The Annals of Statistics
1996, Vol. 24, No. 2, 841–861

ASYMPTOTICALLY UNIFORMLY MOST POWERFUL TESTS IN PARAMETRIC AND SEMIPARAMETRIC MODELS[1]

By Sungsub Choi, W. J. Hall and Anton Schick

McGill University, University of Rochester and Binghamton University

Tests of hypotheses about finite-dimensional parameters in a semiparametric model are studied from Pitman's moving alternative (or local) approach using Le Cam's local asymptotic normality concept. For the case of a real parameter being tested, *asymptotically uniformly most powerful* (AUMP) tests are characterized for one-sided hypotheses, and AUMP *unbiased* tests for two-sided ones. An asymptotic invariance principle is introduced for multidimensional hypotheses, and AUMP *invariant* tests are characterized. These provide optimality for Wald, Rao (score), Neyman–Rao (effective score) and likelihood ratio tests in parametric models, and for Neyman–Rao tests in semiparametric models when constructions are feasible. Inversions lead to *asymptotically uniformly most accurate* confidence sets. Examples include one-, two- and k-sample problems, a linear regression model with unknown error distribution and a proportional hazards regression model with arbitrary baseline hazards. Results are presented in a format that facilitates application in strictly parametric models.

1. Introduction. The first rigorous work to define and construct tests which are asymptotically optimal was by Wald (1943). He argued that maximum likelihood estimators may be asymptotically sufficient for detecting local deviations from the null hypothesis and showed that a test based on them—now called a *Wald test*—is asymptotically *most stringent*: its asymptotic power function is closest to the asymptotic envelope power function in the minimax sense in local neighborhoods of the null hypothesis. He also considered two other definitions of optimality, each achieved by the same test: namely, *asymptotically best average power*, or *constant power*, *over a family of surfaces*. Wald also showed equivalence with the likelihood ratio test; score tests had not yet been introduced. [In an earlier paper, Wald (1941), he showed that a Wald test has a *global* optimality, but only in models with a single real parameter.] However, this work of Wald has not been distilled into textbook form—and his optimality results are not even

Received July 1995.

[1]Partially supported by a Merck Predoctoral Fellowship in Biostatistics at the University of Rochester (Choi), National Cancer Institute Grant NCI 5P30 CA11198-20 (Hall) and NSF Grant DMS-90-03696 (both).

AMS 1991 *subject classifications*. Primary 62F05; secondary 62G20.

Key words and phrases. Local alternatives, effective scores, unbiased tests, invariance, efficient tests, adaptation, asymptotic confidence sets.

quoted in textbooks!—although the test carrying his name has become a standard in modern statistical practice. And as the scope of statistical applications has broadened to problems with infinite-dimensional parameters (nonparametric and semiparametric models), his parametric formulation is no longer adequate.

Almost two decades later, Le Cam (1960) introduced *local asymptotic normality* (LAN) of the log-likelihood ratio along with its asymptotic distribution under local alternatives, the latter being essential in evaluating local power. He also showed that the scores are asymptotically sufficient for local departures. His work was elaborated, extended and distilled by Hájek and Šidák (1967). This paved the way for asymptotic optimality of score-based (Rao) tests, which was dealt with by Roussas (1972) for parametric simple hypothesis problems. About this time, Neyman (1959) introduced his $C(\alpha)$ tests, a forerunner of *effective score tests* [or Neyman–Rao tests—see Hall and Mathiason (1990)] which are a generalization of the popular *score tests*.

Recent books by Strasser (1985), Le Cam (1986), Le Cam and Yang (1990), Fabian and Hannan (1985) and Andersen, Borgan, Gill and Keiding (1993) and papers by Hall and Mathiason (1990) and Wefelmeyer (1987) include some material on large-sample tests, all based on LAN. Strasser imposes *similarity* or *unbiasedness* for tests about real parameters and has only limited results for nuisance functions. Le Cam as well as Le Cam and Yang deals with decision problems in a very general and abstract setting, and devotes little attention to the specifics under study here. Fabian and Hannan and Hall and Mathiason each confine attention to parametric models and define optimality of tests in limited ways, the first by reference to matching the performance of the Wald test and the second within a restricted class of tests. Andersen, Borgan, Gill and Keiding base their development partly on Choi (1989), a forerunner to much of this paper. Wefelmeyer removes the similarity constraint when testing against a particular contiguous alternative, and this implies asymptotic optimality in certain one-sided testing problems. We expand this latter approach here.

Begun, Hall, Huang and Wellner (1983) present a theory of asymptotically efficient estimation in semiparametric models utilizing LAN. We attempt a parallel theory of testing hypotheses about a finite-dimensional parameter, rigorous but not too complex mathematically. Unlike Begun, Hall, Huang and Wellner or Strasser, however, we emphasize a directional approach as in Huang (1982). The paper by Hall and Mathiason (1990) lays the groundwork for our approach by introducing effective score tests. And we have formulated assumptions in parallel with this parametric case; the paper thus provides a theory of optimal testing in parametric models as well as in semiparametric models.

Results presented here parallel those for testing hypotheses about part of a vector μ when observing a normally distributed random variable with mean vector $B\mu$ and known variance matrix B. In large samples, the normal variable is the score vector, B the information and μ the shift in the parameter under local alternatives. When the test is one-sided about a real component μ_1 of $\mu = (\mu_1, \mu_2)$, a uniformly most powerful (UMP) test is possible. This can be derived as a test of Neyman structure using theory presented in Lehmann (1986) by conditioning on a statistic which is sufficient and complete for the null hypothesis. However, this theory can be avoided—thereby avoiding a need for dealing with asymptotic sufficiency and completeness. An alternative derivation is as follows: fix a null value of the nuisance parameter μ_2, say μ_2^0, and first restrict attention to level-α tests of $\mu = (0, \mu_2^0)$ vs. μ in the hyperplane $\{\mu\colon \mu_1 \in \mathscr{R},\ \mu_2 = \mu_2^0 - B_{22}^{-1} B_{21} \mu_1\}$. The Neyman–Pearson lemma [Lehmann (1986)] asserts that a test based on the *effective observation*—the residual from regressing the coordinate of interest on the other coordinates—is UMP. Since this test and its power are free of μ_2^0, it remains UMP when the parameter space is no longer restricted to the hyperplane.

This argument can be extended to handle two-sided alternatives and find a UMP unbiased test. For testing hypotheses about μ_1 of dimension $d \geq 1$ against unrestricted alternatives, a UMP invariant test—invariant under a group of nonsingular linear transformations for the whole vector—can be obtained relying on a maximal invariant statistic as well as on a sufficient and complete statistic [Lehmann (1986)]. Instead, one may deal with the nuisance parameter μ_2 by confining attention to a restricted alternative of a hyperplane—thus considering only effective observations—and requiring invariance only for effective observations.

Asymptotic analogs of each are given here resulting in the notion of asymptotically uniformly most powerful (AUMP), AUMPU (unbiased) and AUMPI (invariant) tests. In Section 2, Le Cam's LAN is presented along with necessary notation and assumptions. And effective scores, effective information and efficient test statistics are defined. We then characterize AUMP one-sided tests in Section 3. As in Wefelmeyer (1987), the common asymptotic similarity restriction is avoided. Characterization is done by stating the asymptotic local power function. Sufficient for optimality is that a test be equivalent to a *canonical effective score test*—an optimal test requiring knowledge of nuisance parameters. Stein's (1956) notion of *adaptation* is discussed briefly at the end: replacing, in tests which are optimal when certain nuisance parameters are known, these parameters by estimates without affecting the asymptotic performance of the test. This is like a large-sample version of Studentization; a variance parameter can be replaced by an estimate without any large-sample penalty.

Two-sided and multidimensional tests are discussed in the next two sections. The approach used in Section 3 does not directly generalize. Instead, we focus on a hyperplane in a certain direction and appeal to the asymptotic power representation in the Appendix, essentially reducing the problem into a simple normal shift problem discussed above. For invariance, it is noted that the testing problem is invariant under locally linear transformation of the parameter of interest, and this motivates requiring standardized effective score tests to be rotation invariant (asymptotically). As in Wald (1943), but in contrast to common linear-model parametric hypothesis testing, invariance is not imposed to deal with the nuisance parameters. Again, optimal tests are equivalent to canonical effective score tests. Obtaining asymptotic confidence sets by inversion is briefly discussed in Section 6.

The construction of asymptotically efficient tests is discussed in Section 7. In the parametric case, likelihood ratio tests [assuming LAN holds uniformly —see Hall and Mathiason (1990)], Wald tests and Rao (score) tests are all asymptotically efficient, as are effective score tests—canonical tests with nuisance parameters replaced by $\sqrt{n}$-consistent estimates. These are typically not feasible in the semiparametric case. We modify Bickel's (1982) and Schick's (1986, 1987) constructions of adaptive estimates to construct effective score (Neyman–Rao) tests when infinite-dimensional nuisance parameters are involved.

The methodology is demonstrated in several semiparametric examples in the final section. In particular, we show that the partial score tests of Cox (1975), popular in survival analysis, are optimal. Some parametric examples of effective score tests appear in Hall and Mathiason (1990); the theory here asserts their optimality, without the restrictive conditions imposed there.

In summary, we characterize, and show how to construct, a large-sample test with a simply stated optimality: either AUMP, AUMPU or AUMPI. Popular tests, such as Wald, score and likelihood ratio tests, are asymptotically equivalent to this test, and hence share this optimality. We thus simplify, clarify and extend the efficiency concepts of large-sample testing, introduced in Wald's fundamental paper of 1943.

2. Local asymptotic normality, effective scores and efficient test statistics. Suppose we are investigating a specific characteristic of a probability measure $P_{n,\theta}$, $\theta \in \Theta$, based on some potential data $\underline{X}_n$. The subscript n is an index of the amount of data, for example, sample size. By adopting reparametrization if necessary, we assume that θ can be partitioned (at least locally) into (ϑ, η) so that the characteristic we are interested in is identifiable solely by ϑ and the hypothesis to be tested is given as H: $\vartheta = \vartheta_0$. The parameter ϑ (of finite dimension $d \geq 1$) is called the *parameter of interest* and the parameter η (of arbitrary dimension) the *nuisance parameter* or *function*.

We confine attention to contiguous alternatives [Le Cam (1960) and Hájek and Šidák (1967)]. Define a $\sqrt{n}$-neighborhood of ϑ_0 as a collection of sequences $\vartheta_n(h_\vartheta) = \vartheta_0 + n^{-1/2}h_\vartheta + n^{-1/2}\delta_{n\vartheta}$ for $h_\vartheta \in \mathscr{H}_\vartheta$ and $\|\delta_{n\vartheta}\| = o(1)$, where the local parameter space $\mathscr{H}_\vartheta$ is a subset of $\mathscr{R}^d$ containing 0. Similarly define a $\sqrt{n}$-neighborhood of η (with η fixed but unknown) as $\eta_n(h_\eta) = \eta + n^{-1/2}h_\eta + n^{-1/2}\delta_{n\eta}$ for $h_\eta \in \mathscr{H}_\eta$ and $\|\delta_{n\eta}\| = o(1)$, where the local nuisance parameter space $\mathscr{H}_\eta$ is a Hilbert space (typically, a subspace of a Cartesian product of copies of $\mathscr{R}$ and/or of $\mathscr{L}_2$, the space of square integrable functions with respect to some fixed measure). Here, $\|\cdot\|$ denotes the norm of the appropriate Hilbert space. We use $\langle \cdot, \cdot \rangle$ to represent inner product in a similar fashion. Let $h = (h_\vartheta, h_\eta)$ be an element in the product space $\mathscr{H}$ of $\mathscr{H}_\vartheta$ and $\mathscr{H}_\eta$, and let $\theta_0 = (\vartheta_0, \eta)$ and $\theta_n = \theta_n(h) = (\vartheta_n(h_\vartheta), \eta_n(h_\eta))$—we omit h

when dependence on it is obvious from the context. The performance of a test is evaluated against these sequences of parameters θ_n, approaching θ_0 from *direction h*.

Let $dP_{n,\theta_n}/dP_{n,\theta_0}$ be the Radon-Nikodym derivative of the absolutely continuous part of P_{n,θ_n} with respect to P_{n,θ_0}. Denote A_n as the support of P_{n,θ_0}. Following Le Cam (1960, 1969), we assume that likelihood ratios of local alternatives to the null hypothesis are asymptotically log-normally distributed. More specifically, for $\underset{\sim}{x}_n$ in A_n, and for each direction h,

$$L_n(h) = \log \frac{dP_{n,\theta_n(h)}}{dP_{n,\theta_0}} = S_n h - \frac{1}{2}\sigma^2(h) + r_n(h), \tag{1}$$

where $S_n = S_n(\underset{\sim}{x}_n) = (S_{n\vartheta}, S_{n\eta})$ is an h-free random linear functional which is asymptotically Gaussian under θ_0 with mean 0 and variance B in the sense that $S_n\tilde{h}$ is asymptotically normally distributed with mean 0 and variance $\sigma^2(\tilde{h}) = \langle \tilde{h}, B\tilde{h} \rangle$ for every $\tilde{h}$, and B is a positive-definite self-adjoint bounded linear operator. Under θ_0, $r_n(h)$ converges in probability to 0 for every h. When $\underset{\sim}{x}_n \notin A_n$, L_n is defined arbitrarily.

This is the analog of the parametric LAN assumption in Hall and Mathiason (1990), and we choose notation to emphasize the parametric case. The full process in h is not needed, although it is implicit due to the linearity of S_n [see Strasser (1985)]. The joint convergence in $S_n h$ and $S_n \tilde{h}$ is sufficient for application of Le Cam's third lemma below. The asymptotic covariance is $\langle h, B\tilde{h} \rangle$. A sufficient condition for LAN in the case of random sampling is given in Begun, Hall, Huang and Wellner (1983), namely, Hellinger differentiability with respect to θ of the marginal density. This condition is easily extended to multisample, regression and censoring models, or these can be accommodated in the iid case as in Begun, Hall, Huang and Wellner (1983). See also Bickel, Klaassen, Ritov and Wellner (1993) for this and Fabian and Hannan (1987) for generalizations to dependent data settings.

Among the immediate consequences of LAN are:

(P1) Contiguity, or Le Cam's first lemma [Le Cam (1960) and Hájek and Šidák (1967)], part of which asserts that $P_{\theta_n(h)}(A_n) \to 1$ for each h. Since $dP_{\theta_n(h)} = \exp\{L_n(h)\}\, dP_{\theta_0}$ on A_n, it also implies that $\exp\{L_n(h)\}$ is uniformly integrable under θ_0.

(P2) The asymptotic distribution of the score under local alternatives is readily available. Le Cam's third lemma [Le Cam (1960), Hájek and Šidák (1967) and Hall and Mathiason (1990)] implies that S_n is asymptotically Gaussian under $\theta_n(h)$ with mean Bh and variance B—that is, the asymptotic distribution of $S_n\tilde{h}$ under $\theta_n(h)$ is normal with mean $\langle \tilde{h}, Bh \rangle$ and variance $\sigma^2(\tilde{h})$ for every $\tilde{h}$.

We have no need for asymptotic sufficiency of the scores here; but see Wald (1943), Le Cam (1960) and Strasser (1985).

Of course, S_n and B depend on θ_0; they are the *score* and *information* at θ_0. Real quantities $S_n \tilde{h}$ and $\sigma^2(\tilde{h})$ are the *directional* score and information in direction $\tilde{h}$ [Huang (1982)]. We call $S_{n\vartheta}$ the score for the parameter of interest and $S_{n\eta}$ the score for the nuisance parameter. The information B may also be partitioned into (B_{ij}), $i, j = 1, 2$, where B_{11} is the information for ϑ, B_{22} the information for η and B_{12} and (B_{21}) the co-information.

We assume that B_{22} has a bounded inverse B_{22}^{-1}. This allows us to define the *effective information* B^* as $B^* = B_{11} - B_{12}B_{22}^{-1}B_{21}$ and the *effective score* S_n^* as $S_n^* a = S_n(a, -B_{22}^{-1}B_{21}a) = S_{n\vartheta}\, a - S_{n\eta}B_{22}^{-1}B_{21}a$, $a \in \mathscr{R}^d$. We also think of B^* as the $d \times d$ matrix defined by $a^T B^* a = \langle a, B^* a\rangle$ and of S_n^* as a d-dimensional random vector defined by $a^T S_n^* = S_n^* a$. Under $\theta_n(h)$, the random vector S_n^* is asymptotically normal with mean $B^* h_\vartheta$ and variance B^* and is asymptotically independent of $S_{n\eta}h_\eta$ for each $h_\eta \in \mathscr{H}_\eta$. Thus the effective information is the asymptotic variance of the effective score, and the random variable $S_n^* a$ is the residual from projection of $S_{n\vartheta}\, a$ onto the space spanned by $S_{n\eta}$—the part of $S_{n\vartheta}\, a$ which is orthogonal (asymptotically uncorrelated) to $S_{n\eta}$. Of course, it is the same as $S_{n\vartheta}\, a$ if $B_{12} = 0$ or if no nuisance parameter is present. The positive definiteness of B implies that of B^*. Thus we can standardize the effective score. To stress the dependence of the standardized effective score $B^{*-1/2}S_n^*$ on the nuisance parameter η, we shall denote it by $\xi_n(\eta)$.

Although it has an explicit algebraic form, the effective score for a specific problem may be difficult to obtain, especially when nuisance functions are involved. Calculation of adjoints or inverses of linear operators is not always straightforward. Frequently, it is easier to minimize (in h_η) the asymptotic variance of the directional score $S_{n\vartheta}h_\vartheta + S_{n\eta}h_\eta$ whose solution is the least favorable direction $\tilde{h}_\eta = -B_{22}^{-1}B_{21}h_\vartheta$. Since $S_n^* h_\vartheta = S_{n\vartheta}h_\vartheta + S_{n\eta}\tilde{h}_\eta$ for every h_ϑ, S_n^* can be recovered from this.

We will find that asymptotically efficient tests are characterized in terms of the standardized effective score $\xi_n(\eta)$. So what is needed is a version of the standardized effective score that is independent of the nuisance parameter, that is, a statistic T_n for which $T_n - \xi_n(\eta)$ converges to 0 in $P_{n,(\vartheta_0,\eta)}$-probability for every η. We call such a statistic T_n an *efficient test statistic*. It has the same asymptotic properties as $\xi_n(\eta)$, namely being $\mathrm{AN}(0, I)$ under H and $\mathrm{AN}(B^{*1/2}h_\vartheta, I)$ under local alternatives (h_ϑ, h_η) for every η. (Alternatively, T_n could be defined as any statistic having these asymptotic normality properties.) We show, in turn, in Sections 3 to 5 that one-sided, two-sided and multidimensional tests based on T_n—as if we were testing that a normally distributed, identity variance, T_n had mean 0—are asymptotically uniformly most powerful in some appropriate sense. Construction of such T_n's is considered in Section 7.

The concept of effective scores was first introduced by Neyman (1959) in a slightly different form; also see Basawa and Koul (1988). Effective scores and information also play a key role in large-sample estimation; see Begun, Hall, Huang and Wellner (1983), Hall and Mathiason (1990) and Bickel, Klaassen, Ritov and Wellner (1993).

3. Asymptotically efficient one-sided tests. We consider testing H: $\vartheta = \vartheta_0$ vs. $K_{(1)}$: $\vartheta > \vartheta_0$ for real ϑ with η unspecified. We first act as if η (but not h_η) is known and consider a local form of these hypotheses, $h_\vartheta = 0$ vs. $h_\vartheta > 0$ with h_η unspecified. It is assumed that $\mathscr{H}_\vartheta$ contains the half-line $[0, \infty)$. (We could enlarge the null hypothesis and its local form to $\vartheta \le \vartheta_0$ and $h_\vartheta \le 0$, but we keep the simpler form as stated.)

Fix $\alpha \in (0, 1)$. A test ψ_n is of *asymptotic level* α at η if

$$\limsup E_{\theta_n(0, h_\eta)} \psi_n \le \alpha \quad \text{for every } h_\eta.$$

Such a test ψ_n is *asymptotically most powerful of level* α *against* $\theta_n(h)$ if $\liminf E_{\theta_n(h)} \psi_n \ge \limsup E_{\theta_n(h)} \psi_n'$ for every other such test ψ_n'. If this is true for every h with $h_\vartheta > 0$, such a test is *asymptotically uniformly most powerful of level* α *at* η, short AUMP(α, η). A test is AUMP(α) if it is AUMP(α, η) for each nuisance parameter η. The word "local" is implicit everywhere.

It should be noted that the asymptotic level requirement is imposed for every h_η. This requirement is crucial and plays the role of restriction to *regular estimates* in estimation theory; see Begun, Hall, Huang and Wellner (1983) and Hall and Mathiason (1990).

Note that, under LAN, the asymptotic power of a test can be evaluated by a computation under the null distribution. Since, for any test ψ_n, $E_{\theta_n(h)} \psi_n = E_{\theta_n(h)} \psi_n \mathbf{1}(A_n) + E_{\theta_n(h)} \psi_n \mathbf{1}(A_n^c)$, and the second part vanishes by (P1) as n increases, we have, for every h,

$$\begin{aligned} E_{\theta_n(h)} \psi_n &= E_{\theta_0} \psi_n \exp\{L_n(h)\} + o(1) \\ &= E_{\theta_0} \psi_n \exp\{S_n h - \tfrac{1}{2} \sigma^2(h) + r_n(h)\} + o(1). \end{aligned} \tag{2}$$

Fix $h_1 = (h_{\vartheta 1}, h_{\eta 1})$, $h_{\vartheta 1} > 0$, temporarily choose $h_{\eta 0}$ and test the simple hypotheses $h_0 = (0, h_{\eta 0})$ versus h_1. Applying the Neyman–Pearson lemma to the right side of (2) [without the $o(1)$ term], we find an optimal test of asymptotic level α to be of the form $\varphi_n = 1$ if

$$\begin{aligned} L_n(h_1) - L_n(h_0) = S_n(h_1 - h_0) - \tfrac{1}{2}\{\sigma^2(h_1) - \sigma^2(h_0)\} \\ + \{r_n(h_1) - r_n(h_0)\} > c_n, \end{aligned}$$

and $\varphi_n = 0$ if $L_n(h_1) - L_n(h_0) < c_n$. The asymptotic distribution of $S_n(h_1 - h_0)$ under $\theta_n(h)$ is normal with mean $\langle h_1 - h_0, Bh \rangle$ and variance $\sigma^2(h_1 - h_0)$. Letting $h = h_0$ (the null hypothesis), we find that $\liminf c_n \ge c = z_\alpha \sigma(h_1 - h_0) - \frac{1}{2}\sigma^2(h_1 - h_0)$, where z_α is the upper α-quantile of the standard normal distribution Φ. Now taking $h = h_1$ (the alternative), it follows that $\limsup E_{\theta_n(h_1)} \psi_n \le 1 - \Phi\{z_\alpha - \sigma(h_1 - h_0)\}$.

Simple algebra shows that the noncentrality $\sigma^2(h_1 - h_0)$ is minimized in $h_{\eta 0}$ when $h_{\eta 1} - h_{\eta 0} = -B_{22}^{-1} B_{21} h_{\vartheta 1}$, which we call the *least favorable direction*. The point $(0, h_\eta^0)$, $h_\eta^0 = h_{\eta 1} + B_{22}^{-1} B_{21} h_{\vartheta 1}$, is the projection of h_1 onto the local null space under the inner product induced by B, namely, $\langle h, k \rangle_B = \langle h, Bk \rangle$, $h, k \in \mathscr{H}$. As the point in the local null space closest to h_1, it is the most difficult one to distinguish from h_1. By plugging in this least favorable direction, we have

$$\limsup E_{\theta_n(h_1)}\psi_n \le 1 - \Phi\big(z_\alpha - B^{*1/2}h_{\vartheta 1}\big) = \Phi\big(B^{*1/2}h_{\vartheta 1} - z_\alpha\big), \tag{3}$$

where $B^* = B_{11} - B_{12}B_{22}^{-1}B_{21}$ is the effective information. Note that this bound depends on h_1 only through $h_{\vartheta 1}$, that part of the departure we are interested in. It is achieved by the test

$$\phi_n = \mathbf{1}(\xi_n(\eta) \ge z_\alpha) = \begin{cases} 1, & \text{if } \xi_n(\eta) \ge z_\alpha, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $\xi_n(\eta)$ is the (real-valued) standardized effective score $B^{*-1/2}S_n^*$. Since the test is free of $h_{\vartheta 1}$ as well as $h_{\eta 1}$, we may claim that ϕ_n is AUMP(α, η). Of course, any equivalent test is also AUMP(α, η). (Two tests are asymptotically equivalent, or simply equivalent, if their difference converges to 0 in $P_{n,(\vartheta_0, \eta)}$-probability.)

Note that no asymptotic version of unbiasedness or similarity constraint has been required, though ϕ_n does have these properties. A similar approach can be found in Wefelmeyer (1987).

Alternatively, we can first go to the limit and then apply the Neyman–Pearson lemma. This is done as follows. Fix h_1 as above and a test ψ_n of asymptotic level α at η. Choose a subsequence n' of n such that $\lim E_{\theta'_n(h_1)}\psi_{n'} = \limsup E_{\vartheta_n(h_1)}\psi_n$. Lemma 1 in the Appendix yields a subsequence n'' such that $\lim E_{\theta_{n''}(h)}\psi_{n''} = \int\varphi(z)\,d\Phi(z - B^{*1/2}h_\vartheta)$ for every $h = (h_\vartheta, h_\eta^0 - B_{22}^{-1}B_{21}h_\vartheta)$ with $h_\vartheta \ge 0$ and some test φ of level α. By the Neyman–Pearson lemma such a test φ satisfies $\int\varphi(z)\,d\Phi(z - B^{*1/2}h_\vartheta) \le \Phi(B^{*1/2}h_\vartheta - z_\alpha)$ and achieves equality if and only if $\varphi(z) = \mathbf{1}(z \ge z_\alpha)$ almost everywhere z. This shows again that $\mathbf{1}(\xi_n(\eta) \ge z_\alpha)$ is AUMP(α, η), and Lemma 2 in the Appendix gives us uniqueness up to equivalence. This approach is extended in Sections 4 and 5.

Let us now summarize our results.

THEOREM 1. *Every test ψ_n of asymptotic level α at η satisfies* (3) *for every $h_1 = (h_{\vartheta 1}, h_{\eta 1}) \in [0,\infty) \times \mathscr{H}_\eta$. The canonical effective score test $\phi_n = \mathbf{1}(\xi_n(\eta) \ge z_\alpha)$, and any equivalent test, is* AUMP(α, η) *for testing H: $\vartheta = \vartheta_0$ versus $K_{(1)}$: $\vartheta > \vartheta_0$ and is unique up to equivalence. Moreover, for each $h_1 \in [0,\infty) \times \mathscr{H}_\eta$, $E_{\theta_n(h_1)}\phi_n$ converges to the right side of* (3).

If there is an efficient test statistic (defined in Section 2), we can achieve efficiency for every η with a global test.

COROLLARY 1. *If T_n is an efficient test statistic, then $\phi_{(1)}(T_n) = \mathbf{1}(T_n \ge z_\alpha)$ is* AUMP(α) *for testing H versus $K_{(1)}$.*

It may be noted, as in the small-sample case, that the form (4) of asymptotically optimal tests depends on α only through the critical value z_α. Hence, asymptotic p-values may be defined.

Before closing this section, note that the asymptotic power function in (3) is an increasing function of the noncentrality $B^{*1/2}h_\vartheta$. If the nuisance parameter η had been known, we could have used a test based on $S_{n\vartheta}$ with power function (of the same form) with noncentrality $B_{11}^{1/2}h_\vartheta$. The ratio B^*/B_{11} is Pitman's *asymptotic relative efficiency* (ARE), relative to what is possible if η were known. The loss of efficiency, $1 - \text{ARE}$, is the price for not knowing the nuisance parameter. We then ask: when can a nuisance parameter be added without causing loss of efficiency?

Enlarge the nuisance parameter as (η, τ). Local parameters and corresponding scores and information are enlarged accordingly. We choose τ as that part of the nuisance parameter (if any), possibly after reparametrization, for which Stein's (1956) orthogonality condition holds: $B_{13} = B_{12}B_{22}^{-1}B_{23}$ ($B_{13} = 0$ if the role of η is vacuous). That is, $S_n^* = S_{n\vartheta} - S_{n\eta}B_{22}^{-1}B_{21}$ and $S_{n\tau}$ are asymptotically independent. A simple sufficient case is when $S_{n\tau}$ is asymptotically independent of all other scores. An even more transparent example is when the likelihood may be factored into two parts containing only (ϑ, η) and τ, respectively. A linear regression problem is a typical example, where the regression coefficients (including the parameter of interest) are separated from the covariate distribution (a nuisance parameter).

According to Stein, τ is "a parameter that makes the problem more difficult" but should not affect the large-sample performance of a test. Suppose we act as if τ were known, and construct a test which is equivalent to the optimal canonical test ϕ_n. If τ is also perturbed, the asymptotic distribution of S_n^* in direction h is normal with mean $B^*h_\vartheta + (B_{13} - B_{12}B_{22}^{-1}B_{23})h_\tau$ and variance B^* and hence is free of h_τ if and only if Stein's condition holds. Thus, no additional loss of efficiency is incurred by τ; such a nuisance parameter is said to be *adaptable*. A corresponding optimal test (if existent) is called an *adaptive test*. A good illustrative example with nontrivial co-informations is the parametric regression problem: $X_i = \eta + \vartheta Z_i + \tau\varepsilon_i$, $i = 1, \dots, n$, where iid Z_i's with known distribution have mean μ and finite positive variance v^2, ε_i's are iid and independent of the Z_i's and have known density f with finite Fisher information I_f. It can be easily shown that it satisfies Stein's condition. To test H: $\vartheta = \vartheta_0$, we may act as if the scale parameter τ is known to get the effective score $S_n^* = \tau^{-1}n^{-1/2}\sum[(Z_i - \mu) \times s\{(X_i - \eta - \vartheta_0 Z_i)/\tau\}]$ with $s = -f'/f$ and the effective information $B^* = \tau^{-2}v^2 I_f$.

4. Asymptotically unbiased two-sided tests. We continue to assume the parameter of interest ϑ to be real, but the alternative hypothesis is now two-sided. The local alternative is $h_\vartheta \neq 0$. We assume that $\mathscr{H}_\vartheta = \mathscr{R}$ and confine attention to tests that are *asymptotically unbiased at* η, namely tests ψ_n for which $\limsup E_{\theta_n(h_0)}\psi_n \leq \liminf E_{\theta_n(h_1)}\psi_n$ for every $h_0 = (0, h_{\eta 0})$ and $h_1 = (h_{\vartheta 1}, h_{\eta 1})$, $h_{\vartheta 1} \neq 0$. A test ψ_n is an *asymptotically uniformly most powerful unbiased level* α *test at* η, short AUMPU(α, η), if ψ_n is asymptotically unbiased at η and of asymptotic level α at η and if for every other such test ψ_n' and each $\theta_n(h)$ with $h_\vartheta \neq 0$, $\liminf E_{\theta_n(h)}\psi_n \geq \limsup E_{\theta_n(h)}\psi_n'$. Again, if a test is AUMPU(α, η) for every η, it is called AUMPU(α).

Fix an arbitrary $h_1 = (h_{\vartheta 1}, h_{\eta 1})$ in $\mathscr{R} \times \mathscr{H}_\eta$ with $h_{\vartheta 1} \neq 0$. Set $h_\eta^0 = h_{\eta 1} + B_{22}^{-1}B_{21}h_{\vartheta 1}$ and $\mathscr{G} = \{(a, h_\eta^0 - B_{22}^{-1}B_{21}a): a \in \mathscr{R}\}$. Then h_1 belongs to $\mathscr{G}$, and every element of $\mathscr{G}$ has the same projection $(0, h_\eta^0)$ on the local null space. Let ψ_n be a test that has asymptotic level α at η and is asymptotically unbiased at η. Choose a subsequence n' that achieves the most power at h_1, that is, $\lim E_{\theta_{n'}(h_1)}\psi_{n'} = \limsup E_{\theta_n(h_1)}\psi_n$. Lemma 1 in the Appendix assures a subsequence n'' of n' and a test φ for which

$$\lim E_{\theta_{n''}(h)}\psi_{n''} = \int \varphi(z)\, d\Phi\big(z - B^{*1/2}h_\vartheta\big)$$

for every h in $\mathscr{G}$. By the properties of ψ_n, the test φ is an unbiased level-α test for the asymptotic testing problem which tests whether the mean of a normal distribution with variance 1 equals 0. A best unbiased level-α test ϕ for this asymptotic testing problem must satisfy $\phi(z) = \mathbf{1}(|z| \geq z_{\alpha/2})$ almost everywhere. Its finite-sample analog $\phi_n = \mathbf{1}(|\xi_n(\eta| \geq z_{\alpha/2})$ is asymptotically most powerful against $\theta_n(h_1)$ among tests which are of asymptotic level α at η and asymptotically unbiased at η. Since this test is free of h_1 and h_1 is arbitrary, we may conclude that it is AUMPU(α, η). In the next section, the same test will be derived based on an invariance principle.

THEOREM 2. *Every test ψ_n that has asymptotic level α at η and is asymptotically unbiased at η for testing H: $\vartheta = \vartheta_0$ against $K_{(2)}$: $\vartheta \neq \vartheta_0$ satisfies*

$$\limsup E_{\theta_n(h)}\psi_n \leq \Phi\big(|B^{*1/2}h_\vartheta| - z_{\alpha/2}\big) + \Phi\big(-|B^{*1/2}h_\vartheta| - z_{\alpha/2}\big) \tag{5}$$

for all $h = (h_\vartheta, h_\eta) \in \mathscr{R} \times \mathscr{H}_\eta$. The two-sided canonical effective score test

$$\phi_n = \mathbf{1}\big(|\,\xi_n(\eta)| \geq z_{\alpha/2}\big) = \begin{cases} 1, & \textit{if } |\xi_n(\eta)| \geq z_{\alpha/2}, \\ 0, & \textit{otherwise}, \end{cases} \tag{6}$$

and any equivalent test, is AUMPU(α, η) *and is unique up to equivalence. Moreover, for each $h = (h_\vartheta, h_\eta) \in \mathscr{R} \times \mathscr{H}_\eta$, $E_{\theta_n(h)}\phi_n$ converges to the right side of* (5).

COROLLARY 2. *If T_n is an efficient test statistic, then $\phi_{(2)}(T_n) = \mathbf{1}(|T_n| \geq z_{\alpha/2})$ is* AUMPU(α) *for testing H versus $K_{(2)}$.*

5. Asymptotically invariant tests of multidimensional hypotheses. The parameter being tested is now of dimension $d \geq 1$, with unrestricted alternatives as in Section 4. We assume that $\mathscr{H}_\vartheta = \mathscr{R}^d$. Since it is obvious that there is no AUMP test against unrestricted alternatives, we will first introduce an invariance principle and consider those tests that satisfy the invariance criterion. The invariance we consider arises from the observation that the score varies with the parametrization.

Let $\zeta = \zeta(\vartheta)$ be a (smooth) reparametrization of ϑ that satisfies $\| \zeta(\vartheta_n(h_\vartheta)) - \zeta_0 - n^{-1/2} J^{-1} h_\vartheta \| = o(1)$ for every $h_\vartheta \in \mathscr{R}^d$, $\zeta_0 = \zeta(\vartheta_0)$ and J is a nonsingular (to retain identifiability) $d \times d$ Jacobian matrix. Further assume that $\zeta_0 = \vartheta_0$, subtracting the difference from ζ if necessary. The family of local alternatives, $\mathscr{R}^d - \{0\}$, is the same. That is, the local parameter space is invariant under reparametrization. Note also that $(0, h_\eta^0)$ is invariant and retains its role as the projection on the local null space in the new parametrization.

On the other hand, the effective score $S^*_{n\zeta}$ under the parametrization ζ must satisfy $S^*_{n\zeta} - J^T S^*_{n\vartheta} \to 0$ in P_{n,θ_0}-probability and can thus be chosen equal to $J^T S^*_{n\vartheta}$. The asymptotic distribution of $S^*_{n\vartheta}$ under $\theta_n(h)$ is normal with mean $B^* h_\vartheta$ and variance B^*, while that of $S^*_{n\zeta}$ is normal with mean $B^*_\zeta h_\zeta$ and variance $B^*_\zeta = J^T B^* J$. We notice that the asymptotic null distributions of the standardized effective scores, $B^{*-1/2} S^*_{n\vartheta}$ and $B_\zeta^{*-1/2} S^*_{n\zeta}$, are standard normal under both parametrizations. Also, the family of asymptotic local alternative distributions is normal with identity variance, restricted only by the mean not being 0. Thus the asymptotic distribution of the standardized effective score is invariant under nonsingular transformations. Hence it would seem desirable for a canonical score test to be invariant, at least asymptotically, to the way the hypothesis and the score are represented.

Further noting that $B_\zeta^{*-1/2} S^*_{n\zeta} = R^T B^{*-1/2} S^*_{n\vartheta}$ and the set of all $R = B^{*1/2} J (J^T B^* J)^{-1/2}$ (orthonormalization) is precisely the set of all orthogonal $d \times d$ matrices $\{R: R^T R = I\}$, we define: a test ψ_n is *asymptotically invariant at* η if, for each hyperplane $\mathscr{G} = \mathscr{G}_{h_\eta^0} = \{(h_\vartheta, h_\eta^0 - B_{22}^{-1} B_{21} h_\vartheta): h_\vartheta \in \mathscr{R}^d\}$ with $h_\eta^0 \in \mathscr{H}_\eta$, every subsequence has a further subsequence, as provided by Lemma 1 in the Appendix, of which the limit test φ is *rotation invariant*, that is, $\varphi(u) = \varphi(R^T u)$ for every $u \in \mathscr{R}^d$ and every rotation R. It is well known [Lehmann (1986)] that a best rotation invariant test of level α for testing the mean shift $h^*_\vartheta = B^{*1/2} h_\vartheta$ in the standard multivariate normal distribution equals almost everywhere the test $\phi(u) = \mathbf{1}(u^T u \ge \chi_d^2(\alpha))$, where $\chi_d^2(\alpha)$ is the upper α-quantile of the chi-square distribution with d degrees of freedom. Following a similar argument as in the previous section, we then see that $\phi(\xi_n(\eta))$ is *asymptotically uniformly most powerful among all tests that are asymptotically invariant at* η *and of asymptotic level* α *at* η, short AUMPI(α, η).

THEOREM 3. *Each test* ψ_n *which is asymptotically invariant at* η *and of asymptotic level* α *at* η *for testing* H: $\vartheta = \vartheta_0$ *against* $K_{(3)}$: $\vartheta \ne \vartheta_0$ *satisfies*

$$\limsup E_{\theta_n(h)} \psi_n \le 1 - G_d\big(\chi_d^2(\alpha); h_\vartheta^T B^* h_\vartheta\big) \tag{7}$$

for every $h = (h_\vartheta, h_\eta) \in \mathscr{R}^d \times \mathscr{H}_\eta$, *where* $G_d(\cdot; \sigma^2)$ *is the noncentral chi-square distribution function with d degrees of freedom and noncentrality* σ^2. *The canonical effective score test of quadratic form*

$$\phi_n = \mathbf{1}\big(\xi_n(\eta)^T \xi_n(\eta) \ge \chi_d^2(\alpha)\big) = \begin{cases} 1, & \textit{if } \xi_n(\eta)^T \xi_n(\eta) \ge \chi_d^2(\alpha), \\ 0, & \textit{otherwise}, \end{cases} \tag{8}$$

and any equivalent test, is AUMPI(α, η) *and is unique up to equivalence. Moreover, for each* $h = (h_\vartheta, h_\eta) \in \mathscr{R}^d \times \mathscr{H}_\eta$, $E_{\theta_n(h)}\phi_n$ *converges to the right side of* (7).

COROLLARY 3. *If* T_n *is an efficient test statistic, then* $\phi_{(3)}(T_n) = \mathbf{1}(T_n^T T_n \geq \chi_d^2(\alpha))$ *is* AUMPI(α) *for testing* H *versus* $K_{(3)}$, *that is,* AUMPI(α, η) *for every* η.

Note that $\xi_n(\eta)^T \xi_n(\eta)$ is the quadratic form $S_n^{*T} B^{*-1} S_n^*$, evaluated at $\theta = (\vartheta_0, \eta)$, and $T_n^T T_n$ serves as an estimate thereof. Note also that no asymptotic unbiasedness or similarity constraints, or invariance with respect to any transformations of η or $S_{n\eta}$, have been required although the test (8) does have these properties. This invariance requirement is not unlike Wald's (1943) requirement of constant power, with respect to the parameter of interest, on certain ellipsoids. The test may be shown to be *asymptotically most stringent* and *asymptotically maximin*. Hence our optimality agrees with that of Wald; see Wald (1943) or Hájek and Šidák (1967).

This invariance methodology does not extend to hypotheses about an infinite-dimensional parameter of interest. By considering every finite-dimensional projection of local departures, we can see that the only test that satisfies the invariance requirement is the trivial test $\psi_n = \alpha$ (or its asymptotic equivalents). Asymptotic properties of other nonparametric tests, such as the Kolmogorov–Smirnov test or the Cramér–von Mises test, may be investigated under criteria less restrictive than invariance; see Strasser (1985).

6. Asymptotic confidence sets. The simplest definition of a confidence set for ϑ, with asymptotic confidence coefficient $1 - \alpha$ (fixed throughout this section), is a random set C_n in the range of ϑ for which

$$\liminf P_{n,\theta_n}(\vartheta \in C_n) \geq 1 - \alpha \tag{9}$$

for all $\theta_n = (\vartheta, \eta_n)$ in the parameter space with $\eta_n = \eta + n^{-1/2} h_\eta + n^{-1/2} \delta_{n\eta}$ as before. With this definition, a family of AUMP(α) tests—one test for each null value $\vartheta = \vartheta_0$—may be inverted to achieve *asymptotically uniformly most accurate* (AUMA) confidence sets. The same holds for families of AUMPU(α) and AUMPI(α) tests resulting in confidence intervals or ellipsoids. The reasoning is the same as in the small-sample case, as presented in Lehmann (1986). However, (9) allows the quality of the approximation to vary with the parameter values.

An alternative definition would insert an "infimum over ϑ" [or even over (ϑ, η)] after the "lim inf" in (9). Still, asymptotically optimal confidence sets result from an inversion of asymptotically optimal effective score tests if uniformity with respect to the appropriate parameter is inserted into the basic assumptions. We omit details. Such remarks have been made by others; see, for example, Le Cam and Yang (1990).

7. Constructing efficient test statistics. In rare cases, some optimal canonical tests identified in the previous sections may be free of nuisance parameters [e.g., Examples 1(b) and 4]. However, as a rule, nuisance parameters do appear, and hence an estimate T_n of the standardized effective score is needed.

It is well known that classical tests such as likelihood ratio tests, Wald tests and score (Rao) tests are efficient for parametric problems. In fact, Hall and Mathiason (1990) showed that all of these tests are pointwise equivalent [up to $o_p(1)$] to the optimal canonical test under certain regularities. They also proposed another type of efficient test—the *Neyman–Rao test* [Mathiason (1982)] which is constructed by replacing nuisance parameters in the effective score by $\sqrt{n}$-consistent estimates and consistently estimating the effective information; see also Basawa and Koul (1988). (An estimator $\tilde{\eta}_n$ for η is $\sqrt{n}$-consistent if $\sqrt{n}\|\tilde{\eta}_n - \eta\|$ is bounded in probability.) Hence the Neyman–Rao test can utilize a wide variety of estimates such as moment or quantile estimates (for nuisance parameters), while the likelihood ratio and Rao tests typically involve maximum likelihood estimates (asymptotically efficient estimates, to be exact) which are sometimes difficult to find.

Wald tests can be used for more general problems if we can find *asymptotically efficient* estimates for the parameters of interest—an estimator $\hat{\vartheta}_n$ with $n^{-1/2}(\hat{\vartheta}_n - \vartheta_n)$ asymptotically $N(0, B^{*-1})$ under every $(\vartheta_n(h_\vartheta), \eta_n(h_\eta))$ [Begun, Hall, Huang and Wellner (1983)]. Asymptotic efficiency of estimates is frequently shown by establishing joint asymptotic normality with the scores under the null hypothesis, and then appealing to Le Cam's third lemma to determine asymptotic distributions under local alternatives.

Frequently for infinite-dimensional nuisance parameters, such as densities or hazard functions, $\sqrt{n}$-consistent estimators are not available, and hence neither the likelihood ratio test nor the Rao test can be used. The Neyman–Rao test above is also inappropriate. Bickel (1982) proposed an estimation method for adaptable semiparametric problems, where the estimators adjust themselves (adapt) to available information about the nuisance parameter. Those estimators are shown to be asymptotically efficient and thus can be used for Wald tests resulting in efficient tests. A remarkable aspect to Bickel's method is that estimators which are only consistent are used for nuisance nonparametric components. His construction was generalized and improved by Schick (1986) and Klaassen (1987) to cover also nonadaptable situations. Bickel's original method requires splitting the sample into two unbalanced parts and estimating the score function based on the first smaller subsample which is then discarded. Schick (1986) proposed a modification that splits the sample into (about-equal sized) halves, and uses both subsamples alternatively in estimating the score function and in evaluating the (effective) score. We may extend the result in a straightforward manner to splitting the sample into a finite number of subsamples. Klaassen (1987) shows that the sufficient conditions in Schick (1986) are also necessary.

Borrowing this method, we can modify the Neyman–Rao test so that it results in efficient tests for many iid semiparametric problems. (It can be applied to nonadaptable problems as well as adaptable problems because the effective scores are orthogonal to the scores for the nuisance parameters by definition.) In the case of iid observations $X_1, X_2, \dots,$ the standardized effective score can be expressed as $\xi_n(\eta) = (nI^*(\eta))^{-1/2}\sum s_\eta^*(X_i)$, where s_η^* is the *effective score function* for one observation and $I^*(\eta)$ is the covariance matrix of $s_\eta^*(X_1)$ under the parameter (ϑ_0, η). Of course, $I^*(\eta)$ is the matrix associated with B^*.

PROPOSITION. *Let F_η be the distribution function of a single observation under $\theta_0 = (\vartheta_0, \eta)$. Suppose $\hat{s}_n^*(\cdot) = S_n(\cdot; X_1, \dots, X_n)$ is an estimate of the effective score function s_η^*. If $\hat{s}_n^*$ satisfies*

$$\text{(i)} \qquad n^{1/2}\int \hat{s}_n^*(x)\, dF_\eta(x) \to 0$$

and

$$\text{(ii)} \qquad \int \left\| \hat{s}_n^*(x) - s_\eta^*(x) \right\|^2 dF_\eta(x) \to 0$$

in $P_{n,(\vartheta_0,\eta)}$-probability for every η, then there exists an efficient test statistic T_n.

A possible construction is as follows. Let m denote the integer part of $n/2$ and set $\hat{s}_{n,j}^* = S_{n-m}(X_j; X_{m+1}, \dots, X_n)$ for $j = 1, \dots, m$, and $\hat{s}_{n,j}^* = S_m(X_j; X_1, \dots, X_m)$ for $j = m+1, \dots, n$. Then $T_n = \{\sum \hat{s}_{n,j}^*(\hat{s}_{n,j}^*)^T\}^{-1/2}\sum \hat{s}_{n,j}^*$ works. See Schick (1986) for the proof. Under the additional assumptions of Schick (1987), the sample splitting scheme can be avoided altogether, and we can use $\hat{s}_n^*(X_j)$ instead of $\hat{s}_{n,j}^*$.

Condition (ii) is a consistency requirement. It is easily satisfied if s_n^* is continuous in η since there are many consistent estimates (e.g., kernel estimates) even for densities or hazard functions. For some problems (e.g., Example 2), condition (i) is satisfied naturally for reasonable score function estimates. For others, special considerations are necessary [e.g., use of symmetrized kernels in Example 1(a)], which vary from problem to problem. See Schick (1993, 1994) for explicit constructions in semiparametric regression models.

8. Examples.

EXAMPLE 1(a) (Testing the point of symmetry). Let X_i, $i = 1, \dots, n$, be iid with density $f(\cdot - \vartheta)$, where f is symmetric at 0 and has finite Fisher information I_f. Testing H: $\vartheta = 0$ is one of the classical problems in statistics, dating back to Stein (1956) and Hájek (1962). As in Begun, Hall, Huang and Wellner (1983), we treat the root density $f^{1/2}$ as a nuisance parameter since it is square integrable.

Let $\vartheta_n = n^{-1/2}h_\vartheta$ and consider a sequence of symmetric root densities $f_n^{1/2} = f^{1/2} + n^{-1/2}h_f + n^{-1/2}\delta_{nf}$, where h_f is a symmetric, orthogonal (to $f^{1/2}$) and square integrable function, and $\|\delta_{nf}\| = o(1)$. By simple expansions we get $S_{n\vartheta} = n^{-1/2}\sum s(x_i)$, $s = -f'/f$ and $S_{nf}h_f = 2n^{-1/2}\sum(h_f f^{-1/2})(x_i)$ with $\sigma^2(h) = I_f + 4\|h_f\|^2$. Since f' is antisymmetric, there is zero covariance between the two scores, implying that $S_{n\vartheta}$ is the effective score and the problem is adaptable. The score function s can be estimated by the symmetrized kernel method [Bickel (1982) and Schick (1986, 1987)].

We may then treat the resulting $s_i = \hat{s}_n(x_i)$, $i = 1,\dots,n$, as if it is a normal random sample; $T_n = n^{1/2}\bar{s}_n/\hat{\sigma}_n$, where $\bar{s}_n$ and $\hat{\sigma}_n$ are the sample mean and standard deviation of the s_i's, estimates the standardized effective score, and the test $\phi_n = \mathbf{1}(T_n \geq z_\alpha)$ is thus an adaptive efficient [AUMP(α)] test. The two-sided version is AUMPU(α) and AUMPI(α) (under sign changes of ϑ and $S_{n\vartheta}$).

Inverting two-sided tests of $\vartheta = \vartheta_0$, we find that the set $\{\vartheta\colon n^{1/2}|\bar{s}_n(\vartheta)| \leq z_{\alpha/2}\hat{\sigma}_n(\vartheta)\}$ has asymptotic confidence coefficient $1-\alpha$ and is AUMA unbiased and invariant; here $\bar{s}_n(\vartheta)$ and $\hat{\sigma}_n(\vartheta)$ are as before after subtracting ϑ from all of the x_i's. An equivalent Wald interval is $\hat{\vartheta}_n \pm n^{1/2}z_{\alpha/2}\hat{\sigma}_n$, based on Bickel's estimate of ϑ and any consistent estimate $\hat{\sigma}_n$ (e.g., the one above).

EXAMPLE 1(b) (Testing the median without assuming symmetry). Suppose that, instead of the density being symmetric, it is median-centered at 0. Then we get the same scores as in Example 1(a) but a different directional information [in direction $h = (1, h_f)$], $\sigma^2(h) = 4\|g - h_f\|^2$, where $g = (f^{1/2})'$. Note that $h_f^{(+)}(x) = h_f(x)\mathbf{1}(x \geq 0)$ and $h_f^{(-)}(x) = h_f(x)\mathbf{1}(x < 0)$ are necessarily orthogonal to $f^{1/2}(x)$. Under this constraint, $\|g^{(+)} - h_f^{(+)}\|^2$ is minimized by a simple projection

$$h_f^{0(+)} = g^{(+)} - \frac{\langle g^{(+)}, f^{(+)1/2}\rangle}{\|f^{(+)1/2}\|^2} f^{(+)1/2} = g^{(+)} + f(0)f^{(+)1/2}.$$

Similarly, we get $h_f^{0(-)} = g^{(-)} - f(0)f^{(-)1/2}$, and hence $h_f^0(x) = g(x) + f(0)\operatorname{sign}(x)f^{1/2}(x)$. Thus the effective score is $S_n^* = 2f(0)n^{-1/2}\sum \operatorname{sign}(x_i)$ with effective information $4f^2(0)$ (assumed positive), and the standardized effective score $n^{-1/2}\sum \operatorname{sign}(x_i)$ does not depend on the nuisance parameter.

Hence the standard sign test $\psi_n = \mathbf{1}(n^{-1/2}\sum \operatorname{sign}(x_i) \geq z_\alpha)$ is an efficient one-sided test for H: $\vartheta = 0$. An equivalent Wald test statistic is $2n^{1/2}\hat{f}_n(0)\operatorname{median}(x_i)$ using a consistent estimate $\hat{f}_n(0)$. Loss of information for not knowing f is $4\|h_f^0\|^2$, which becomes 0 when f is the density of a Laplace distribution, as expected. The ARE, relative to an optimal parametric test when f is known, is $1 - 4\|h_f^0\|^2/I_f = 4f^2(0)/I_f$.

EXAMPLE 2 (Testing homogeneity of medians). Consider two independent random samples X_i, $i = 1,\dots,n_X$, and Y_j, $j = 1,\dots,n_Y$, from distributions with densities $f(\cdot - \mu)$ and $g(\cdot - \mu - \vartheta)$, where f and g each have median 0 and finite Fisher information. To test H: $\vartheta = 0$, let $\vartheta_n = n^{-1/2}h_\vartheta$ and $\mu_n =$

$\mu + n^{-1/2}h_\mu$, $n = n_X + n_Y$. Also construct local neighborhoods of f and g as before. Then we get $S_{n\vartheta} = n^{-1/2}\Sigma(-g'/g)(y_j - \mu)$, $S_{n\mu} = n^{-1/2}\{\Sigma(-f'/f) \times(x_i - \mu) + \Sigma(-g'/g)(y_j - \mu)\}$, $S_{nf}h_f = 2n^{-1/2}\Sigma(h_f f^{-1/2})(x_i - \mu)$ and $S_{ng}h_g = 2n^{-1/2}\Sigma(h_g g^{-1/2})(y_j - \mu)$. Assuming $n_X/n \to p$ for some $0 < p < 1$ as n gets large, the information [in direction $(1, h_\mu, h_f, h_g)$] is

$$p\|2h_f - h_\mu f' f^{-1/2}\|^2 + q\|2h_g - (1 + h_\mu)g' g^{-1/2}\|^2, \qquad q = 1 - p.$$

As in Example 1(b), the least favorable directions are $2h_f^0(x) = h_\mu\{f'f^{-1/2}(x) + 2f(0)\text{sign}(x)f^{1/2}(x)\}$ and $2h_g^0(y) = (1 + h_\mu)\{g'g^{-1/2}(y) + 2g(0)\text{sign}(y)g^{1/2}(y)\}$ for any h_μ. The resulting quantity $4pf^2(0)h_\mu^2 + 4qg^2(0)(1 + h_\mu)^2$ is again minimized by $h_\mu^0 = -qg^2(0)[pf^2(0) + qg^2(0)]^{-1}$, and that gives the effective information $4pqf^2(0)g^2(0)/[pf^2(0) + qg^2(0)]$.

The standardized effective score is thus $\xi_n(\eta) = \{npq(pf^2(0) + qg^2(0))\}^{-1/2}\{pf(0)\Sigma\,\text{sign}(y_j - \mu) - qg(0)\Sigma\,\text{sign}(x_i - \mu)\}$. We can construct an efficient score test by estimating $f(0)$, $g(0)$, p and the common median μ. Alternatively, we can use an equivalent Wald test based on the difference between sample medians, and also requiring estimation of $f(0)$ and $g(0)$, or otherwise estimating the variances of sample medians.

EXAMPLE 3 (Linear regression and one-way ANOVA). Let (X_i, Z_i), $i = 1,\dots,n$, be iid where X_i given $Z_i = z_i$ has density $f(\cdot - \vartheta^T z_i)$. The error density f is arbitrary except that it has finite Fisher information I_f. Also assume that the covariate has mean μ and finite positive definite covariance matrix V.

Let $\vartheta_n = \vartheta + n^{-1/2}h_\vartheta$. Let also $f_n^{1/2} = f^{1/2} + n^{-1/2}h_f + n^{-1/2}\delta_{nf}$, where $\|\delta_{nf}\| = o(1)$. The square integrable function h_f is orthogonal to $f^{1/2}$. Then it is easy to see that $S_{n\vartheta} = n^{-1/2}\Sigma z_i s(e_i)$ and $S_{nf}h_f = 2n^{-1/2}\Sigma(h_f f^{-1/2})(e_i)$, where $s = -f'/f$, $e_i = x_i - \vartheta^T z_i$. The directional information is $I_f h_\vartheta^T V h_\vartheta + 4E\|h_f + \frac{1}{2}(h_\vartheta^T\mu)sf^{1/2}\|^2$. Obviously, it is minimized by $h_f^0 = -\frac{1}{2}(h_\vartheta^T\mu)sf^{1/2}$ for a given h_ϑ, resulting in the effective score $S_n^* = n^{-1/2}\Sigma(z_i - \mu)s(e_i)$ with the effective information $I_f V$.

Again construction of an efficient test is possible. Obtain a kernel-based estimate $\hat{s}_n$ of s as in Example 1(a) based on residuals e_i, without symmetrizing. (Use the hypothesized value for ϑ. If ϑ is multidimensional and we are testing the first coordinate, estimate the other coordinates of ϑ using least squares or a robust variation under the null hypothesis.) Then apply standard least squares methods to (s_i, z_i), $i = 1,\dots,n$, where $s_i = \hat{s}_n(e_i)$, to get an efficient test. The test is adaptive for f, compared to the regression problem with known error distribution and arbitrary intercept. Note that there is no need to assume symmetry of the error distribution. See Schick (1987).

The above is also valid for nonrandom covariates as long as $\bar{z}_n = n^{-1}\Sigma z_i \to \mu$, $n^{-1}\Sigma(z_i - \bar{z}_n)(z_i - \bar{z}_n)^T \to V$ and $n^{-1/2}\max_{1\le i\le n}\|z_i\| \to 0$. Linear regression with censoring is treated in Choi (1989).

The problem of comparing locations of k samples (labeled $1,\dots,k$, with sample sizes $n_1,\dots,n_k$, $n = n_1 + \cdots + n_k$) for homogeneity can be handled by considering the distribution of the jth group, $j = 1,\dots,k-1$, as that of the last group shifted by ϑ_j, and thus testing H: $\vartheta_1 = \cdots = \vartheta_{k-1} = 0$ with $\eta = (\vartheta_k, f^{1/2})$. Assuming that $n_j/n \to p_j$, $0 < p_j < 1$, for $j = 1,\dots,k$, an efficient test can be obtained by estimating the score function as previously described and applying the standard one-way analysis of variance method to $s_{ij} = \hat{s}_n(x_{ij})$'s, where x_{ij} is the ith observation in the jth group. The limiting null distribution of the test statistic is the chi-square distribution with $k-1$ degrees of freedom, and the test is AUMP invariant—under linear transformation of the location shifts and the standardized effective scores.

As a parametric homogeneity analog, consider k independent samples from Weibull distributions with common shape β and scale parameters $\exp(\vartheta_j)$, $j = 1,\dots,k$. Now the score s need not be estimated, only the shape β, which, if estimated efficiently, leads to the same kind of analysis of variance as just described, with the same kind of asymptotic optimality. Specifically, the F-test of one-way analysis of variance applied to $y_{ij} = x_{ij}^{\hat{\beta}_n}$, $i = 1,\dots,n_j$, $j = 1,\dots,k$, is asymptotically efficient.

EXAMPLE 4 (Proportional hazards regression with time-dependent covariates and censoring). We observe triples $(t_i, d_i, \{z_i(s),\ 0 \le s \le t_i\})$, $i = 1,\dots,n$, where t_i is survival ($d_i = 1$) or censoring ($d_i = 0$) time of the ith subject with covariate process $z_i(s)$. The hazard function of the subject at time t is assumed to be

$$\lambda\big(t|\{z_i(s), 0 \le s \le t\}\big) = \lambda\big(t|z_i(t)\big) = \lambda(t)\exp\big\{\vartheta^T z_i(t)\big\}$$

for some unknown baseline hazard function λ. To avoid an identifiability problem, let $\sum z_i(0) = 0$, recentering if necessary. The censoring distribution is arbitrary other than being noninformative and conditionally independent of survival times given covariates.

Let $\vartheta_n = \vartheta + n^{-1/2} h_\vartheta$ and $f_n^{1/2} = f^{1/2} + n^{-1/2} h_f + n^{-1/2}\delta_{nf}$, where f is the density associated with λ, h_f is a square integrable and orthogonal (to $f^{1/2}$) function and $\|\delta_{nf}\| = o(1)$. The corresponding hazard function is $\lambda_n = \lambda(1 + n^{-1/2}h_\lambda) + n^{-1/2}\delta_{n\lambda}$ for some $\delta_{n\lambda} = o(1)$, where

$$h_\lambda(t) = 2h_f(t)f^{-1/2}(t) + 2\int_0^t h_f(s) f^{1/2}(s)\,ds \int_t^\infty f(s)\,ds.$$

Hence the contribution of a single observation $(t_i, d_i, \{z_i(s), 0 \le s < t_i\})$ to the log-likelihood ratio of (ϑ_n, λ_n) to (ϑ, λ) is

$$n^{-1/2} d_i h_\vartheta^T z_i(t_i) + d_i \log\left\{\frac{\lambda_n(t_i)}{\lambda(t_i)}\right\} - \int\left[\frac{\lambda_n(s)}{\lambda(s)}\exp\{n^{-1/2}h_\vartheta^T z_i(s)\} - 1\right] Y_i(s)\,d\Lambda_i(s),$$

where $Y_i(s) = \mathbf{1}(s \le t_i)$ and $\Lambda_i(s) = \int_0^s \lambda(s)\exp\{\vartheta^T z_i(s)\}\, ds$. By simple expansions, we obtain

$$S_{n\vartheta} = n^{-1/2} \sum \int z_i(s)\, dW_i(s), \qquad S_{n\lambda} h_\lambda = n^{-1/2} \sum \int h_\lambda(s)\, dW_i(s),$$

and information

$$n^{-1} E\left[\sum \int \{h_\vartheta^T z_i(s) + h_\lambda(s)\}^2 Y_i(s)\, d\Lambda_i(s) \right], \tag{10}$$

where $W_i(s) = N_i(s) - \int_0^s Y_i(u)\, d\Lambda_i(u)$ and $N_i(s) = \mathbf{1}(s \ge t_i)\, d_i$.

Note that the quantity inside the expectation in (10) may be expanded as

$$\begin{aligned} &\int h_\vartheta^T m_2(s) h_\vartheta\, d\Lambda(s) + 2\int h_\vartheta^T m_1(s) h_\lambda(s)\, d\Lambda(s) + \int m_0(s) h_\lambda^2(s)\, d\Lambda(s) \\ &\quad = \int h_\vartheta^T \{m_2(s) - m_1(s) m_1^T(s)/m_0(s)\} h_\vartheta\, d\Lambda(s) \\ &\qquad + \int \{h_\vartheta^T m_1(s)/m_0(s) + h_\lambda(s)\}^2 m_0(s)\, d\Lambda(s), \end{aligned}$$

where $\Lambda(s) = \int_0^s \lambda(u)\, du$ and $m_k(s) = \sum z_i^k(s) Y_i(s) \exp\{\vartheta^T z_i(s)\}$ for $k = 0, 1, 2$, with $z_i^2(s)$ shorthand for $z_i(s) z_i(s)^T$. Hence it is obvious that (10) is minimal when $h_\lambda^0 = -h_\vartheta^T m_1/m_0$, resulting in the effective information

$$n^{-1} E\left[\int \{m_2(s) - m_1(s) m_1^T(s)/m_0(s)\}\, d\Lambda(s) \right] \tag{11}$$

and the effective score

$$\begin{aligned} S_n^* = S_{n\vartheta} - S_{n\lambda}(m_1/m_0) &= n^{-1/2} \sum \int \{z_i(s) - m_1(s)/m_0(s)\}\, dW_i(s) \\ &= n^{-1/2} \sum \int \{z_i(s) - m_1(s)/m_0(s)\}\, dN_i(s) \\ &= n^{-1/2} \sum \{z_i(t_i) - m_1(t_i)/m_0(t_i)\}, \end{aligned}$$

which is commonly known as the Cox partial score [Cox (1972, 1975)]. A natural estimator for the effective information is obtained by looking at the martingale $\sum W_i(s)$, equating its increments to 0 and thus replacing $d\Lambda(s)$ in (11) by $m_0^{-1}(s)\, d\{\sum N_i(s)\}$, which results in the Cox partial information $n^{-1} \sum d_i [m_2(t_i)/m_0(t_i) - m_1(t_i) m_1^T(t_i)/m_0^2(t_i)]$.

Therefore, tests based on the Cox partial score and information are AUMP; likewise, confidence intervals and sets are AUMA. Checking the least favorable direction, we also notice that tests are adaptive for baseline hazards when (and only when) $m_1(t) = 0$ for all t; a simple sufficient case is that $\vartheta = 0$ and the censoring time does not depend on the covariate (or no censoring at all). In general, the ARE of the Cox test compared to an optimal parametric test with known baseline is the proportion of the effective information (11) in the full information $n^{-1} E[\int m_2(s)\, d\Lambda(s)]$. These results formalize those of Efron (1977) and Oakes (1977).

APPENDIX

Suppose the assumptions introduced in Section 2 hold. Fix $h_\eta^0 \in \mathscr{H}_\eta$ and set $\mathscr{G} = \{(h_\vartheta, h_\eta^0 - B_{22}^{-1}B_{21}h_\vartheta)\colon h_\vartheta \in \mathscr{H}_\vartheta\}$. Note that the power of a test ψ_n for $h \in \mathscr{G}$ under LAN can be expressed as [from (2)]

$$E_{\theta_n(h)}\psi_n = E_{\theta_0}\psi_n \exp\left\{S_n^* h_\vartheta + S_{n\eta}h_\eta^0 - \tfrac{1}{2}\sigma^2(h) + r_n(h)\right\} + o(1).$$

We develop a limiting expression for the right side under the fixed null value θ_0, and recognize it to be the expectation of a limiting test function times a limiting likelihood ratio. Consequently, it can be expressed simply as the expectation of a test function under alternative hypothesis conditions. A final step integrates out the limiting version of $S_{n\eta}h_\eta^0$, which has a parameter-free distribution when confined to the specific hyperplane. We thereby achieve an asymptotic power representation on the hyperplane as the power of a test based on a limiting version Z of the standardized effective score, that is, $E\varphi(Z + B^{*1/2}h_\vartheta)$ for some test φ and some standard normal random vector Z.

LEMMA 1. *For every test ψ_n and every subsequence n', there exists a subsequence n'' of n' and a test function φ from $\mathscr{R}^d$ to* $[0,1]$ *such that*

$$\lim E_{\theta_{n''}(h)}\psi_{n''} = \int \varphi(z)\, d\Phi_d\left(z - B^{*1/2}h_\vartheta\right) \tag{12}$$

for every $h = (h_\vartheta, h_\eta^0 - B_{22}^{-1}B_{21}h_\vartheta)$ in $\mathscr{G}$, where Φ_d denotes the d-dimensional standard normal distribution.

PROOF. Since $Y_n = (\psi_n, S_n^*, S_{n\eta}h_\eta^0)$ are tight under P_{n,θ_0}, we can choose a subsequence n'' so that $Y_{n''}$ converges in distribution to some (ϕ, S^*, W) under P_{n'',θ_0}, where S^* and W are independent, S^* is normal with mean 0 and variance B^* and W is normal with mean 0 and variance $\sigma_2^2 = \langle h_\eta^0, B_{22}h_\eta^0\rangle$. Therefore, for each $h \in \mathscr{G}$,

$$\begin{aligned} E_{\theta_{n''}(h)}\psi_{n''} &= E_{\theta_0}\psi_{n''}\exp\left\{L_{n''}(h)\right\} + o(1) \\ &\to E\phi\exp\left\{S^{*T}h_\vartheta + W - \tfrac{1}{2}\sigma^2(h)\right\}, \end{aligned} \tag{13}$$

since $\psi_n \exp\{L_n(h)\}$ is uniformly integrable [see (P1)].

Replace ϕ in (13) by a function $\psi(S^*, W)$ by taking conditional expectation given (S^*, W). Also note that $\sigma^2(h) = h_\vartheta^T B^* h_\vartheta + \sigma_2^2$. Now the exponential factor in the limit expression in (13) may be recognized as the likelihood ratio of multinormal distributions with means $\mu = (B^*h_\vartheta, \sigma_2^2)$ and 0 and common variance V. Hence this limit expression is the same as $E\psi(S^* + B^*h_\vartheta, W + \sigma_2^2) = E\psi_0(S^* + B^*h_\vartheta)$, say, after integrating out W (which is independent of S^* and has an h_ϑ-free distribution). The final expression in (12) results from the standardization $Z = B^{*-1/2}S^*$. $\square$

We expect limiting test functions to be indicators of the limiting version Z of the standardized effective score $\xi_{n''}(\eta)$ being in an appropriate "rejection region" C. We now show that this occurs if and only if $\psi_{n''}$ is asymptotically equivalent to $\mathbf{1}(\xi_{n''}(\eta) \in C)$.

LEMMA 2. *Let φ and n'' be as in Lemma* 1, *and let C be a measurable subset of $\mathscr{R}^d$. Then $\varphi(z) = \mathbf{1}(z \in C)$ almost everywhere z if and only if $\psi_{n''} - \mathbf{1}(\xi_{n''}(\eta) \in C) \to 0$ in $P_{n'',(\vartheta_0,\eta)}$-probability.*

PROOF. The limiting variable ϕ in the previous proof can be represented as $\phi = t(S^*, W, U)$, where U is uniformly distributed on $(0, 1)$ and is independent of the pair (S^*, W) and where t is a measurable function from $\mathscr{R}^d \times \mathscr{R} \times (0, 1)$ to $[0, 1]$. Indeed, if $F(\cdot|S^*, W)$ denotes the conditional distribution function of the random variable ϕ given S^* and W, then (ϕ, S^*, W) has the same distribution as $(F^{-1}(U|S^*, W), S^*, W)$, where $F^{-1}(u|S^*, W) = \inf\{v: F(v|S^*, W) \geq u\}$ for $0 < u < 1$; thus we can take $\phi = t(S^*, W, U) = F^{-1}(U|S^*, W)$. In this case, the limiting test φ becomes $\varphi(z) = E(t(B^{*1/2}z, W + \sigma_2^2, U))$.

Since t is $[0, 1]$-valued, $\varphi(z) = 1$ $(= 0$, respectively) if and only if $t(B^{*1/2}z, W + \sigma_2^2, U) = 1$ $(= 0$, respectively) almost surely. Thus $\varphi(z) = \mathbf{1}(z \in C)$ almost everywhere z if and only if $t(B^{*1/2}z, w, u) = \mathbf{1}(z \in C)$ for almost all $(z, w, u) \in \mathscr{R}^d \times \mathscr{R} \times (0, 1)$. As $\psi_{n''} - \mathbf{1}(\xi_{n''}(\eta) \in C)$ converges in distribution to $t(B^{*1/2}Z, W, U) - \mathbf{1}(Z \in C)$, the latter is equivalent to $\psi_{n''} - \mathbf{1}(\xi_{n''}(\eta) \in C) \to 0$ in $P_{n'',(\vartheta_0,\eta)}$-probability. $\square$

Acknowledgments. This paper is based in part on the Ph.D. dissertation of the first author, supervised by the second author, with later additional collaborative work. Many improvements were brought about through the efforts of the third author, who served as a referee of earlier versions of the manuscript. We thank the Editors for facilitating the final collaboration.

REFERENCES

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

BASAWA, I. V. and KOUL, H. L. (1988). Large-sample statistics based on quadratic dispersion. *Internat. Statist. Rev.* **56** 199–219.

BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

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

CHOI, S. (1989). On asymptotically optimal tests. Ph.D. dissertation, Dept. Statistics, Univ. Rochester.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–279.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

FABIAN, V. and HANNAN, J. (1985). *Introduction to Probability and Mathematical Statistics*. Wiley, New York.

FABIAN, V. and HANNAN, J. (1987). Local asymptotic behavior of densities. *Statist. Decisions* **5** 105–138. [Correction (1988) **6** 195.]

HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

HÁJEK, J. and ŠIDÁK, J. (1967). *Theory of Rank Tests*. Academic Press, New York.

HALL, W. J. and MATHIASON, D. (1990). On large-sample estimation and testing in parametric models. *Internat. Statist. Rev.* **58** 77–97.

HUANG, W.-M. (1982). Parametric estimation when there are nuisance functions. Ph.D. dissertation, Dept. Statistics, Univ. Rochester.

KLAASSEN, C. A. S. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** 1548–1562.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 37–98.

LE CAM, L. (1969). *Théorie Asymptotique de la Décision Statistique*. Univ. Montréal Press.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.

MATHIASON, D. (1982). Large sample test procedures in the presence of nuisance parameters. Ph.D. dissertation, Dept. Statistics, Univ. Rochester.

NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics* (*Harald Cramér Volume*) (U. Grenander, ed.) 212–234. Wiley, New York.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64** 441–448.

ROUSSAS, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics*. Cambridge Univ. Press.

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151.

SCHICK, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16** 89–105. [Correction (1989) **22** 269–270.]

SCHICK, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486–1521. [Correction and addendum (1995) **23** 1862–1863.]

SCHICK, A. (1994). Efficient estimation in regression models with unknown scale functions. *Math. Methods Statist.* **3** 171–212.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187–195. Univ. California Press, Berkeley.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

WALD, A. (1941). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Statist.* **12** 1–19.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

WEFELMEYER, W. (1987). Testing hypotheses on independent, not identically distributed models. In *Mathematical Statistics and Probability Theory* (M. L. Puri et al., eds.) **A** 267–282. Reidel, Berlin.

S. CHOI
DEPARTMENT OF EPIDEMIOLOGY
AND BIOSTATISTICS
MCGILL UNIVERSITY
1020 PINE AVENUE WEST
MONTRÉAL, QUÉBEC
CANADA H3A 1A2

W. J. HALL
DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NEW YORK 14627

A. SCHICK
DEPARTMENT OF MATHEMATICAL SCIENCES
BINGHAMTON UNIVERSITY
BINGHAMTON, NEW YORK 13902-6000