1998).

5. On the choice of the ψ-process. In this section we discuss general characteristics of a desirable choice of the ψ-process. Note at the outset that this process determines the class of alternative hazard functions when the class $\mathscr{C} = \{\lambda_0(\cdot, \eta)\colon \eta \in \Gamma\}$ does not hold. Let $\lambda_0(\cdot)$ denote the true but unknown hazard rate function and define the mapping from Γ into $\mathscr{K}$ via

$$\eta \mapsto \kappa(\cdot\,;\eta) = \log\left\{\frac{\lambda_0(\cdot\,;\eta)}{\lambda_0(\cdot)}\right\},$$

with the convention that $0/0 = 1$, and where $\mathscr{K} = \{\kappa\colon \int_0^\tau \kappa(t)^2 y(t)\lambda_0(t)\,dt < \infty\}$. Endow $\mathscr{K}$ with the inner product $\langle \kappa_1, \kappa_2\rangle = \int_0^\tau \kappa_1(t)\kappa_2(t)y(t)\lambda_0(t)\,dt$, and norm $\|\kappa\| = \sqrt{\langle \kappa, \kappa\rangle}$. Let $\mathscr{K}_\Gamma$ be the closure of the subspace of $\mathscr{K}$ whose members are $\kappa(\cdot; \eta)$, $\eta \in \Gamma$. For a given $\lambda_0(\cdot)$, let $\eta_0 \in \Gamma$ be such that $\|\kappa(\cdot; \eta_0)\| \le \|\kappa(\cdot; \eta)\|$ for every $\kappa(\cdot; \eta) \in \mathscr{K}_\Gamma$, so $\lambda_0(\cdot; \eta_0)$ is the member of $\mathscr{C}$ closes to $\lambda_0(\cdot)$. Then, for each $\eta \in \Gamma$, $\kappa(\cdot; \eta) = \kappa(\cdot; \eta_0) + \log\{\lambda_0(\cdot; \eta)/\lambda_0(\cdot; \eta_0)\}$ and $\|\kappa(\cdot; \eta)\|^2 = \|\kappa(\cdot; \eta_0)\|^2 + \|\log\{\lambda_0(\cdot; \eta)/\lambda_0(\cdot; \eta_0)\}\|^2$. Furthermore, $\log\{\lambda_0(\cdot; \eta)/\lambda_0(\cdot; \eta_0)\} = (\eta - \eta_0)^t\rho(\cdot; \eta_0) + o(\|\eta - \eta_0\|^2)$, where $\rho(\cdot; \eta) = (\partial/\partial n)\log \lambda_0(\cdot; \eta)$. In view of these considerations, for $\mathscr{A}_k$ in (1.1) to be a suitable class of alternatives for $\mathscr{C}$, the ψ-process should ideally span the linear space of $\kappa(\cdot; \eta_0)$ which is generated as one varies $\lambda_0(\cdot)$ in the space $\mathscr{A}$ of hazard rate functions which could arise when the hypothesized class $\mathscr{C}$ is false. It may also be more informative if the components of ψ are orthogonal and each indicates certain meaningful and interpretable departures from $\lambda_0(\cdot; \eta)$. One possible choice of the ψ-process is for the components to be (asymptotically) orthonormal in the sense of satisfying

$$\int_0^\tau \psi^{(0)}(t)^{\otimes 2} y(t)\lambda_0(t)\,dt = \mathbf{I}_k, \tag{5.1}$$

where $y(\cdot)$ is the true limiting function of $Y(\cdot)/a_n^2$. In the context of the local alternatives studied in the preceding section, this condition becomes $\mathbf{\Sigma}_{11}(\tau; \eta_0) = \int_0^\tau \psi^{(0)}(t; \eta_0)^{\otimes 2} y(t; \eta_0)\lambda_0(t; \eta_0)\,dt = \mathbf{I}_k$, or equivalently,

$$\int_0^{\Lambda_0(\tau;\eta_0)} \phi^{(0)}(t;\eta_0)\, y\big[\Lambda_0^{-1}(t;\eta_0);\eta_0\big]\,dt = \mathbf{I}_k, \tag{5.2}$$

where $\phi^{(0)}$ is defined according to $\psi^{(0)}(t; \eta) = \phi^{(0)}[\Lambda_0(t; \eta); \eta]$. Thus, one may choose the $\psi^{(0)}$-process, or equivalently, the $\phi^{(0)}$-process, such that $\phi^{(0)}$ has components which are orthonormal with respect to $y[\Lambda_0^{-1}(\cdot; \eta_0)]$ over the region $[0, \Lambda_0(\tau; \eta_0)]$.

However, since the limiting function $y(\cdot)$ involves elements of the mechanism that leads to incomplete observations, this program of choosing $\psi^{(0)}$ to be orthogonal may only be feasible in practice if stringent conditions are imposed on the incompleteness mechanism. To circumvent this potential difficulty we instead advocate that the $\psi^{(0)}$-process be chosen, not with the primary goal of having orthogonal components, but rather that it should span the linear space of $\kappa(\cdot; \eta_0)$ generated by varying $\lambda_0(\cdot)$ in the space of hazard functions which arise when $\mathscr{C}$ does not hold. If feasible, one may then orthogonalize the components and then use the orthogonalized components. As will be seen in the next section, in terms of local asymptotic power, one does not gain any advantage by using an orthogonal ψ-process. Such a $\psi^{(0)}$

may be of polynomial form, for example, $\psi_i^{(0)}(\Lambda_0^{-1}(t;\eta);\eta) = t^{i-1}$, $i = 1,\dots,k$, or it could be chosen at the outset to be well-known orthogonal polynomials such as Laguerre polynomials. This idea of using polynomial basis functions was the one considered by Neyman (1937) and expounded in Rayner and Best (1989) in classical settings. Other possibilities are ψ-processes based on trigonometric basis functions, and possibly, wavelets such as in Fan (1996). This latter possibility will be explored in future research since it possesses the promise of being able to focus the smooth goodness-of-fit tests on local differences between the class $\mathscr{C}$ and the true $\lambda_0(\cdot)$.

6. Local asymptotic power. Let us examine the asymptotic local power of the test in (3.2) if $\psi^{(0)}(t;\eta) = \phi^{(0)}[\Lambda_0(t;\eta);\eta]$ is chosen to satisfy (5.2). The sequence of true hazard rate functions consists of

$$\lambda_0^{(n)}(t) = \lambda_0(t;\eta_0)\exp\{\theta_n^{\mathrm{t}}\phi^{(n)}[\Lambda_0(t;\eta_0);\eta_0]\}, \qquad n = 1,2,\dots,$$

where $\theta_n = a_n^{-1}\gamma(1+o(1))$ and $\phi^{(n)}$ and $\phi^{(0)}$ satisfy condition (IV) stated in Section 11. Without loss of generality and for standardization purposes, assume that $|\gamma|^2 = \gamma^{\mathrm{t}}\gamma = 1$. Using Theorem 3.1, the test statistic $\overline{S}_k(\tau;\hat{\eta})$ in (3.2) converges in distribution to a noncentral chi-squared distribution with degrees-of-freedom $k^* = \mathrm{rank}[\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0)]$ and noncentrality parameter $\delta^2 = \gamma^{\mathrm{t}}\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0)\gamma$, where, with $\tau^0 = \Lambda_0(\tau;\eta_0)$, $y^{(0)}(s;\eta) = y[\Lambda_0^{-1}(s;\eta);\eta]$ and $\rho^{(0)}(s;\eta) = \rho[\Lambda_0^{-1}(s;\eta);\eta]$,

$$\boldsymbol{\Sigma}_{11.2}(\tau;\eta_0) = \mathbf{I}_k - \boldsymbol{\Sigma}_{12}(\tau;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0);$$

$$\boldsymbol{\Sigma}_{12}(\tau;\eta_0) = \int_0^{\tau^0}\phi^{(0)}(s;\eta_0)\rho^{(0)}(s;\eta_0)^{\mathrm{t}}y^{(0)}(s;\eta_0)\,ds;$$

and

$$\boldsymbol{\Sigma}_{22}(\tau;\eta_0) = \int_0^{\tau^0}\rho^{(0)}(s;\eta_0)^{\otimes 2}y^{(0)}(s;\eta_0)\,ds.$$

Thus, the noncentrality parameter can be written as

$$\delta^2 = 1 - \left|\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1/2}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)\gamma\right|^2.$$

The limiting local power (ALP) of the α-level test in (3.2) is $\mathrm{ALP}(\tau;\eta_0) = \mathbf{P}\{\chi^2_{k^*}(\delta^2) \geq \chi^2_{k^*;\alpha}\}$, where $\chi^2_{k^*}(\delta^2)$ is a noncentral chi-squared distributed variable with k^* degrees-of-freedom and noncentrality parameter δ^2. This power is maximized whenever $\boldsymbol{\Sigma}_{12}(\tau;\eta_0) = 0$, and in such a case, $k^* = k$ and $\delta^2 = 1$. Recall from Section 4 that $\boldsymbol{\Sigma}_{21}(\tau;\eta_0) = 0$ leads to the orthogonality between $\phi^{(0)}$ and $\rho^{(0)}$ and enables the substitution of $\hat{\eta}$ for η_0 while retaining the same asymptotic distributions for $U_1(\cdot;\eta_0)$ and $U_1(\cdot;\hat{\eta})$.

Denote by $\phi_0^{(0)}$ the limiting ϕ-process which is orthogonal to $\rho^{(0)}$ so it satisfies (5.2) and $\int_0^{\tau^0}\phi_0^{(0)}(s;\eta_0)\rho^{(0)}(s;\eta_0)^{\mathrm{t}}y^{(0)}(s;\eta_0)\,ds = \mathbf{0}$. In principle, such an orthogonalization can be done using a Gram–Schmidt orthogonalization technique [see the earlier version of this manuscript, Peña (1996), Lemma 5.1], though implementing the orthogonalization process may be tedious or require restrictive assumptions on the incompleteness mechanisms. Another approach, implemented in an example in Section 8, is to start with some appropriate $\phi^{(0)}$, and the desired $\phi_0^{(0)}$ is obtained via

$$\phi_0^{(0)} = \left[\langle \phi^{(0)}, (\phi^{(0)})^{t}\rangle - \langle \phi^{(0)}, (\rho^{(0)})^{t}\rangle\langle \rho^{(0)}, (\rho^{(0)})^{t}\rangle^{-1}\langle \rho^{(0)}, (\phi^{(0)})^{t}\rangle\right]^{-1/2}$$
$$\times\left[\phi^{(0)} - \langle \phi^{(0)}, (\rho^{(0)})^{t}\rangle\langle \rho^{(0)}, (\rho^{(0)})^{t}\rangle^{-1}\rho^{(0)}\right],$$

where $\langle f, g\rangle = \int_0^{\tau^0} f(s)g(s)y^{(0)}(s;\eta_0)\,ds$, and for vectors $f = (f_1, \dots, f_k)^{t}$ and $g = (g_1, \dots, g_k)^{t}$, $\langle f, g^{t}\rangle$ is the matrix consisting of elements $\langle f_i, g_j\rangle$. For the sequence of alternative hazard rate functions

$$\lambda_0^{(n)}(\cdot) = \lambda_0(\cdot;\eta_0)\exp\left\{a_n^{-1}\gamma^{t}\phi_0^{(n)}[\Lambda_0(\cdot;\eta_0);\eta_0]\right\}, \tag{6.1}$$

with $\gamma^{t}\gamma = 1$, the asymptotic local power of the test which rejects $H_0\colon \lambda(\cdot) \in \mathscr{C}$ whenever $\overline{S}_k^{(0)}(\tau) > \chi^2_{k;\alpha}$, where

$$\overline{S}_k^{(0)}(\tau) = \frac{1}{a_n^2}\sum_{i=1}^{k}\left[\int_0^{\tau}\phi_{0i}^{(n)}[\Lambda_0(s;\hat{\eta});\hat{\eta}]\,dM(s;\hat{\eta})\right]^2 \tag{6.2}$$

is $\mathbf{P}\{\chi_k^2(1) \ge \chi^2_{k;\alpha}\}$. Note the simple form of the test statistic which arises because of the assumed asymptotic orthonormality of the $\phi_0^{(n)}$ and its asymptotic orthogonality with $\rho^{(0)}$.

Suppose now that $\phi_1^{(0)}$ is another limiting $\phi^{(0)}$-process which is not orthogonal to $\rho^{(0)}$ and representable via

$$\phi_1^{(0)}(\cdot;\eta_0) = \mathbf{A}(\eta_0)\phi_0^{(0)}(\cdot;\eta_0) + \mathbf{B}(\eta_0)\rho^{(0)}(\cdot;\eta_0), \tag{6.3}$$

where $\mathbf{A}(\eta_0)$ is a $k \times k$-nonsingular matrix. We inquire whether the asymptotic local power of the test based on $\phi_1^{(0)}$ or its associated $\{\phi_1^{(n)}\}$ sequence is different from the ALP of the test based on $\phi_0^{(0)}$ under the *same* local alternatives in (6.1). With respect to the sequence $\{\phi_1^{(n)}\}$, the local alternatives in (6.1) become

$$\lambda_0^{(n)}(\cdot) = \lambda_0(\cdot;\eta_0)\exp\left\{a_n^{-1}\gamma_*^{t}\,\phi_*^{(n)}[\Lambda_0(\cdot;\eta_0);\eta_0]\right\}, \tag{6.4}$$

where

$$\phi_*^{(n)}(\cdot;\eta_0) = \left((\phi_1^{(n)}(t;\eta_0))^{t}, \rho^{(0)}(\cdot;\eta_0)^{t}\right)^{t};$$
$$\gamma_* = \left(\gamma^{t}\mathbf{A}(\eta_0)^{-1}, -\gamma^{t}\mathbf{A}(\eta_0)^{-1}\mathbf{B}(\eta_0)^{t}\right)^{t}.$$

It is then easy to show that the limiting covariance matrices associated with the $\{\phi_*^{(n)}\}$ sequence are

$$\boldsymbol{\Sigma}_{11}^*(\tau;\eta_0) = \int_0^{\tau^0}\left(\phi_*^{(0)}(s;\eta_0)\right)^{\otimes 2} y^{(0)}(s;\eta_0)\,ds = \begin{bmatrix} \mathbf{A}^{\otimes 2} + \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{B}^{t} & \mathbf{B}\boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{22}\mathbf{B}^{t} & \boldsymbol{\Sigma}_{22} \end{bmatrix};$$
$$\boldsymbol{\Sigma}_{12}^*(\tau;\eta_0) = \int_0^{\tau^0}\phi_*^{(0)}(s;\eta_0)\left(\rho^{(0)}(s;\eta_0)\right)^{t} y^{(0)}(s;\eta_0)\,ds = \begin{bmatrix} \mathbf{B}\boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Therefore,

$$\begin{aligned}\mathbf{\Sigma}^*_{11.2}(\tau;\eta_0) &= \mathbf{\Sigma}^*_{11}(\tau;\eta_0) - \mathbf{\Sigma}^*_{12}(\tau;\eta_0)\mathbf{\Sigma}^*_{22}(\tau;\eta_0)^{-1}\mathbf{\Sigma}^*_{21}(\tau;\eta_0)\\ &= \begin{bmatrix}\mathbf{A}(\eta_0)^{\otimes 2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\end{aligned}$$

which has rank k. The noncentrality parameter of the limiting chi-squared distribution of the $\{\phi_1^{(n)}\}$-based test statistic given by

$$\begin{aligned}\bar{S}^{(1)}_{k+q}(\tau) = \frac{1}{a_n^2}&\left[\int_0^\tau \phi_*^{(n)}[\Lambda_0(t;\hat{\eta});\hat{\eta}]\,dM(t;\hat{\eta})\right]^{\mathrm{t}}\bar{\mathbf{\Sigma}}^*_{11.2}(\tau;\hat{\eta})^{-}\\ &\times\left[\int_0^\tau \phi_*^{(n)}[\Lambda_0(t;\hat{\eta});\hat{\eta}]\,dM(t;\hat{\eta})\right]\end{aligned}$$

becomes $\delta_*^2 = \gamma_*^{\mathrm{t}}\mathbf{\Sigma}^*_{11.2}(\tau;\eta_0)\gamma_* = \gamma^{\mathrm{t}}\mathbf{A}^{-1}\mathbf{A}^{\otimes 2}(\mathbf{A}^{-1})^{\mathrm{t}}\gamma = \gamma^{\mathrm{t}}\gamma = 1$. The asymptotic local power of the $\{\phi_1^{(n)}\}$-based test therefore coincides with that of the $\{\phi_0^{(n)}\}$-based test with respect to the same sequence of local alternatives. This shows that it is not necessary to choose the $\phi^{(0)}$-process to be orthogonal to $\rho^{(0)}$ to achieve optimum power. However, if one is able to feasibly choose the $\phi^{(0)}$-process to be orthogonal to the $\rho^{(0)}$-process, aside from satisfying condition (5.2), the test statistic becomes parsimonious as it is just a sum of squared components [see the test statistics in (6.2)]. The ith ($i = 1,\dots,k$) component of the test statistic, given by

$$S_{k,i}(\tau) = \frac{1}{a_n^2}\left[\int_0^\tau \phi_{0i}^{(n)}[\Lambda_0(s;\hat{\eta});\hat{\eta}]\,dM(s;\hat{\eta})\right]^2, \tag{6.5}$$

may be used to develop a *directional* test for the null hypothesis H_0: $\theta = 0$ versus the alternative hypothesis H_1: $\theta_i \neq 0$; $\theta_j = 0$ ($j \neq i$). If the $\phi_0^{(0)}$ are properly chosen, then results of such directional tests could indicate the type of departure of the true hazard rate $\lambda_0(\cdot)$ from the class $\mathscr{C}$ if H_0 is rejected. In addition, for such a choice of $\phi_0^{(0)}$, the unknown nuisance parameter η does not pose any difficulties since it can be replaced by its restricted MLE and *no* adjustments are needed.

7. Efficiency loss from misspecified models. We explore in this section the loss of efficiency under two types of model misspecification: oversmoothing and undersmoothing. As a consequence of the results of Section 5, without loss of generality, we may assume that $\phi^{(0)} = (\phi_1^{(0)},\dots,\phi_k^{(0)})^{\mathrm{t}}$ satisfies

$$\begin{aligned}\int_0^\tau \phi^{(0)}(t)^{\otimes 2}y(t)\lambda_0(t;\eta_0)\,dt &= \mathbf{I},\\ \int_0^{\tau^{(0)}} \phi^{(0)}(t)\rho(t)^{\mathrm{t}}y(t)\lambda_0(t;\eta_0)\,dt &= \mathbf{0}\end{aligned} \tag{7.1}$$

and, under the sequence of local alternatives $\{H_{1n}\}$,

$$a_n \log\left\{\frac{\lambda_0^{(n)}(t)}{\lambda_0(t;\eta)}\right\} = \gamma^{\mathrm{t}}\phi^{(0)}(t) + \xi^{\mathrm{t}}\rho(t) + o_p(1)$$

for some $\gamma \in \Re^k$ and $\xi \in \Re^q$. Denote by $\phi^{(n)} = (\phi_1^{(n)}, \dots, \phi_k^{(n)})^{\mathrm{t}}$ the observable process satisfying equation (IV) in the Appendix (with $\psi^{(n)}$ replaced by $\phi^{(n)}$, and $\psi^{(0)}$ replaced by $\phi^{(0)}$).

The first type of misspecification is "oversmoothing." In this situation suppose that there is a k_1 with $1 \le k_1 < k$ such that, under $\{H_{1n}\}$, $\gamma_2 = (\gamma_{k_1+1}, \dots, \gamma_k)^{\mathrm{t}} = \mathbf{0}$, where $\gamma = (\gamma_1^{\mathrm{t}}, \gamma_2^{\mathrm{t}})^{\mathrm{t}}$. Equivalently, $\phi_2^{(0)} = (\phi_{k_1+1}^{(0)}, \dots, \phi_k^{(0)})^{\mathrm{t}}$ is not needed in describing the alternative space. Suppose we use the test which rejects H_0 whenever

$$S_k^{\mathrm{ov}}(\tau) = \frac{1}{a_n^2}\left|\int_0^\tau \phi^{(n)}(t)\, dM(t;\hat{\eta})\right|^2 \ge \chi^2_{k;\alpha}. \tag{7.2}$$

From Theorem 3.1, and by virtue of the condition in (7.1), the limiting distribution of the test statistic $S_k^{\mathrm{ov}}(\tau)$, under $\{H_{1n}\}$, is $\chi_k^2(\delta_{\mathrm{ov}}^2)$, where $\delta_{\mathrm{ov}}^2 = \gamma^{\mathrm{t}}\gamma = \gamma_1^{\mathrm{t}}\gamma_1 = |\gamma_1|^2$ since $\gamma_2 = \mathbf{0}$. Under the assumed conditions the appropriate test should reject H_0 whenever

$$S_{k_1}^{\mathrm{op1}}(\tau) = \frac{1}{a_n^2}\left|\int_0^\tau \phi_1^{(n)}(t)\, dM(t;\hat{\eta})\right|^2 \ge \chi^2_{k_1;\alpha}, \tag{7.3}$$

where $\phi_1^{(n)} = (\phi_1^{(n)}, \dots, \phi_{k_1}^{(n)})^{\mathrm{t}}$. The limiting distribution of $S_{k_1}^{\mathrm{op1}}(\tau)$, under $\{H_{1n}\}$, is $\chi_k^2(\delta_{\mathrm{op1}}^2)$, where $\delta_{\mathrm{op1}}^2 = \gamma_1^{\mathrm{t}}\gamma_1 = |\gamma_1|^2$. Employing the local asymptotic relative efficiency (LARE) criterion of Woolson and Sen (1974), we find the LARE of the oversmoothed test in (7.2) with respect to the test in (7.3) to be LARE($\{S_k^{\mathrm{ov}}\}$: $\{S_{k_1}^{\mathrm{op1}}\}$) $= R(k_1, k, \alpha)$ where, for $k_1, k_2 \in \{1, 2, \dots\}$, $\alpha \in (0,1)$,

$$R(k_1, k_2, \alpha) = \frac{\mathbf{P}\left\{\chi^2_{k_2+2} \ge \chi^2_{k_2;\alpha}\right\} - \alpha}{\mathbf{P}\left\{\chi^2_{k_1+2} \ge \chi^2_{k_1;\alpha}\right\} - \alpha}.$$

Values of $R(k_1, k_2, \alpha)$ were tabulated in Woolson and Sen (1974) for $\alpha \in \{0.10, 0.05, 0.01, 0.005\}$ and $k_1, k_2 \in \{1, 2, \dots, 10\}$, or can be easily obtained using a statistical computer package. Since $k > k_1$, it is immediate that LARE($\{S_k^{\mathrm{ov}}\}$: $\{S_{k_1}^{\mathrm{op1}}\}$) < 1, and, examining Tables 1 and 2 in Woolson and Sen (1974), there could be a substantial loss in efficiency with excessive oversmoothing. For example, with $\alpha = 0.05$, if $k = 8$ and $k_1 = 3$, then LARE($\{S_k^{\mathrm{ov}}\}$: $\{S_{k_1}^{\mathrm{op1}}\}$) $= 0.55$.

The second type of misspecification is "undersmoothing." In this situation we utilize a test which rejects H_0 whenever

$$S_{k_1}^{\mathrm{un}}(\tau) = \frac{1}{a_n^2}\left|\int_0^\tau \phi_1^{(n)}(t)\, dM(t;\hat{\eta})\right|^2 \ge \chi^2_{k_1;\alpha}, \tag{7.4}$$

where $k_1 < k$, when in reality $\gamma_2 = (\gamma_{k_1+1}, \dots, \gamma_k)^t$ is not equal to $\mathbf{0}$. But the appropriate test should reject H_0 whenever

$$S_k^{op2}(\tau) = \frac{1}{a_n^2}\left|\int_0^\tau \phi^{(n)}(t)\, dM(t;\hat{\eta})\right|^2 \geq \chi^2_{k;\alpha}. \tag{7.5}$$

The limiting distribution of $S_k^{op2}(\tau)$, under $\{H_{1n}\}$, is $\chi_k^2(\delta_{op2}^2)$, where $\delta_{op2}^2 = \gamma^t\gamma = |\gamma|^2$. On the other hand, the limiting distribution of $S_{k_1}^{un}(\tau)$, under H_{1n}, is $\chi_k^2(\delta_{un}^2)$, with $\delta_{un}^2 = \gamma_1^t\gamma = |\gamma_1|^2$. Consequently, the LARE of the undersmoothed test in (7.4) relative to that in (7.5) is LARE($\{S_k^{un}\}$: $\{S_{k_1}^{op2}\}$) $= R(k, k_1, \alpha)(|\gamma_1|^2/|\gamma|^2)$. Since $k > k_1$, then $R(k, k_1, \alpha) > 1$, and since $|\gamma_1|^2/|\gamma|^2 < 1$, it is possible for LARE($\{S_k^{un}\}$: $\{S_{k_1}^{op2}\}$) to exceed unity for a specific direction of approach, determined by γ, of the local alternatives to the null specification. For example, if $k = 2$, $k_1 = 1$, and $\alpha = 0.05$, then LARE($\{S_k^{un}\}$: $\{S_{k_1}^{op2}\}$) > 1 whenever $\gamma_1^2/\gamma_2^2 > 0.53$. The main problem, however, with the undersmoothed test is that it could be inconsistent and extremely inefficient for some other directions γ. This is evident by noting that when $\gamma_1 = \mathbf{0}$, and since $\gamma_2 \neq \mathbf{0}$, then LARE($\{S_k^{un}\}$: $\{S_{k_1}^{op2}\}$) $= 0$. This is a manifestation of the fact that for such a direction, the test is not even consistent. By comparing the consequences of these two types of misspecification based on their LAREs, undersmoothing is therefore a more serious problem than oversmoothing. On the other hand, if the directions or departures from the null specification determined by γ_2 and/or ϕ_2 are not of primary interest, then such undesirable consequences of undersmoothing may be ignored. Thus, it is evident that the determination of the smoothing order k is a very important problem, and this issue will be addressed in future work. It has been the subject of recent papers by Bickel and Ritov (1992), Kallenberg and Ledwina (1995) and Ledwina (1994) in classical settings, and Fan (1996) in relation to the use of wavelet transforms.

8. Concrete examples. *A total-time-on-test specification.* We illustrate the proposed smooth goodness-of-fit tests by considering specific choices of the ψ-process and testing the class of constant hazard rates and the two-parameter Weibull class of hazard rate functions. The first specification is one-dimensional with $\psi_1^{(0)}$ defined via

$$\psi_1^{(0)}(t) = \frac{r^{(0)}(t)}{r^{(0)}(\tau^0)} - \frac{1}{2}, \tag{8.1}$$

where $r^{(0)}(t) = \int_0^t y^{(0)}(s)\,ds$, $t \in [0, \tau^0]$ and $\tau^0 = \Lambda_0(\tau;\eta)$. For this choice, noting that $dr^{(0)}(t) = y^{(0)}(t)\,dt$, then

$$\langle \psi_1^{(0)}, \psi_1^{(0)}\rangle = \|\psi_1^{(0)}\|^2 = \int_0^{\tau^0} \left[\psi_1^{(0)}(t)\right]^2 y^{(0)}(t)\, dt = r^{(0)}(\tau^0)/12.$$

With $\langle f, g\rangle = \int_0^{\tau^0} fg\, dr^{(0)}$, let $\gamma_1 = \langle \psi_1^{(0)}, (\rho^{(0)})^t\rangle$ and $\mathbf{\Psi}^{-1} = (\langle \rho^{(0)}, (\rho^{(0)})^t\rangle)^{-1}$. Applying the orthogonalization method in (6.1) to $\psi_1^{(0)}(\cdot)$ yields $\phi_1^{(0)}(\cdot)$, which is of unit norm and is orthogonal to $\rho^{(0)}(\cdot)$ and given by

$$\phi_1^{(0)}(t) = \frac{\left[r^{(0)}(t)/r^{(0)}(\tau^0) - \frac{1}{2}\right] - \gamma_1 \boldsymbol{\Psi}^{-1}\rho^{(0)}(t)}{\sqrt{r^{(0)}(\tau^0)/12 - \gamma_1 \boldsymbol{\Psi}^{-1}\gamma_1^{t}}}. \tag{8.2}$$

The empirical version, $\phi_1^{(n)}(t)$, of $\phi_1^{(0)}(t)$, is obtained by replacing $r^{(0)}(t)$ by $R_0^R(t)/a_n^2$, where with $Y_0^R(t) = Y[\Lambda_0^{-1}(t;\hat{\eta})]$, $R_0^R(t) = \int_0^t Y_0^R(s)\,ds$; $y^{(0)}(t)$ by $Y_0^R(t)/a_n^2$; η by $\hat{\eta}$ and $\rho^{(0)}(t)$ by $\hat{\rho}^{(0)}(t) = \rho[\Lambda_0^{-1}(t;\hat{\eta});\hat{\eta}]$. Let $\hat{\tau}^0 = \Lambda_0(\tau;\hat{\eta})$. The asymptotic α-level smooth goodness-of-fit test of H_0: $\lambda_0(\cdot) \in \mathscr{C} = \{\lambda_0(\cdot;\eta)$: $\eta \in \Gamma\}$ generated by the ψ-process in (8.1) rejects H_0 whenever $\overline{S}(\tau;\hat{\eta}) \geq \chi^2_{1;\alpha}$, where with $N_0^R(t) = N[\Lambda_0^{-1}(t;\hat{\eta})]$ and noting that $\int_0^{\hat{\tau}^0}\hat{\rho}^{(0)}[dN_0^R - Y_0^R\,ds\} = 0$ by definition of $\hat{\eta}$,

$$\begin{aligned}\overline{S}(\tau;\hat{\eta}) &= \frac{1}{a_n^2}\left[\int_0^{\hat{\tau}^0}\frac{\left\{R_0^R(s)/R_0^R(\hat{\tau}^0) - \frac{1}{2}\right\}}{\sqrt{\left[R_0^R(\hat{\tau}^0)/\left(12a_n^2\right) - \hat{\gamma}_1\hat{\boldsymbol{\Psi}}^{-1}\hat{\gamma}_1^{t}\right]}}\right.\\ &\qquad\qquad \left.\times\left\{dN_0^R(s) - Y_0^R(s)\,ds\right\}\right]^2\\ &= \left[\frac{N_0^R(\hat{\tau}^0)}{R_0^R(\hat{\tau}^0)}\right]\frac{\left[Q^R(\hat{\tau}^0)\right]^2}{\left[1 - 12\hat{\Delta}(\hat{\tau}^0)\right]},\end{aligned} \tag{8.3}$$

where

$$Q^R(\hat{\tau}^0) = \sqrt{12N_0^R(\hat{\tau}^0)}\left\{\frac{1}{N_0^R(\hat{\tau}^0)}\int_0^{\hat{\tau}^0}\frac{R_0^R(t)}{R_0^R(\hat{\tau}^0)}\,dN_0^R(t) - \frac{1}{2}\right\}; \tag{8.4}$$

$$\begin{aligned}\hat{\Delta}(\hat{\tau}^0) &= \frac{\hat{\gamma}_1\hat{\boldsymbol{\Psi}}^{-1}\hat{\gamma}_1^{t}}{a_n^{-2}R_0^R(\hat{\tau}^0)}\\ &= \left(\int_0^{\hat{\tau}^0}\psi_1^{(n)}\hat{\rho}^{(0)}\,d\psi_1^{(n)}\right)^{t}\left(\int_0^{\hat{\tau}^0}\left(\hat{\rho}^{(0)}\right)^{\otimes 2}\,d\psi_1^{(n)}\right)^{-1}\\ &\qquad\times\left(\int_0^{\hat{\tau}^0}\psi_1^{(n)}\hat{\rho}^{(0)}\,d\psi_1^{(n)}\right).\end{aligned} \tag{8.5}$$

The first term in (8.3) converges in probability to one. The statistic $Q^R(\hat{\tau}^0)$ in (8.4) is a generalization of the normalized spacings test statistic introduced by Barlow, Bartholomew, Bremmer and Brunk (1972) [see Doksum and Yandell (1984)] when applied to the generalized residual processes (N_0^R, Y_0^R). On the other hand, the term $[1 - 12\hat{\Delta}(\hat{\tau}^0)]^{-1}$ represents the variance adjustment which is needed due to the estimation of the unknown nuisance parameter η by $\hat{\eta}$. It is interesting to note that, through the framework of this paper, the generalized normalized spacings test can be viewed as a score and a smooth goodness-of-fit test. Recall that such a test in classical settings was shown by Barlow, Bartholomew, Bremmer and Brunk (1972) and Doksum and Yandell (1984) to have good power against increasing failure rate (IFR) alternatives.

EXAMPLE 8.1. Suppose interest is in testing that $\lambda_0(\cdot)$ belongs to the constant hazard class, equivalently, the class of exponential distributions, $\mathscr{C}_E = \{\lambda_0(t;\eta) = \eta\colon \eta \in \Re\}$. Thus, $\rho^{(0)}(t;\eta) = \eta^{-1}$. Then

$$\gamma_1 = (1/\eta)\int_0^{\tau^0}\left[r^{(0)}(t)/r^{(0)}(\tau^0) - \tfrac{1}{2}\right]y^{(0)}(t)\,dt = 0,$$

hence in the test statistic in (8.3) we could set $\hat{\Delta}(\hat{\tau}^0)$ to its true value of zero. For this constant hazards model, the estimation of η by $\hat{\eta}$ has *no* effect on the asymptotic distribution. This adaptiveness property was also observed in Aban and Peña (1998) and Baltazar–Aban and Peña (1995). The resulting smooth goodness-of-fit test is therefore just the generalized normalized spacings test, applied to the generalized residual processes (N_0^R, Y_0^R).

EXAMPLE 8.2. Suppose instead that one is interested in testing that $\lambda_0(\cdot)$ belongs to the two-parameter Weibull class of hazard rate functions $\mathscr{C}_{2W} = \{\lambda_0(t;\alpha,\eta) = (\alpha\eta)(\eta t)^{\alpha-1}\colon\ \alpha > 0,\ \eta > 0\}$. Straightforward calculations yield $\rho^{(0)}(t;\alpha,\eta) = [\alpha/\eta, (1/\alpha)(1+\log t)]^{\mathrm{t}}$, $\gamma_1 = (r^{(0)}(\tau^0)/\alpha)[0, C(\tau^0)]^{\mathrm{t}}$ and

$$\mathbf{\Psi} = r^{(0)}(\tau^0)\begin{bmatrix} \dfrac{\alpha^2}{\eta^2} & \dfrac{1}{\eta}E^{(1)}(\tau^0) \\ \dfrac{1}{\eta}E^{(1)}(\tau^0) & \dfrac{1}{\alpha^2}E^{(2)}(\tau^0)\end{bmatrix},$$

where, with $D^{(j)}(\tau^0) = \int_0^{\tau^0}(\log w)^j\,d\psi_1^{(0)}(w)$, $(j = 0,1,2)$, $E^{(1)}(\tau^0) = D^{(0)}(\tau^0) + D^{(1)}(\tau^0)$, $E^{(2)}(\tau^0) = D^{(0)}(\tau^0) + 2D^{(1)}(\tau^0) + D^{(2)}(\tau^0)$ and $C(\tau^0) = \int_0^{\tau^0}(1 + \log w)\psi_1^{(0)}(w)\,d\psi_1^{(0)}(w)$. Also, define $V(\tau^0) = E^{(2)}(\tau^0) - [E^{(1)}(\tau^0)]^2$. The empirical versions of these quantities are $\hat{D}^{(j)}(\hat{\tau}^0)$'s, $\hat{E}^{(j)}(\hat{\tau}^0)$'s, $\hat{C}(\hat{\tau}^0)$ and $\hat{V}(\hat{\tau}^0)$, which are found by replacing $\psi_1^{(0)}(\cdot)$ with $\psi_1^{(n)}(\cdot) = R_0^R(\cdot)/R_0^R(\hat{\tau}^0) - 1/2$ in their definitions. The test statistic is then obtained from (8.3) by setting $\hat{\Delta}(\hat{\tau}^0) = \hat{C}(\hat{\tau}^0)^2/\hat{V}(\hat{\tau}^0)$.

It is interesting to see what becomes of $\bar{S}(\tau;\hat{\alpha},\hat{\eta})$ in the *complete* data situation where $y(t) = \exp\{-\Lambda(t)\}$ so $y^{(0)}(t) = \exp\{-t\}$, and when $\tau = \tau^0 = \infty$. In this case $r^{(0)}(t) = 1 - \exp\{-t\}$, $\psi_1^{(0)}(t) = 1/2 - \exp\{-t\}$, and by simple calculations and with $\gamma = -0.5772\ldots$ being Euler's constant, $D^{(0)}(\infty) = 1$, $D^{(1)}(\infty) = -\gamma$, $D^{(2)}(\infty) = \pi^2/6 + \gamma^2$, $E^{(1)}(\infty) = 1 - \gamma$, $E^{(2)}(\infty) = \pi^2/6 + (1-\gamma)^2$, $C(\infty) = (\log 2)/2$ and $V(\infty) = \pi^2/6$. Consequently, one finds that for this complete data setting, the true variance adjustment factor in (8.3) is

$$[1 - 12\Delta(\infty)]^{-1} = \left[1 - \frac{18(\log 2)^2}{\pi^2}\right]^{-1} = 8.0802\ldots;$$

clearly a nonnegligible factor. For this Weibull model, an adjustment on the variance is therefore needed when one applies the generalized normalized spacings test on the generalized residual processes (N_0^R, Y_0^R), in contrast with the case of constant hazard rate functions.

A polynomial-type specification. The next specification is $\psi_k^{(n)}(t) = \psi_k^{(0)}(t) = \mathbf{P}_k = (1, t, \dots, t^{k-1})^{\mathrm{t}}$. In contrast to the preceding specification where an orthogonalized version was utilized, for this polynomial-type specification we forego such orthogonalization. The resulting tests from these specifications were used in the numerical studies reported in Section 9.

Recalling that $N_0^R(t) = N[\Lambda_0^{-1}(t;\hat{\eta})]$, $Y_0^R(t) = Y[\Lambda_0^{-1}(t;\hat{\eta}]$, $\hat{\rho}^{(0)}(t) = \rho[\Lambda_0^{-1}(t;\hat{\eta});\hat{\eta}]$ and $\hat{\tau}^{(0)} = \Lambda_0(\tau;\hat{\eta})$, let $dM_0^R = dN_0^R - Y_0^R\,dt$, $dL_0^R = \frac{1}{2}[dN_0^R + Y_0^R\,dt]$, and $\overline{\mathbf{\Sigma}}_{11.2}(\tau;\hat{\eta}) = (1/a_n^2)\int_0^{\hat{\tau}^0}\mathbf{P}_k^{\otimes 2}\,dL_0^R - \overline{\Delta}(\hat{\tau}^0)$ where

$$\hat{\Delta}(\hat{\tau}^0) = \frac{1}{a_n^2}\left(\int_0^{\hat{\tau}^0}\mathbf{P}_k(\hat{\rho}^{(0)})^{\mathrm{t}}\,dL_0^R\right)\left(\int_0^{\hat{\tau}^0}(\hat{\rho}^{(0)})^{\otimes 2}\,dL_0^R\right)^{-1} \times\left(\int_0^{\hat{\tau}^0}\hat{\rho}^{(0)}(\mathbf{P}_k)^{\mathrm{t}}\,dL_0^R\right).$$

Then the test statistic in (3.2) becomes

$$(8.6)\qquad \overline{S}_k(\tau;\hat{\eta}) = \frac{1}{a_n^2}\left[\int_0^{\hat{\tau}^0}\mathbf{P}_k\,dM_0^R\right]^{\mathrm{t}}\left[\overline{\mathbf{\Sigma}}_{11.2}(\tau;\hat{\eta})\right]^{-}\left[\int_0^{\hat{\tau}^0}\mathbf{P}_k\,dM_0^R\right],$$

and the asymptotic α-level smooth goodness-of-fit test rejects H_0: $\lambda_0(\cdot) \in \mathscr{C} = \{\lambda_0(\cdot;\eta): \eta \in \Gamma\}$ whenever $\overline{S}_k(\tau;\hat{\eta}) \geq \chi^2_{\hat{k}^*;\alpha}$, where $\hat{k}^* = \operatorname{rank}[\overline{\mathbf{\Sigma}}_{11.2}(\tau;\hat{\eta})]$.

To achieve more concreteness, assume that the incomplete data mechanism is the random censorship model, where $T_1, T_2, \dots, T_n$ are i.i.d. failure times with common hazard rate function $\lambda_0(\cdot)$, and $C_1, C_2, \dots, C_n$ are i.i.d. censoring variables with some common hazard rate function. With $Z_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i \leq C_i\}$ $(i = 1, \dots, n)$, the observable processes are $N(t) = \sum_{i=1}^n I\{Z_i \leq t,\ \delta_i = 1\}$ and $Y(t) = \sum_{i=1}^n I\{Z_i \geq t\}$. For testing H_0: $\lambda_0 \in \mathscr{C} = \{\lambda_0(\cdot;\eta): \eta \in \Gamma\}$, denote by $R_i = \Lambda_0(Z_i;\hat{\eta})$, $(i = 1, \dots, n)$, the Cox–Snell generalized residuals [Cox and Snell (1968)]. The residual processes become

$$N_0^R(t) = N\left[\Lambda_0^{-1}(t;\hat{\eta})\right] = \sum_{i=1}^n I\{R_i \leq t,\ \delta_i = 1\};$$
$$Y_0^R(t) = Y\left[\Lambda_0^{-1}(t;\hat{\eta})\right] = \sum_{i=1}^n I\{R_i \geq t\}.$$

For this random censorship model, the sequence of normalizing constants $\{\alpha_n\colon n = 1, 2, \dots\}$ has $a_n^2 = n$, $n = 1, 2, \dots$. If τ is such that $Z_{(n)} \leq \tau$, or equivalently, $R_{(n)} \le \hat{\tau}^0$, it is straightforward to see that

$$\int_0^{\hat{\tau}^0} \mathbf{P}_k \, dM_0^R = \left[\sum_{i=1}^n R_i^{m-1} \left(\delta_i - \frac{R_i}{m} \right) \right]_{m=1,2,\dots,k}; \tag{8.7}$$

$$\begin{aligned} &\int_0^{\hat{\tau}^0} (\mathbf{P}_k)^{\otimes 2} \, dL_0^R \\ &\quad = \frac{1}{2} \left[\sum_{i=1}^n R_i^{m_1+m_2-2} \left(\delta_i + \frac{R_i}{m_1+m_2-1} \right) \right]_{m_1, m_2 = 1,2,\dots,k}. \end{aligned} \tag{8.8}$$

On the other hand, the other two terms in the covariance matrix may be in closed forms depending on the form of $\hat{\rho}^{(0)}$.

EXAMPLE 8.3. If $\mathscr{C} = \mathscr{C}_E$, the constant hazards class in Example 8.1, then $\hat{\rho}^{(0)}(t) = 1/\hat{\eta}$ and

$$\int_0^{\hat{\tau}^0} \mathbf{P}_k \left(\hat{\rho}^{(0)} \right)^{\mathrm{t}} dL_0^R = \frac{1}{\hat{\eta}} \frac{1}{2} \sum_{i=1}^n \left[R_i^{m-1} \left(\delta_i + \frac{R_i}{m} \right) \right]_{m=1,\dots,k};$$

$$\int_0^{\hat{\tau}^0} \left(\hat{\rho}^{(0)} \right)^{\otimes 2} dL_0^R = \frac{1}{\hat{\eta}^2} \frac{1}{2} \sum_{i=1}^n (\delta_i + R_i).$$

Thus, under this random censorship model and when testing the constant hazards class, the relevant estimator of the covariance matrix is the $k \times k$ matrix

$$\begin{aligned} \overline{\boldsymbol{\Sigma}}_{11.2} = \frac{1}{2a_n^2} \Bigg\{ & \sum_{i=1}^n \left[R_i^{m_1+m_2-2} \left(\delta_i + \frac{R_i}{m_1+m_2-1} \right) \right]_{m_1, m_2 = 1,\dots,k} \\ & - \left[\sum_{i=1}^n (\delta_i + R_i) \right]^{-1} \left[\sum_{i=1}^n R_i^{m-1} \left(\delta_i + \frac{R_i}{m} \right) \right]_{m=1,\dots,k}^{\otimes 2} \Bigg\}. \end{aligned} \tag{8.9}$$

The test statistic $\overline{S}_k(\tau; \hat{\eta})$ in (8.6) is then obtained using (8.7) and (8.9). It is interesting to note that this statistic is just a function of the generalized residuals (R_i, δ_i), $(i = 1, \dots, n)$, and depends on the estimator of the nuisance parameter, $\hat{\eta}$, only through the residuals. Furthermore, there is no need to estimate the nuisance parameters associated with the censoring mechanism as this was circumvented through the use of the optional variation and the predictable variation processes in estimating the covariance matrix.

EXAMPLE 8.4. If $\mathscr{C} = \mathscr{C}_{2W}$, the two-parameter Weibull class considered in Example 8.2, the smooth goodness-of-fit test statistic generated by the polynomial specification $\psi^{(n)} = \mathbf{P}_k$ is also in computable form. For $m = 1, \dots, k$

and $l = 0, 1, 2$, let

$$\hat{D}^{(m,l)} = \int_0^{\hat{\tau}^0} t^{m-1} (\log t)^l \, dL_0^R;$$

$$\hat{E}^{(m,l)} = \int_0^{\hat{\tau}^0} t^{m-1} (1 + \log t)^l \, dL_0^R = \sum_{j=0}^{l} \binom{l}{j} \hat{D}^{(m,j)}.$$

Then it is routine to show that

$$\int_0^{\hat{\tau}^0} \mathbf{P}_k (\hat{\rho}^{(0)})^{\mathrm{t}} \, dL_0^R = [(\hat{\gamma}_m)_{m=1,\dots,k}] = \left[\left(\frac{\hat{\alpha}}{\hat{\eta}} \hat{E}^{(m,0)}, \frac{1}{\hat{\alpha}} \hat{E}^{(m,1)} \right)_{m=1,\dots,k} \right];$$

$$\int_0^{\hat{\tau}^0} (\hat{\rho}^{(0)})^{\otimes 2} \, dL_0^R = \hat{\mathbf{\Psi}} = \begin{bmatrix} \left(\frac{\hat{\alpha}}{\hat{\eta}} \right)^2 \hat{E}^{(1,0)} & \frac{1}{\hat{\eta}} \hat{E}^{(1,1)} \\ \frac{1}{\hat{\eta}} \hat{E}^{(1,1)} & \frac{1}{\hat{\alpha}^2} \hat{E}^{(1,2)} \end{bmatrix}.$$

The $k \times k$ covariance adjustment term is therefore

$$\overline{\Delta}(\hat{\tau}^0) = \left[\left(\hat{\gamma}_i \hat{\mathbf{\Psi}}^{-1} \hat{\gamma}_j^{\mathrm{t}} \right)_{i,j=1,\dots,k} \right], \tag{8.10}$$

and explicit expressions for computing the $\hat{\gamma}_i \hat{\mathbf{\Psi}}^{-1} \gamma_j^{\mathrm{t}}$'s are

$$\hat{\gamma}_i \hat{\mathbf{\Psi}}^{-1} \hat{\gamma}_j^{\mathrm{t}} = \Big[\hat{E}^{(i,0)} \hat{E}^{(1,2)} \hat{E}^{(j,0)} - \hat{E}^{(i,0)} \hat{E}^{(1,1)} \hat{E}^{(j,1)} - \hat{E}^{(i,1)} \hat{E}^{(1,1)} \hat{E}^{(j,0)} + \hat{E}^{(i,1)} \hat{E}^{(1,0)} \hat{E}^{(j,1)} \Big] \Big[\hat{E}^{(1,0)} \hat{E}^{(1,2)} - [\hat{E}^{(1,1)}]^2 \Big]^{-1}.$$

Furthermore, recalling that the generalized residuals in this Weibull case are defined via $R_i = (\hat{\eta} Z_i)^{\hat{\alpha}}$, if $\hat{\tau}^0 \geq R_{(n)}$, then straightforward calculations and integration-by-parts yield the computational forms for the $\hat{D}^{(m,l)}$'s given by

$$\begin{aligned} \hat{D}^{(m,l)} &= \frac{1}{2} \sum_{i=1}^{n} \left\{ \delta_i R_i^{m-1} (\log R_i)^l + \frac{(-1)^l}{m^{l+1}} \int_{-\log(R_i^m)}^{\infty} u^l \exp(-u) \, du \right\} \\ &= \frac{1}{2} \sum_{i=1}^{n} R_i^{m-1} \left\{ \delta_i (\log R_i)^l + \frac{(-1)^l l! R_i}{m^{l+1}} \sum_{j=0}^{l} \frac{[-m \log(R_i)]^j}{j!} \right\}. \end{aligned}$$

The test statistic in (8.6) can then be formed from (8.7), (8.8) and (8.10). Again, note that the smooth goodness-of-fit test is just a function of the right-censored generalized residuals (R_i, δ_i)'s, and depends on the estimator $(\hat{\alpha}, \hat{\eta})$ of the nuisance parameter vector (α, η) only through the generalized residuals.

9. Numerical studies of levels and powers. *Achieved levels of tests.* Simulation studies were performed to ascertain the levels and powers, for small and moderate sample sizes, of the smooth goodness-of-fit tests arising from the specifications $\mathbf{P}_k = (1, t, \dots, t^{k-1})^t$. The null distributions considered in the simulations were the exponential and the Weibull distributions. The right-censored data were generated by assuming that the censoring hazard function was proportional to the failure time hazard function, the so-called Koziol–Green model of random censorship [Koziol and Green (1976) and Chen, Hollander and Langberg (1982)]. Theoretical proportions of 0.75 and 0.50 uncensored values were specified. Tests considered in the simulations were those based on $\overline{S}_2$, $\overline{S}_3$, $\overline{S}_4$ and $\overline{S}_5$, arising from the specification $\mathbf{P}_k$ and described in the preceding section. Three sample sizes were considered: $n = 20$, $n = 50$ and $n = 100$, while the levels of the tests were set to 1%, 5% and 10%. The computer code was written in Fortran, and random number generators and a subroutine for obtaining generalized inverses from the IMSL Library (1987) were utilized. The programs were ran on a Silicon Graphics Power Challenge workstation with four processors running the SGI Irix 6.2 operating system at Bowling Green State University. For each combination of simulation parameters, 2000 replications were performed.

Tables 1 and 2 summarize the results of the achieved levels of the tests for the exponential null and the Weibull null distributions, respectively. To conserve space, only those associated with asymptotic levels of 5% and 10% are presented. For each of these null distributions, two sets of parameter values were specified. Examining these tables, one finds that the achieved

TABLE 1

Simulated levels of the asymptotic smooth goodness-of-fit tests based on the specification $\psi_k^{(n)} = \mathbf{P}_k = (1, t, \dots, t^{k-1})^t$ *under an exponential null distribution**

Null dist.		**Exponential(η)**							
Parameters		**$\eta = 2$**				**$\eta = 5$**			
% Uncensored		**75%**		**50%**		**75%**		**50%**	
n	**Level *k***	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
20	2	4.30	10.10	6.40	13.15	6.25	12.10	6.55	11.35
	3	4.20	9.90	4.95	10.85	5.20	11.30	4.45	10.65
	4	6.55	12.40	4.95	12.70	6.15	12.60	4.70	12.60
	5	5.45	12.25	4.05	10.15	5.25	11.90	3.30	9.20
50	2	4.65	9.75	6.65	12.50	5.00	9.45	5.60	11.30
	3	5.45	10.60	5.50	11.25	4.35	9.75	4.95	11.35
	4	6.55	11.40	5.50	10.75	5.10	10.90	5.85	11.50
	5	6.40	12.10	5.00	11.10	5.20	11.45	4.20	11.10
100	2	4.90	9.65	4.75	9.60	4.45	8.90	4.35	9.45
	3	4.55	9.35	4.35	9.75	4.65	9.50	4.25	9.30
	4	5.70	10.80	5.30	10.10	5.10	10.85	4.80	9.95
	5	5.75	12.15	4.90	10.35	5.30	9.95	4.75	9.45

* The number of replications was $m = 2000$, and the right-censoring model is the Koziol–Green model.

TABLE 2

Simulated levels of the asymptotic smooth goodness-of-fit tests based on the specification $\psi_k^{(n)} = \mathbf{P}_k = (1, t, \ldots, t^{k-1})^{\mathrm{t}}$ *under a two-parameter Weibull null distribution**

Null dist.		Weibull(α,η)							
Parameters		$(\alpha,\eta)=(2,1)$				$(\alpha,\eta)=(3,2)$			
% Uncensored		75%		50%		75%		50%	
n	Level k	5%	10%	5%	10%	5%	10%	5%	10%
20	2	3.80	8.55	6.20	12.35	5.05	10.00	6.85	13.40
	3	5.95	12.65	5.60	11.80	6.60	14.30	6.65	14.40
	4	5.05	11.90	3.70	10.50	6.45	13.85	5.40	12.30
	5	3.90	11.30	2.75	7.35	5.55	12.75	3.55	9.45
50	2	4.30	9.10	4.80	10.05	4.60	9.20	6.20	11.10
	3	5.40	12.25	5.15	11.00	6.30	13.20	5.70	12.00
	4	4.80	10.85	4.80	11.25	6.10	12.85	5.30	11.35
	5	5.25	12.05	3.45	8.75	6.70	13.70	4.60	9.85
100	2	3.90	8.75	4.95	9.60	4.20	8.15	4.80	9.85
	3	5.75	11.20	4.65	11.00	5.15	10.60	5.25	10.80
	4	5.55	11.00	4.15	9.35	5.00	10.20	4.30	10.10
	5	6.00	11.75	4.65	10.55	5.80	11.20	5.30	10.20

* The number of replications was $m = 2000$, and the right-censoring model is the Koziol–Green model.

levels are consistent with the specified asymptotic levels, especially for $n = 50$ and $n = 100$. The achieved levels for $n = 20$ under the Weibull distribution with $(\alpha, \eta) = (3, 2)$ at the 75% and 50% uncensored levels were somewhat anticonservative, but overall it can be concluded that the asymptotic approximations is acceptable for moderate sample sizes, at least for the exponential and Weibull-distributed failure times. It should be mentioned that the use of the combined estimator of the covariance matrix $\overline{\Sigma}$ in (3.1), which is a convex combination of the estimator of the predictable quadratic variation process and the estimator of the optional variation process, considerably improved the approximations. In previous simulations [see simulation results in Peña (1998) when testing for simple hypotheses] using either of the estimators, but not a combined estimator, the finite sample behavior of tests were quite anticonservative for small to moderate sample sizes, possibly owing to the instability of both estimators. Thus, the use of the combined estimator, even in other settings, seems worthy of serious consideration and warrants further theoretical and empirical investigations. We intend to provide a theoretical justification for the use of this combined estimator in future work.

Achieved powers of tests. Simulations were also ran to examine the achieved powers of the tests against specific alternatives and to partially address the issue of an appropriate smoothing order k. In particular, the null hypothesis of an exponential distribution was tested when the failure times were generated by a Weibull distribution and a gamma distribution. Also, the test for the Weibull distribution was examined when the underlying distribution of the failure times was a gamma distribution. Except for the generation of the failure times, the computer program and the parameters for the level and power simulations were similar. The smooth goodness-of-fit tests considered were those based on $\bar{S}_k$ ($k = 2, 3, 4, 5$), arising from the specification $\mathbf{P}_k$, the sample size utilized was $n = 100$, the number of replications was 2000 and the achieved powers of the asymptotic 5%-level tests are reported here, although the achieved powers of the 1%-level and 10%-level asymptotic tests were also available from the simulation outputs. Summaries of the achieved powers of the tests are presented in Table 3 for testing exponentiality when the true failure time distribution is Weibull with shape parameter α and scale parameter $\eta = 1$; Table 4 for testing exponentiality when the true failure time distribution is gamma with shape parameter α and scale parameter $\eta = 1$ and Table 5 for testing the Weibull distribution when the true failure time distribution is gamma with shape parameter α and scale parameter $\eta = 1$.

Examination of these tables reveals that the appropriate smoothing order k depends on the type of alternatives being considered. For instance, when testing the exponential distribution but with the true failure time distribution being Weibull with shape parameter α, then if $\alpha < 1$ the tests based on $\bar{S}_5$ and $\bar{S}_3$ are most preferable among the four tests considered, while if

TABLE 3

*Simulated powers of the asymptotic 5%-level smooth goodness-of-fit tests based on the specification $\psi_k^{(n)} = \mathbf{P}_k = (1, t, \ldots, t^{k-1})^t$ for testing an exponential null distribution when the true distribution is Weibull with shape parameter α and scale parameter $\eta = 1$**

	Test statistic			
α	$\bar{S}_2$	$\bar{S}_3$	$\bar{S}_4$	$\bar{S}_5$
0.60	97.20	99.25	99.30	99.70
0.70	77.50	89.15	86.40	90.00
0.80	37.95	50.90	44.70	49.65
0.85	21.50	31.25	26.75	30.35
0.90	10.90	14.95	12.45	16.20
0.95	5.65	7.60	7.50	7.60
(H_0) 1.00	4.60	4.45	5.25	5.70
1.05	7.85	5.65	7.10	6.00
1.10	15.15	10.40	11.05	9.90
1.15	27.70	19.90	20.60	18.30
1.20	41.50	31.85	32.85	27.20
1.35	81.90	76.70	77.00	70.95
1.50	97.05	95.70	95.95	94.80
1.75	99.95	99.95	100.00	99.95

* The number of replications was $m = 2000$.

TABLE 4

*Simulated powers of the asymptotic 5%-level smooth goodness-of-fit tests based on the specification $\psi_k^{(n)} = \mathbf{P}_k = (1, t, \ldots, t^{k-1})^t$ for testing an exponential null distribution when the true distribution is gamma with shape parameter α and scale parameter $\eta = 1$**

	Test statistic			
α	$\bar{S}_2$	$\bar{S}_3$	$\bar{S}_4$	$\bar{S}_5$
0.50	83.00	95.50	94.60	97.45
0.60	57.80	80.05	75.50	81.95
0.80	14.30	22.90	19.50	22.65
0.90	6.20	8.80	8.50	9.55
(H_0) 1.00	5.25	5.65	5.65	6.00
1.05	5.20	4.45	6.25	5.00
1.10	9.55	6.60	8.80	7.80
1.15	12.10	9.30	11.65	10.05
1.20	17.65	11.75	15.30	12.45
1.35	35.00	29.45	35.95	30.70
1.50	58.55	55.05	61.85	56.50
1.75	84.50	86.00	90.30	86.90
2.00	95.25	96.65	98.40	97.80
4.00	100.00	100.00	100.00	100.00

* The number of replications was $m = 2000$.

TABLE 5

*Simulated powers of the asymptotic 5%-level smooth goodness-of-fit tests based on the specification $\psi_k^{(n)} = \mathbf{P}_k = (1, t, \ldots, t^{k-1})^t$ for testing a Weibull null distribution when the true distribution is gamma with shape parameter α and scale parameter $\eta = 1$**

	Test statistic			
α	$\bar{S}_2$	$\bar{S}_3$	$\bar{S}_4$	$\bar{S}_5$
0.40	32.00	22.35	17.45	14.60
0.50	18.15	12.15	10.20	9.60
0.75	6.90	5.40	4.85	5.15
(H_0) 1.00	4.45	6.00	5.00	5.95
1.50	6.70	10.30	8.60	9.40
2.00	8.60	14.80	13.95	14.40
4.00	22.00	35.40	30.00	31.50
6.00	26.55	41.80	35.50	36.30
8.00	30.25	48.30	41.05	43.25
10.00	34.55	53.00	46.85	49.25
12.00	36.75	55.75	49.15	51.50
15.00	40.95	59.90	52.80	54.20
20.00	42.35	63.05	56.45	58.50
25.00	44.35	65.15	57.70	61.25

* The number of replications was $m = 2000$.

$\alpha > 1$, then the test based on $\bar{S}_2$ is preferable. The test based on $\bar{S}_3$ however may serve as a viable omnibus test if one does not possess a good prior knowledge of the value of α relative to one. If the failure times on the other hand have a gamma distribution with shape parameter α, then the tests based on $\bar{S}_5$ and $\bar{S}_3$ have the most power when $\alpha < 1$, while when $\alpha > 1$ then the $\bar{S}_4$-based test achieved the most power; but if one does not possess a good prior knowledge of the value of α relative to one, then the $\bar{S}_4$-based test could serve as an omnibus test. When testing for the Weibull distribution with the failure times having a gamma distribution, the $\bar{S}_2$-based test is most preferred for $\alpha < 1$, and the $\bar{S}_3$-based test achieved the most power when $\alpha > 1$. If one does not have a good idea on the value of α relative to one, then the $\bar{S}_3$-based test could serve as an omnibus test. Through the results of these modest simulation studies, it is therefore seen that the proposed class of smooth goodness-of-fit tests, which is simple to implement and depends only on the generalized residuals in the case of the random censorship model, provides a class of omnibus tests. Further power comparisons of these tests with smooth tests arising from other specifications of the ψ-process and other existing tests are however warranted in future studies to be able to make definitive conclusions concerning their potential.

10. Applications to real data sets. To illustrate the usefulness of the proposed smooth goodness-of-fit tests in applied work, we examined three data sets that have been considered in the literature. The first data set consisted of 20 uncensored operational lifetimes of bearings (in hours) which was presented and analyzed in Angus (1982) [see also Rayner and Best (1989, page 90)]. Using the smooth goodness-of-fit tests arising from the specification $\mathbf{P}_k$ $(k = 2, 3, 4, 5)$, we tested the null hypothesis that the operational times were generated by an exponential distribution. The values of the test statistics, together with their associated p-values were $\bar{S}_2 = 10.42(p = 0.0012)$, $\bar{S}_3 = 10.72(p = 0.0047)$, $\bar{S}_4 = 11.21(p = 0.0107)$ and $\bar{S}_5 = 12.60(p = 0.0134)$, so the hypothesis of exponential distribution is rejected. The conclusion is consistent with the results of Angus (1982) and those of Rayner and Best (1989), pages 90 and 91. The Weibull model null hypothesis for the underlying distribution of those operational times was also tested. The maximum likelihood estimates of the shape and scale parameters under the Weibull model were $\hat{\alpha} = 2.103575$ and $\hat{\eta} = 0.000103$, and the resulting values of the test statistics are $\bar{S}_2 = 0.66(p = 0.4166)$, $\bar{S}_3 = 0.71(p = 0.7012)$, $\bar{S}_4 = 0.94(p = 0.8154)$ and $\bar{S}_5 = 5.05\ (p = 0.2821)$; hence the Weibull null hypothesis could not be rejected. The Weibull model may therefore be a viable model for these operational times, especially so since the p-values are quite high for all four tests. The Weibull model was not tested in Rayner and Best (1989) because the smooth goodness-of-fit test arising from the usual density-based formulation was not yet available [see page 139 of Rayner and Best (1989)]. However, using their component analysis of their smooth goodness-of-fit test for exponentiality, they suggested that an alternative model for the operational times is a chi-squared distribution with seven degrees-of-freedom.

The second data set considered consisted of randomly right-censored failure times (in months) arising from a study of cisplatin-based chemotherapy on lung cancer patients found in Gatsonis, Hsieh and Korwar (1985) [see also Peña (1998) for the results of tests on simple null hypotheses]. There were 86 observations with 23 of them right-censored. The hypothesis of exponentiality with an unspecified scale parameter was tested using the tests based on $\bar{S}_k$ ($k = 2, 3, 4, 5$). Under the exponential model, the maximum likelihood estimate of the scale parameter was $\hat{\eta} = 0.06635$, so the estimate of the mean failure time was 15.07 months. The values of the test statistics, together with their p-values for the exponentiality test were $\bar{S}_2 = 1.92(p = 0.1661)$, $\bar{S}_3 = 1.94(p = 0.3788)$, $\bar{S}_4 = 7.56(p = 0.0561)$ and $\bar{S}_5 = 12.85(p = 0.0121)$. Note that if one is to use either of the tests based on $\bar{S}_2$, $\bar{S}_3$ or $\bar{S}_4$, then the exponential model will not be rejected, but if one utilizes the test based on $\bar{S}_5$, then it will be rejected at the 5% level. These results seem to suggest that the true hazard rate function is quite close to a constant function, but possesses nonnegligible high frequency terms. In particular, note the steep increase in values from $\bar{S}_4$ to $\bar{S}_5$ relative to the change in values from $\bar{S}_2$ to that of $\bar{S}_4$. The Weibull distribution was also tested for this data set. Based on the results of the tests [e.g., $\bar{S}_3 = 8.35(p = 0.0153)$ and the observed values of $\bar{S}_4$ and $\bar{S}_5$ both indicate rejection of the null hypothesis], the Weibull distribution is not a viable model for this failure time data.

Finally, the two-sample data set arising from autologous (auto) and allogeneic (allo) bone marrow transplants described and presented in Section 1.9 of Klein and Moeschberger (1997) was analyzed using the smooth goodness-of-fit tests in this paper. This data set was extensively analyzed in Klein and Moeschberger (1997). In particular, it was utilized to illustrate fitting parametric models (their Example 12.1 on page 377) and diagnostic plots for checking parametric regression models (their Section 12.5). The "auto" sample consisted of 51 observations of which 23 were right-censored, while the "allo" sample consisted of 50 observations of which 28 were right-censored. For each of these sample data, the hypothesis of exponentiality of the underlying failure time distributions were tested. The values of the smooth goodness-of-fit test statistics are as follows: for the "allo" group, $\bar{S}_2 = 22.33(p \approx 0)$, $\bar{S}_3 = 24.54(p \approx 0)$, $\bar{S}_4 = 24.58(p \approx 0)$ and $\bar{S}_5 = 24.65(p \approx 0)$, so the exponential model is not viable for the "allo" group failure times. For the "auto" group, the values of the test statistics were $\bar{S}_2 = 2.36(p = 0.1247)$, $\bar{S}_3 = 2.96(p = 0.2274)$, $\bar{S}_4 = 11.98(p = 0.0075)$ and $\bar{S}_5 = 12.37(p = 0.0148)$. Thus, on the basis of the observed value of $\bar{S}_3$, the exponential model is not also a viable model for describing the failure times for this "auto" group.

The viability of the Weibull distribution as a model for the failure times of each group was also tested. For the "allo" group, the observed values of the test statistics were $\bar{S}_2 = 8.34(p = 0.0039)$, $\bar{S}_3 = 0.12(p = 0.0105)$, $\bar{S}_3 = 10.16(p = 0.0173)$ and $\bar{S}_5 = 10.16(p = 0.0378)$. These results show that the Weibull model is not viable. However, relative to the exponential distribution, the Weibull model provides a better fit to the data. For the "auto" group, the test for the Weibull distribution resulted in the values $\bar{S}_2 = 2.78(p = 0.0952)$, $\bar{S}_3 = 5.41(p = 0.0670)$, $\bar{S}_4 = 10.72(p = 0.0133)$ and $\bar{S}_5 = 12.07(p = 0.0168)$. These values, especially those for $\bar{S}_4$ and $\bar{S}_5$, indicate that the Weibull model is not a viable model of the failure times for this "auto" group, though comparing the values of the test statistics for the tests of exponentiality and Weibullness suggests that, *ignoring other possible distributional models*, the exponential model provides as good a fit as the Weibull distribution [see a similar conclusion stated by Klein and Moeschberger (1997), page 377]. It should be mentioned also that through hazard plotting, Klein and Moeschberger [(1997), pages 390 and 391] concluded that the exponential, Weibull and lognormal models are reasonable models for the two groups with the exception of the exponential model for the "allo" group. Their conclusions were based on the fact that the hazard plots appear to be linear. It should be pointed out, however, that because model nuisance parameters are being estimated, these hazard plots have a tendency to be more linear if no adjustments are made for the effect of such estimation [Baltazar–Aban and Peña (1995)]. From a theoretical point of view, the tendency of such diagnostic plots to become more linear is manifested by the theoretically observed decrease in the variance of test statistics when applied to quantities (such as the generalized residuals) with estimated parameters [see also the results of Durbin (1975) and Stephens (1976) in more classical settings]. Thus, formal goodness-of-fit tests, such as those provided by the smooth goodness-of-fit tests, should be performed and sole reliance on diagnostic plots using generalized residuals, especially when nuisance parameters were estimated, should be avoided to prevent misleading conclusions.

11. Summary and concluding remarks. For a counting process $\{N(t),\ t \in \mathscr{T}\}$ with compensator process $\{A(t),\ t \in \mathscr{T}\}$ satisfying $A(t) = \int_0^t Y(s)\lambda(s)\,ds$, where $\{Y(t),\ t \in \mathscr{T}\}$ is an observable predictable process and $\lambda(\cdot)$ is an unknown hazard rate function, the problem of testing whether $\lambda(\cdot)$ belongs to some parametric family of hazard rate function $\mathscr{C} = \{\lambda_0(\cdot;\eta)$: $\eta \in \Gamma \subseteq \Re^q\}$ was considered. A procedure for extending Neyman's (1937) smooth goodness-of-fit test was formulated and developed, with the formulation being more suitable for models specified in terms of hazard or intensity functions, now typical in survival and reliability models. The extension was obtained by embedding $\mathscr{C}$ in the class $\mathscr{A}_k = \{\lambda_k(\cdot;\theta,\eta)\colon \eta \in \Gamma, \theta \in \Re^k\}$, where $\lambda_k(\cdot;\theta,\eta) = \lambda_0(\cdot;\eta)\exp\{\theta^{\mathrm{t}}\psi(\cdot;\eta)\}$, and $\psi(\cdot;\eta) = (\psi_1(\cdot;\eta),\dots,\psi_k(\cdot;\eta))^{\mathrm{t}}$ is a vector of processes satisfying certain regularity conditions. The test statistics were quadratic forms of the statistic

$$U_1(\tau;\hat{\eta}) = \int_0^\tau \psi(s;\hat{\eta})\,dM(s;\hat{\eta}),$$

where $M(t;\eta) = N(t) - \int_0^t Y(s)\lambda_0(s;\eta)\,ds$ and $\hat{\eta}$ is the restricted maximum likelihood estimator of η based on $\{(N(t),Y(t)),\ t \in \mathscr{T}\}$. The resulting test statistics can also be viewed as functions of generalized residual processes, which are processes typically used in model validation and diagnostics. Asymptotic properties of the test statistics were obtained under a sequence of local alternatives. By examining the asymptotic properties of $U_1(\tau; \eta)$ and $U_1(\tau; \hat{\eta})$, where η_0 is the true parameter value if the hypothesized class holds, the effect of estimating η_0 by $\hat{\eta}$ was determined. Conditions in order for this substitution to have *no* effect, an adaptiveness property, were obtained. The asymptotic local powers of the tests were also examined for the purpose of providing guidelines on how to choose ψ. It turns out that if $\psi(\cdot; \eta)$ is asymptotically orthonormal and orthogonal to $\rho(\cdot; \eta) = (\partial/\partial n)\log \lambda_0(\cdot; \eta)$ when considered as elements of an appropriate Hilbert space, then the adaptiveness property is satisfied and the resulting tests have simple forms aside from having the potential of providing information on how $\lambda_0(\cdot)$ differs from $\mathscr{C}$ if the hypothesis is rejected. However, choosing the ψ-process to be orthonormal and to be orthogonal to ρ may not be feasible in general situations since the orthogonalization process requires specific knowledge of the incompleteness mechanism. Thus, from a practical point of view, k and ψ should be chosen such that (ψ, ρ) should asymptotically span the space of functions obtained by varying $\lambda_0(\cdot)$ in a scaled version of $\log\{\lambda_0(\cdot)/\lambda_0(\cdot; \eta)\}$, without requiring that ψ be orthonormal and orthogonal to ρ. The consequences in terms of efficiency losses of misspecifying k, either by oversmoothing or undersmoothing, were examined. In terms of the criterion of local asymptotic relative efficiency, it was found that undersmoothing is a more serious type of problem than oversmoothing, since the resulting test in such a misspecified model may not have any sensitivity in detecting some types of alternatives. The general procedure was applied to several concrete situations, and, in particular, the exact forms of the tests were obtained under the random censorship model. Furthermore, the specific smooth goodness-of-fit tests for testing the class of constant hazard rate functions and the class of two-parameter Weibull hazard rate functions were explicitly expressed. The results of simulation studies showing that for small to moderate sample sizes the achieved levels of the tests were acceptable were also presented. Furthermore, the achieved powers of the tests for specific null and alternative models were described. Finally, the tests for exponentiality and for Weibullness associated with a polynomial-type of specification of the ψ-process were used to reanalyze three data sets that have appeared in the literature. These applications indicate the potential of the proposed smooth goodness-of-fit tests in applied work.

We conclude with some remarks. The general methodology for developing smooth goodness-of-fit tests described in this paper is promising for generating omnibus, and possibly directional, tests. However, further studies need to be undertaken in future work. One of the main problems is how to choose, maybe dynamically, the smoothing order k. A sequential type of procedure for deciding the value of k, possibly in the spirit of Bickel and Ritov (1992), Fan (1996) and/or Ledwina (1994), needs to be explored. Another problem is to explore other possibilities for the ψ-process, in particular to examine the potential of using wavelet bases which could lead to procedures which are sensitive to detecting local departures from the specified null model $\mathscr{C}$ [Fan (1996)]. Furthermore, more extensive power studies using different ψ specifications and under a variety of alternative models should be undertaken to see the performance of these smooth goodness-of-fit tests relative to existing goodness-of-fit tests such as Pearson, Kolmogorov–Smirnov, Cramér–von Mises, and more specialized types of goodness-of-fit tests.

APPENDIX

In this Appendix we state the regularity conditions and present and prove the asymptotic results.

A. Regularity conditions. The following are regularity conditions needed for the asymptotic results. For a vector $\mathbf{v} = (v_1, \ldots, v_m)^{\mathrm{t}}$, $|\mathbf{v}| = \sqrt{\mathbf{v}^{\mathrm{t}}\mathbf{v}}$, while for a matrix $\mathbf{V} = (v_{ij})$, we let $\|\mathbf{V}\| = \max_{i,j}|v_{ij}|$.

(I) There exists a neighborhood $\Gamma_0 \subset \Gamma$ of η_0 such that on $\mathscr{T} \times \Gamma_0$, $\lambda_0(t;\eta) > 0$, and the partial derivatives $(\partial/\partial\eta_i)\lambda_0(s;\eta)$, $(\partial^2/\partial\eta_i\,\partial\eta_j)\lambda_0(s;\eta)$ and $(\partial^3/\partial\eta_i\,\partial\eta_j\,\partial\eta_k)\lambda_0(s;\eta)$ exist and are continuous at $\eta = \eta_0$.

(II) On $\mathscr{T} \times \Gamma_0$, the log-likelihood process

$$l(t;\theta,\eta) = \int_0^t \log[Y(s)\lambda(s;\theta,\eta)]\,dN(s) - \int_0^t Y(s)\lambda(s;\theta,\eta)\,ds$$

can be differentiated three times with respect to η and with the order of the differentiation and integration operations being interchangeable.

(III) On $\mathscr{T} \times \Gamma_0$, the partial derivatives $(\partial/\partial\eta_i)\psi(s;\eta)$, $(\partial^2/\partial\eta_i\,\partial\eta_j)\psi(s;\eta)$ and $(\partial^3/\partial\eta_i\,\partial\eta_j\,\partial\eta_k)\psi(s;\eta)$ exist and are continuous at $\eta = \eta_0$ and with the processes $\{\psi(t;\eta_0)\colon t \in \mathscr{T}\}$ and $\{(\partial/\partial\eta_i)\psi(t;\eta)\colon t \in \mathscr{T}\}$ being locally bounded and predictable.

(IV) There exist deterministic functions $y(\cdot)$ defined on $\mathscr{T}$, and $\psi^{(0)}(\cdot;\cdot)$ defined on $\mathscr{T} \times \Gamma_0$, such that for $(t,\eta) \in \mathscr{T} \times \Gamma_0$ and as $n \to \infty$,

$$\left\| a_n^{-2}\int_0^t \psi(s;\eta)^{\otimes 2} Y(s)\lambda_0(s;\eta)\,ds - \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2} y(s)\lambda_0(s;\eta)\,ds \right\| \to 0 \quad \text{in probability;}$$

$$\left\| a_n^{-2}\int_0^t \psi(s;\eta)\rho(s;\eta)^{\mathrm{t}} Y(s)\lambda_0(s;\eta)\,ds - \int_0^t \psi^{(0)}(s;\eta)\rho(s;\eta)^{\mathrm{t}} y(s)\lambda_0(s;\eta)\,ds \right\| \to 0 \quad \text{in probability;}$$

$$\left\| a_n^{-2} \int_0^t \rho(s;\eta)^{\otimes 2} Y(s) \lambda_0(s;\eta)\, ds - \int_0^t \rho(s;\eta)^{\otimes 2} y(s) \lambda_0(s;\eta)\, ds \right\| \to 0 \quad \text{in probability.}$$

(V) Defining the matrices of functions on $\mathscr{T} \times \Gamma_0$ given by

$$\mathbf{\Sigma}_{11}(t;\eta) = \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2} y(s) \lambda_0(s;\eta)\, ds;$$

$$\mathbf{\Sigma}_{12}(t;\eta) = \mathbf{\Sigma}_{21}(t;\eta)^{\mathrm{t}} = \int_0^t \psi^{(0)}(s;\eta) \rho(s;\eta)^{\mathrm{t}} y(s) \lambda_0(s;\eta)\, ds;$$

$$\mathbf{\Sigma}_{22}(t;\eta) = \int_0^t \rho(s;\eta)^{\otimes 2} y(s) \lambda_0(s;\eta)\, ds,$$

the $(k+q) \times (k+q)$ matrix

$$\mathbf{\Sigma}(t;\eta_0) = \begin{bmatrix} \mathbf{\Sigma}_{11}(t;\eta_0) & \mathbf{\Sigma}_{12}(t;\eta_0) \\ \mathbf{\Sigma}_{21}(t;\eta_0) & \mathbf{\Sigma}_{22}(t;\eta_0) \end{bmatrix}$$

have finite elements for each $t \in \mathscr{T}$, and $\mathbf{\Sigma}(\tau;\eta_0)$ is positive definite.

(VI) For each $\varepsilon > 0$ and $t \in \mathscr{T}$, as $n \to \infty$,

$$a_n^{-2} \int_0^t |\psi(s;\eta_0)|^2 I\{|\psi(s;\eta_0)| > \varepsilon a_n\} Y(s) \lambda_0(s;\eta_0)\, ds \to 0 \quad \text{in probability;}$$

$$a_n^{-2} \int_0^t |\rho(s;\eta_0)|^2 I\{|\rho(s;\eta_0)| > \varepsilon a_n\} Y(s) \lambda_0(s;\eta_0)\, ds \to \quad \text{in probability.}$$

(VII) There exist functions G_1 and H_1 defined on $\mathscr{T}$ such that for each $t \in \mathscr{T}$,

$$\sup_{\eta \in \Gamma_0} \left| \frac{\partial^3}{\partial\eta_i\, \partial\eta_j\, \partial\eta_k} \lambda_0(t;\eta) \right| \le G_1(t);$$

$$\sup_{\eta \in \Gamma_0} \left| \frac{\partial^3}{\partial\eta_i\, \partial\eta_j\, \partial\eta_k} \log \lambda_0(t;\eta) \right| \le H_1(t),$$

and such that, as $n \to \infty$,

$$a_n^{-2} \int_0^\tau G_1(s) Y(s)\, ds \to \int_0^\tau G_1(s) y(s)\, ds \quad (\text{in probability}) < \infty;$$

$$a_n^{-2} \int_0^\tau H_1(s) Y(s) \lambda_0(s;\eta_0)\, ds \to \int_0^\tau H_1(s) y(s) \lambda_0(s;\eta_0)\, ds \quad (\text{in probability}) < \infty;$$

$$a_n^{-2} \int_0^\tau H_1^2(s) Y(s) \lambda_0(s;\eta_0)\, ds \to \int_0^\tau H_1^2(s) y(s) \lambda_0(s;\eta_0)\, ds \quad (\text{in probability}) < \infty.$$

(VII) There exists a predictable process $G_2(\cdot)$ with $\sup_{\eta\in\Gamma_0}\|(\partial^2/\partial\eta\partial\eta^{\mathrm{t}})\psi_i(s;\eta)\| \le G_2(s)$ for each $i = 1,\dots,k$, and such that, as $n\to\infty$,

$$a_n^{-2}\int_0^\tau \left\| \frac{\partial}{\partial\eta}\psi(s;\eta_0)\right\|^2 Y(s)\lambda_0(s;\eta_0)\,ds = O_p(1) \quad \text{and}$$

$$a_n^{-2}\int_0^\tau G_2(s)^2 Y(s)\lambda_0(s;\eta_0)\,ds \to \text{finite number in probability.}$$

B. Proofs. Before proving Theorem 3.1 we establish two intermediate results needed in its proof. For $\mathscr{I}$ an interval in $\Re$, $\mathscr{D}(\mathscr{I})^p$ denotes the product space $\mathscr{D}(\mathscr{I})\times\cdots\times\mathscr{D}(\mathscr{I})$ endowed with Skorohod's product topology [Billingsley (1968)].

PROPOSITION B.1. *Under a sequence of models with true parameter values* $(\theta^{(n)},\eta^{(n)}) = (a_n^{-1}\gamma(1+o(1)),\eta_0)$, $n=1,2,\dots,$ *if conditions* (I)–(VI) *are satisfied then as* $n\to\infty$,

$$\frac{1}{a_n}\begin{bmatrix} U_1(\cdot;\eta_0)\\ U_2(\cdot;\eta_0)\end{bmatrix} = \frac{1}{a_n}\int_0^{\cdot}\begin{bmatrix}\psi(s;\eta_0)\\ \rho(s;\eta_0)\end{bmatrix}\{dN(s) - Y(s)\lambda_0(s;\eta_0)\,ds\}$$

converges weakly on $\mathscr{D}[0,\tau]^{k+q}$ *to* $Z(\cdot;\eta_0) = [Z_1(\cdot;\eta_0)^{\mathrm{t}}, Z_2(\cdot;\eta_0)^{\mathrm{t}}]^{\mathrm{t}}$, *a Gaussian process with mean function*

$$\mu(\cdot;\eta_0) = \mathbf{E}\begin{bmatrix} Z_1(\cdot;\eta_0)\\ Z_2(\cdot;\eta_0)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Sigma}_{11}(\cdot;\eta_0)\\ \boldsymbol{\Sigma}_{21}(\cdot;\eta_0)\end{bmatrix}\gamma$$

and covariance matrix function $\mathrm{Cov}\{Z(t_1;\eta_0), Z(t_2;\eta_0)\} = \boldsymbol{\Sigma}(t_1\wedge t_2;\eta_0)$, $t_1, t_2 \in\mathscr{T}$.

PROOF. Note that

$$\frac{1}{a_n}\begin{bmatrix} U_1(\cdot;\eta_0)\\ U_2(\cdot;\eta_0)\end{bmatrix} = \frac{1}{a_n}\int_0^{\cdot}\begin{bmatrix}\psi(s;\eta_0)\\ \rho(s;\eta_0)\end{bmatrix} dM\big(s;\theta^{(n)},\eta_0\big)$$
$$+\frac{1}{a_n^2}\int_0^{\cdot}\begin{bmatrix}\psi(s;\eta_0)\\ \rho(s;\eta_0)\end{bmatrix} Y(s)\left[\frac{\partial}{\partial\theta}\lambda(s;\theta^*,\eta_0)\right]^{\mathrm{t}}\big(a_n\theta^{(n)}\big)\,ds,$$

where $M(s;\theta^{(n)},\eta_0) = N(s) - \int_0^s \lambda(s;\theta^{(n)},\eta_0)\,ds$, and with $\theta^*\in[0,\theta^{(n)}]$, so $\theta^*\to 0$ as $n\to\infty$. Since $(\partial/\partial\theta)\lambda(s;\theta,\eta_0) = \lambda_0(s;\eta_0)\psi(s;\eta_0)\exp\{\theta^{\mathrm{t}}\psi(s;\eta_0)\}$, then

$$\begin{bmatrix}\psi(s;\eta_0)\\ \rho(s;\eta_0)\end{bmatrix}\left[\frac{\partial}{\partial\theta}\lambda(s;\theta^*,\eta_0)\right]^{\mathrm{t}}$$
$$= \begin{bmatrix}\psi(s;\eta_0)^{\otimes 2}\\ \rho(s;\eta_0)\psi(s;\eta_0)^{\mathrm{t}}\end{bmatrix}\exp\big\{(\theta^*)^{\mathrm{t}}\psi(s;\eta_0)\big\}\lambda_0(s;\eta_0),$$

so by condition (IV) and since $\theta^{(n)} = a_n^{-1}\gamma(1 + o(1))$,

$$\frac{1}{a_n^2}\int_0^t \begin{bmatrix} \psi(s;\eta_0) \\ \rho(s;\eta_0) \end{bmatrix} Y(s) \left[\frac{\partial}{\partial\theta}\lambda(s;\theta^*,\eta_0)\right]^{\mathrm{t}} \left(a_n\theta^{(n)}\right) ds$$

converges in probability to $\begin{bmatrix} \Sigma_{11}(t;\eta_0) \\ \Sigma_{21}(t;\eta_0) \end{bmatrix}\gamma$. Under the model $(\theta^{(n)}, \eta^{(n)}) = (a_n^{-1}\gamma \cdot (1 + o(1)), \eta_0)$, $\{M(t;\theta^{(n)},\eta_0)\colon t \in \mathscr{T}\}$ is a local square-integrable martingale with quadratic variation process $\langle M(\cdot;\theta^{(n)},\eta_0)\rangle(t) = \int_0^t Y(s)\lambda(s;\theta^{(n)},\eta_0)\,ds$. Therefore, $a_n^{-1}\int_0^{\cdot}[\psi(s;\eta_0)^{\mathrm{t}}, \rho(s;\eta_0)^{\mathrm{t}}]^{\mathrm{t}}\,dM(s;\theta^{(n)},\eta_0)$ is a local square-integrable martingale with quadratic variation process given by

$$\frac{1}{a_n^2}\int_0^{\cdot} \begin{bmatrix} \psi(s;\eta_0) \\ \rho(s;\eta_0) \end{bmatrix}^{\otimes 2} Y(s)\,\lambda\left(s;\theta^{(n)},\eta_0\right) ds,$$

which by condition (IV) converges in probability to $\mathbf{\Sigma}(\cdot;\eta_0)$. By condition (VI), the Lindeberg-type condition of Rebolledo's martingale central limit theorem [Andersen, Borgan, Gill and Keiding (1993)] also holds, so it follows that $a_n^{-1}\int_0^{\cdot}[\psi(s;\eta_0)^{\mathrm{t}}, \rho(s;\eta_0)^{\mathrm{t}}]^{\mathrm{t}}\,dM(s;\theta^{(n)},\eta_0)$ converges weakly on $\mathscr{D}[0,\tau]^{k+q}$ to a zero-mean Gaussian martingale with variance matrix function $\mathbf{\Sigma}(\cdot;\eta_0)$. Combining the two results above, we conclude that $a_n^{-1}[U_1(\cdot;\eta_0)^{\mathrm{t}}, U_2(\cdot;\eta_0)^{\mathrm{t}}]^{\mathrm{t}}$ converges weakly on $\mathscr{D}[0,\tau]^{k+q}$ to $Z(\cdot;\eta_0)$, where the latter is defined in the statement of the proposition.

PROPOSITION B.2. *If conditions* (I)–(VIII) *are satisfied, then under* H_{1n}, *the restricted* MLE $\hat\eta \equiv \hat\eta(\theta = 0)$ *is asymptotically normal with asymptotic mean* $\eta_0 + \mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}\mathbf{\Sigma}_{21}(\tau;\eta_0)\gamma/a_n$ *and asymptotic variance* $\mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}/a_n^2$; *and it has the asymptotic representation* $a_n(\hat\eta - \eta_0) = \mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}a_n^{-1}U_2(\tau;\eta_0) + o_p(1)$. *In particular*, $\hat\eta$ *is* a_n-*consistent for* η_0.

PROOF. Since $\hat\eta$ is a solution of $U_2(\tau;\eta) = 0$, then by a first-order Taylor expansion

$$0 = a_n^{-1}U_2(\tau;\eta_0) + a_n^{-2}\left\{\frac{\partial}{\partial\eta}U_2(\tau;\eta)\right\}a_n(\hat\eta - \eta_0), \tag{B.1}$$

where η^* lies in the line segment connecting η_0 and $\hat\eta$, and

$$\frac{\partial}{\partial\eta}U_2(\tau;\eta) = \int_0^\tau \dot\rho(s;\eta)\,dM(s;\eta) - \int_0^\tau \rho(s;\eta)^{\otimes 2}Y(s)\,\lambda_0(s;\eta)\,ds,$$

where $\rho(s;\eta) = (\partial/\partial\eta)\log\lambda_0(s;\eta)$ and $\dot\rho(s;\eta) = (\partial/\partial\eta^{\mathrm{t}})\rho(s;\eta)$. Under H_{0n}, Borgan (1984) has shown that there exists a unique sequence $\{\hat\eta\}$ such that $\hat\eta \to \eta_0$ in probability, and furthermore, for any $\eta^* \to \eta_0$ in probability,

$$-\frac{\partial}{\partial\eta}U_2(\tau;\eta^*) \to \mathbf{\Sigma}_{22}(\tau;\eta_0) \quad \text{in probabililty.} \tag{B.2}$$

Consequently, under H_{0n}, $a_n(\hat\eta - \eta_0) = \mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}a_n^{-1}U_2(\tau;\eta_0) + o_p(1)$. Interest, though, is with respect to H_{1n}. Denote by $\mathbf{P}_{(\theta^{(n)},\eta_0)}$ the probability measure under H_{1n} and by $\mathbf{P}_{\eta_0}$ the probability measure under H_{0n}. Then

$$\log \frac{d\mathbf{P}_{(\theta^{(n)},\,\eta_0)}}{d\mathbf{P}_{\eta_0}}\bigg|_{\mathscr{F}_\tau} = l\big(\tau;\theta^{(n)},\eta_0\big) - l(\tau;0,\eta_0)$$

$$= \int_0^\tau \big(\theta^{(n)}\big)^{\mathrm{t}} \psi(s;\eta_0)\, dN(s) - \int_0^\tau \Big(\exp\Big\{\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)\Big\} - 1\Big) Y(s)\lambda_0(s;\eta_0)\, ds.$$

Since

$$\exp\Big\{\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)\Big\} - 1 = \big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0) + \tfrac{1}{2}\exp\{\xi_n^*(s)\}\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)^{\otimes 2}\big(\theta^{(n)}\big),$$

with $\xi_n^*(s) \in [0, (\theta^{(n)})^{\mathrm{t}}\psi(s;\eta_0)]$, then

$$\log \frac{d\mathbf{P}_{(\theta^{(n)},\,\eta_0)}}{d\mathbf{P}_{\eta_0}}\bigg|_{\mathscr{F}_\tau}$$

$$= \int_0^\tau \big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)\, dM(s;\eta_0) - \frac{1}{2}\int_0^\tau \exp\{\xi_n^*(s)\}\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)^{\otimes 2}\big(\theta^{(n)}\big)Y(s)\lambda_0(s;\eta_0)\, ds.$$

In addition, since $\theta^{(n)} = a_n^{-1}\gamma(1 + o(1))$, then

$$\log \frac{d\mathbf{P}_{(\theta^{(n)},\,\eta_0)}}{d\mathbf{P}_{\eta_0}}\bigg|_{\mathscr{F}_\tau} = \frac{1}{a_n}\int_0^\tau \gamma^{\mathrm{t}}\psi(s;\eta_0)\, dM(s;\eta_0) - \frac{1}{2a_n^2}\int_0^\tau \exp\{\xi_n^*(s)\}\big(\gamma^{\mathrm{t}}\psi(s;\eta_0)^{\otimes 2}\gamma\big) \times Y(s)\lambda_0(s;\eta_0)\, ds + o_p(1).$$

Consequently,

$$\begin{bmatrix} \log \dfrac{d\mathbf{P}_{(\theta^{(n)},\,\eta_0)}}{d\mathbf{P}_{\eta_0}}\bigg|_{\mathscr{F}_\tau} \\ a_n^{-1}(\hat{\eta} - \eta_0) \end{bmatrix}$$

$$= \frac{1}{a_n}\int_0^\tau \begin{bmatrix} \gamma^{\mathrm{t}}\psi(s;\eta_0) \\ \mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}\rho(s;\eta_0) \end{bmatrix} dM(s;\eta_0) - \frac{1}{2a_n^2}\int_0^\tau \begin{bmatrix} \exp\{\xi_n^*(s)\}\big(\gamma^{\mathrm{t}}\psi(s;\eta_0)^{\otimes 2}\gamma\big) \\ 0 \end{bmatrix} Y(s)\lambda_0(s;\eta_0)\, ds, \tag{B.3}$$

which, under H_{0n}, converges in distribution to a $(1+q)$-dimensional normal distribution with mean vector $\begin{bmatrix} -\frac{1}{2}\gamma^t\boldsymbol{\Sigma}_{11}(\tau;\eta_0)\gamma \\ 0 \end{bmatrix}$ and covariance matrix

$$\text{(B.4)} \qquad \begin{bmatrix} \gamma^{t}\boldsymbol{\Sigma}_{11}(\tau;\eta_0)\gamma & \gamma^{t}\boldsymbol{\Sigma}_{12}(\tau;\eta_0)\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1} \\ \boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)\gamma & \boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1} \end{bmatrix}.$$

By LeCam's third lemma [cf. Andersen, Borgan, Gill and Keiding (1993)], it follows that under H_{1n}, (B.3) converges in distribution to a $(1+q)$-dimensional normal distribution with covariance matrix given by (B.4) and mean vector

$$\begin{bmatrix} \dfrac{1}{2}\gamma^{t}\boldsymbol{\Sigma}_{11}(\tau;\eta_0)\gamma \\ \boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)\gamma \end{bmatrix}.$$

Thus, under H_{1n}, $a_n^{-1}(\hat{\eta}-\eta_0) \to_d N_q(\boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}\boldsymbol{\Sigma}_{21}(\tau;\eta_0)\gamma, \boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1})$, hence $\hat{\eta} \to \eta_0$ in probability. From (B.1) and (B.2), it follows that under H_{1n}, $a_n(\hat{\eta}-\eta_0) = \boldsymbol{\Sigma}_{22}(\tau;\eta_0)^{-1}a_n^{-1}U_2(\tau;\eta_0) + o_p(1)$.

PROOF OF THEOREM 3.1. By a first-order Taylor expansion, $a_n^{-1}U_1(\cdot;\hat{\eta}) = a_n^{-1}U_1(\cdot;\eta_0) + \{a_n^{-2}(\partial/\partial\eta)U_1(\cdot;\eta^*)\}a_n(\hat{\eta}-\eta_0)$. Since $U_1(\cdot;\eta) = \int_0^{\cdot}\psi(s;\eta)\{dN(s) - Y(s)\lambda_0(s;\eta)\,ds\}$, then

$$\frac{\partial}{\partial\eta^{t}}U_1(\cdot;\eta) = \int_0^{\cdot}\dot{\psi}(s;\eta)\{dN(s) - Y(s)\lambda_0(s;\eta)\,ds\} - \int_0^{\cdot}\psi(s;\eta)\rho(s;\eta)^{t}Y(s)\lambda_0(s;\eta)\,ds,$$

where $\dot{\psi}(s;\eta) = (\partial/\partial\eta^t)\psi(s;\eta)$. By conditions (I), (IV) and (V) and by continuity of $\psi(s;\eta)$ with respect to η, and since $\eta^* \to \eta_0$ in probability,

$$a_n^{-2}\int_0^{\cdot}\psi(s;\eta^*)\rho(s;\eta^*)^{t}Y(s)\lambda_0(s;\eta^*)\,ds \to \boldsymbol{\Sigma}(\cdot;\eta_0) \quad \text{in probability.}$$

On the other hand,

$$\begin{aligned} &\left\| a_n^{-2}\int_0^{\cdot}\dot{\psi}(s;\eta^*)\{dN(s) - Y(s)\lambda_0(s;\eta^*)\,ds\}\right\| \\ &\quad\le \left\| a_n^{-2}\int_0^{\cdot}\dot{\psi}(s;\eta^*)\,dM\big(s;\theta^{(n)},\eta_0\big)\right\| \\ &\qquad + \left\| a_n^{-2}\int_0^{\cdot}\dot{\psi}(s;\eta^*)Y(s)\big[\lambda\big(s;\theta^{(n)},\eta_0\big) - \lambda_0(s;\eta^*)\big]\,ds\right\|. \end{aligned}$$

By a first-order Taylor expansion, $\dot{\psi}(\cdot;\eta^*) = \dot{\psi}(\cdot;\eta_0) + \ddot{\psi}(\cdot;\eta^{**})(\eta^* - \eta_0)$ with $\eta^{**} \in [\eta^*, \eta_0]$, so

$$\begin{aligned}
&a_n^{-2}\int_0^{\cdot} \dot{\psi}(s;\eta^*)\, dM\big(s;\theta^{(n)},\eta_0\big)\\
&\qquad = a_n^{-2}\int_0^{\cdot} \dot{\psi}(s;\eta_0)\, dM\big(s;\theta^{(n)},\eta_0\big)\\
&\qquad\quad + \left(a_n^{-2}\int_0^{\cdot} \ddot{\psi}(s;\eta^{**})\, dM\big(s;\theta^{(n)},\eta_0\big)\right)(\eta^* - \eta_0).
\end{aligned}$$

By condition (III), $\int_0^{\cdot} \dot{\psi}(s;\eta_0)\, dM(s;\theta^{(n)},\eta_0)$ is a local square-integrable martingale, and by Lenglart's inequality,

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{t\in\mathscr{T}}\left\| a_n^{-2}\int_0^t \dot{\psi}(s;\eta_0)\, dM\big(s;\theta^{(n)},\eta_0\big)\right\| \ge \varepsilon\right\}\\
&\qquad \le \frac{\delta}{\varepsilon^2} + \mathbf{P}\bigg\{a_n^{-4}\int_0^{\tau}\big\|\dot{\psi}(s;\eta_0)\big\|^2 Y(s)\,\lambda_0(s;\eta_0)\\
&\qquad\qquad\qquad \times\exp\Big\{\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)\Big\}\, ds \ge \delta\bigg\},
\end{aligned}$$

so by condition (VIII) the right-hand side can be made arbitrarily small. Thus, $a_n^{-2}\int_0^{\cdot}\dot{\psi}(s;\eta_0)\, dM(s;\theta^{(n)},\eta_0) = o_p(1)$. On the other hand,

$$\begin{aligned}
&\left\| a_n^{-2}\int_0^{\cdot} \ddot{\psi}(s;\eta^{**})\, dM\big(s;\theta^{(n)},\eta_0\big)\right\|\\
&\quad \le a_n^{-2}\int_0^{\tau}\big\|\ddot{\psi}(s;\eta^{**})\big\|\, dN(s)\\
&\qquad + a_n^{-2}\int_0^{\tau}\big\|\ddot{\psi}(s;\eta^{**})\big\| Y(s)\,\lambda_0(s;\eta_0)\exp\Big\{\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)\Big\}\, ds\\
&\quad \le a_n^{-2}\int_0^{\tau} G_2(s)\, dN(s)\\
&\qquad + a_n^{-2}\int_0^{\tau} G_2(s) Y(s)\,\lambda_0(s;\eta_0)\exp\Big\{\big(\theta^{(n)}\big)^{t}\psi(s;\eta_0)\Big\}\, ds.
\end{aligned}$$

By condition (VIII) and the Cauchy–Schwarz inequality, the last term can be shown to converge in probability to a finite quantity. Next, since

$$\begin{aligned}
\lambda\big(s;\theta^{(n)},\eta_0\big) - \lambda_0(s;\eta^*) &= \lambda_0(s;\eta_0)\exp\{\xi^*(s)\}\big(\theta^{(n)}\big)^{\mathrm{t}}\psi(s;\eta_0)\\
&\quad - \left[\frac{\partial}{\partial\eta}\lambda_0(s;\eta^{**})\right]^{\mathrm{t}}(\eta^* - \eta_0)\\
&= a_n^{-1}\gamma^{\mathrm{t}}\lambda_0(s;\eta_0)\exp\{\xi^*(s)\}\psi(s;\eta_0)(1 + o(1))\\
&\quad - \left[\frac{\partial}{\partial\eta}\log\lambda_0(s;\eta^{**})\right]^{\mathrm{t}}(\eta^* - \eta_0)\,\lambda_0(s;\eta^{**}),
\end{aligned}$$

then $\|a_n^{-2}\int_0^{\cdot}\dot{\psi}(s;\eta^*)Y(s)[\lambda(s;\theta^{(n)},\eta_0)-\lambda_0(s;\eta^*)]\,ds\| = o_p(1)$. Therefore, $a_n^{-2}(\partial/\partial\eta)U_1(\cdot;\eta^*)$ equals $-\mathbf{\Sigma}_{12}(\cdot;\eta_0)+o_p(1)$, hence $a_n^{-1}U_1(\cdot;\hat{\eta}) = a_n^{-1}U_1(\cdot;\eta_0) - [\mathbf{\Sigma}_{12}(\cdot;\eta_0)+o_p(1)]a_n(\hat{\eta}-\eta_0)$. Using the representation of $a_n(\hat{\eta}-\eta_0)$ in Proposition B.2, we obtain

$$\begin{aligned} a_n^{-1}U_1(\cdot;\hat{\eta}) &= a_n^{-1}U_1(\cdot;\eta_0) \\ &\quad -\Big\{\mathbf{\Sigma}_{12}(\cdot;\eta_0)+o_p(1)\Big\}\Big\{\mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}a_n^{-1}U_2(\tau;\eta_0)+o_p(1)\Big\} \\ &= a_n^{-1}U_1(\cdot;\eta_0) - \mathbf{\Sigma}_{12}(\cdot;\eta_0)\mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}a_n^{-1}U_2(\tau;\eta_0)+o_p(1) \\ &= \Big[\mathbf{I} - \mathbf{\Sigma}_{12}(\cdot;\eta_0)\mathbf{\Sigma}_{22}(\tau;\eta_0)^{-1}\Big]\begin{bmatrix} a_n^{-1}U_1(\cdot;\eta_0) \\ a_n^{-1}U_2(\tau;\eta_0)\end{bmatrix} + o_p(1). \end{aligned}$$

From Proposition B.1, it follows that $a_n^{-1}U_1(\cdot;\hat{\eta})$ converges weakly to a Gaussian process $\tilde{Z}_1(\cdot;\eta_0)$ with mean function $\mu_1(\cdot;\eta_0)$ and covariance matrix function $\tilde{\mathbf{\Sigma}}_1(t_1,t_2;\eta_0)$, where these vector–matrix functions are defined in the statement of the theorem. □

Acknowledgments. The author thanks the two referees, Associate Editor and Editors for their comments and suggestions which led to improvements in the paper. He thanks Professors Zenia Agustin and Robert Strawderman for discussions.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ABAN, I. and PEÑA, E. (1998). Properties of test statistics applied to residuals in failure time models, *J. Statist. Plann. Inference*. To appear.

AKRITAS, M. (1988). Pearson-type goodness-of-fit tests: the univariate case. *J. Amer. Statist. Assoc.* **83** 222–230.

ANDERSEN, P., BORGAN, O., GILL, R. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

ANDERSEN, P. and GILL, R. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

ANGUS, J. (1982). Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation. *J. Statist. Plann. Inference* **6** 241–251.

BALTAZAR–ABAN, I. and PEÑA, E. (1995). Properties of hazard-based residuals and implications in model diagnostics. *J. Amer. Statist. Assoc.* **90** 185–197.

BARLOW, R., BARTHOLOMEW, D., BREMMER, J. and BRUNK, H. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

BARLOW, W. and PRENTICE, R. (1988). Residuals for relative risk regression. *Biometrika* **75** 65–74.

BHAT, B. and NAGNUR, B. (1965). Locally most stringent tests and Lagrangian multiplier tests of linear hypotheses. *Biometrika* **52** 459–468.

BICKEL, P., KLAASEN, C., RITOV, Y. and WELLNER, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

BICKEL, P. and RITOV, Y. (1992). Testing for goodness of fit: a new approach. In *Nonparametric Statistics and Related Topics* (A. K. Md. E. Saleh, ed.) 51–57. North-Holland, Amsterdam.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BORGAN, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* **11** 1–16.

CHEN, Y., HOLLANDER, M. and LANGBERG, N. (1982). Small-sample results for the Kaplan–Meier estimator. *J. Amer. Statist. Assoc.* **77** 141–144.

CHERNOFF, H. and LEHMANN, E. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25** 579–586.

CHOI, S., HALL, W. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861.

COX, D. (1972). Regression models and life tables (with discussion). *J. Royal Statist. Soc. Ser. B* **34** 187–220.

COX, D. and REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Royal Statist. Soc. Ser. B* **49** 1–39.

COX, D. and SNELL, E. (1968). A general definition of residuals. *J. Royal Statist. Soc. Ser. B* **30** 248–275.

DOKSUM, K. and YANDELL B. (1984). Tests for exponentiality. In *Handbook of Statistics 4: Nonparametric Methods* (P.R. Krishnaiah and P. K. Sen, eds.) 579–611. North-Holland, Amsterdam.

DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, Philadelphia.

DURBIN, J. (1975). Kolmogorov–Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika* **62** 5–22.

EUBANK, R. and HART, J. (1992). Testing goodness-of-fit regression via order selection criteria. *Ann. Statist.* **20** 1412–1425.

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688.

FLEMING, T. and HARRINGTON, D. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

GATSONIS, C., HSIEH, H. and KORWAR, R. (1985). Simple nonparametric tests for a known standard survival based on censored data. *Comm. Statist. Theory Methods* **14** 2137–2162.

GRAY, R. and PIERCE, D. (1985). Goodness-of-fit for censored survival data. *Ann. Statist.* **13** 552–563.

HABIB, M. and THOMAS, D. (1986). Chi-squared goodness-of-fit tests for randomly censored data. *Ann. Statist.* **14** 759–765.

HJORT, N. (1990). Goodness-of-fit tests in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18** 1221–1258.

IMSL LIBRARY (1987). *User's Manual: Stat/Library—Fortran Subroutines for Statistical Analysis*. IMSL, Houston.

INGLOT, T., KALLENBERG, W. and LEDWINA, T. (1994). Power approximations to and power comparison of smooth goodness-of-fit tests. *Scand. J. Statist.* **21** 131–145.

KALLENBERG, W. and LEDWINA, T. (1995). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23** 1594–1608.

KHMALADZE, E. (1981). A martingale approach in the theory of goodness-of-fit tests. *Teor. Veroyatnost. i Primenen.* **26** 246–265 (English translation in *Theory Probab. Appl.* **26** 240–257.)

KHMALADZE, E. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829.

KIM, J. (1993). Chi-square goodness-of-fit tests for randomly censored data. *Ann. Statist.* **21** 1621–1639.

KLEIN, J. and MOESCHBERGER, M. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York.

KOZIOL, J. and GREEN, S. (1976). A Cramér–von Mises statistic for randomly censored data. *Biometrika* **63** 139–156.

LAGAKOS, S. (1981). The graphical evaluation of explanatory variables in proportional hazard regression models. *Biometrika* **68** 93–98.

LEDWINA, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89** 1000–1005.

LI, G. and DOSS, H. (1993). Generalized Pearson–Fisher chi-square goodness-of-fit tests, with applications to models with life history data. *Ann. Statist.* **21** 772–797.

LOYNES, R. (1980). The empirical distribution function of residuals from generalised regression. *Ann. Statist.* **8** 285–298.

NEYMAN, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuarietidskr.* **20** 149–199.

NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Harald Cramér Volume* (U. Grenander, ed.) 213–234. Wiley, New York.

PEÑA, E. (1995). Residuals from type II censored samples. In *Recent Advances in Life-Testing and Reliability* (N. Balakrishnan, eds.) 523–543. CRC Press, Boca Raton.

PEÑA, E. (1996). Smooth goodness-of-fit tests for composite hypothesis in hazard based models. Technical report, Dept. Mathematics and Statistics, Bowling Green State University.

PEÑA, E. (1998). Smooth goodness-of-fit tests for the baseline hazard Cox's Proportional Hazards Model. *J. Amer. Statist. Assoc.* **93** 673–692.

PIERCE, D. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10** 475–478.

RANDLES, R. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10** 462–474.

RANDLES, R. (1984). On tests applied to residuals. *J. Amer. Statist. Assoc.* **79** 349–354.

RAO, K. and ROBSON, D. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. *Comm. Statist.* **3** 1139–1153.

RAYNER, J. and BEST, D. (1989). *Smooth Tests of Goodness of Fit*. Oxford Univ. Press.

RAYNER, J. and BEST, D. (1990). Smooth tests of goodness of fit: an overview. *Internat. Statist. Rev.* **58** 9–17.

STEPHENS, M. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann. Statist.* **4** 357–369.

STEPHENS, M. (1992). Introduction to Kolmogorov (1933). On the empirical determination of a distribution. In *Breakthroughs in Statistics* **2**: *Methodology and Distribution* (S. Kotz and N. Johnson, eds.) Springer, New York.

THERNEAU, T., GRAMBSCH, P. and FLEMING, T. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–160.

THOMAS, D. and PIERCE, D. (1979). Neyman's smooth goodness-of-fit test when the hypothesis is composite. *J. Amer. Statist. Assoc.* **74** 441–445.

WOOLSON, R. and SEN, P. K. (1974). Asymptotic comparison of a class of multivariate multiparameter tests. *Comm. Statist.* **3** 813–828.

DEPARTMENT OF MATHEMATICS AND STATISTICS
BOWLING GREEN STATE UNIVERSITY
BOWLING GREEN, OHIO 43403
E-MAIL: pena@stochos.bgsu.edu