University of North Carolina at Chapel Hill
The University of North Carolina at Chapel Hill Department of Biostatistics Technical Report Series

Year 2007 *Paper* 1

On Asymptotically Optimal Tests Under Loss of Identifiability in Semiparametric Models

Rui Song* Michael R. Kosorok†
Jason P. Fine‡

*University of North Carolina at Chapel Hill, rsong@bios.unc.edu
†University of North Carolina-Chapel Hill, kosorok@unc.edu
‡University of Wisconsin, jfine@bios.unc.edu

On Asymptotically Optimal Tests Under Loss of Identifiability in Semiparametric Models

Rui Song, Michael R. Kosorok, and Jason P. Fine

Abstract

We consider tests of hypotheses when the parameters are not identifiable under the null in semiparametric models, where regularity conditions for profile likelihood theory fail. Exponential average tests based on integrated profile likelihood are constructed and shown to be asymptotically optimal under a weighted average power criterion with respect to a prior on the nonidentifiable aspect of the model. These results extend existing results for parametric models, which involve more restrictive assumptions on the form of the alternative than do our results. Moreover, the proposed tests accomodate models with infinite dimensional nuisance parameters which either may not be identifiable or may not be estimable at the usual parametric rate. Examples include tests of the presence of a change-point in the Cox model under current status data, tests of regression parameters in odds-rate models and tests of the number of mixture components in two-component mixture models. Optimal tests have not prevously been studied for these scenarios. We study the asymptotic distribution of the proposed tests under the null, fixed contiguous alternatives and random contiguous alternatives. We also propose a weighted bootstrap procedure for computing the critical values of the test statistics.

Submitted to the Annals of Statistics

ON ASYMPTOTICALLY OPTIMAL TESTS UNDER LOSS OF IDENTIFIABILITY IN SEMIPARAMETRIC MODELS

By Rui Song, Michael R. Kosorok* and Jason P. Fine†

University of North Carolina and University of Wisconsin

We consider tests of hypotheses when the parameters are not identifiable under the null in semiparametric models, where regularity conditions for profile likelihood theory fail. Exponential average tests based on integrated profile likelihood are constructed and shown to be asymptotically optimal under a weighted average power criterion with respect to a prior on the nonidentifiable aspect of the model. These results extend existing results for parametric models, which involve more restrictive assumptions on the form of the alternative than do our results. Moreover, the proposed tests accommodate models with infinite dimensional nuisance parameters which either may not be identifiable or may not be estimable at the usual parametric rate. Examples include tests of the presence of a change-point in the Cox model under current status data, tests of regression parameters in odds-rate models and tests of the number of mixture components in two-component mixture models. Optimal tests have not previously been studied for these scenarios. We study the asymptotic distribution of the proposed tests under the null, fixed contiguous alternatives and random contiguous alternatives. We also propose a weighted bootstrap procedure for computing the critical values of the test statistics.

1. Introduction. In this paper we investigate nonstandard testing problems involving a family of probability distributions $\{P_\theta, \theta \in \Theta\}$, known up to a parameter θ, in a parameter space Θ. The parameter space Θ is assumed to be a subset of an infinite-dimensional metric space. The null and alternative hypotheses are:

$$H_0: \ \theta \in \Theta_0 \ \text{ vs. } \ H_1 : \theta \in \Theta \backslash \Theta_0,$$

where Θ_0 is a subset of Θ and contains at least two elements. In the usual testing framework, the parameters are unique under the null, so that identifiability is not an issue. While we allow multiple values of θ satisfying the

*Supported in part by Grant CA075142 from the National Cancer Institute.
†Supported in part by Grant CA094893 from the National Cancer Institute.
AMS 2000 subject classifications: Primary 62A01, 62G10; secondary 62G20, 62C99
Keywords and phrases: Change-point model, contiguous alternatives, empirical process, exponential average test, mixture model, nonstandard testing problem, odds-rate model, optimal test, power, profile likelihood.

imsart-aos ver. 2007/09/18 file: idensemi16.tex date: October 1, 2007

null, we assume that the null distribution, denoted by P_0, is unique, where $\Theta_0 = \{\theta \in \Theta : P_\theta = P_0\}$.

Under this set-up, the true value of θ is not identifiable under the null, since for any $\theta \neq \theta'$ in Θ_0, $P_\theta = P_{\theta'} = P_0$. Such loss of identifiability occurs in diverse applications in the social, biological, physical and medical sciences. We consider three examples in detail; see Section 2 for an overview of work on these special cases of the general paradigm described above.

The statistical literature contains numerous precedents on the nonidentifiability problem in parametric models, see Chernoff (1954), Chernoff and Lander (1995), Dacunha-Castelle and Gassiat (1999), Liu and Shao (2003). Among others, Dacunha-Castelle and Gassiat (1999) proposed a locally conic parametrization approach to enable asymptotic expansions of the likelihood ratio test under loss of identifiability under the null. Liu and Shao (2003) derived a quadratic approximation of the loglikelihood ratio function by using Hellinger distance. Most authors directly study the approximation of the log-likelihood ratio function in some neighborhood and obtain its asymptotic null distribution. However, the asymptotic optimality properties of the classical likelihood ratio tests (LRT) do not hold anymore (Lindsay, 1995) and Wald and score tests are not even well defined in these nonstandard problems. To our knowledge, all results for testing nonidentifiable P_0 using likelihood based tests are for parametric models. In this paper, we investigate the construction of optimal likelihood based tests for semiparametric models.

A key question which arises in the nonidentifiable context (Dacunha-Castelle and Gassiat, 1999) is: since the parameter is not identifiable, around which point can an expansion be made? To address this question, we assume the existence of a "full rank" reparameterization which contains all the information of the null model and in which all parameters are identifiable. To be specific, we partition $\theta \equiv (\psi, \zeta)$ and $\psi \equiv (\beta, \eta)$, where $\beta \in \mathbb{R}^p$ is a parameter of interest, $\zeta \in \mathbb{R}^q$ and η is a parameter defined on an arbitrary parametric space, $\mathcal{H}_\eta$. We assume that the information in the null model can be absorbed into the parameter space of η, through this full rank reparameterization. This is made precise in Section 3.

When the models involved are parametric, a special case when η does not depend on ζ under the null, that is, ζ is only present under the alternative, has been studied extensively by Andrews and Ploberger (1994); Davies (1977, 1987); Hansen (1996); King and Shively (1993), and others. Davies (1977) showed that the likelihood ratio test is optimal in the sense that as the significance level of the test tends to zero, its power function approaches that of the optimum test when ζ is given. These optimality results are very weak and do not provide any guidance regarding the performance of the test in practical applications, where the significance level is fixed, eg. at level .05 (Andrews, 1999). Andrews and Ploberger (1994) studied optimal tests for parametric models using the weighted average power criterion originally introduced by Wald (1943) when studying the likelihood ratio test under regularity conditions, where the model is identifiable under the null. Under loss of identifiability, the likelihood ratio test is generally less powerful than the optimal test in Andrews and Ploberger (1994). These optimal tests possess a Bayesian interpretation, where the weight corresponds to a prior on the nonidentifiable parameter, and are asymptotically equivalent to a Bayesian posterior odds ratio.

In this paper we adapt the weighted average power criterion (Andrews and Ploberger, 1994; Wald 1943) to construct optimal tests in semiparametric models under loss of identifiability. We extend the results of Andrews and Ploberger (1994) in at least four directions.

First, Andrews and Ploberger (1994) address only parametric models, as is the case for most literature on testing problems with nonidentifiability under the null. Our optimality results are available for semiparametric models, where η may be infinite dimensional and ζ may not be estimable at the usual parametric rate under both the null and the alternative. A semiparametric profile likelihood is adopted to reduce the infinite-dimensional model to a finite-dimensional uniformly least-favorable submodel; see Murphy and van der Vaart (2000) for a discussion of profile likelihood in regular settings. We note that, however, uniformly least favorable submodel is a new concept, which is not discussed in Murphy and van der Vaart (2000). The development of this concept is both nontrivial and critical to establishing an appropriate optimality criterion for general semiparametric models under loss of identifiability.

Second, the results of Andrews and Ploberger (1994) are applicable for tests where a nuisance parameter (namely ζ) is present only under the alternative. This may not be true in our situation, where a nondegenerate reparameterization may be needed to make ζ vanish under the null. Furthermore, our tests and the optimality results do not depend on the reparameterization.

Third, Andrews and Ploberger (1994) establish that their test is optimal with respect to local alternatives for ψ involving a multivariate normal prior with singular covariance matrix. In our approach, it is only necessary to specify the prior in the direction of β, the parameter of interest, and no prior is needed on the remaining parameter η. This enables us to avoid the singular covariance issue in Andrews and Ploberger (1994).

Fourth, we develop a simple and effective Monte Carlo method of inference for the proposed test statistics.

Adopting a profile likelihood approach has several advantages. First, under the identifiable submodel, the MLE for η may converge at a slower rate than the usual $\sqrt{n}$ rate, such as the change-point Cox model with current status data. This makes the theoretical justification based on Taylor expansion of the full likelihood fail. Second, even if the MLE of the nonparameric component converges at the $\sqrt{n}$ rate, semiparametric likelihoods may not be suitably "differentiable," in particular, when such a likelihood contains certain empirical terms, as with, for example, the odds-rate model. Third, handling the remainder terms in a Taylor type expansion is challenging, owing to the presence of the infinite dimensional parameters, and a delicate Banach space analysis is required. Employing the profile likelihood enables us to address these issues rigorously.

The remainder of the paper is organized as follows. In section 2, we present three rather different examples for which loss of identifiability occurs under the null. The examples include a two-component mixture model, an odds-rate model with right censored data and a change-point Cox model with current status data. In section 3, we present the generic testing problem and the model and data assumptions. The optimality results are given in Section 4. We verify that the results hold for the three examples in Section 5. The detailed proofs are given in Section 6.

2. Examples. This section contains three examples. For each example, we discuss how loss of identifiability arises under the null. We also present previous work on testing under such nonidentifiability in order to clarify the potential contributions of the results in this paper.

2.1. *A two-component mixture model.* Finite mixture models arise in many applications. For simplicity, we consider the simplest case of a two-component mixture with density $g(\rho, \mu_1, \mu_2, \eta) = \rho f(\mu_1, \eta) + (1-\rho) f(\mu_2, \eta)$, where $f(\mu, \eta)$ is a parametric p.d.f. with parameters $\mu \in \mathbb{R}^p$, $\eta \in \mathbb{R}^q$, such as a location-scale family. Let $\beta = \mu_2 - \mu_1$, $\theta = (\rho, \beta', \mu_1', \eta')'$, and the hypothesis of interest is $\beta = 0$; that is, there is only a single component in the mixture. For convenience, we assume the mixing proportion $\rho \in (0, 1]$ and $\mu_1 = \mu_2 = \mu_0$ under the null.

Under the null, the lack of identifiability of ρ, μ_1, and μ_2 complicates statistical inference. The difficulties have been widely studied in the context of testing for homogeneity in finite mixture models (Chen and Chen, 2003; Chen et al., 2004; Chernoff and Lander, 1995; Dacunha-Castelle and Gassiat, 1999; Lindsay, 1995; Liu and Shao, 2003; Zhu and Zhang, 2004). To our knowledge, all existing results focus on the limiting distribution of the likelihood ratio tests (and other associated likelihood based tests) under the null. These properties are rather complicated and may require that nuisance parameters are known under the null, where a reparameterization is often used (Liu and Shao, 2003; Zhu and Zhang, 2004). When the nuisance parameters are unknown under both H_0 and H_1, limited results have been developed. Furthermore, using such reparameterizations may yield models with unbounded second derivatives and usual first order Taylor approximation may not be possible, as in Dacunha-Castelle and Gassiat (1999), where a 5th order expansion was required.

In this paper, we address the test of homogeneity for mixture models with general parametric densities and do not assume that nuisance parameters are known under the null. To our knowledge, optimality issues for homogeneity testing in mixtures has not been addressed in the statistical literature.

2.2. *An odds-rate model with right censored data.* We consider right censored survival data generated from an odds-rate model. Let T be a nonnegative random variable representing the failure time, C be the independent censoring time, $V \equiv \min(T, C)$ and $Z \equiv Z(\cdot)$ be a corresponding p-dimensional covariate process. The observed data $\{X_i = (V_i, \Delta_i, Z_i), i = 1, ..., n\}$ consists of n i.i.d. realizations of $X = (V, \Delta, Z)$, where $\Delta \equiv 1\{T \leq V\}$, $1\{\cdot\}$ is the indicator function. In this model, the hazard function of the survival time T given covariates Z is

$$\lambda\{t; Z(t), W\} = \eta(t) W \exp\{\beta^T Z(t)\}, \tag{1}$$

where t is the time index, W is an unobserved gamma frailty with mean 1 and variance ζ, β is a $p-$dimensional regression parameter and $\eta(\cdot)$ is a completely unspecified baseline hazard function.

When β is not zero, the odds-rate model has been treated extensively; see Kosorok et al. (2004); Murphy et al. (1997); Murphy and van der Vaart (1997, 2000); Parner (1998), among others. Scharfstein et al. (1998) considered semiparametric efficient estimation in the setting where the covariates are time-independent, ζ is assumed known and $\eta(\cdot)$ is assumed to be absolutely continuous. Bagdonavičius and Nikulin (1999) considered estimation for a class of proportional hazards model, which includes the odds-rate model with ζ unspecified, based on a modified partial likelihood. Kosorok et al. (2004) considered robust inference for odds-rate models when the frailty distribution and regression covariates may be misspecified. To our knowledge, problems associated with testing the null $\beta = 0$ when the frailty parameter is unknown have not been previously considered in the statistical literature.

It has been shown that ζ and $\eta(\cdot)$ are not identifiable under the null (Kosorok et al., 2004). Intuitively, when $\beta = 0$, the covariate process Z provides no information for the failure time process. The frailty W and the baseline hazard $\eta(\cdot)$ are not distinguishable from each other, hence ζ and $\eta(\cdot)$ are not identifiable. Thus, the testing problem described above is nonregular and standard asymptotic results are not applicable. In this paper, we propose an optimal test of $\beta = 0$.

2.3. *A change-point Cox model with current status data.* Change-point models have been studied extensively and have proven to be popular in clinical research. In many settings, a change-point effect is realistic and can be much easier to interpret than a quadratic or more complex nonlinear effect (Chappell, 1989). Change-point Cox models have been widely used in survival applications, as in Luo and Boyett (1997); Luo et al. (1997); Pons (2003), where likelihood ratio tests were investigated. However, to our knowledge, the issue of optimal testing has not been explored for such models.

Under current status censoring, a subject is examined once at a random observation time V and at that time it is observed whether the event time $T \leq V$ or not. The observed data $\{X_i = (V_i, \Delta_i, Z_i), i = 1, ..., n\}$ consists of n i.i.d. realizations of $X = (V, \Delta, Z)$, where $\Delta \equiv 1\{T \leq V\}$ and Z is a $d-$dimensional covariate. Here we let $d = 1$ for simplicity. In this example, we assume that the time to event T satisfies a change-point Cox model conditionally on the covariate Z. That is, the density of X is given by:

$$p_\theta(x) = \left(1 - e^{-e^{r_\gamma(z)}\Lambda(v)}\right)^{\Delta} \left(e^{-e^{r_\gamma(z)}\Lambda(v)}\right)^{1-\Delta} f_{V,Z}(v, z), \tag{2}$$

with $r_\gamma(z) = \alpha z + (\beta_1 + \beta_2 z)1\{z > \zeta\}$, where α, β_1 and β_2 are scalar regression parameters, ζ is the change-point parameter and $\Lambda(\cdot)$ is the cumulative baseline hazard function. We also define the collected parameters $\beta \equiv (\beta_1, \beta_2)$, $\xi \equiv (\beta, \alpha)$, $\gamma \equiv (\xi, \zeta)$ and $\eta \equiv (\alpha, \Lambda)$. We are particularly interested in the hypothesis test of the existence of change-points for regression parameters in Cox models based on the above current status data, that is, $H_0: \ \beta = 0$.

Although Cox regression with current status data was discussed by Huang (1996) and others, change-point Cox regression has not been studied with current status data. The development of optimal tests in the current status setting is further complicated by the fact that the nuisance parameter Λ cannot be estimated at the parametric rate, unlike with right censored data.

In model (2), the change-point parameter is present only under the alternative. The theoretical results presented in the sequel will yield optimal tests with both right censored data, where the nuisance parameter is root-n estimable, and current status data, where it is not root-n estimable.

3. The hypotheses tests and assumptions.

3.1. *The optimal tests.* In this subsection we formulate the tests of hypotheses when the parameters are not identifiable under the null. Let P_θ denote the probability measure, based on observed data $\tilde{X}_n \equiv (X_1, X_2, ...X_n)$, where $\theta \in \Theta$ and the subscript n is the sample size. As mentioned previously, the parameters $\theta \in \Theta_0$ under the null hypothesis are not identifiable. We assume that θ can be partitioned as (ψ, ζ), with ζ q-dimensional and ψ of arbitrary dimension. We further assume that ψ can be partitioned as (β, η) so that the null hypothesis can be stated in terms of β, with the nuisance parameter η having arbitrary dimension. The likelihood function of the data is given by $l_n(\theta)$ and the profile likelihood for β and ζ is defined as $pl_n(\beta, \zeta) = \sup_\eta l_n(\beta, \eta, \zeta)$. For the semiparametric model $\{P_{(\beta,\eta,\zeta)}\}$ on a sample space $\mathcal{X}$, we assume $\beta \in \mathbb{R}^p$, $\zeta \in \Xi$, a compact subset of $\mathbb{R}^q$ and $\eta \in \mathcal{H}_\eta$, which is a subset of a Banach space.

The hypotheses to be tested are:

$$H_0: \quad \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0. \tag{3}$$

When $\beta = \beta_0$, the null distribution P_0 is unique and the likelihood for a single observation under the null is abbreviated as l^0. Let $\pi \equiv (\eta, \zeta)$. The null set of π is Π_0 and its cardinality is the same as that of Ξ. $\Theta_0 = \{\beta_0\} \times \Pi_0$. For each $\zeta \in \Xi$, $\eta_0(\zeta) \equiv \{t \in \mathcal{H}_\eta : (t, \zeta) \in \Pi_0\}$ is an interior point of $\mathcal{H}_\eta$. Let $\psi_0(\zeta) \equiv (\beta_0, \eta_0(\zeta))$, and $\theta_0(\zeta) \equiv (\psi_0(\zeta), \zeta)$. Thus, Θ_0 can be represented as $\Theta_0 = \{\theta_0(\zeta) : \zeta \in \Xi\}$.

We denote $\dot{l}_\beta \in L_2^0(P_\theta)$ as the derivative of $\log l_1(\theta)$ with respect to β and $\ddot{l}_\beta$ is the second derivative of $\log l_1(\theta)$ with respect to β. $L_2^0(P_\theta)$ refers to the class of square integrable functions under the measure P_θ with mean 0. The score operator for η is defined as $\dot{l}_\eta$, which is a bounded linear map from $\mathcal{H}_\eta$ to $L_2^0(P_\theta)$ with adjoint operator $\dot{l}_\eta^\star$: $L_2^0(P_\theta) \mapsto \overline{\mathcal{H}}_\eta$, where $\overline{\mathcal{H}}_\eta$ is the closed linear span of $\mathcal{H}_\eta$. The information operator is $\dot{l}_\eta^\star \dot{l}_\eta : L_2^0(P_\theta) \mapsto L_2^0(P_\theta)$. The efficient score for β is the ordinary score function $\dot{l}_\beta$ minus its orthogonal projection onto the closed linear span of the score operator $\dot{l}_\eta$. The efficient information for β is $\tilde{I}_\beta = \int \tilde{l}_\beta \tilde{l}_\beta' dP_\theta$, which is the asymptotic variance of the efficient score function.

We use the notations $\mathbb{P}_n$ and $\mathbb{G}_n$ for the empirical distribution and the empirical process of the observations. That is, for every measurable function

f and probability measure P,

$$\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i), \quad Pf = \int f dP, \;\; \mathbb{G}_n f = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i) - P(f)).$$

We note that although simultaneous estimation of β and ζ fails under the null due to nonidentifiability, estimation results for $\hat{\beta}_n(\zeta)$, the MLE of β at a fixed value of ζ, are usually valid under the null. This suggests making inference about β using $\hat{\beta}_n(\zeta)$, which is the approach taken in the previous literature on testing with nonidentifiable models. For fixed $\zeta \in \Xi$, the score, Wald and likelihood ratio test statistics for testing H_0 against H_1 are given by

$$\begin{aligned} R_n(\zeta) &= \mathbb{P}_n \dot{l}_\beta(\hat{\theta}_0(\zeta))' \left\{\mathbb{P}_n \dot{l}_\beta \dot{l}_\beta'(\hat{\theta}_0(\zeta))\right\}^{-1} \mathbb{P}_n \dot{l}_\beta(\hat{\theta}_0(\zeta)), \\ W_n(\zeta) &= (\hat{\beta}_n(\zeta) - \beta_0)' \mathbb{P}_n \left\{\dot{l}_\beta \dot{l}_\beta'(\hat{\theta}_n(\zeta))\right\} (\hat{\beta}_n(\zeta) - \beta_0), \text{ and} \\ LR_n(\zeta) &= -2\left\{l_n(\hat{\theta}_0(\zeta)) - l_n(\hat{\theta}_n(\zeta))\right\}, \end{aligned}$$

where $\hat{\theta}_n(\zeta) \equiv (\hat{\beta}_n(\zeta), \hat{\eta}_n(\zeta), \zeta)$ is the unrestricted MLE of θ at a fixed value of ζ and $\hat{\theta}_0(\zeta) \equiv (\beta_0, \hat{\eta}_0(\zeta), \zeta)$ is the restricted MLE of θ for a fixed value of ζ under the null. $\mathbb{P}_n \dot{l}_\beta(\hat{\theta}_0(\zeta)) = \mathbb{P}_n \dot{l}_\beta(\beta_0, \hat{\eta}_0(\zeta), \zeta)$ is the empirical score function of β evaluated at the restricted MLE $\hat{\theta}_0(\zeta)$. $\mathbb{P}_n \dot{l}_\beta(\hat{\theta}_n(\zeta)) = \mathbb{P}_n \dot{l}_\beta(\hat{\beta}_n(\zeta), \hat{\eta}_n(\zeta), \zeta)$ is the empirical score function of β evaluated at the unrestricted MLE $\hat{\theta}_n(\zeta)$. The inverse matrix of $\mathbb{P}_n \left\{\dot{l}_\beta \dot{l}_\beta'(\hat{\theta}_n(\zeta))\right\}$ estimates the covariance matrix of $\hat{\beta}_n(\zeta)$.

The optimal tests we propose take the form:

$$\begin{aligned} \mathrm{ER}_n &= (1+c)^{-\frac{p}{2}} \int \exp\left(\frac{1}{2}\frac{c}{1+c} R_n(\zeta)\right) dJ(\zeta), \\ \mathrm{EW}_n &= (1+c)^{-\frac{p}{2}} \int \exp\left(\frac{1}{2}\frac{c}{1+c} W_n(\zeta)\right) dJ(\zeta), \text{ and} \\ \mathrm{ELR}_n &= (1+c)^{-\frac{p}{2}} \int \exp\left(\frac{1}{2}\frac{c}{1+c} LR_n(\zeta)\right) dJ(\zeta), \end{aligned}$$

where $c > 0$ is a known constant and $J(\cdot)$ is a pre-selected integrable prior on ζ. Their optimality will be discussed in section 4. We note that, in semiparametric settings, the computation of the score and the information may involve high dimensional maximization and nonparametric smoothing (Huang, 1996). The tests ER_n and EW_n may be computationally harder than ELR_n. Thus the likelihood ratio based test ELR_n appears more attractive in these situations.

3.2. *The assumptions.* To derive asymptotically optimal tests of H_0, we consider local alternatives to H_0 of the form $l_n(\beta_0+h/\sqrt{n},\eta,\zeta)$ with ζ and η unspecified. The optimality criterion will involve a weighted average power criterion, where the averaging is with respect to an integrable prior $Q^c_\zeta(h)$ on the values of h in $\mathbb{R}^p$ defining local alternatives and an integrable prior $J(\zeta)$ on ζ. Before formally stating the optimality criterion, we give assumptions on the data and the parameter spaces. The first two assumptions postulate the existence of the prior on local alternatives, $Q^c_\zeta(h)$.

A1 For $\zeta \in \Xi$, the efficient information function of β evaluated at $\theta_0(\zeta)$, $\tilde{I}_\beta(\theta_0(\zeta))$, is uniformly continuous in β and ζ over $B_0 \times \Xi$, where B_0 is some neighborhood of β_0. Furthermore, $\tilde{I}_\beta(\theta_0(\zeta))$ is uniformly positive definite over $\zeta \in \Xi$, that is $\inf_{\zeta\in\Xi} \lambda_{\min}\left\{\tilde{I}_\beta(\theta_0(\zeta))\right\} > 0$, where $\lambda_{\min}(C)$ is the smallest eigenvalue of the matrix C.

A2 Q^c_ζ is a normal measure with mean β_0 and variance $c\tilde{I}^{-1}_\beta(\theta_0(\zeta))$ for $\zeta \in \Xi$, where $c > 0$ is a scalar constant.

Assumptions A1 and A2 are analogous to assumptions 1(e), 1(f) and 4 of Andrews and Ploberger (1994), although there are fundamental differences. Andrews and Ploberger (1994) work directly by building on the full parametric likelihood and their assumptions refer to the information matrix for all parameters. Furthermore, their optimality results are defined in terms of local alternatives for ψ, where the prior is a multivariate normal with singular covariance matrix. Our assumptions A1 and A2 are only for the parameter of interest, β, with no prior assumptions needed for η under either the null or the alternative.

The next assumption posits the existence of a full rank reparameterization.

B There exists a map $\phi_\zeta : \mathcal{H}_\eta \mapsto \mathcal{H}_\eta$, which is one-to-one and uniformly Hadamard-differentiable at η tangentially to $\mathcal{H}_\eta$ over $\zeta \in \Xi$, i.e.,

$$\sup_{(\eta+t_nh_n(\zeta),\zeta)\in\Pi_0} \left\| \frac{\phi_\zeta(\eta+t_nh_n(\zeta))-\phi_\zeta(\eta)}{t_n} - \dot{\phi}_\zeta(\eta)(h(\zeta)) \right\| \to 0,$$

as $\sup_{\zeta\in\Xi} \|h_n(\zeta)-h(\zeta)\| \to 0$, and $t_n \to 0$, where $h(\zeta)$ is in the tangent space of $\mathcal{H}_\eta$ for all $\zeta \in \Xi$ and $\|\cdot\|$ denotes the norm of $\mathcal{H}_\eta$. Its derivative $\dot{\phi}_\zeta$ is one-to-one and continuously invertible uniformly over $\zeta \in \Xi$. That is, there exists a positive constant k such that $\|\dot{\phi}_\zeta(\eta_1(\zeta)-\eta_2(\zeta))\| \geq k\|\eta_1(\zeta)-\eta_2(\zeta)\|$ for every $\eta_1(\zeta)$ and $\eta_2(\zeta)$ in $\mathcal{H}_\eta$ for all $\zeta \in \Xi$. Let $\overline{\eta} \equiv \phi_\zeta(\eta)$, $\ell_1(\beta_0,\overline{\eta},\zeta)(x) \equiv l_1(\beta_0,\phi^{-1}_\zeta(\overline{\eta}),\zeta)(x) = l^0(x)$, where ζ vanishes under the null, for all elements x in $\mathcal{X}$.

This reparameterization does not change the likelihood, that is, the equality $l_1(\beta, \eta(\zeta), \zeta)(x) = \ell_1(\beta, \overline{\eta}, \zeta)(x)$ holds both under the null and the alternative. Under the null, the likelihood $l_1(\beta_0, \eta_0(\zeta), \zeta) = \ell_1(\beta_0, \overline{\eta}_0, \zeta)$ for a specific $\overline{\eta}_0$, which does not depend on ζ, and ζ disappears in the null likelihood. We thus reduce the parameter dimension of the null space from Π_0 to $\mathcal{H}_\eta$.

The reason we assume the existence of such a full rank reparameterization is to eliminate the dependence between parameters η and ζ. The issue is that the results are with respect to a perturbation of the parameter η, which is not well defined in the original space, due to the dependence between parameters η and ζ. Subsequent assumptions are built on the new parameterization $\overline{\theta} \equiv (\beta, \overline{\eta}, \zeta)$. However, we note that the results still hold for the original parameterization, since the efficient score and efficient information are invariant under such reparameterization, as summarized in the following lemma:

LEMMA 1. *Under assumption B, $\tilde{l}_\beta(\theta) = \tilde{\ell}_\beta(\overline{\theta})$, where $\tilde{\ell}_\beta(\overline{\theta})$ is the efficient score of β under the new reparameterization. The efficient information matrix is also invariant to these reparameterizations.*

REMARK 1. *The full rank reparameterization in assumption B may not be unique. We will show later in the proof of theorem 2 that the optimal tests proposed in this paper are invariant to the choice of the full rank reparameterization.*

The next set of conditions assumes the existence of a uniformly least-favorable submodel. This submodel can be viewed as a "uniform" version of the least favorable submodel discussed in Murphy and van der Vaart (2000): the convergence rate of the nuisance parameter now is in the "uniform" sense, and the efficient score and the efficient information possess Donsker and Glivenko-Cantelli properties with "larger" index sets, respectively. When the set of ζ, Ξ, is a singleton, this new submodel concept reduces to the ordinary least favorable submodel. The development of this concept is critical to establishing an appropriate optimality criterion for general semiparametric models under loss of identifiability. Here are the needed assumptions:

C1 There exists a map $t \mapsto \overline{\eta}_t$ from a fixed neighborhood of β_0 into $\mathcal{H}_\eta$, such that the map $t \mapsto \ell(t, \overline{\theta})$ defined by $\ell(t, \overline{\theta}) \equiv \ell_1(t, \overline{\eta}_t, \zeta)$ is twice continuous differentiable. Let $\dot{\ell}(t, \overline{\theta})$ and $\ddot{\ell}(t, \overline{\theta})$ denote the derivatives with respect to t. The submodel with parameters $(t, \overline{\eta}_t, \zeta)$ passes through $\overline{\eta}$ at $t = \beta$, that is, $\eta_\beta(\beta, \overline{\eta}, \zeta) = \overline{\eta}$ for all $\zeta \in \Xi$.

C2 The submodel is uniformly least-favorable at $\overline{\psi}_0 \equiv (\beta_0, \overline{\eta}_0)$ and ζ for estimating β_0 in the sense that $\dot{\ell}(\beta_0, \overline{\psi}_0, \zeta) = \tilde{\ell}_\beta(\overline{\psi}_0, \zeta)$. As $(t, \beta, \overline{\eta}) \to (\beta_0, \beta_0, \overline{\eta}_0)$, we assume that $\sup_{\zeta \in \Xi} \|\dot{\ell}(t, \overline{\psi}, \zeta) - \tilde{\ell}_\beta(\overline{\psi}_0, \zeta)\| = o_{P_0}(1)$ and $\sup_{\zeta \in \Xi} \|\ddot{\ell}(t, \overline{\psi}, \zeta) - \ddot{\ell}(\beta_0, \overline{\psi}_0, \zeta)\| = o_{P_0}(1)$. In the sequel, we let o_P^{Ξ} denote a quantity going to zero in probability, under P, uniformly over the set Ξ.

C3 We assume that $\hat{\overline{\psi}}_0$, the restricted MLE of $\overline{\psi}$ under the null, satisfies $\hat{\overline{\psi}}_0 = \overline{\psi}_0 + o_{P_0}(1)$. The unrestricted MLE $\hat{\overline{\psi}}_n(\zeta) = \overline{\psi}_0 + o_{P_0}^{\Xi}(1)$. Moreover, let $\hat{\overline{\eta}}_\beta(\zeta) \equiv \mathrm{argmax}_{\overline{\eta}} \ell_n(\beta, \overline{\eta}, \zeta)$, that is $p\ell_n(\beta, \zeta) = \ell_n(\beta, \hat{\overline{\eta}}_\beta(\zeta), \zeta)$. Assume that for any random sequences $\tilde{\beta}_n \to_{P_0} \beta_0$, we have $\hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta) = \overline{\eta}_0 + o_{P_0}^{\Xi}(1)$ and the uniform "no-bias" condition:

$$P_0 \dot{\ell}(\beta_0, \tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) = o_{P_0}^{\Xi}(\|\tilde{\beta}_n - \beta_0\| + n^{-1/2}). \tag{4}$$

C4 There exist neighborhoods U of β_0 and V of $\overline{\psi}_0$, such that the class of functions $\{\dot{\ell}(t, \overline{\psi}, \zeta) : t \in U, \overline{\psi} \in V, \zeta \in \Xi\}$ is P_0-Donsker with square integrable envelope function and the class of functions $\{\ddot{\ell}(t, \overline{\psi}, \zeta) : t \in U, \overline{\psi} \in V, \zeta \in \Xi\}$ is P_0-Glivenko-Cantelli and is bounded in $L_1(P_0)$, where $L_1(P_\theta)$ refers to the class of integrable functions under P_θ.

Assumptions C1–C4 set the stage for the quadratic expansion of the profile likelihood and the derivation of the optimality properties of the proposed tests. Note that these assumptions can also be built on the original parameterization, but we use the new parameterization for ease of presentation. Since our formulation includes parametric models as special cases, the existence of a uniformly least-favorable submodel in our set-up covers all situations considered by Andrews and Ploberger (1994).

Compared with Andrews and Ploberger (1994), we have a stronger form of the unbiasedness condition and stronger requirements on the consistency of the estimators for the expansion of the profile likelihood. This is partly due to the more general structure of the semiparametric model. As in assumption C3, we require that if $\tilde{\beta}_n$ is any sequence of estimators consistent for β_0, $\hat{\overline{\eta}}_\zeta(\tilde{\beta}_n)$ must be consistent for $\overline{\eta}_0$, the true value of the nuisance parameter $\overline{\eta}$, uniformly over Ξ. In Andrews and Ploberger (1994), consistency is only needed for the unconstrained MLE (assumption 2) and the constrained MLE under the null hypothesis (assumption 3).

To evaluate the local asymptotic distribution of the proposed tests, we require differentiability in quadratic mean (DQM) of the parameters $\overline{\psi}$, as stated in the following assumption D, which is commonly used to evaluate the local power. It will be verified for the three examples presented in Section 2. Unlike in assumptions C1–C4, the full rank reparameterization is indispensable in assumption D:

D Differentiability in quadratic mean of the parameter $\overline{\psi}$. A perturbation of $\overline{\psi}$ in its domain is $\overline{\psi}_t = \overline{\psi}_0 + th + o(1)$, where $h \equiv (h_\beta, h_{\overline{\eta}})$, $h_\beta \in \mathbb{R}^p$ and $h_{\overline{\eta}} \in \overline{\mathcal{H}}_\eta$. The DQM condition for $\overline{\psi}_0$ with respect to the collection of paths $\{\overline{\psi}_t\}$ is :

$$\int \left[\frac{(dP_{\overline{\psi}_t,\zeta})^{1/2} - (dP_0)^{1/2}}{t} - \frac{1}{2}(A_\zeta h) dP_0^{1/2} \right]^2 \to 0, \text{as } t \to 0,$$

for all $\zeta \in \Xi$, where A_ζ is a bounded linear operator defined on $\mathbb{R}^p \times \overline{\mathcal{H}}_\eta$ and takes values in $L_2^0(P_\theta)$.

Differentiability in quadratic mean implies that the range of A_ζ is contained in $L_2^0(P_\theta)$. Note that $A_\zeta h = \partial/\partial t \ell_1(\overline{\psi}_t, \zeta)|_{t=0}$, following similar arguments as in Kosorok and Song (2007), where $h = (h_\beta, h_{\overline{\eta}})$. We define A_ζ to be given by $A_\zeta(h_\beta, h_{\overline{\eta}}) = \dot{\ell}'_\beta(\overline{\psi}, \zeta)h_\beta + \dot{\ell}_{\overline{\eta}}(\overline{\psi}, \zeta)h_{\overline{\eta}}$, where $\dot{\ell}_\beta$ and $\dot{\ell}_{\overline{\eta}}$ are the score operators for β and $\overline{\eta}$ respectively. Moreover, $\mathbb{R}^p \times \overline{\mathcal{H}}_\eta$ is a Hilbert space with $\|\cdot\|$ denoting its norm and $\langle \cdot, \cdot \rangle$ denoting its inner product. Since in parametric settings, twice continuously differentiability implies DQM (Pollard, 1995), this assumption is weaker than Assumption 1(c) in Andrews and Ploberger (1994).

4. Main results. This section includes several main results. The first one is about the asymptotic null distribution of the proposed tests.

4.1. *The distributions of the test statistics under the null.* To establish the asymptotic null distribution of the test statistics, a key result about the uniform profile likelihood expansion is summarized in the following lemma.

LEMMA 2. *Under assumptions A–C, for any random sequence $\tilde{\beta}_n \to_{P_0} \beta_0$,*

$$\begin{aligned} (5) \qquad \log pl_n(\tilde{\beta}_n, \zeta) &= \log pl_n(\beta_0, \zeta) + n(\tilde{\beta}_n - \beta_0)' \mathbb{P}_n \tilde{l}_\beta(\theta_0(\zeta)) \\ &\quad - \frac{1}{2} n(\tilde{\beta}_n - \beta_0)' \tilde{I}_\beta(\theta_0(\zeta))(\tilde{\beta}_n - \beta_0) + o_{P_0}^{\Xi}(\sqrt{n}\|\tilde{\beta}_n - \beta_0\| + 1)^2. \end{aligned}$$

Lemma 2 enables us to establish the asymptotic equivalence of these test statistics and their asymptotic distributions:

THEOREM 1. *Under assumptions A–C and* $H_0 : \beta = \beta_0$, $ELR_n = EW_n + o_{P_0}(1) = ER_n + o_{P_0}(1) \to_d e\chi(c)$, *where*

$$e\chi(c) = (1+c)^{-\frac{p}{2}} \int \exp\left(\frac{1}{2}\frac{c}{1+c}\mathbb{G}'(\theta_0(\zeta))\tilde{I}_\beta^{-1}(\theta_0(\zeta))\mathbb{G}(\theta_0(\zeta))\right) dJ(\zeta),$$

and $\mathbb{G}(\theta_0(\zeta))$ *is the limiting process of* $\mathbb{G}_n\tilde{l}_\beta(\theta_0(\zeta))$, *which is a mean zero Gaussian process with variance function* $\sigma^2(\zeta) = \tilde{I}_\beta(\theta_0(\zeta))$ *indexed by* ζ *and with covariance function* $\sigma^2(\zeta_1, \zeta_2) = P_0\left\{\tilde{l}_\beta(\theta_0(\zeta_1))\tilde{l}_\beta(\theta_0(\zeta_2))'\right\}$, *indexed by* ζ_1 *and* ζ_2, ζ, ζ_1 *and* $\zeta_2 \in \Xi$.

REMARK 2. *We note that when* $J(\cdot)$ *does not equal a prior on* ζ, *corresponding rather to a weight function, the results in Theorem 1 will generally hold, although the test may no longer possess the optimality discussed in the sequel. Theorem 1 should also hold if* $Q_\zeta^c(h)$ *is not a prior distribution, corresponding rather to a weight function on local alternatives for* β. *This robustness indicates that the tests are generally valid for testing the null under loss of identifiability, yielding a large class of test statistics, with the optimal test being a member of this class.*

We also note that the optimal tests depend on the weight function $Q_\zeta^c(\cdot)$ only through the scalar c. The larger c is, the more weight is given to alternatives for which β is large. For example, for a test of the change-point model, larger values of c correspond to greater weight being given to larger changes. In the special case where $J(\zeta)$ is a pointmass at a single value ζ_0, the optimal test rejects if and only if $LR(\zeta_0)$ exceeds some constant (i.e., the optimal test equals the standard score test for fixed ζ_0) and the optimal test is independent of c. When $J(\zeta)$ is not a pointmass distribution, however, the optimal test ELR_n depends on c. The larger c is, the more power is directed at alternatives for which β is large.

The limit as $c \to 0$ of the $2(ELR_n - 1)/c$ statistic is equal to the "average score" statistic $\int LR_n(\zeta)dJ(\zeta)$, which is the limit of the ELR statistics that are designed for alternatives that are very close to the null hypothesis. At the other extreme, the limit as $c \to \infty$ is $\log \int \exp(LR_n(\zeta)/2)dJ(\zeta)$. Thus for testing against more distant alternatives the optimal test statistic is still of an average exponential form.

If the constant $c/(1+c)$ which appears in the definition of ELR_n is replaced by a constant $r > 0$, then the limit as $r \to \infty$ of ELR_n is the likelihood ratio test, equivalently, the "sup score" statistic studied in Kosorok and Song (2007). Hence, the sup score test is designed for distant alternatives, but is of a more extreme form than the optimal exponential test, since the latter requires $r < 1$. It can be easily shown as a corollary to theorem 1 that the usual likelihood ratio, Wald and score tests have the following distribution:

COROLLARY 1. *Under the null hypotheses and assumptions A-D,* $\sup_\zeta LR_n(\zeta) = \sup_\zeta W_n(\zeta) + o_{P_0}(1) = \sup_\zeta R_n(\zeta) + o_{P_0}(1) \to_d \chi$, *with* $\chi = \sup_\zeta \mathbb{G}'(\theta_0(\zeta))\tilde{I}_\beta^{-1}(\theta_0(\zeta))\mathbb{G}(\theta_0(\zeta))$.

4.2. *Optimality of the proposed tests.* The second main result of this paper is the optimality property of the proposed tests. Following assumptions in section 3, we consider local alternatives $\beta = \beta_0 + h_\beta/\sqrt{n} + o(n^{-1/2})$ for $h_\beta \in \mathbb{R}^p$ with prior distribution $Q_\zeta^c(h_\beta)$ on the local alternative direction h_β and prior distribution $J(\zeta)$ on the nonidentifiable parameter ζ. The optimality result is as follows:

THEOREM 2. *Under assumptions A–D, the test statistics in theorem 1 are asymptotically uniformly most powerful for testing* $H_0 : \beta = \beta_0$ *against the contiguous alternative*

$$\int dP^n_{\overline{\psi}_0 + h/\sqrt{n} + o(n^{-1/2}),\zeta} dQ_\zeta^c(h_\beta) dJ(\zeta),$$

where $h \equiv (h_\beta, h_\eta(\zeta))$, $h_\eta(\zeta) \equiv \tilde{q}_\zeta' h_\beta$ *and where* $\tilde{q}_\zeta \equiv -(\dot{\ell}_\eta^\star \dot{\ell}_\eta)^- \dot{\ell}_\eta^\star \dot{\ell}_\beta(\overline{\psi}, \zeta)$ *is the uniformly least-favorable direction indexed by* ζ. *Moreover, this optimality result is invariant under the choice of* ϕ_ζ *in assumption B.*

Theorem 2 also implies that the proposed tests have the greatest weighted average power asymptotically in the class of all tests of asymptotic significance level α, against the alternative $P^n_{\overline{\psi}_0 + h/\sqrt{n} + o(n^{-1/2}),\zeta}$. That is, they maximize

$$\overline{\lim_{n\to\infty}} \int P(\phi_n \text{ rejects} | \overline{\psi}_0 + h/\sqrt{n} + o(n^{-1/2}), \zeta) dQ_\zeta^c(h_\beta) dJ(\zeta)$$

over all tests ϕ_n of asymptotic level α.

Our optimality results are under alternatives $\beta_0 + h_\beta/\sqrt{n} + o(n^{-1/2})$, with nonsingular normal weights on h_β. Our weights on h_β are precisely Andrews and Ploberger's [2] weights projected onto the parameter space that is of interest. Thus our results and Andrews and Ploberger's are consistent.

We now discuss the choice of the direction q_ζ, the priors $Q_\zeta^c(\cdot)$ and $J(\cdot)$. By the Neyman-Pearson lemma, for any appropriate prior distributions $Q_\zeta^c(\cdot)$ and $J(\cdot)$ and any known directions q_ζ, a UMP test for testing $H_0 : \beta =$

β_0 against the contiguous alternative $\int dP^n_{\overline{\psi}_0+h/\sqrt{n}+o(n^{-1/2}),\zeta} dQ^c_\zeta(h_\beta)dJ(\zeta)$, where $h \equiv (h_\beta, h_\eta(\zeta))$, $h_\eta(\zeta) = q'_\zeta h_\beta$ is defined by

$$\gamma_n = \begin{cases} 1, \text{ if } QLR_n > k_{\alpha n}, \\ \lambda_n, \text{ if } QLR_n = k_{\alpha n}, \\ 0, \text{ if } QLR_n < k_{\alpha n}, \end{cases}$$

where $k_{\alpha n} > 0$, $\lambda_n \in [0,1]$ are constants such that the rejection probability is α under the null and

$$\mathrm{QLR}_n = \frac{\int l_n\left(\overline{\psi}_0 + h/\sqrt{n} + o(n^{-1/2}), \zeta\right) dQ^c_\zeta(h_\beta)dJ(\zeta)}{l^0_n}.$$

We have the following result:

Corollary 2. *Under assumptions A–D, the null hypothesis and the contiguous alternatives,*

$$QLR_n = (1+c)^{-\frac{p}{2}} \int \exp\left(\frac{1}{2}\frac{c}{1+c} LR_n(\zeta)\right) W(q_\zeta, \zeta)dJ(\zeta) + o_p(1),$$

where $W(q_\zeta, \zeta) \le 1$ is defined in (17). When $q_\zeta = \tilde{q}_\zeta$, $W(\tilde{q}_\zeta, \zeta) = 1$ and $QLR_n = ELR_n + o_{P_0}(1)$.

As the alternatives we consider are contiguous to the null, in each direction q_ζ, which indexes QLR_n, there exists a consistent estimator $\tilde{\eta}_n(q_\zeta)$ of $\eta_0(\zeta)$ by the convolution theorem, provided certain conditions hold. The optimal tests can thus be built on $\tilde{\eta}_n(q_\zeta)$.

In applications with composite hypotheses where q_ζ is unknown, there may not exist a direction which can maximize the power over all directions (Bickel et al., 2006). In a regular testing problem where all parameters are identifiable, it can be shown that the likelihood ratio test, which is built on the uniformly least-favorable direction, will maximize the minimum power of all directions of the alternatives, over all the test based directions. In our nonregular testing problem, the situation is further complicated, since the power depends on the covariance structure of $\mathbb{G}(\theta_0(\zeta))$. It is not clear if the maxmin property still holds in our problem. We note that, however, our tests can be interpreted as the "maximum direction" test. Moreover, since the power of the test is not affected by multiplying by a constant in QLR_n, we can standardize $W(q_\zeta, \zeta)dJ(\zeta)$ to obtain $d\tilde{J}(\zeta)$, which is a probability measure on ζ. Then the question of the optimal choice of both q_ζ and $J(\zeta)$ reduces to the question of the optimal choice of $\tilde{J}(\zeta)$. Hence, without loss of generality we can replace q_ζ with $\tilde{q}_\zeta$. For this reason, we should choose $q_\zeta = \tilde{q}_\zeta$ and focus on the choice of $Q^c_\zeta(\cdot)$ and $\tilde{J}(\cdot)$ for optimization.

The main reason we use the normal weight for Q^c_ζ in this paper is to facilitate a comparison with Andrews and Ploberger (1994). Using the normal prior with covariance matrix proportional to the efficient information matrix also leads to a significant simplification of the representation of the test statistics, since many terms cancel in the proof of theorem 1. However we note that the choice of $Q^c_\zeta(\cdot)$ is not limited to the normal weight studied in this paper, as indicated in the proof of theorem 2. More general choices of the priors $Q^c_\zeta(\cdot)$ and $J(\cdot)$ merit future consideration, but this is beyond the scope of the current paper.

The optimality of the likelihood ratio statistics with loss of identifiability under the null for semiparametric models is of potential interest. Similar to the likelihood ratio test under loss of identifiability with parametric models (Andrews and Ploberger, 1994), in the semiparametric setting, the profile likelihood ratio statistic is not of the optimal average exponential form. It can be shown to be a limit of an average exponential test, but only if one considers the limit as a parameter being pushed beyond an admissible boundary, similar to Andrews and Ploberger (1995).

4.3. *The distributions of the test statistics under local alternatives.* To gain insight into the power of the optimal tests in practice, it is worthwhile to study their asymptotic distributions under local alternatives. In the following two theorems, theorem 3 gives the asymptotic distribution for fixed local alternatives $P^n_{\psi_0+h/\sqrt{n}+o(n^{-1/2}),\zeta_1}$, while theorem 4 gives the asymptotic distribution for random local alternatives $\int dP^n_{\psi_0+h/\sqrt{n}+o(n^{-1/2}),\zeta} dQ^c_\zeta(h_\beta)dJ(\zeta)$. As shown in the theorems, the distributions depend on the form of the alternative, which will depend in part on the specifics of the application. These results also usually depend on the prior distributions $J(\cdot)$ and $Q^c_\zeta(\cdot)$, for both fixed alternatives and random alternatives, in different manners though.

THEOREM 3. *Under local alternatives $P^n_{\psi_0+h/\sqrt{n}+o(n^{-1/2}),\zeta_1}$ and assumptions A–D, $ELR_n = EW_n + o_p(1) = ER_n + o_p(1) \to_d f\chi(c)$, with*

$$f\chi(c) = (1+c)^{-\frac{p}{2}} \int \exp\left[\frac{1}{2}\frac{c}{1+c}\left\{\mathbb{G}(\theta_0(\zeta)) + \nu_\star(h_\beta,\zeta,\zeta_1)\right\}' \tilde{I}_\beta^{-1}(\theta_0(\zeta))\left\{\mathbb{G}(\theta_0(\zeta)) + \nu_\star(h_\beta,\zeta,\zeta_1)\right\}\right] dJ(\zeta),$$

where $\nu_\star(h_\beta,\zeta,\zeta_1) \equiv P_0\tilde{l}_\beta(\theta_0(\zeta))\tilde{l}_\beta(\theta_0(\zeta_1))' h_\beta$.

Now we establish the asymptotic distribution of the test statistics under the alternative $\int dP^n_{\psi_0+h/\sqrt{n}+o(n^{-1/2}),\zeta} dQ^{c_1}_\zeta(h_\beta)dJ(\zeta)$:

THEOREM 4. *Under assumptions A–D and the local alternative* $\int dP^n_{\psi_0+h/\sqrt{n}+o(n^{-1/2}),\zeta} dQ^{c_1}_\zeta(h_\beta)dJ(\zeta)$, $ELR_n = EW_n + o_p(1) = ER_n + o_p(1) \to_d r\chi(c, c_1)$, *where* $r\chi(c, c_1)$ *is a real random variable such that its cumulative distribution function* $Pr(r\chi(c, c_1) \le t) = P_0\left[1\{e\chi(c) \le t\}e\chi(c_1)\right]$, *where* $e\chi(\cdot)$ *is as defined in theorem 1, and both* $e\chi(c)$ *and* $e\chi(c_1)$ *are computed from the same realization of the process* $\mathbb{G}(\theta_0(\cdot))$.

4.4. *Monte Carlo computation and inference.* Although we have obtained the asymptotic distributions of the test statistics, these distributions generally have complicated analytic forms which depend on the values of unknown nuisance parameters. We now introduce a weighted bootstrap method to obtain the asymptotically valid critical values of $e\chi(c)$. This method does not require explicit evaluation of the limit distribution, thereby avoiding the numerical difficulties inherent to such evaluation.

We first generate n i.i.d. positive random variables $\kappa_1, \ldots, \kappa_n$, with mean $0 < \mu_\kappa < \infty$, variance $0 < \sigma^2_\kappa < \infty$ and with $\int_0^\infty \sqrt{P(\kappa_1 > u)}du < \infty$. Next, we divide each weight by the sample average of the weights $\bar{\kappa}$, to obtain "standardized weights" $\kappa^\circ_1, \ldots, \kappa^\circ_n$ which sum to n. For a real, measurable function f, define the weighted empirical measure $\mathbb{P}^\circ_n f \equiv n^{-1}\sum_{i=1}^n \kappa^\circ_i f(X_i)$. Let $\hat{\psi}^\circ_n(\zeta) = (\hat{\beta}^\circ_n(\zeta), \hat{\eta}^\circ_n(\zeta))$ denote the maximizer of $l^\circ_n(\psi, \zeta)$ over $\psi \in \Psi$ at fixed $\zeta \in \Xi$, where l°_n is obtained by replacing $\mathbb{P}_n$ with $\mathbb{P}^\circ_n$ in the definition of l_n. Similarly, let $\hat{\psi}^\circ_0(\zeta) = (\hat{\beta}^\circ_0(\zeta), \hat{\eta}^\circ_0(\zeta))$ denote the maximizer of $(l^0_n)^\circ(\psi, \zeta)$ over $\psi \in \Psi$ at fixed $\zeta \in \Xi$, where $(l^0_n)^\circ$ is obtained by replacing $\mathbb{P}_n$ with $\mathbb{P}^\circ_n$ in the definition of l^0_n, the log likelihood under the null. Now repeat the bootstrap procedure a large number of times $\tilde{M}_n$ and compute the differences of the bootstrapped unrestricted MLE and restricted MLE of β: $d\hat{\beta}^\circ_k(\zeta) = \hat{\beta}^\circ_{n,k}(\zeta) - \hat{\beta}^\circ_{0,k}(\zeta)$, $k = 1, ..., M_n$, as processes of ζ. Note that we are allowing the number of bootstraps to depend on n. Define $\zeta \mapsto \hat{\mu}_n(\zeta) \equiv \tilde{M}_n^{-1}\sum_{k=1}^{\tilde{M}_n} d\hat{\beta}^\circ_k(\zeta)$ and let

$$\zeta \mapsto \hat{V}_n(\zeta) = \tilde{M}_n^{-1}\sum_{k=1}^{\tilde{M}_n}\left(d\hat{\beta}^\circ_{1,k}(\zeta) - \hat{\mu}_n(\zeta)\right)\left(d\hat{\beta}^\circ_{1,k}(\zeta) - \hat{\mu}_n(\zeta)\right)'.$$

To estimate critical values, we compute the standardized bootstrap test statistics $T^\circ_{n,k} \equiv (1+c)^{-\frac{p}{2}} \times$

$$\int \exp\left[\frac{1}{2}\frac{c}{1+c}\left\{\left(d\hat{\beta}^\circ_{1,k}(\zeta) - \hat{\mu}_n(\zeta)\right)'\hat{V}_n^{-1}(\zeta)\left(d\hat{\beta}^\circ_{1,k}(\zeta) - \hat{\mu}_n(\zeta)\right)\right\}\right] dJ(\zeta),$$

for $1 \leq k \leq \tilde{M}_n$. For a test of size α, we compare the observed test statistics with the $(1-\alpha)$th quantile of the corresponding $\tilde{M}_n$ standardized bootstrap statistics. The reason we subtract off the mean is to ensure that we obtain a valid approximation to the null distribution when the null hypothesis may not be true. If not, then there may be loss of power, although the type I error rate will still be controlled when the null is true. The proof of the bootstrap validity can be built upon the proof of theorems 7–8 in Kosorok and Song (2007). We omit the details.

5. Examples, continued. In this section, we study the examples introduced in Section 2 to illustrate the variation in nonidentifiability settings, to show the different ways that the full rank reparameterizations and uniformly least-favorable submodels can be defined.

5.1. *A two-component mixture model.* We begin by noting that Andrews and Ploberger (1994) is not applicable to this example since their assumption 1(f) is violated here. To be concrete, the information $I(\theta)$ for β and μ_1 evaluated at the null is:

$$I(\theta)|_{(\beta,\mu_1)=(0,\mu_0)} = P\left[\left.\frac{\partial \log g^2(\theta)}{\partial(\beta,\mu_1)^2}\right|_{(0,\mu_0)}\right] = \begin{pmatrix} \rho^2 C_{11}(\mu_0,\eta) & \rho C_{11}(\mu_0,\eta) \\ \rho C_{11}(\mu_0,\eta) & C_{11}(\mu_0,\eta) \end{pmatrix},$$

with

$$C_{11}(\mu_0,\eta) = P\left[\left.\frac{\partial^2 \log f(\mu,\eta)}{\partial \mu^2}\right|_{\mu=\mu_0}\right].$$

Thus $I(\theta)$ is not positive definite at any value of ρ, which is due to the fact that μ_1 and μ_2 are mutually indistinguishable under the null. To account for this, we choose $\zeta = (\mu_1', \rho)'$ and $\psi = (\beta, \eta)$. In this example, ζ disappears under the null and we may take $\phi(u) = u$, the identity map. In contrast with Andrews and Ploberger (1994), we are able to enlarge the "nuisance" parameter space under the null such that the parameters in their orthogonal space are identifiable. The identifiability of this full rank submodel permits the use of conventional techniques for deriving optimality results, such as Taylor expansion.

Assumption C is easily satisfied since the existence of the least favorable submodel is self-evident for parametric models. Assumption A is true if the information of μ and η in $f(\mu,\eta)$ is positive definite and continuous in μ and η, since the efficient information $\tilde{I}_\beta(\theta_0(\zeta)) = \rho^2\left[C_{11} - C_{12}C_{22}^{-1}C_{21}\right](\theta_0(\zeta))$, where

$$C_{11} = \frac{\partial^2 \log f(\mu,\eta)}{\partial \mu^2}, C_{21} = C_{12}' = \frac{\partial^2 \log f(\mu,\eta)}{\partial \mu \partial \eta}, \quad \text{and } C_{22} = \frac{\partial^2 \log f(\mu,\eta)}{\partial \eta^2}.$$

The continuous differentiability of g with respect to ψ will imply the DQM of $\psi = (\beta', \eta')'$. The proof is similar to that in Liu and Shao (2003).

5.2. *An odds-rate model with right censored data.* The odds-rate model we consider in this paper posits that the hazard function has the form (1). We define $g_\zeta(s) \equiv (1+\zeta s)^{-1/\zeta}$, for $\zeta > 0$, and $g_0(s) \equiv \lim_{\zeta \downarrow 0} g_\zeta(s) = \exp(-s)$. Let $S_Z(\cdot)$ denote the survival function of T given Z, after integrating over W, $S_Z(t)$ becomes $g_\zeta\left(\int_0^t e^{\beta' Z(u)} d\eta(u)\right)$, where the cumulative baseline hazard function $\eta(\cdot)$ is a non-negative, monotone increasing cadlag (right-continuous with left-hand limits) function. We will argue later that assumptions A–D can be checked under the following assumptions. The true null survival function is unique and denoted S_0. The censoring time C is independent of T given Z and uninformative of ζ and β. Moreover, for a finite time point τ, $P_0 1\{C \geq \tau\} = P_0 1\{C = \tau\} > 0$ almost surely. $\zeta \in \Xi \equiv [0, K_0]$ for some known $K_0 < \infty$. The null value $\beta_0 = 0$ is an interior point of a known compact set $B_0 \in \mathbb{R}^p$. The parameter space for η, $\mathcal{H}_\eta$, is a Banach space consisting of continuous and monotone increasing functions on the interval $[0,\tau]$ equipped with the total variation norm $\|\cdot\|_v$. Its closed linear span is denoted as $\overline{\mathcal{H}}_\eta$. The function $\eta(\cdot) \in \mathcal{H}_\eta$ satisfies $\eta(0) = 0$ and $\eta(\tau) < \infty$. The covariate process $Z(\cdot)$ is uniformly bounded in total variation on $[0,\tau]$ and $\text{var}[Z(0+)]$ is positive definite.

The true values of $\pi \equiv (\eta, \zeta)$ are not unique under the null, since the null set Π_0 contains all pairs of (η,ζ) satisfying, for $t \in [0,\tau]$, $(1+\zeta\eta(t))^{-1/\zeta} = S_0(t)$, when $\zeta \in (0, K_0]$; and $\exp(-\eta(t)) = S_0(t)$, when $\zeta = 0$. In this example, ζ appears both under the null and the alternative. Equivalently, for any fixed $\zeta \in (0,K_0]$, $\eta_0(t)(\zeta) = \left(S_0(t)^{-\zeta} - 1\right)/\zeta$ and for $\zeta = 0$, $\eta_0(t)(\zeta) = -\log(S_0(t)),\ t \in [0,\tau]$. Hence $\Pi_0 = \{(\zeta, \eta_0(\zeta)) : \zeta \in \Xi\}$. Let $\overline{\eta} = \phi_\zeta(\eta) \equiv (1+\eta\zeta)^{1/\zeta} - 1$, for $\zeta > 0$; and $\overline{\eta} = \lim_{\zeta \to 0} \phi_\zeta(\eta) = \exp(\eta) - 1$. It can be easily checked that $\overline{\eta} \in \mathcal{H}_\eta$. The following arguments reveal that the map $\phi_\zeta(\eta) : \mathcal{H}_\eta \mapsto \mathcal{H}_\eta$ satisfies assumption B.

The log likelihood function with the new parameter $\overline{\theta} = (\beta, \overline{\eta}, \zeta)$ is

$$(6) \qquad \ell_n(\overline{\theta}) = \mathbb{P}_n \Big[\delta\{\log a_1(v) + (\zeta - 1)\log(\overline{\eta}(v)+1)\} + \beta' z(v) + (1+\delta\zeta)\log g_\zeta\left\{\int_0^v e^{\beta' z(s)} (\overline{\eta}(s)+1)^{\zeta-1} d\overline{\eta}(s)\right\}\Big],$$

where $a_1(\cdot)$ is the derivative of $\overline{\eta}(\cdot)$. We will replace $a_1(\cdot)$ with $n\Delta\overline{\eta}(\cdot)$ in the sequel, since this form of the empirical log-likelihood function is asymptotically equal to the true log-likelihood function.When $\beta = 0$, it is clear that ζ vanishes since (6)=$\mathbb{P}_n\{\delta \log \Delta\overline{\eta}(v) - (\delta+1)\log(1+\overline{\eta}(v))\}$. The odds-rate model with new parameterization $\overline{\psi} \equiv (\beta, \overline{\eta})$ is identifiable under the null, since the null survival function $S_0(t|z) = (1+\overline{\eta})^{-1}$ is a strictly monotone function of $\overline{\eta}$ and unique.

The Gâteaux derivative of $\phi_\zeta(\eta)$ at $\eta \in \mathcal{H}_\eta$ exists and is obtained by differentiating $\phi_\zeta(\eta)$ along the submodels $t \mapsto \eta + th$. This derivative is $\dot{\phi}_\zeta(\eta)(h) \equiv \partial/\partial t \phi_\zeta(\eta+th)|_{t=0} = (1+\zeta\eta)^{1/\zeta-1}h$ for $\zeta > 0$ and $\exp(\eta)h$ for $\zeta = 0$.

The Gâteaux differentiability of $\phi_\zeta(\eta)$ pointwisely in ζ can be strengthened to uniform Fréchet differentiability by noticing that

$$\lim_{t\downarrow 0} \sup_{\zeta\in\Xi} \sup_{\|h\|_v \le r, h\in\overline{\mathcal{H}}_\eta} \left| \int_0^1 \left\{ \dot{\phi}_\zeta(\eta + sth(\zeta)) - \dot{\phi}_\zeta(\eta) \right\} ds \right| = 0,$$

for any $r > 0$. Thus $\sup_{\zeta\in\Xi}\sup_{\|h\|_v\le r, h\in\overline{\mathcal{H}}_\eta} \left\| \phi_\zeta(\eta + h(\zeta)) - \phi_\zeta(\eta) - \dot{\phi}_\zeta(\eta)(h(\zeta)) \right\|_v / \|h(\zeta)\|_v = o(1)$, as $\|h(\zeta)\|_v \to 0$ uniformly over $\zeta \in \Xi$, which we will hereafter refer to as "uniform Fréchet differentiability." Since $\dot{\phi}_\zeta(\eta)(h)$ is uniformly bounded and Lipschitz in h, by checking the definition, we can show that $\dot{\phi}_\zeta$ is one-to-one and continuously invertible uniformly over $\zeta \in \Xi$.

To define a uniformly least-favorable submodel, we calculate scores for β and $\overline{\eta}$. Let $\mathcal{H}$ denote the space of elements $h = (h_1, h_2)$ such that $h_1 \in \mathbb{R}^p$ and $h_2 \in \mathcal{H}_\eta$. Consider the one-dimensional submodel defined by the map $t \mapsto \overline{\psi}_t \equiv \overline{\psi} + t(h_1, \int_0^{(\cdot)} h_2(u)d\overline{\eta}(u))$, $h \in \mathcal{H}$. The derivative of $\log \ell_n(\overline{\psi}_t, \zeta)$ with respect to t evaluated at $t = 0$ yields score operators $\dot{\ell}_n(\overline{\psi}, \zeta)(h) \equiv (\dot{\ell}_{n\beta}(h_1), \dot{\ell}_{n\overline{\eta}}(h_2))$, where

$$\begin{aligned}
\dot{\ell}_{n\beta}(\overline{\psi},\zeta)(h_1) &= \mathbb{P}_n \dot{\ell}_\beta(h_1) \\
&= \mathbb{P}_n \left\{ \delta h_1' Z(X) - (1+\delta\zeta) \frac{\int_0^\tau h_1' Z(u) Y(u) e^{\beta' Z(u)} (\overline{\eta}(u)+1)^{\zeta-1} d\overline{\eta}(u)}{1 + \zeta \int_0^\tau h_1' Z(u) Y(u) e^{\beta' Z(u)} (\overline{\eta}(u)+1)^{\zeta-1} d\overline{\eta}(u)} \right\},
\end{aligned}$$

and

$$\begin{aligned}
\dot{\ell}_{n\overline{\eta}}(\overline{\psi},\zeta)(h_2) &= \mathbb{P}_n \dot{\ell}_{\overline{\eta}}(h_2) \\
&= \mathbb{P}_n \Bigg\{ \int_0^\tau \left(h_2(u) + \frac{(\zeta-1)\int_0^u h_2(s) d\overline{\eta}(s)}{\overline{\eta}(u)+1} \right) dN(u) - (1+\delta\zeta) \times \\
&\quad \frac{\int_0^\tau Y(u) e^{\beta' Z(u)} (\overline{\eta}(u)+1)^{\zeta-2} \left[(\zeta-1) \int_0^u h_2(s) d\overline{\eta}(s) + h_2(u)(1+\overline{\eta}(u)) \right] d\overline{\eta}(u)}{1 + \zeta \int_0^\tau Y(u) e^{\beta' Z(u)} (\overline{\eta}(u)+1)^{\zeta-1} d\overline{\eta}(u)} \Bigg\},
\end{aligned}$$

with $Y(u) \equiv 1\{V \geq u\}$.

To obtain the information operator, we consider the two-dimensional submodel defined by the map $(s,t) \mapsto \overline{\psi}_{st} \equiv \overline{\psi} + s(h_1, \int_0^{(\cdot)} h_2(u)d\overline{\eta}(u)) + t(\tilde{h}_1, \int_0^{(\cdot)} \tilde{h}_2(u)d\overline{\eta}(u))$, where $h, \tilde{h} \in \mathcal{H}$. Define $\mathcal{H}_\infty = \{h \in \mathcal{H} : \|h\|_{\mathcal{H}} < \infty\}$. The information operator $\overline{\sigma}_{\overline{\theta}}(h) : \mathcal{H}_\infty \mapsto \mathcal{H}_\infty$ is given by $-P_0\partial/\partial s\partial t\ell_1(\overline{\psi}_{st})\Big|_{s,t=0} = \overline{\psi}(\overline{\sigma}_{\overline{\theta}}(h))$. We will show $\overline{\sigma}_{\overline{\theta}}$ is one-to-one, continuously invertible and onto uniformly over $\zeta \in \Xi$, via part (1) of lemma 3 below for which it suffices to show that the information operator for the original parameterization σ_θ is one-to-one, continuously invertible and onto uniformly over $\zeta \in \Xi$:

Lemma 3. *(1) Assume $\phi_\zeta : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}_\psi \subset \mathbb{E}$ is one-to-one, continuously invertible and onto and $\psi_\zeta : \mathbb{E}_\psi \subset \mathbb{E} \mapsto \mathbb{F}$ is one-to-one, continuously invertible and onto, then $\psi_\zeta \circ \phi_\zeta : \mathbb{D}_{\phi_\zeta} \mapsto \mathbb{F}$ is one-to-one, continuously invertible and onto. (2). Assume $\phi_\zeta : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}_\psi \subset \mathbb{E}$ is uniformly Fréchet differentiable at $\theta \in \mathbb{D}_\psi$ and $\psi_\zeta : \mathbb{E}_\psi \subset \mathbb{E} \mapsto \mathbb{F}$ is uniformly Fréchet differentiable at $\phi_\zeta(\theta)$ over $\zeta \in \Xi$. Then $\psi_\zeta \circ \phi_\zeta : \mathbb{D}_\phi \mapsto \mathbb{F}$ is uniformly Fréchet differentiable at θ with derivative $\psi'_\zeta(\phi_\zeta(\theta)) \circ \phi'_\zeta(\theta)$.*

With the same derivation of $\overline{\sigma}_{\overline{\theta}}$, $\sigma_\theta : \mathcal{H}_\infty \mapsto \mathcal{H}_\infty$ takes the form

$$\sigma_\theta(h) = \begin{pmatrix} \sigma_\theta^{11} & \sigma_\theta^{12} \\ \sigma_\theta^{21} & \sigma_\theta^{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

where

$$\begin{aligned}
\sigma_\theta^{11}(h_1) &= -P_0 S(\theta) \int_0^\tau h_1' Z(u) Y(u) e^{\beta' Z(u)} d\eta_0(u), \\
\sigma_\theta^{12}(h_2) &= -P_0 S(\theta) \int_0^\tau h_2(u) Z(u) Y(u) e^{\beta' Z(u)} d\eta_0(u), \\
\sigma_\theta^{21}(h_1) &= -P_0 S(\theta)(1 + \zeta\eta(T \wedge \tau)(\zeta)) h_1' Z(u) Y(u) e^{\beta' Z(u)} \\
&\quad -\zeta P_0 S(\theta) Y(u) \int_0^\tau h_1' Z(u) Y(u) e^{\beta' Z(u)} d\eta_0(u), \\
\sigma_\theta^{22}(h_2) &= -P_0 S(\theta)(1 + \zeta\eta(T \wedge \tau)(\zeta)) h_2(u) Y(u) e^{\beta' Z(u)} \\
&\quad -\zeta P_0 S(\theta) Y(u) \int_0^\tau h_2(u) Y(u) e^{\beta' Z(u)} d\eta_0(u),
\end{aligned}$$

with $S(\theta) = -(1 + \delta\zeta)/(1 + \zeta\overline{\eta}(\tau))^2$.

All of the operators σ_θ^{ij}, $1 \leq i, j \leq 2$ are uniformly compact and bounded over $\zeta \in \Xi$. With a similar argument as in Kosorok et al. (2004), the linear operator $\sigma_\theta : \mathcal{H}_\infty \mapsto \mathcal{H}_\infty$ is one-to-one, continuously invertible and onto uniformly over $\zeta \in \Xi$ by verifying the conditions of lemma 4 below. Thus a uniformly least-favorable submodel for estimating β in the presence of $\overline{\eta}$ and ζ is $\overline{\eta}_t(\beta,\overline{\eta},\zeta) = \left(1 + (\beta - t)'\nu_{\overline{\theta}}\right) d\overline{\eta}$, where $\nu_{\overline{\theta}} : \mathbb{R} \mapsto \mathbb{R}^p$ is the uniformly least-favorable direction at $(\beta_0, \overline{\eta}, \zeta)$ defined by $h'\nu_{\overline{\theta}} = (\overline{\sigma}^{22}_{\overline{\theta}})^{-1}\overline{\sigma}^{21}_{\overline{\theta}}h,\ h \in \mathbb{R}^p$. This leads to $\ell(t,\beta,\overline{\eta},\zeta) = \ell_1(\beta,\overline{\eta}_t(\theta),\zeta)$. Because $\overline{\eta}_\beta(\beta,\overline{\eta},\zeta) = \overline{\eta}$, C1 is satisfied. Since $\partial/\partial t|_{t=\beta_0}\ell(t,\beta_0,\overline{\eta}_0,\zeta) = \dot{\ell}_\beta(\beta_0,\overline{\psi}_0,\zeta) = \tilde{\ell}_\beta(\overline{\psi}_0,\zeta)$, where $\tilde{\ell}_\beta(x) = \dot{\ell}_\beta - \dot{\ell}_{\overline{\eta}}\nu_\theta$ is the efficient score for β, C2 is satisfied due to the continuity of the involved functions with respect to $\overline{\psi}$ and the fact that Ξ is compact. The efficient information for β is $\tilde{\mathcal{I}}_\beta = P_0\tilde{\ell}_\beta\tilde{\ell}'_\beta$. That $\{\dot{\ell}(t,\overline{\psi},\zeta) :\ t \in U, \overline{\psi} \in V, \zeta \in \Xi\}$ is P_0-Donsker and $\{\ddot{\ell}(t,\overline{\psi},\zeta) :\ t \in U, \overline{\psi} \in V, \zeta \in \Xi\}$ is P_0-Glivenko-Cantelli for some neighborhoods U and V follows from standard empirical process arguments.

LEMMA 4. *Let $A_\zeta = T_\zeta + K_\zeta :\ \mathbb{D} \mapsto \mathbb{E}$ be a linear operator between Banach spaces, where T_ζ is onto and there exists $c_1 > 0$, such that $\|T_\zeta h\| \geq c_1\|h\|$ for all $h \in \mathbb{D}$ and $\zeta \in \Xi$, and K_ζ is uniformly compact, i.e., $\{\zeta \in \Xi, \|h\| \leq 1 : \cup_{\|h\|\leq 1} K_\zeta h\}$ is compact. Then if $N(A_\zeta) = \{0\}$ for all $\zeta \in \Xi$, then A_ζ is onto and there exists $c_2 > 0 :\ \|A_\zeta h\| \geq c_2\|h\|$, $\forall \zeta \in \Xi$ and all $h \in \mathbb{D}$.*

It follows from corollary 8.1.3 in Golub and Van Loan (1983) that the set of eigenvalues is a continuous function of the elements of $\tilde{\mathcal{I}}_\beta(\overline{\theta})$, which are continuous functions of ζ. The set of eigenvalues is therefore a continuous function of ζ. Thus $\inf_\zeta \lambda_{\min}\{\tilde{I}_\beta(\theta_0(\zeta))\} > 0$ by the compactness of Ξ, and assumption A1 is satisfied.

To show the consistency of the restricted MLE $\hat{\overline{\psi}}_0$ and the unrestricted MLE $\hat{\overline{\psi}}_n(\zeta)$, let $\hat{\overline{\theta}}_\beta(\zeta) \equiv (\beta, \hat{\overline{\eta}}_\beta(\zeta), \zeta)$. The score function $\dot{\ell}_{\overline{\eta}}(h_2)$ is equal to zero when evaluated at $\hat{\overline{\theta}}_\beta(\zeta)$. Since h_2 in $\dot{\ell}_{\overline{\eta}}(h_2)$ is arbitrary, we can choose $h_2(u) = 1\{u \leq t\}$ and equate the resulting expression to zero, which yields

$$\begin{aligned}&\mathbb{P}_n\left\{\int\left(\zeta + \frac{1-\zeta}{\overline{\eta}(u)+1}\right)dN(u)\right\}\\ &\quad = \mathbb{P}_n\left\{\frac{(1+\delta\zeta)\int_0^\tau Y(u)\exp^{\beta' Z(u)}(\overline{\eta}+1)^{\zeta-2}(\zeta\overline{\eta}(u)+1)d\overline{\eta}}{1+\zeta\int_0^v \exp^{\beta' z(u)}(\overline{\eta}+1)^{\zeta-1}d\overline{\eta}}\right\}.\end{aligned}$$

Let $\tilde{Q}_{n,\zeta} \equiv \mathbb{P}_n\left\{\int\left(\zeta + \frac{1-\zeta}{\overline{\eta}(u)+1}\right)dN(u)\right\}$ and define

$$W(u;\overline{\theta}) \equiv \frac{(1+\delta\zeta)\int_0^\tau Y(u)\exp^{\beta' Z(u)}(\overline{\eta}+1)^{\zeta-2}(\zeta\overline{\eta}(u)+1)d\overline{\eta}}{1+\zeta\int_0^v \exp^{\beta' z(u)}(\overline{\eta}+1)^{\zeta-1}d\overline{\eta}}.$$

From the above arguments, $\hat{\overline{\eta}}_\beta(\zeta)$ satisfies the following equation:

$$\hat{\overline{\eta}}_\beta(\zeta)(t) = \int_0^t \{\mathbb{P}_n W(u; \hat{\overline{\theta}}_\beta(\zeta))\}^{-1} d\tilde{Q}_{n,\zeta}(u).$$

Noting that $\tilde{Q}_{n,\zeta}(\cdot)$ is a monotone counting process at every fixed ζ, the consistency of $\hat{\overline{\eta}}_0$ follows from arguments similar to the proof of theorem 3 in Kosorok et al. (2004). The uniform consistency of $\hat{\overline{\eta}}_n$ over $\zeta \in \Xi$ also follows from the continuity of $\ell_n(\overline{\theta})$ in ζ and the fact that $(\dot{\ell}_{\overline{\psi}}(\overline{\theta}) : t \in [0,\tau], \overline{\psi} \in \Psi_0, \zeta \in \Xi)$ is P_0-Donsker, where $\dot{\ell}_{\overline{\psi}} \equiv (\dot{\ell}_\beta, \dot{\ell}_{\overline{\eta}})$ and Ψ_0 is some neighborhood of $\overline{\psi}_0$.

To verify the uniform no-bias condition (4), it suffices to show that $\sup_{\zeta\in\Xi} \|\hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta) - \overline{\eta}_0\|_\infty = O^\star_{P_0}(\|\tilde{\beta}_n - \beta_0\| + n^{-1/2})$, for any sequence $\tilde{\beta}_n \to \beta_0$, and * denotes the outer expectation. By verifying conditions in lemma 5 below, we have

$$\sup_{\zeta\in\Xi} (\mathbb{P}_n - P_0) \left\{ \dot{\ell}_{\overline{\psi}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) - \dot{\ell}_{\overline{\psi}}(\beta_0, \overline{\eta}_0, \zeta) \right\} = o^\star_{P_0}(n^{-1/2}).$$

Together with the fact that $\mathbb{P}_n \dot{\ell}_{\overline{\psi}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) = P_0 \dot{\ell}_{\overline{\psi}}(\beta_0, \overline{\eta}_0, \zeta) = 0$, we obtain $P_0 \left\{ \dot{\ell}_{\overline{\psi}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) - \dot{\ell}_{\overline{\psi}}(\beta_0, \overline{\eta}_0, \zeta) \right\} =$

$$P_0 \dot{\ell}_{\overline{\psi}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) - \mathbb{P}_n \dot{\ell}_{\overline{\psi}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) = -(\mathbb{P}_n - P_0) \dot{\ell}_{\overline{\psi}}(\beta_0, \overline{\eta}_0, \zeta) + o^\star_{P_0}(n^{-1/2}),$$

uniformly over $\zeta \in \Xi$.

Let $\dot{l}_\psi(h) \equiv (\dot{l}_\beta(h_1), \dot{l}_\eta(h_2))$ denote the score operator of ψ with the original parameterization. It was shown in Kosorok et al. (2004) that the operator $\psi \mapsto \dot{l}_\psi$ is Fréchet differentiable with derivative $\psi(\sigma_\theta(h))$, and it can be strengthened to uniform Fréchet differentiablity due to the smoothness of the involved functions. Since ϕ_ζ is uniformly Fréchet differentiable, by part (2) of lemma 3, the chain rule for uniform Fréchet differentiability, $\dot{\ell}_{\overline{\psi}} \equiv (\dot{\ell}_\beta, \dot{\ell}_{\overline{\eta}})$ is uniformly Fréchet differentiable with derivative $\sigma_{\phi_\zeta^{-1}(\overline{\theta})} \circ \dot{\phi}_\zeta^{-1}(\overline{\theta})$.

By the uniform Fréchet differentiability of $\dot{\ell}_{\overline{\psi}}$, $\overline{\sigma}_{\overline{\theta}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta) - \overline{\eta}_0) =$

$$P_0 \left\{ \dot{\ell}_{\overline{\psi}}(\tilde{\beta}_n, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) - \dot{\ell}_{\overline{\psi}}(\beta_0, \overline{\eta}_0, \zeta) \right\} + o^{\Xi}_{P_0}(\|\tilde{\beta}_n - \beta_0\| + \|\hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta) - \overline{\eta}_0\|_\infty).$$

Since $\overline{\sigma}_{\overline{\theta}}$ is linear, the first term on the right-hand side is of the order $O_{P_0}(n^{-1/2})$. It follows that $\sup_{\zeta\in\Xi} \|\hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta) - \overline{\eta}_0\|_\infty = O^\star_{P_0}(\|\tilde{\beta}_n - \beta_0\| + n^{-1/2})$, since $\overline{\sigma}_{\overline{\theta}}$ is uniformly continuous invertible over $\zeta \in \Xi$.

LEMMA 5. *Suppose that* $U_n(\psi,\zeta)(h) = \mathbb{P}_n\nu(\psi,\zeta)(h)$ *and* $U(\psi,\zeta)(h) = P\nu(\psi,\zeta)(h)$ *for given* P*-measurable functions* $\nu(\psi,\zeta)(h)$ *indexed by* $\Psi \times \Xi$ *and an arbitrary index set* $\mathcal{H}_\eta$. *Provided* $\psi = \psi_0$, $\nu(\psi_0,\zeta)(h) = \nu(\psi_0)(h)$. *If* $\hat{\psi}_n(\zeta) = \psi_0 + o_P^{\Xi}(1)$, *the class of functions* $\{\nu(\psi,\zeta)(h) - \nu(\psi_0)(h) : \|\psi - \psi_0\| < \delta, h \in \mathcal{H}_\eta, \zeta \in \Xi\}$ *is* $P-$*Donsker for some* $\delta > 0$ *and* $\sup_{\zeta\in\Xi,h\in\mathcal{H}_\eta} P_0(\nu(\psi,\zeta)(h) - \nu(\psi_0)(h))^2 \to 0$, *as* $\psi \to \psi_0$, *then* $\sup_{\zeta\in\Xi} \|\sqrt{n}(U_n - U)(\hat{\psi}_n(\zeta),\zeta) - \sqrt{n}(U_n - U)(\psi_0,\zeta)\| = o_P^\star(1 + \sqrt{n}\|\hat{\psi}_n(\zeta) - \psi_0\|)$.

5.3. *A change-point Cox model with current status data.* In the change-point Cox model with current status data, a test of the existence of a threshold effect corresponds to a test of the null $H_0 : \beta = 0$. The change-point parameter ζ is present only under the alternative. Hence it suffices to take ϕ_ζ as the identity map.

We make the following assumptions and will argue that the assumptions A, C and D in Section 3 can be checked under these assumptions. Given Z, T and V are independent. Z belongs to a compact subset of $\mathbb{R}$. The change-point parameter $\zeta \in [a,b]$, for some known $-\infty < a < b < \infty$ with $Pr(Z < a) > 0$ and $Pr(Z > b) > 0$. $P(\mathrm{Var}(Z|V)) > 0$, which guarantees that, as we will show later, the efficient information $\tilde{I}_\beta(\theta_0(\zeta))$ is positive definite uniformly over $\zeta \in \Xi$. The Lebesgue density of V is positive and continuous on its support $[\sigma,\tau]$ with $0 < \sigma < \tau < \infty$. The baseline hazard function Λ is continuously differentiable at $[\sigma,\tau]$, with derivative that is bounded away from 0 and satisfies $\Lambda_0(\sigma) > 0$, $\Lambda_0(\tau) < M$, for some known M. We let $\mathcal{H}_\Lambda$ denote a set of non-decreasing cadlag functions Λ on $[\sigma,\tau]$ with $\Lambda(\tau) \le M$.

The likelihood function equals (2) with $f_{V,Z}(v,z)$ removed, because it can be absorbed into the underlying measure on the sample space. The log-likelihood for a single observation $\log l_1(\theta)$ takes the form $\log l_1(\theta) = \delta \log[1 - \exp\{-\Lambda(v)\exp(r_\gamma(z))\}] - (1-\delta)\exp(r_\gamma(z))\Lambda(v)$. Define $Z(\zeta) \equiv (1\{Z > \zeta\}, Z1\{Z > \zeta\}, Z)$ and note that with such a data representation we can adopt much material in the literature and hence simplify our arguments.

To define a uniformly least-favorable submodel in β, we take two steps. For step 1, we calculate scores for ξ and Λ. The score function for ξ is $\dot{l}_\xi(x) = z(\zeta)\Lambda(v)Q(x;\theta)$ with

$$Q(x;\theta) = e^{r_\gamma(z)}\left[\delta \frac{e^{-e^{r_\gamma(z)}\Lambda(v)}}{1 - e^{-e^{r_\gamma(z)}\Lambda(v)}} - (1-\delta)\right].$$

The score operator for Λ along $\Lambda_t = \Lambda + th$ with $t \ge 0$ and h a non-decreasing non-negative right continuous function, is given by

$$\dot{l}_\Lambda(h)(x) = \frac{\partial}{\partial t} \log p(x; \gamma, \Lambda_t)\Big|_{t=0} = h(v)Q(x; \theta).$$

We project $\dot{l}_\xi(X)$ onto the space generated by $\dot{l}_\Lambda$. That is, we need to find a function $h^\star_\zeta(V) \in H_\Lambda$ such that $\dot{l}_\xi - \dot{l}_\Lambda(h^\star_\zeta) \perp \dot{l}_\Lambda(h)$, for all $h \in \mathcal{H}_\Lambda$, which is equivalent to solving the least square problem $P_\theta \|\dot{l}_\xi - \dot{l}_\Lambda h\|^2$. The solution under the null is $h^\star_\zeta(V) \equiv \Lambda_0(V)h^{\star\star}_\zeta(V)$, where
$h^{\star\star}_\zeta = P(Z(\zeta)Q^2(X;\psi))/P(Q^2(X;\theta))$, which is assumed to possess a version that is differentiable componentwise with the derivatives being bounded on $[\sigma, \tau]$ uniformly over $\zeta \in \Xi$. It can be shown that $\Lambda_t(\theta)$ is indeed a hazard function when t is sufficiently close to ξ.

The uniformly least-favorable direction for ξ is $\Lambda_t(\theta) = \Lambda + (\xi - t)'\varphi(\Lambda)h^{\star\star}_\zeta \circ \Lambda_0^{-1} \circ \Lambda$. Here φ is a function mapping $[0, M]$ into $[0, \infty)$ such that $\varphi(y) = y$ on $[\Lambda_0(\sigma), \Lambda_0(\tau)]$ and the function $y \mapsto \varphi(y)/y$ is Lipschitz and $\varphi(y) \leq c(\min(y, M - y))$ for a sufficiently large constant c. The efficient score for ξ for this uniformly least-favorable submodel is given by:

$$\tilde{l}_\xi(x; t, \theta) = \left[z - \frac{\varphi(\Lambda)(v)}{\Lambda_t(\theta)(v)} h^{\star\star}_\zeta \circ \Lambda_0^{-1} \circ \Lambda(v) \right] \Lambda_t(\theta)(v) Q(x; t, \Lambda_t(\theta)).$$

Λ_0^{-1} may be extended to $[0, \infty)$ by setting $\Lambda_0^{-1}(u) = \sigma$ for $u \leq \Lambda_0(\sigma)$ and $\Lambda_0^{-1}(u) = \tau$ for $u > \Lambda_0(\tau)$.

For step 2, we next project $\dot{l}_\beta(x)$ onto the space generated by $\tilde{l}_\xi$. The efficient score function for β, $\tilde{l}_\beta$, is the first two coordinates of $\tilde{l}_\xi$ minus its projection on the remaining coordinates of $\tilde{l}_\xi$. Since $\tilde{l}_\xi$ lies in a finite-dimensional space, the projection path has a matrix representation. The efficient information for ξ, $\tilde{I}_\xi$ can be partitioned as a two-by-two block matrix, with $\tilde{I}^{11}_\xi(\theta)$ denote its first two-by-two principle submatrix, and so on. We define $\nu'_\theta = (1, -(\tilde{I}^{22}_\xi)^{-1}\tilde{I}^{21}_\xi)$, and $\xi_t(\theta) = \xi - (\beta - t)\nu_\theta$. We also refine $\Lambda_t(\theta) = \Lambda + (\xi_t(\theta) - t)'\varphi(\Lambda)h^{\star\star}_\zeta \circ \Lambda_0^{-1} \circ \Lambda$.

Now we use the uniformly least-favorable path $t \mapsto (\xi_t(\theta), \Lambda_t(\theta))$ in the parameter space for the nuisance parameter $\eta \equiv (\alpha, \Lambda)$. This leads to $l(t, \beta, \alpha, \Lambda) = \log l(\xi_t(\theta), \Lambda_t(\theta))$. This submodel is least favorable at (ξ_0, Λ_0) uniformly over $\zeta \in \Xi$ since $\partial/\partial t\Big|_{t=\beta_0} l(t, \beta_0, \alpha, \Lambda) = \nu'_\theta \tilde{l}_\xi$, whereas $\nu'_\theta \tilde{l}_\xi = \tilde{l}_\beta$. The efficient information matrix for β, $\tilde{I}_\beta = \tilde{I}^{11}_\xi - \tilde{I}^{11}_\xi \left(\tilde{I}^{22}_\xi\right)^{-1} \tilde{I}^{21}_\xi(\theta)$. The remainder of assumption C4 can be verified by standard empirical process arguments.

To verify assumption A1 in section 3, it suffices to show that $\tilde{I}_\xi$ is uniformly positive definite over $\zeta \in \Xi$, which can be achieved by checking that

$\inf_{\zeta\in\Xi} \lambda_{\min}\{P_0(\mathrm{Cov}(Z(\zeta)|V))\} > 0$. We first show that the random vector $(Z, 1\{Z > \zeta\}, Z1\{Z > \zeta\})$ is linearly independent given V pointwisely in $\zeta \in \Xi$. Suppose that given V,

$$aZ + b1\{Z > \zeta\} + cZ1\{Z > \zeta\} = 0, \tag{7}$$

for some constants a, b and c. Our aim is to show $a = b = c = 0$. When $Z \leq \zeta$, (7) becomes $aZ1\{Z \leq \zeta\} = 0$. Since $\mathrm{Var}(Z|V) > 0$ and $P(Z \leq \zeta|V) > 0$, for every $\zeta \in \Xi$, $\mathrm{Var}(Z|Z \leq \zeta, V) > 0$, and therefore $a = 0$. When $Z > \zeta$, (7) becomes $(b + cZ)1\{Z > \zeta\} = 0$. If $c \neq 0$, $Z = -b/c$, which is contradicted with the fact that $\mathrm{Var}(Z|Z > \zeta, V) > 0$. Thus we conclude that $c = 0$ and $b = 0$ as a consequence. That $P(\mathrm{Cov}(Z(\zeta)|V))$ is uniformly positive definite over $\zeta \in \Xi$ follows since $P(\mathrm{Cov}(Z(\zeta)|V))$ is a continuous function of ζ and Ξ is compact.

The profile likelihood estimator $\hat{\psi}_n(\zeta)$ can be shown to be consistent for (β_0, Λ_0) by a similar proof as used for the full maximum likelihood estimator in Huang (1996). The following lemma shows the uniform consistency of $\hat{\psi}_n(\zeta)$ under the null.

LEMMA 6. $\hat{\psi}_n(\zeta) - \psi_0 = o^{\Xi}_{P_0}(1)$.

To verify the uniform no-bias condition (4), we need the following result about the uniform rate of convergence:

LEMMA 7. *Suppose that $d(\eta, \eta_1) : \eta, \eta_1 \in \mathcal{H}_\eta$ is the metric defined on $\mathcal{H}_\eta$, and C_1, C_2 and C_3 are positive constants,*

$$P_0(m_{\beta,\eta,\zeta} - m_{\beta,\eta_0,\zeta}) \leq -C_1 d^2(\eta, \eta_0) + C_2\|\beta - \beta_0\|^2, \tag{8}$$

$$P_0^\star \sup_{\beta\in B,\eta\in\mathcal{H}_\eta,\|\beta-\beta_0\|<\delta,d(\eta,\eta_0)<\delta,\zeta\in\Xi} |\mathbb{G}_n(m_{\beta,\eta,\zeta} - m_{\beta,\eta_0,\zeta})| \leq C_3\phi_n(\delta), \tag{9}$$

for functions ϕ_n such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ and sets $B\times\mathcal{H}_\eta\times\Xi$ such that under the null $Pr(\tilde{\beta}_n \in B, \hat{\eta}_{\tilde{\beta}_n}(\zeta) \in \mathcal{H}_\eta, \zeta \in \Xi) \to 1$. Then $\sup_{\zeta\in\Xi} r_n d(\hat{\eta}_{\tilde{\beta}_n}(\zeta), \eta_0) \leq O^\star_{P_0}(1 + r_n\|\tilde{\beta}_n - \beta_0\|)$ for any sequence of positive numbers r_n such that $r_n^2\phi_n(1/r_n) \leq \sqrt{n}$ for every n.

We apply lemma 7 with $\eta = (\alpha, \Lambda)$, $\mathcal{H}_\eta = \mathbb{R}\times\overline{\mathcal{H}}_\Lambda$, where $\overline{\mathcal{H}}_\Lambda$ is the closed linear span of $\mathcal{H}_\Lambda$, $d(\eta, \eta_1) = \|\alpha - \alpha_1\| + \|\Lambda - \Lambda_1\|_2$ and

$$m_{\beta,\eta,\zeta} = \begin{cases} \log \frac{p_{\beta,\eta,\zeta}}{p_{\beta_0,\eta_0}}, & \text{if } \eta = \eta_0, \\ 2\log \frac{p_{\beta,\eta,\zeta} + p_{\beta_0,\eta_0}}{p_{\beta_0,\eta_0}}, & \text{otherwise.} \end{cases}$$

Condition (8) can be established by the Taylor expansion and the uniform boundedness on the derivatives of the loglikelihood. Condition (9) can be verified using lemma 3.3 of Murphy and van der Vaart (1999), with the choice $\phi_n(\delta) = \delta^{1/2}\left(1 + M\delta^{-3/2}/\sqrt{n}\right)$, where $M \geq \|m_{\beta,\eta,\zeta}\|_\infty$ is a constant. These conditions imply that $\|\hat{\alpha}_{\tilde{\beta}_n}(\zeta) - \alpha_0\| + \|\hat{\Lambda}_{\tilde{\beta}_n}(\zeta) - \Lambda_0\|_2 = O_p^{\Xi}(\|\tilde{\beta}_n - \beta_0\| + n^{-1/3})$, for any sequence $\tilde{\beta}_n \to 0$. Now we only need to verify

$$P_0\dot{\ell}(\beta_0, \beta_0, \hat{\overline{\eta}}_{\tilde{\beta}_n}(\zeta), \zeta) = o_{P_0}^{\Xi}(\|\tilde{\beta}_n - \beta_0\| + n^{-1/2}), \tag{10}$$

which is equivalent to (4) under regularity conditions. We further decompose (10) as (17) in Murphy and van der Vaart (2000), which could be easily verified by the Taylor expansion and the uniform boundedness on the first and second derivatives of the loglikelihood.

It is not difficult to see that $\{p_{\xi,\Lambda}(\zeta)\}$ is differentiable in quadratic mean at (ψ_0, ζ) with respect to the set of directions $\{\xi_0 + th_1, \Lambda_0 + th_2\}$, where $h_1 \in \mathbb{R}^3$, and h_2 is a non-decreasing non-negative right continuous function.

Remark 3. *In the Cox model with current status data, the cumulative hazard function Λ is estimable at the $n^{1/3}$ rate. In the Cox model with threshold covariates for right censored data, the MLE of the change-point parameter converges with rate n. It can be shown that the MLE of the parameters in the change-point Cox model with current status data have three different convergence rates. While this interesting phenomenon merits future research, we note that it does not restrict the optimality results here, since estimation of ζ is unnecessary under the null. The key condition for the optimality results to hold is that the convergence rate of the baseline cumulative hazard function as a process in ζ under the null is sufficiently fast to permit the expansion of the profile likelihood, uniformly in ζ.*

6. Proofs.

Proof of lemma 1. Since $\dot{\phi}_\zeta$ is linear, continuously invertible and one-to-one, the tangent set for η and $\overline{\eta}$ are identical. By the chain rule, $\dot{\ell}_{\overline{\eta}}(\gamma) = \dot{l}_\eta \dot{\phi}_\zeta^{-1}(\gamma)$ for any γ in the tangent set of $\overline{\eta}$. The efficient score for β with the parameter (β, η, ζ) is: $\tilde{l}_\beta(\beta, \eta, \zeta) = (I - \dot{l}_\eta(\dot{l}_\eta^\star l_\eta)^{-1}\dot{l}_\eta^\star)\dot{l}_\beta(\psi, \zeta)$ and with the parameter $(\beta, \overline{\eta}, \zeta)$ is: $(I - \dot{\ell}_{\overline{\eta}}(\dot{\ell}_{\overline{\eta}}^\star \ell_{\overline{\eta}})^{-1}\dot{\ell}_{\overline{\eta}}^\star)\dot{\ell}_\beta(\psi, \zeta)$. The efficient score function is invariant under such reparameterizations since

$$\begin{aligned} I - \dot{\ell}_{\overline{\eta}}(\ell_{\overline{\eta}}^\star \ell_{\overline{\eta}})^{-1}\dot{\ell}_{\overline{\eta}}^\star(\overline{\psi}, \zeta) &= I - \dot{l}_\eta \dot{\phi}_\zeta^{-1}(\dot{\phi}_\zeta^{-1\star}\dot{l}_\eta^\star \dot{l}_\eta \dot{\phi}_\zeta^{-1})^{-1}\dot{\phi}_\zeta^\star \dot{l}_\eta^\star(\psi, \zeta) \\ &= I - \dot{l}_\eta \dot{\phi}_\zeta^{-1}\dot{\phi}_\zeta(\dot{l}_\eta^\star \dot{l}_\eta)^{-1}\dot{\phi}_\zeta^\star(\dot{\phi}_\zeta^\star)^{-1}\dot{l}_\eta^\star(\psi, \zeta) = I - \dot{l}_\eta(\dot{l}_\eta^\star l_\eta)^{-1}\dot{l}_\eta^\star(\psi, \zeta), \end{aligned}$$

and $\dot{\ell}_\beta(\overline{\psi}, \zeta) = \dot{l}_\beta(\psi, \zeta)$. That the efficient information matrix is invariant under reparameterizations thus follows from its definition. □

PROOF OF LEMMA 2. It suffices to show that under the full rank reparameterization, for any random sequence $\tilde{\beta}_n \rightarrow_{P_0} \beta_0$,

$$\log p\ell_n(\tilde{\beta}_n,\zeta) = \log p\ell_n(\beta_0,\zeta) + n(\tilde{\beta}_n-\beta_0)'\mathbb{P}_n\tilde{\ell}_\beta(\overline{\psi}_0,\zeta) \\ -\frac{1}{2}n(\tilde{\beta}_n-\beta_0)'\tilde{\mathcal{I}}_\beta(\overline{\psi}_0,\zeta)(\hat{\beta}_n-\beta_0) + o^{\Xi}_{P_0}(\sqrt{n}\|\tilde{\beta}_n-\beta_0\|+1)^2. \tag{11}$$

By assumption C2, C4 and the dominated convergence theorem, for every $(\tilde{t},\tilde{\beta},\tilde{\overline{\eta}}) - (\beta_0,\beta_0,\overline{\eta}_0) \rightarrow 0$, we have $P_0(\dot{\ell}(\tilde{t},\tilde{\beta},\tilde{\overline{\eta}},\zeta) - \tilde{\ell}_\beta(\overline{\psi}_0,\zeta))^2 = o^{\Xi}(1)$. Similarly, we have $P_0\ddot{\ell}(\tilde{t},\tilde{\beta},\tilde{\overline{\eta}},\zeta) - P_0\ddot{\ell}(\beta_0,\beta_0,\overline{\eta}_0,\zeta) = o^{\Xi}(1)$. The derivative of the function $t \mapsto \log \ell(t,\overline{\psi}_0,\zeta)$ satisfies $P_0\ddot{\ell}(\beta_0,\overline{\psi}_0,\zeta) = -\tilde{\mathcal{I}}_\beta(\overline{\psi}_0,\zeta)$. These facts, together with the empirical process conditions, imply that for every random sequence $(\tilde{t},\tilde{\beta},\tilde{\overline{\eta}}) \rightarrow (\beta_0,\beta_0,\overline{\eta}_0)$, $\mathbb{G}_n\dot{\ell}(\tilde{t},\tilde{\beta},\tilde{\overline{\eta}}) - \mathbb{G}_n\tilde{\ell}_\beta(\overline{\psi}_0,\zeta) = o^{\Xi}_{P_0}(1)$ and $\mathbb{P}_n\ddot{\ell}(\tilde{t},\tilde{\beta},\tilde{\overline{\eta}},\zeta)+\tilde{\mathcal{I}}_\beta(\overline{\psi}_0,\zeta) = o^{\Xi}_{P_0}(1)$. The subsequent steps of the proof are similar to those used in the proof of theorem 1 in Murphy and van der Vaart (2000). $\square$

PROOF OF THEOREM 1. The proof takes several steps. We first show the asymptotic equivalence of these statistics, which is summarized in lemma 8 below. With a small abuse of notation, let $PLR_n \equiv \int pl_n(\beta + h/\sqrt{n},\zeta)dQ^c_\zeta(h)dJ(\zeta)/pl_n(\beta_0,\zeta)$. It is the profile likelihood ratio of the alternative over the null and it can be approximated by

$$\overline{PLR}_n \equiv \int \exp\left\{\frac{1}{2}\overline{\beta}_n(\theta_0(\zeta))'\tilde{I}_\beta(\theta_0(\zeta))\overline{\beta}_n(\theta_0(\zeta))\right\} \\ \times \int \exp\left\{-\frac{1}{2}(\overline{\beta}_n(\theta_0(\zeta))-h)'\tilde{I}_\beta(\theta_0(\zeta))(\overline{\beta}_n(\theta_0(\zeta))-h)\right\} dQ^c_\zeta(h)dJ(\zeta),$$

with the linear statistic $\overline{\beta}_n(\theta_0(\zeta)) \equiv \sqrt{n}\tilde{I}^{-1}_\beta(\theta_0(\zeta))\mathbb{P}_n\tilde{l}_\beta(\theta_0(\zeta))$. An approximate exponential Wald statistic $\overline{EW}_n$ is defined as

$$\overline{EW}_n = (1+c)^{-\frac{p}{2}} \int \exp\left(\frac{1}{2}\frac{c}{1+c}\overline{W}_n(\zeta)\right) dJ(\zeta),$$

where $\overline{W}_n(\zeta) = \overline{\beta}_n(\theta_0(\zeta))'\tilde{I}_\beta(\theta_0(\zeta))\overline{\beta}_n(\theta_0(\zeta))$.

Now we show the asymptotic distribution of these tests under the null hypothesis. Assume without loss of generality that $\hat{\beta}_n$ and $\hat{\overline{\psi}}_n$ take their values in U and V as defined in assumption C4, respectively. Following lemma 3.2 in Murphy and van der Vaart (1997), we have $\mathbb{G}_n(\dot{\ell}(\hat{\beta}_n,\hat{\overline{\psi}}_n,\zeta) - \tilde{\ell}_\beta(\overline{\psi}_0,\zeta)) \rightarrow_{P_0} 0$. Thus $\tilde{\ell}_\beta(\psi_0,\zeta) = \tilde{l}_\beta(\theta_0(\zeta))$ is P_0-Donsker as a class indexed by $\zeta \in \Xi$ and $\overline{EW}_n \rightarrow_d e\chi(c)$ by the continuous mapping theorem. Lemma 8 below then gives the desired results of theorem 1. $\square$

LEMMA 8. *Under the null hypothesis and assumptions A–D, (1)* $PLR_n - \overline{PLR}_n \to_{P_0} 0$*, (2)* $\overline{PLR}_n = \overline{EW}_n$*, (3)* $\overline{EW}_n - EW_n \to_{P_0} 0$*, (4)* $\overline{EW}_n - ER_n \to_{P_0} 0$ *and (5)* $ER_n - ELR_n \to_{P_0} 0$.

PROOF OF LEMMA 8. For notational simplicity, let $\overline{\beta}_n = \overline{\beta}_n(\theta_0(\zeta))$ and $\tilde{I}_0 = \tilde{I}_\beta(\theta_0(\zeta))$.

We first show (1). For $0 < M < \infty$, define

$$PLR_n(M) = \int_{\zeta\in\Xi}\int_{\|h\|\le M} pl_n\left(\beta_0 + h/\sqrt{n}, \zeta\right) dQ^c_\zeta(h) dJ(\zeta)/pl_n(\beta_0,\zeta),$$

and

$$\begin{aligned}\overline{PLR}_n(M) &= \int_{\zeta\in\Xi}\exp\left(\frac{1}{2}\overline{\beta}'_n\tilde{I}_\beta\overline{\beta}_n\right)\\ &\quad\times\int_{\|h\|\le M}\exp\left(-\frac{1}{2}(\overline{\beta}_n - h)'\tilde{I}_0(\overline{\beta}_n - h)\right) dQ^c_\zeta(h)dJ(\zeta).\end{aligned}$$

Note that for any $M > 0$,

$$\begin{aligned}|PLR_n - \overline{PLR}_n| &\le |PLR_n - PLR_n(M)| + |PLR_n(M) - \overline{PLR}_n(M)|\\ &\quad+ |\overline{PLR}_n - \overline{PLR}_n(M)|.\end{aligned}$$

Hence it suffices to show that (i) $|PLR_n - PLR_n(M)| \to_{P_0} 0$, (ii) $|\overline{PLR}_n - \overline{PLR}_n(M)| \to_{P_0} 0$ and (iii) $|PLR_n(M) - \overline{PLR}_n(M)| \to_{P_0} 0$, as $n \to \infty$ and $\forall M:\ 0 < M < \infty$. To show (i), for any $\varepsilon > 0$,

$$Pr(|PLR_n - PLR_n(M)| > \varepsilon) \le \varepsilon^{-1}P_0|PLR_n - PLR_n(M)|$$

$$= \varepsilon^{-1}P\int_{\zeta\in\Xi}\int_{\|h\|>M}\frac{pl_n(\beta_0 + h/\sqrt{n},\zeta)}{pl_n(\beta_0,\zeta)}dQ^c_\zeta(h)dJ(\zeta) \tag{12}$$

$$\le \varepsilon^{-1}P\int_{\zeta\in\Xi}\int_{\|h\|>M}\frac{pl_n(\hat{\beta}_n(\zeta),\zeta)}{pl_n(\beta_0,\zeta)}dQ^c_\zeta(h)dJ(\zeta) \tag{13}$$

$$\to \varepsilon^{-1}P\int_{\zeta\in\Xi}\int_{\|h\|>M}(1 + o_p(1))dQ^c_\zeta(h)dJ(\zeta) \tag{14}$$

$$= \varepsilon^{-1}\int_{\zeta\in\Xi}\int_{\|h\|>M}dQ^c_\zeta(h)dJ(\zeta) + o(1), \tag{15}$$

where (12) uses assumption C and (13) holds by definition of the profile likelihood. (14) holds by assumption C3 and lemma 2. (15) holds by Fubini's theorem. The right hand side of (15) can be made arbitrarily small for all n by taking M large enough, since Q^c_ζ is a uniformly tight measure.

For (ii), we have

$$
\begin{aligned}
(16)\quad & \left|\overline{PLR}_n - \overline{PLR}_n(M)\right| \\
&= \int_{\zeta\in\Xi} \exp\left(\frac{1}{2}\mathbb{P}_n\tilde{l}_\beta(\theta_0(\zeta))'\tilde{I}_0^{-1}\mathbb{P}_n\tilde{l}_\beta(\theta_0(\zeta))\right) \\
&\times \int_{\|h\|>M} \exp\left(-\frac{1}{2}(\overline{\beta}_n - h)'\tilde{I}_0(\overline{\beta}_n - h)\right) dQ_\zeta^c(h)dJ(\zeta) \\
&\le \exp\left(\frac{1}{2}\sup_{\zeta\in\Xi}\|\mathbb{P}_n\tilde{l}_\beta(\theta_0(\zeta))\|^2 \sup_{\zeta\in\Xi}\|\tilde{I}_0^{-1}\|\right) \times \int_{\zeta\in\Xi}\int_{\|h\|>M} dQ_\zeta^c(h)dJ(\zeta).
\end{aligned}
$$

In the inequality, $\|\mathbb{P}_n\tilde{l}_\beta(\theta_0(\zeta))\|^2 = O_{P_0}^{\Xi}(1)$ follows from assumption C4. The fact that $\|\tilde{I}_0^{-1}\| = O_{P_0}^{\Xi}(1)$ follows from assumption A. The last term $\int_{\zeta\in\Xi}\int_{\|h\|>M} dQ_\zeta^c(h)dJ(\zeta) \to 0$, as $M\to\infty$. Hence (16)$=o_p(1)$, as $M\to\infty$.

Now we show (iii). For contiguous sequences $\beta_0 + h/\sqrt{n} \to_{P_0} \beta_0$ and $\|h\| \le M$, lemma 2 yields the following expansion of the profile likelihood under the null:

$$
\begin{aligned}
&\log pl_n(\beta_0 + h/\sqrt{n}, \zeta) = \log pl_n(\beta_0,\zeta) + \sqrt{n}h'\mathbb{P}_n\tilde{l}_\beta(\theta_0(\zeta)) - \frac{1}{2}h'\tilde{I}_0 h + o_{P_0}^{\Xi}(1) \\
&= \frac{1}{2}\overline{\beta}_n'\tilde{I}_0\overline{\beta}_n - \frac{1}{2}(\overline{\beta}_n - h)'\tilde{I}_0(\overline{\beta}_n - h) + o_{P_0}^{\Xi}(1),
\end{aligned}
$$

therefore,

$$
\begin{aligned}
&PLR_n(M) = \iint_{\|h\|\le M} \left(pl_n(\beta_0 + h/\sqrt{n},\zeta) - pl_n(\beta_0,\zeta)\right) dQ_\zeta^c(h)dJ(\zeta) \\
&= \iint_{\|h\|\le M} \exp\left(\frac{1}{2}\overline{\beta}_n'\tilde{I}_0\overline{\beta}_n - \frac{1}{2}(\overline{\beta}_n - h)'\tilde{I}_0(\overline{\beta}_n - h) + o_p^{\Xi}(1)\right) dQ_\zeta^c(h)dJ(\zeta) \\
&= \overline{PLR}_n(M) + o_p(1),
\end{aligned}
$$

where the last equality follows from $\overline{PLR}_n(M) = O_p(1)$, by arguments analogous to those used in (16) above. The proof for part (1) is now completed.

For part (2), since $h \sim Q_\zeta^c = N(0, c\tilde{I}_0^{-1})$,

$$
\overline{PLR}_n = \int_{\zeta\in\Xi} \xi_n(\zeta)dJ(\zeta),
$$

with

$$\begin{aligned}
\xi_n(\zeta) &= \int \exp\left(\frac{1}{2}\overline{\beta}_n'\tilde{I}_0\overline{\beta}_n - \frac{1}{2}(\overline{\beta}_n - h)'\tilde{I}_0(\overline{\beta}_n - h)\right) dQ_\zeta^c(h) \\
&= (2\pi)^{-p/2}\det{}^{1/2}(\tilde{I}_0/c) \\
&\quad \times \int \exp\left[\frac{1}{2}\left\{\overline{\beta}_n'\tilde{I}_0\overline{\beta}_n - (h-\overline{\beta}_n)'\tilde{I}_0(h-\overline{\beta}_n) - \frac{h'\tilde{I}_0 h}{c}\right\}\right] dh \\
&= (2\pi)^{-p/2}\det{}^{1/2}(\tilde{I}_0/c) \\
&\quad \times \int \exp\left[\frac{1}{2}\left\{\frac{c}{1+c}\overline{\beta}_n'\tilde{I}_0\overline{\beta}_n - (h - \frac{c}{1+c}\overline{\beta}_n)'\frac{1+c}{c}\tilde{I}_0(h - \frac{c}{1+c}\overline{\beta}_n)\right\}\right] dh \\
&= (1+c)^{-p/2}\exp\left(\frac{1}{2}\frac{c}{1+c}\overline{\beta}_n'\tilde{I}_0\overline{\beta}_n\right),
\end{aligned}$$

where the last equality holds by integrating out a normal density.

For part (3), it follows from lemma 2 and assumption C3 that $\sqrt{n}\|\hat{\beta}_n(\zeta) - \beta_0\| = O_{P_0}^{\Xi}(1)$, and reapplication of lemma 2 and the argmax theorem yields $\sqrt{n}(\hat{\beta}_n(\zeta) - \beta_0) = \tilde{I}_\beta(\theta_0(\zeta))^{-1}\sqrt{n}\mathbb{P}_n\tilde{I}_\beta(\theta_0(\zeta)) + o_{P_0}^{\Xi}(1)$. Part (3) now follows.

For the proof of part (4) and part (5), it suffices to show that $W_n(\zeta) - R_n(\zeta) = o_{P_0}^{\Xi}(1)$ and $R_n(\zeta) - LR_n(\zeta) = o_{P_0}^{\Xi}(1)$. These results follow from Donsker properties and standard arguments. We omit the details. The proof of lemma 8 is thus completed. □

Proof of corollary 1. The proof is similar to the proof of theorem 1. We omit the details. □

Proof of corollary 2. The proof follows the same lines as the proof of part (2)(iii) of lemma 8, with

(17)

$$W(q_\zeta, \zeta) = (2\pi)^{-p/2}\det{}^{1/2}\left(\frac{1+c}{c}\tilde{I}_0\right) \times$$
$$\int \exp\left[-\frac{1+c}{2c}(\lambda - \frac{c}{1+c}\overline{\beta}_n)'\tilde{I}_0(\lambda - \frac{c}{1+c}\overline{\beta}_n) - \lambda'\langle q_\zeta - \tilde{q}_\zeta, P\dot{l}_\eta^\star\dot{l}_\eta(q_\zeta - \tilde{q}_\zeta)'\rangle_\eta\lambda\right] d\lambda,$$

where $\langle\cdot,\cdot\rangle_\eta$ is the inner product defined on $\overline{\mathcal{H}}_\eta$, and $W(q_\zeta, \zeta) \le 1$ since $\langle q_\zeta - \tilde{q}_\zeta, P\dot{l}_\eta^\star\dot{l}_\eta(q_\zeta - \tilde{q}_\zeta)'\rangle_\eta$ is nonnegative definite. □

Before giving the proof of theorem 2, we need the following lemma:

Lemma 9. *Under assumptions A–D, for any $0 < c < \infty$ and any $\zeta \in \Xi$, the densities $\ell_n(\overline{\psi}_0 + h/\sqrt{n}, \zeta)$ and $\int \ell_n(\overline{\psi}_0 + h/\sqrt{n}, \zeta)dQ_\zeta^c(h)dJ(\zeta)$ are contiguous to the densities l_n^0. As a consequence, the results of lemma 8 still hold under local alternatives $\{P_{\overline{\psi}_0 + h/\sqrt{n}, \zeta}\}$ and $\{\int P_{\overline{\psi}_0 + h/\sqrt{n}, \zeta}dQ_\zeta^c(h)dJ(\zeta)\}$.*

PROOF OF LEMMA 9. Assumption D implies that a LAN (local asymptotic normal) expansion for the log-likelihood ratio holds immediately by lemma 3.10.11 of van der Vaart and Wellner (1996):

$$\Lambda_{n\zeta} \equiv \log\left(\frac{dP^n_{\psi_0+h/\sqrt{n},\zeta}}{dP^n_{\psi_0,\zeta}}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_\zeta h(X_i) - \frac{1}{2}\|A_\zeta h\|^2 + o_{P_0}(1).$$

It follows from LAN that $\Lambda_{n\zeta} \to_d W_\zeta$, where $W_\zeta \sim N(-1/2\|A_\zeta h\|^2, \|A_\zeta h\|^2)$, under P_0. Therefore, under P_0,

$$\exp(\Lambda_{n\zeta}) \equiv \frac{dP^n_{\overline{\psi}_0+h/\sqrt{n},\zeta}}{dP^n_0} \to_d \exp W_\zeta.$$

$P_0(\exp(W_\zeta)) = 1$, using the formula for the moment generating function of the normal distribution. By Le Cam's first lemma (van der Vaart, 1996, page 88), we conclude that the sequences of probability measures $\{P_{\overline{\psi}_0+h/\sqrt{n},\zeta}\}$ and $\{P_0\}$ are contiguous, for every $\zeta \in \Xi$. Consequently the convergence in probability that hold under P_0 also hold under $\{P_{\overline{\psi}_0+h/\sqrt{n},\zeta}\}$ and vice-versa. Similarly, since $P(e\chi) = 1$ using the formula for the moment generating function of the χ^2 distribution, we conclude that the sequences $\{\int P^n_{\overline{\psi}_0+h/\sqrt{n},\zeta} dQ^c_\zeta(h) dJ(\zeta)\}$ and P^n_0 are contiguous. □

PROOF OF THEOREM 2. We define a $\sqrt{n}$-neighborhood of β_0 as a collection of sequences $\beta_n(h_\beta) = \beta_0 + h_\beta/\sqrt{n} + o(n^{-1/2})$, for $h_\beta \in \mathbb{R}^p$. A $\sqrt{n}$ neighborhood of η is similarly defined as $\eta_n(h_\eta) = \eta + h_\eta/\sqrt{n} + o(n^{-1/2})$, for $h_\eta \in \mathcal{H}_\eta$. With a small abuse of notation, a local form of the hypotheses can be written as:

$$H_0: \ \overline{\psi} = \overline{\psi}_0 \ \text{ vs. } \ H_1: \ \overline{\psi} = \overline{\psi}_0 + h_1/\sqrt{n}, \tag{18}$$

where $h_1 \in \mathbb{R}^p \times \mathcal{H}_\eta$ which takes the value $(h_{\beta 1}, h_{\eta 1})$, with $h_{\eta 1} = \tilde{q}'_\zeta h_{\beta 1}$. We note that the least favorable direction $\tilde{q}_\zeta$ is invariant under the choice of ϕ_ζ in assumption B, and thus the contiguous alternative H_1 is also invariant under the choice of ϕ_ζ.

Define

$$LR_n \equiv \frac{\int \ell_n\left(\overline{\psi}_0 + h_1/\sqrt{n}, \zeta\right) dQ^c_\zeta(h_{\beta 1}) dJ(\zeta)}{\ell^0_n}. \tag{19}$$

A test defined by LR_n is

$$\tilde{\gamma}_n = \begin{cases} 1, \text{ if } LR_n > \tilde{k}_{\alpha n}, \\ \tilde{\lambda}_n, \text{ if } LR_n = \tilde{k}_{\alpha n}, \\ 0, \text{ if } LR_n < \tilde{k}_{\alpha n}, \end{cases}$$

where $\tilde{k}_{\alpha n} > 0$, $\tilde{\lambda}_n \in [0,1]$ are constants such that the rejection probability is α under the null. For notational simplicity, let $P_1^n = \int P^n_{\overline{\psi}_0+h_1/\sqrt{n},\zeta} dQ^c_\zeta(h_{\beta 1})dJ(\zeta)$. By the Neyman-Pearson lemma, for all $n \geq 1$ and any test ϕ_n with level α, with a small abuse of notation,

$$\begin{aligned}
(20)\quad & \lim_{n\to\infty} \int \phi_n \left\{ \int \ell_n(\overline{\psi}_0 + h_1/\sqrt{n}, \zeta) dQ^c_\zeta(h_{\beta 1}) dJ(\zeta) \right\} dP_1^n \\
& \leq \lim_{n\to\infty} \int \tilde{\gamma}_n \left\{ \int \ell_n(\overline{\psi}_0 + h_1/\sqrt{n}, \zeta) dQ^c_\zeta(h_{\beta 1}) dJ(\zeta) \right\} dP_1^n \\
(21)\quad & = \lim_{n\to\infty} \int I(LR_n > \tilde{k}_{\alpha n}) \left\{ \int \ell_n(\overline{\psi}_0 + h_1/\sqrt{n}, \zeta) dQ^c_\zeta(h_{\beta 1}) dJ(\zeta) \right\} dP_1^n \\
(22)\quad & = \lim_{n\to\infty} \int \left\{ \int I(LR_n > \tilde{k}_{\alpha n}) dP^n_{\overline{\psi}_0+h_1/\sqrt{n},\zeta} \right\} dQ^c_\zeta(h_{\beta 1}) dJ(\zeta) \\
(23)\quad & = \lim_{n\to\infty} \int \left\{ \int I(PLR_n > \tilde{k}_{\alpha n}) dP^n_{\overline{\psi}_0+h_1/\sqrt{n},\zeta} \right\} dQ^c_\zeta(h_{\beta 1}) dJ(\zeta) \\
(24)\quad & = \lim_{n\to\infty} \int \left\{ \int I(EW_n > \tilde{k}_{\alpha n}) dP^n_{\overline{\psi}_0+h_1/\sqrt{n},\zeta} \right\} dQ^c_\zeta(h_{\beta 1}) dJ(\zeta),
\end{aligned}$$

where (21) follows since LR_n has an absolutely continuous asymptotic distribution under the contiguous alternative H_1 and (22) follows by Fubini's theorem. (23) follows since $PLR_n - LR_n = o_P(1)$ under H_1, which can be established at the end of the proof. (24) follows from lemma 9. The results for ER_n and ELR_n also follow from lemma 9. By Fubini's theorem, we obtain $\limsup_{n\to\infty} \int \left\{ \phi_n(P^n_{\overline{\psi}_0+h_1/\sqrt{n},\zeta}) \right\} dQ^c_\zeta(h_{\beta 1}) dJ(\zeta) \leq$
$\lim_{n\to\infty} \int \left\{ \int I(EW_n > \tilde{k}_{\alpha_n}) dP^n_{\overline{\psi}_0+h_1/\sqrt{n},\zeta} \right\} dQ^c_\zeta(h_{\beta 1}) dJ(\zeta)$, which implies that the proposed tests have the greatest weighted average power asymptotically in the class of all tests of asymptotic significance level α, against the alternative $P^n_{\overline{\psi}_0+h/\sqrt{n},\zeta}$.

To show $PLR_n - LR_n = o_P(1)$ under H_1, it suffices to show $PLR_n - LR_n = o_P(1)$ under the null by lemma 9. Define $LR_n(M) \equiv$
$\int_{\zeta\in\Xi} \int_{\|h\|\leq M} \ell_n\left(\overline{\psi}_0 + h_1/\sqrt{n}, \zeta\right) dQ^c_\zeta(h) dJ(\zeta)/\ell_n(\overline{\psi}_0,\zeta)$, note that $\forall M: \ 0 < M < \infty$, $|PLR_n - LR_n| \leq |PLR_n - PLR_n(M)| + |PLR_n(M) - LR_n(M)| + |LR_n - LR_n(M)|$. Hence it suffices to show that (i) $|PLR_n - PLR_n(M)| \to_{P_0} 0$, (ii) $|\overline{LR}_n - \overline{LR}_n(M)| \to_{P_0} 0$ and (iii) $|PLR_n(M) - \overline{LR}_n(M)| \to_{P_0} 0$, as $n \to \infty$. Part (i) was shown in lemma (8). Part (ii) can be similarly established by taking M large enough and applying assumption A.

To show part (iii), we take a Taylor expansion of $\log \ell_n\left(\overline{\psi}_0 + h_1/\sqrt{n}, \zeta\right)$ at $(\overline{\psi}_0, \zeta)$ with respect to h_β along the direction $\tilde{q}_\zeta$, which leads to the following expansion in the least favorable submodel:

$$\log \ell_n(\overline{\psi}_0 + \frac{h_1}{\sqrt{n}}, \zeta) = \log \ell_n(\overline{\psi}_0, \zeta) + \sqrt{n} h'_{\beta 1} \mathbb{P}_n \dot{\ell}(\beta_0, \overline{\psi}_0, \zeta) + \frac{1}{2} h'_{\beta 1} \mathbb{P}_n \ddot{\ell}(\tilde{\beta}, \tilde{\psi}, \zeta) h_{\beta 1}.$$

On the right-hand side, we can replace $\mathbb{P}_n \dot{\ell}(\beta_0, \overline{\psi}_0, \zeta)$ by $\mathbb{P}_n \dot{\ell}_\beta(\overline{\psi}_0, \zeta) + o^{\Xi}_{P_0}(1)$, and $\mathbb{P}_n \ddot{\ell}(\tilde{\beta}, \tilde{\psi}, \zeta)$ by $-\tilde{I}_\beta(\overline{\psi}_0, \zeta) + o^{\Xi}_{P_0}(1)$, according assumption C2. Comparing the above display and lemma 2 with $\tilde{\beta}_n \equiv h_{\beta 1}/\sqrt{n}$, we obtain part (iii). □

Proof of theorem 3. The equivalence of the three tests under local alternatives is shown in lemma 9. To show their asymptotic distribution, a key step is to establish that $\overline{\beta}_n$ converges under $P^n_{\overline{\psi}_0 + h/\sqrt{n}, \zeta_1}$ in distribution to the process $\zeta \mapsto \mathbb{G}(\theta_0(\zeta)) + \nu_\star(h_\beta, \zeta, \zeta_1)$, where $\nu_\star(h_\beta, \zeta, \zeta_1) \equiv P_0 \tilde{l}_\beta(\theta_0(\zeta)) \tilde{l}_\beta(\theta_0(\zeta_1))' h_\beta$, by theorem 3.10.12 in van der Vaart and Wellner (1996). The result follows by lemma 9 and the continuous mapping theorem. □

Proof of theorem 4. The equivalence of the three tests under local alternatives is shown in lemma 9. Since the sequences of densities $\int \ell_n(\overline{\psi}_0 + h/\sqrt{n}, \zeta) dQ^{c_1}_\zeta(h) dJ(\zeta)$ are contiguous to the density l^0_n, we have

$$\left(ELR_n, \frac{\int dP^n_{\overline{\psi}_0 + h/\sqrt{n}, \zeta} dQ^{c_1}_\zeta(h) dJ(\zeta)}{dP^n_0} \right) \rightsquigarrow_d \left(e\chi(c), e\chi(c_1) \right),$$

under P_0. Then $ELR_n \rightarrow_d r\chi(c, c_1)$ under $\int dP^n_{\overline{\psi}_0 + h/\sqrt{n}, \zeta} dQ^{c_1}_\zeta(h) dJ(\zeta)$, by Le Cam's third lemma. □

Proof of lemma 3. For part (1), it suffices to note that $\|\dot{\psi}_\zeta(\phi_\zeta(\theta))(\dot{\phi}_\zeta(\theta)(h))\| \geq c_1 \|\dot{\phi}_\zeta(\theta)(h)\| \geq c_1 c_2 \|h\|$. For part (2), we note that, $\psi_\zeta \circ \phi_\zeta(\theta + th) - \psi_\zeta \circ \phi_\zeta(\theta) = \psi_\zeta(\phi_\zeta(\theta) + tk_t) - \psi_\zeta(\phi_\zeta(\theta))$, where $k_t = \{\phi_\zeta(\theta + th) - \phi_\zeta(\theta)\}/t$. So we rewrite the uniform Fréchet difference as $\psi_\zeta(\phi_\zeta(\theta + h))(\cdot) - \psi_\zeta(\phi_\zeta(\theta))(\cdot) = \dot{\psi}_\zeta(\phi_\zeta(\theta))(\phi_\zeta(\theta + h) - \phi_\zeta(\theta)) + o^{\Xi}(\|\phi_\zeta(\theta + h) - \phi_\zeta(\theta)\|) = \dot{\psi}_\zeta(\phi_\zeta(\theta))\dot{\psi}_\zeta(\theta)(h) + o^{\Xi}(\|h\|)$. □

Proof of lemma 4. Since for an arbitrary random sequence ζ_n, $T^{-1}_{\zeta_n}$ is continuous, the operator $T^{-1}_{\zeta_n} K : \mathbb{E} \mapsto \mathbb{D}$ is compact. Hence $I + T^{-1}_{\zeta_n} K_{\zeta_n}$ is one-to-one and therefore also onto be a result of Riesz for compact operators. Thus $T_{\zeta_n} + K_{\zeta_n}$ is also onto. We will be done if we can show that $I + T^{-1}_{\zeta_n} K_{\zeta_n}$ is continuously invertible, since that would imply that $(T_{\zeta_n} + K_{\zeta_n})^{-1} = (I + T^{-1}_{\zeta_n} K_{\zeta_n})^{-1} T^{-1}_{\zeta_n}$ is bounded. The remainder of the proof follows the proof of lemma 6.17 in Kosorok (To appear). □

Proof of lemma 5. This is a "uniform" version of lemma 3.3.5 in van der Vaart and Wellner (1996). Let $\Psi_\delta \equiv \{\psi : \|\psi - \psi_0\| < \delta\}$ and define an extraction function $f : \ell^\infty(\Psi_\delta \times \Xi \times \mathcal{H}_\eta) \times \Psi_\delta \mapsto \ell^\infty(\mathcal{H}_\eta \times \Xi)$ as $f(z, \psi, \zeta)(h) \equiv z(\psi, \zeta, h)$, where $z \in \ell^\infty(\Psi_\delta \times \mathcal{H}_\eta \times \Xi)$. Since f is continuous at every point (z, ψ_1, ζ), we have $\sup_{h \in \mathcal{H}_\eta, \zeta \in \Xi} |z(\psi, \zeta, h) - z(\psi_1, \zeta, h)| \to 0$ as $\psi \to \psi_1$. Define the stochastic process $Z_n(\psi, \zeta, h) \equiv \mathbb{G}_n(\nu(\psi, \zeta)(h) - \nu(\psi_0, \zeta)(h))$ indexed by $\Psi_\delta \times \Xi \times \mathcal{H}_\eta$. By assumption, Z_n converges weakly in $\ell^\infty(\Psi_\delta \times \Xi \times \mathcal{H}_\eta)$ to a tight Gaussian process Z_0 with continuous sample paths with respect to the metric ρ_ζ defined by $\rho_\zeta^2((\psi_1, \zeta, h_1), (\psi_2, \zeta, h_2)) = P(\nu(\psi_1, \zeta)(h_1) - \nu(\psi_0, \zeta)(h_1) - \nu(\psi_2, \zeta)(h_2) + \nu(\theta_0, \zeta)(h_2))^2$, at fixed ζ. Since as assumed, $\sup_{h \in \mathcal{H}_\eta, \zeta \in \Xi} \rho_\zeta((\psi, h), (\psi_0, h)) \to 0$, we have that f is continuous at almost all sample paths of Z_0 uniformly over $\zeta \in \Xi$. The result now follows by Slutksy's theorem and the continuous mapping theorem. $\square$

Proof of lemma 6. Let $F = \exp(-\Lambda)$ denote the survival probability function. The likelihood as a function of (γ, F) can be formed as $p_\zeta(\xi, F; x) = p_1 1\{z \le \zeta\} + p_2 1\{z > \zeta\}$, where

$$p_1 = \delta(1 - \overline{F}(v)^{\exp(\alpha z)}) + (1 - \delta)\overline{F}(v)^{\exp(\alpha z)},$$

and

$$p_2 = \delta(1 - \overline{F}(v)^{\exp(\beta_1 + (\alpha + \beta_2) z)}) + (1 - \delta)\overline{F}(v)^{\exp(\beta_1 + (\alpha + \beta_2) z)}.$$

Let $0 < \alpha < 1$ be a fixed constant. If $(\xi, F) \ne (\xi_0, F_0)$, then by concavity of the function $u \to \log u$, and by Jensen's inequality,

$$P\left\{\log\left[1 + \alpha\left(\frac{p_\zeta(\xi, F; X)}{p_0(\xi, F; X)} - 1\right)\right]\right\} < 0.$$

It follows from the continuity of the left-hand-side quantity in ζ and the compactness of Ξ that

$$\sup_{\zeta \in \Xi} P\left\{\log\left[1 + \alpha\left(\frac{p_\zeta(\xi, F; X)}{p_0(\xi, F; X)} - 1\right)\right]\right\} < 0.$$

Meanwhile, we notice that for any arbitrary, possibly random sequence $\{\zeta_n\}$,

$$n\mathbb{P}_n\left\{\log\left[1 + \alpha\left(\frac{p_{\zeta_n}(\hat{\xi}_n(\zeta_n), \zeta_n, \hat{F}_n(\zeta_n); X)}{p_0(\xi_0, F_0; X)} - 1\right)\right]\right\} \ge 0.$$

Using a minor adaptation of the Wald type argument used in the proof of theorem 3.2 in Huang (1996), we can show that $\hat{\xi}_n(\zeta_n) - \xi_0 = o_{P_0}(1)$ and $\sup_{v \in [\sigma, \tau]} |\hat{F}(\zeta_n) - F_0| = o_{P_0}(1)$. The assertion of lemma 6 follows from the arbitrariness of the sequence ζ_n and Slutsky's theorem. $\square$

PROOF OF LEMMA 7. For each $n \in \mathbb{N}$, $j \in \mathbb{Z}$ and $M > 0$, define "peels" $S_{n,j,M} = \Big\{(\beta, \eta, \zeta) \in B \times \mathcal{H}_\eta \times \Xi : \ 2^{j-1} < r_n d(\eta, \eta_0) \le 2^j,$ $\|\beta - \beta_0\| \le 2^{-M} d(\eta, \eta_0)\Big\}$. The event $\Big\{\tilde\beta_n \in B, \ \hat\eta_{\tilde\beta_n}(\zeta) \in \mathcal{H}_\eta, \ \zeta \in \Xi,$ $r_n d(\hat\eta_{\tilde\beta_n}(\zeta), \eta_0) \ge 2^M(1 + r_n\|\tilde\beta_n - \beta_0\|)\Big\}$ is contained in one of the peels $\Big\{(\tilde\beta_n, \hat\eta_{\tilde\beta_n}(\zeta), \zeta) \in S_{n,j,M}\Big\}$ over $j \ge M$. By the definition of $\hat\eta_{\tilde\beta_n}(\zeta)$, the variable $\sup_{(\beta,\eta,\zeta)\in S_{n,j,M}} \mathbb{P}_n(m_{\beta,\eta,\zeta} - m_{\beta,\eta_0,\zeta})$ is non-negative on the event $\Big\{(\tilde\beta_n, \hat\eta_{\beta_n}(\zeta), \zeta) \in S_{n,j,M}\Big\}$. We conclude that for any possibly random sequence ζ_n,

$$P^\star\Big(r_n d(\hat\eta_{\tilde\beta_n}(\zeta_n), \eta_0) \ge 2^M(1 + r_n\|\tilde\beta_n - \beta_0\|), \tilde\beta_n \in B, \hat\eta_{\tilde\beta_n}(\zeta_n) \in \mathcal{H}_\eta, \zeta_n \in \Xi\Big) \le \sum_{j \ge M} P^\star\big(\sup_{(\beta,\eta,\zeta)\in S_{j,n,M}} \mathbb{P}_n(m_{\beta,\eta,\zeta} - m_{\beta,\eta_0,\zeta}) \ge 0\big).$$

The remainder of the proof is a straightforward adaptation of the proof of theorem 3.2 given in Murphy and van der Vaart (1999), which details we omit. □

REFERENCES

ANDREWS, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, **67** 1341–1383.

ANDREWS, D. W. K. and PLOBERGER, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative (STMA V36 3473). *Econometrica*, **62** 1383–1414.

ANDREWS, D. W. K. and PLOBERGER, W. (1995). Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *The Annals of Statistics*, **23** 1609–1629.

BAGDONAVIČIUS, V. B. and NIKULIN, M. S. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis*, **5** 329–350.

BICKEL, P. J., RITOV, Y. and STOKER, T. M. (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *The Annals of Statistics*, **34** 721–741.

CHAPPELL, R. (1989). Fitting bent lines to data, with applications to allometry. *Journal of Theoretical Biology*, **138** 235–256.

CHEN, H. and CHEN, J. (2003). Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, **13** 351–365.

CHEN, H., CHEN, J. and KALBFLEISCH, J. D. (2004). Testing for a finite mixture model with two components. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **66** 95–115.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Annals of Methematical Statistics*, **25** 573–578.

CHERNOFF, H. and LANDER, E. (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *Journal of Statistical Planning and Inference*, **43** 19–40.

DACUNHA-CASTELLE, D. and GASSIAT, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *The Annals of Statistics*, **27** 1178–1209.

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64** 247–254.

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74** 33–43.

GOLUB, G. H. and VAN LOAN, C. F. (1983). *Matrix Computations*. Johns Hopkins University Press.

HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis (STMA V37 4551). *Econometrica*, **64** 413–430.

HUANG, J. (1996). Efficient estimation for the proporitional hazards model with interval censoring. *The Annals of Statistics*, **24** 540–568.

KING, M. L. and SHIVELY, T. S. (1993). Locally optimal testing when a nuisance parameter is present only under the alternative. *The Review of Economics and Statistics*, **75** 1–7.

KOSOROK, M. R. (To appear). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

KOSOROK, M. R., LEE, B. L. and FINE, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics*, **32** 1448–1491.

KOSOROK, M. R. and SONG, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics*, **35** 957–989.

LINDSAY, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications*. Institute of Mathematical Statistics.

LIU, X. and SHAO, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics*, **31** 807–832.

LUO, X. and BOYETT, J. M. (1997). Estimations of a threshold parameter in Cox regression. *Communications in Statistics: Theory and Methods*, **26** 2329–2346.

LUO, X., TURNBULL, B. W. and CLARK, L. C. (1997). Likelihood ratio tests for a changepoint with survival data. *Biometrika*, **84** 555–565.

MURPHY, S. A., ROSSINI, A. J. and VAN DER VAART, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, **92** 968–976.

MURPHY, S. A. and VAN DER VAART, A. W. (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics*, **25** 1471–1509.

MURPHY, S. A. and VAN DER VAART, A. W. (1999). Observed information in semiparametric models. *Bernoulli*, **5** 381–412.

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood (C/R: P466-485). *Journal of the American Statistical Association*, **95** 449–465.

PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics*, **26** 183–214.

POLLARD, D. (1995). Another look at differentiability in quadratic mean. URL `http://www.stat.yale.edu/~pollard/LeCamFest/Pollard.pdf`.

PONS, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics*, **31** 442–463.

SCHARFSTEIN, D. O., TSIATIS, A. A. and GILBERT, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*, **4** 355–391.

VAN DER VAART, A. W. (1996). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, **54** 426–482.

Zhu, H. T. and Zhang, H. P. (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **66** 3–16.

Rui Song
Department of Biostatistics,
University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599-7420, U.S.A.
E-mail: rsong@bios.unc.edu

Michael R. Kosorok
Departments of Biostatistics
and Statistics & Operation Research,
University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599-7420, U.S.A.
E-mail: kosorok@unc.edu

Jason P. Fine
Departments of Statistics
and Biostatistics & Medical Informatics,
University of Wisconsin at Madison,
Madison, WI 53706, U.S.A.
E-mail: fine@biostat.wisc.edu