The Annals of Statistics
1996, Vol. 24, No. 1, 336–352

A ROBUST ADJUSTMENT OF THE PROFILE LIKELIHOOD[1]

BY JAMES E. STAFFORD

University of Western Ontario

Under mild misspecifications of model assumptions, maximum likelihood estimates often remain consistent and asymptotically normal. Asymptotic normality will often hold for the signed root of the likelihood ratio statistic and the score statistic as well. However, standard estimates of asymptotic variance are usually inconsistent. This occurs when Bartlett's second identity fails. In the manner of McCullagh and Tibshirani, a variance correction may be used to adjust the profile likelihood so this identity obtains. The resulting likelihood yields the robust versions of the signed root, Wald and score statistic suggested by Kent and Royall.

Assuming model correctness, asymptotic expansions for the first three cumulants of each robust statistic are derived. It is seen that bias and skewness are not severely affected by using a robust statistic. An invariant expression derived for the asymptotic relative efficiency of a robust method allows assessment in numerous examples considered. Even for moderately large sample sizes, losses in efficiency are significant, making the misuse of a robust variance estimate potentially costly. Computer algebra is used in many of the calculations reported in this paper.

1. Introduction. Consider the situation where data $x_1, \dots, x_n$ are assumed to be independent realizations of a distribution that belongs to the parametric family $\{f_\theta(X),\ \theta \in \Theta\}$, where $f_\theta(X)$ is the joint density for the data and θ may be partitioned into a scalar parameter of interest ψ and a nuisance parameter λ. Inference for ψ_0, the true value of ψ, may be based on the signed square root of the likelihood ratio statistic $r(\psi)$, the score function $u(\psi)$ or the maximum likelihood estimate $\hat{\psi}$. Standard asymptotic results show

$$\begin{aligned} r(\psi_0) &\to N(0, v_r), \\ u(\psi_0)/\sqrt{n} &\to N(0, v_u), \\ \sqrt{n}\left(\hat{\psi} - \psi_0\right) &\to N(0, v_{\hat{\psi}}). \end{aligned} \tag{1}$$

Typically, $v_r = 1$ and expressions for $v_u, v_{\hat{\psi}}$ simplify so they may be consistently estimated by quantities that depend solely on the observed information. However, under a mild model misspecification, where v_r, v_u and $v_{\hat{\psi}}$ do

Received November 1994; revised August 1995.
[1]Research supported by the Natural Sciences and Engineering Research Council of Canada.
AMS 1991 *subject classifications*. 62F35, 62F05
Key words and phrases. Asymptotic relative efficiency, Bartlett identities, computer algebra, cumulants, profile likelihood, robustness.

not simplify, estimates of variance based on the observed information will be inconsistent.

Under model misspecification, it is the failure of Bartlett's second identity, which equates the variance of the score and the expected Fisher information, that does not allow simplification in expressions for asymptotic variance. A robust adjustment of the profile log-likelihood to correct this identity is motivated in the manner of McCullagh and Tibshirani (1990) and leads to a simple adjustment involving $\hat{v}_r$, a sample version of v_r. The resulting robust profile likelihood is invariant under a broad class of reparameterizations and has observed information that provides the usual robust estimates of variance given in Kent (1982). This is discussed in detail in Section 2.

The main results of this paper may be found in Section 3. Here, we compare asymptotic expansions of the first three cumulants of the signed root, score and maximum likelihood estimate, standardized by model-based and model-robust variance estimates. An implicit assumption of the paper is the preference of robust methods when a model misspecification occurs. This is sensible given that robust methods remain asymptotically correct while model-based methods typically do not. Therefore cumulant comparisons are made assuming model correctness. This allows us to assess the effect of using a robust variance estimate when it is unnecessary and provides insight into the importance of model assumptions to improve the accuracy of inference. If comparisons were favourable, then robust variance estimates could be used at all times.

Expressions for bias and skewness are found to agree to order $n^{-3/2}$, and hence, robust variance estimate has a limited effect on these cumulants. Model-robust estimates of variance tend to be more variable than their model-based counterparts. For example, the model-based estimate of variance for the signed root is simply 1. One would then expect this to result in a more variable test statistic that may fall into a rejection region more often. A measure of asymptotic relative efficiency Υ, that may be used to make variance comparisons, is derived. Numerous examples are considered in Section 4, where computer algebra methods are used to evaluate Υ. For these examples, robust methods appear to be much less efficient than their model-based competitors, as one might suspect.

Throughout this paper, we assume broad conditions like those found in Huber (1967) or White (1982) that ensure the consistency of the maximum likelihood estimate and the asymptotic normality of $r(\psi)$, $u(\psi)$ and $\hat{\psi}$. Under a model misspecification, we denote the limit of $\hat{\theta}$ by θ^* with components ψ^* and λ^*, and under model correctness, by θ_0 with components ψ_0 and λ_0. In some special cases, ψ^* and ψ_0 will coincide [Gould and Lawless (1988); Lin and Wei (1989)]. However, in general this will not be the case, and one must assume that under a model misspecification ψ^* still has a meaningful scientific interpretation since this is the only quantity for which we may conduct inference.

The remainder of this introduction is devoted to the development of notation. Let $l(\theta)$ denote the usual log-likelihood function with components

$l_i(\theta)$, $i = 1, \dots, n$. We will occasionally use the alternative representation $l(\psi, \lambda)$ for the log-likelihood and, similarly, $l_i(\psi, \lambda)$ for a component. Let $\hat{\lambda}_\psi$ be the constrained maximum likelihood estimate for λ holding ψ fixed. When λ is replaced by $\hat{\lambda}_\psi$ in the log-likelihood, the result is called the profile log-likelihood $l_p(\psi)$. Letting $\partial_\psi = \partial/(\partial\psi)$, the signed square root of the likelihood ratio statistic and the score function are defined as

$$r(\psi) = \operatorname{sgn}(\hat{\psi} - \psi)\left[2\left\{l_p(\hat{\psi}) - l_p(\psi)\right\}\right]^{1/2},$$
$$u(\psi) = \dot{l}_p(\psi) = \partial_\psi l_p(\psi).$$

Let $\hat{\psi}$ and $\hat{\lambda}$ be the ψ and λ components of the global maximum likelihood estimate $\hat{\theta}$. The parameter θ has length $p + 1$; a component of θ is denoted with the use of a subscript θ_r. Derivatives of $l(\theta)$ or $l_i(\theta)$ are also indicated by the use of subscripts. Let $\partial_r = \partial/(\partial\theta_r)$,

$$l_{rst} = \partial_r \partial_s \partial_t l(\theta), \qquad l_{i;rs} = \partial_r \partial_s l_i(\theta).$$

Expected values of sums of products of derivatives of the components of the log-likelihood function are called expected information quantities,

$$I_{rst} = n^{-1} E[l_{rst}], \qquad I_{r,st} = n^{-1} E\left[\sum_i l_{i;r} l_{i;st}\right].$$

Sums that are centered and scaled so they are bounded in probability are denoted as

$$z_{rst} = n^{-1/2} l_{rst} - n^{1/2} I_{rst}, \qquad z_{r,s} = n^{-1/2} \sum_{i=1}^{n} l_{i;r} l_{i;s} - n^{1/2} I_{r,s}.$$

These are typically referred to as $O_p(1)$ random variables. Occasionally, in our notation we require explicit dependence of expressions on ψ and λ and, hence, we subscript terms by these parameters. The nuisance parameter is assumed to be a vector and hence it is subscripted as well to denote a component of that vector:

$$l_{i;\psi\lambda_r\lambda_s} = \partial_\psi \partial_{\lambda_r} \partial_{\lambda_s} l_i(\theta),$$
$$I_{\lambda_r, \lambda_s\psi} = E\left[l_{i;\lambda_r} l_{i;\lambda_s\psi}\right],$$
$$z_{\lambda_r, \psi} = n^{-1/2} \sum_{i=1}^{n} l_{i;\lambda_r} l_{i;\psi} - n^{1/2} I_{\lambda_r, \psi}.$$

The notation I^{rs} and $I^{\lambda_r\lambda_s}$ is used to denote the matrix inverses $-I_{rs}^{-1}$ and $-I_{\lambda_r\lambda_s}^{-1}$, respectively. The (ψ, ψ) and (ψ, λ_r) components of I^{rs} are denoted $I^{\psi\psi}$ and $I^{\psi\lambda_r}$. On occasion, we will estimate an expected information quantity by an observed quantity. For this we use the notation

$$J_{\psi\lambda_r} = n^{-1} l_{\psi\lambda_r}(\hat{\theta}), \qquad J_{\lambda_r, \lambda_s\psi} = n^{-1} \sum_i l_{i;\lambda_r}(\hat{\theta}) l_{i;\lambda_s\psi}(\hat{\theta}), \qquad J^{\lambda_r\lambda_s} = -J_{\lambda_r\lambda_s}^{-1}.$$

Finally, since we are in a multiparameter setting we use the summation convention to represent inner products, where a subscripted index repeated as a superscript represents an implicit sum over that index. For example,

$$I^{rs}z_s = \sum_{s=1}^{p} I^{rs}z_s.$$

2. Bartlett's second identity and estimates of variance. The Bartlett identities result from the density being normed,

$$\int f_\theta(x)\,dx = 1.$$

Successive differentiation of this integral with respect to the components of θ yields identities. Let $f_\theta^r(x) = \partial_r f_\theta(x)$ and $f_\theta^{rs}(x) = \partial_r \partial_s f_\theta(x)$. Then

$$I_r = \int l_r f_\theta(x)\,dx = \int \frac{f_\theta^r(x)}{f_\theta(x)} f_\theta(x)\,dx = \int f_\theta^r(x)\,dx = \partial_r \int f_\theta(x)\,dx = 0$$

and similarly,

$$\begin{aligned}
-I_{rs} &= -\int l_{rs} f_\theta(x)\,dx = -\int \left\{ \partial_s \left[\frac{f_\theta^r(x)}{f_\theta(x)} \right] \right\} f_\theta(x)\,dx \\
&= \int \frac{f_\theta^r(x)}{f_\theta(x)} \frac{f_\theta^s(x)}{f_\theta(x)} f_\theta(x)\,dx - \int \frac{f_\theta^{rs}(x)}{f_\theta(x)} f_\theta(x)\,dx \\
&= \int l_r l_s f_\theta(x)\,dx - \partial_r \partial_s \int f_\theta(x)\,dx = I_{r,s}.
\end{aligned}$$

Now consider a model misspecification where the true density for the data $g(x)$ does not belong to the family $\{f_\theta(X),\ \theta \in \Theta\}$. The cancellations necessary to derive the above identities will not occur since

$$\frac{f_\theta^r(x)}{f_\theta(x)} g(x) \neq f_\theta^r(x), \qquad \frac{f_\theta^{rs}(x)}{f_\theta(x)} g(x) \neq f_\theta^{rs}(x),$$

and hence the identities cannot be obtained.

The maximum likelihood estimate $\hat{\theta}$ will converge to θ^*, the value of the parameter that minimizes the Kullback–Leibler distance between $f_\theta(x)$ and $g(x)$:

$$\begin{aligned}
d(f_\theta, g) &= \int \log\left\{ \frac{g(x)}{f_\theta(x)} \right\} g(x)\,dx \\
&= \int \log\{g(x)\} g(x)\,dx - \int \log\{f_\theta(x)\} g(x)\,dx.
\end{aligned}$$

Since $d(f_\theta, g)$ is minimized at θ^* it is the case that

$$\partial_r d(f_\theta, g)\big|_{\theta=\theta^*} = \int l_r g(x)\,dx\big|_{\theta=\theta^*} = 0$$

and so the expected value of the score is zero at θ^*, even if the model is misspecified. However, only in very special cases will the second identity hold.

That is, we usually have

$$-\int l_{rs} g(x)\, dx \neq \int l_r l_s g(x)\, dx, \qquad \forall \theta \in \Theta.$$

The failure of Bartlett's second identity has implications for estimates of asymptotic variance. Standard asymptotic calculations like those given in Kent (1982) show

$$v_r = \left(I^{\psi r} I_{r,s} I^{\psi s}\right)/I^{\psi\psi}, \qquad v_u = I^{\psi r} I_{r,s} I^{\psi s}/\left(I^{\psi\psi}\right)^2, \qquad v_{\hat{\psi}} = I^{\psi r} I_{r,s} I^{\psi s},$$

which simplify to $v_r = 1$, $v_u = 1/I^{\psi\psi}$ and $v_{\hat{\psi}} = I^{\psi\psi}$ when $I_{r,s} = -I_{rs}$. This allows variance estimates to be based on the observed information alone. However, when a model is misspecified and such simplifications cannot occur, then such estimates will be inconsistent. (Note that under model correctness, asymptotic variance may also be consistently estimated by the expected Fisher information evaluated at $\hat{\theta}$. Efron and Hinkley (1978) show, however, that the observed information is preferable.)

In developing robust methods that address model misspecifications where the second Bartlett identity fails, it is natural to consider correcting the identity. McCullagh and Tibshirani (1990) adjusted the profile likelihood so the first two Bartlett identities obtain to higher order. Their adjustment was motivated by an effort to reduce the effect of nuisance parameter estimation when the model is correctly specified and belongs to a class of adjustments for this purpose [Barndorf-Neilsen (1983); Cox and Reid (1987, 1993)] which have been extensively studied by DiCiccio and Stern (1994a, b). They begin by showing that under model correctness,

$$E\left[n^{-1} u(\psi)\right] = O(n^{-1}), \qquad E\left[n^{-1}\{u(\psi)\}^2\right] - E\left[n^{-1}\dot{u}(\psi)\right] = O(n^{-1}).$$

That is, these identities hold asymptotically, but are only approximate for a finite sample size. To improve these approximations, a bias and variance correction of the profile score is developed by solving for $e(\psi)$ and $v(\psi)$ such that

$$E_{\hat{\lambda}_\psi}\left[u_a(\psi)\right] = 0, \qquad \mathrm{Var}_{\hat{\lambda}_\psi}\left[u_a(\psi)\right] = -E_{\hat{\lambda}_\psi}\left[\partial_\psi u_a(\psi)\right],$$

where $u_a(\psi) = v(\psi)\{u(\psi) - e(\psi)\}$. The adjusted profile log-likelihood is then $l_a(\psi) = \int^\psi u_a(t)\, dt$. The solutions are simply

$$e(\psi) = E_{\hat{\lambda}_\psi}\left[u(\psi)\right], \qquad v(\psi) = \frac{-E_{\hat{\lambda}_\psi}\left[\partial_\psi\{u(\psi) - e(\psi)\}\right]}{\mathrm{Var}_{\hat{\lambda}_\psi}\left[u(\psi)\right]}.$$

Now rather than adjusting the profile likelihood for the effect of nuisance parameter estimation, we consider the use of the McCullagh and Tibshirani (1990) adjustment for robust purposes. In the case of a model misspecification, only the second Bartlett identity fails and so we may let $e(\psi) = 0$. Given the true model is unknown, we cannot compute the expected values required to implement $v(\psi)$ and so replace it by the asymptotically equivalent quantity $\hat{v}(\psi) = i(\psi)/j(\psi)$, where $i(\psi) = -\partial_\psi u(\psi)$ and $j(\psi) = \sum_{i=1}^n \{u_i(\psi)\}^2$. The adjusted profile likelihood then becomes

$$\int^{\psi} \hat{v}(t) u(t)\, dt.$$

To elucidate implementation, integration can be avoided by replacing $\hat{v}(\psi)$ by the asymptotically equivalent quantity $\hat{v}(\hat{\psi})$ without affecting the desired property that the second Bartlett identity obtains. The result is the robust profile likelihood $l_r(\psi) = \hat{v}(\hat{\psi}) l_p(\psi)$ which has the following properties, which are demonstrated below:

1. The second Bartlett identity obtains asymptotically.
2. It is invariant under transformations of the form $\{\psi, \lambda\} \to \{\tau(\psi), \omega(\psi, \lambda)\}$.
3. Just as the standard statistics result from $l_p(\psi)$, the signed root, score and Wald statistics based on $l_r(\psi)$ are simply the robust standardizations of $r(\psi)$, $u(\psi)$ and $\hat{\psi}$.
4. If there is no nuisance parameter present, then $l_r(\psi) = \hat{v}(\hat{\psi}) l(\psi)$, where $i(\psi) = -\ddot{l}(\psi)$ and $j(\psi) = \sum_{i=1}^n \{\dot{l}_i(\psi)\}^2$.

The first property results from the construction of $l_r(\psi)$ and can be verified by differentiating $l_r(\psi)$ and noting the result. In particular, note $E[n^{-1}\dot{l}_r(\psi)] = O(n^{-1})$, and $n^{-1}\{\dot{l}_r(\psi)\}^2$ and $n^{-1}\ddot{l}_r(\psi)$ differ by $O(n^{-1/2})$. The fourth property is entirely straightforward.

Invoking standard calculations, we can show $i(\hat{\psi}) = (J^{\psi\psi})^{-1}$ and similarly $j(\hat{\psi}) = \{J^{\psi r} J_{r,s} J^{\psi s}/(J^{\psi\psi})^2\}$. Hence we have

$$\hat{v}(\hat{\psi}) = \left\{ \frac{J^{\psi r} J_{r,s} J^{\psi s}}{J^{\psi\psi}} \right\}^{-1} = \frac{1}{\hat{v}_r},$$

where $\hat{v}_r$ is simply a sample version of v_r. So the signed root, standard score and Wald statistics based on $l_r(\psi)$ are simply $r(\psi)/\hat{v}_r^{1/2}$, $u(\psi)/(\sqrt{n}\, \hat{v}_u^{1/2})$ and $\sqrt{n}(\hat{\psi} - \psi)/\hat{v}_{\hat{\psi}}^{1/2}$, where

$$\hat{v}_u = J^{\psi r} J_{r,s} J^{\psi s}/(J^{\psi\psi})^2, \qquad \hat{v}_{\hat{\psi}} = J^{\psi r} J_{r,s} J^{\psi s}.$$

That is, they are simply the robust statistics suggested by Kent (1982) based on the asymptotic results (1). The Wald statistic is standardized by $\hat{v}_{\hat{\psi}}$, the well known "sandwich" estimator that occurs naturally in the theory of M-estimation and was evaluated for many parametric settings in Royall (1986). Barndorff-Nielsen and Sorensen (1994) generalize its use to martingale theory.

To establish the invariance properties of $l_r(\psi)$, consider a reparameterization to η. The log-likelihood for η is $l\{\theta(\eta)\}$ and similarly the log-likelihood for θ is $l\{\eta(\theta)\}$. Note $\hat{\eta} = \eta(\hat{\theta})$ and let $\hat{\theta}_r^i = \{\theta_r^i\}|_{\eta=\hat{\eta}} = \{\partial\theta_i/\partial\eta_r\}|_{\eta=\hat{\eta}}$ with inverse $\hat{\eta}_i^r = \{\eta_i^r\}|_{\theta=\hat{\theta}} = \{\partial\eta_r/\partial\theta_i\}|_{\theta=\hat{\theta}}$. From this we have

$$J_{\eta_r\eta_s} = \hat{\theta}_r^i J_{ij} \hat{\theta}_s^j, \qquad J_{\eta_r,\eta_s} = \hat{\theta}_r^i J_{i,j} \hat{\theta}_s^j, \qquad J^{\eta_r\eta_s} = \hat{\eta}_i^r J^{ij} \hat{\eta}_j^s.$$

Assuming η_1 is the parameter of interest, then we have under reparameterization

$$\hat{v}(\hat{\eta}_1) = \left\{ \frac{J^{\eta_1 \eta_i} J_{\eta_i, \eta_j} J^{\eta_j \eta_1}}{J^{\eta_1 \eta_1}} \right\}^{-1} = \frac{\hat{\eta}_i^1 J^{ij} J_{j,k} J^{kl} \hat{\eta}_l^1}{\hat{\eta}_i^1 J^{ij} \hat{\eta}_j^1},$$

which equals $\hat{v}(\hat{\psi})$ if η_i^r is block diagonal; that is, if $\eta_i^1 = 0$, $i > 1$. An important class of transformations under which $l_r(\psi)$ is invariant are those that preserve ψ and hence its interpretability. In the context of the adjusted profile likelihood of Cox and Reid (1987), transformations of this type, that also achieve parameter orthogonality, reduce the effect of nuisance parameter estimation on inference for ψ. In the present context, the effect of nuisance parameter estimation is not a concern, but Kent (1982) does show that parameter orthogonality under the true model simplifies the robust variance estimates. However, in the case of a model misspecification, parameter orthogonality typically cannot be obtained because the true model is unknown.

3. General cumulant expressions. For each of the signed root, score and Wald statistics, we have the choice between a model-robust standardization s_r versus a simpler model-based version s_m. The cautious user may prefer to use s_r more often than is actually necessary (in M-estimation s_r is always used). In this section, we study the effect of redundant use of s_r by comparing, under model correctness, asymptotic expansions for the first three cumulants of each statistic standardized by s_m and then by s_r. The explicit details of how the expansions are derived are lengthy and therefore deferred to the Appendix. An outline is provided here.

Differences in the expansions are due to differences in the expansions for s_r and s_m. In each case we will have

$$s_m = s_0 + n^{-1/2} s_{11} + n^{-1} s_{21} + O_p(n^{-3/2}),$$

$$s_r = s_0 + n^{-1/2}(s_{11} + s_{12}) + n^{-1}(s_{21} + s_{22}) + O_p(n^{-3/2}),$$

where $s_{11}, s_{12}, \dots$ represent the coefficients of the powers of n. The expansion of s_r shares all the terms in the expansion of s_m. Hence, the additional terms s_{12} and s_{22} in the expansion of s_r are of interest since they will effect all differences in all subsequent calculations. Allowing T to be either $r(\psi)$, $u(\psi)$ or $\hat{\psi} - \psi$, T has an expansion of the form

$$T = T_0 + n^{-1/2} T_1 + n^{-1} T_2 + O_p(n^{-3/2}).$$

Letting $T_r = Ts_r$ and $T_m = Ts_m$ we then have

$$\begin{aligned} T_m &= T_0 s_0 + n^{-1/2}(T_1 s_0 + T_0 s_{11}) \\ &\quad + n^{-1}(T_0 s_{21} + T_2 s_0 + T_1 s_{11}) + O_p(n^{-3/2}), \\ T_r &= T_0 s_0 + n^{-1/2}(T_1 s_0 + T_0 s_{11} + T_0 s_{12}) \\ &\quad + n^{-1}(T_0 s_{21} + T_0 s_{22} + T_2 s_0 + T_1 s_{11} + T_1 s_{12}) + O_p(n^{-3/2}). \end{aligned}$$

Letting $d = T_r - T_m$, then

$$d = n^{-1/2} s_{12} T_0 + n^{-1}(T_0 s_{22} + T_1 s_{12}) + O_p(n^{-3/2}).$$

Calculations in the Appendix show $E[T_0 s_{12}] = 0$ for each pair of competing statistics. Also, $T_0 s_{22}, T_1 s_{12}$ involve odd powers of $O_p(1)$ random variables that will have expected values with order $n^{-1/2}$. Hence,

$$\delta_1 = E[d] = O(n^{-3/2}).$$

The case of the skewness is almost as simple. For any random variable X, we may write the skewness ρ_X in terms of its moments,

$$\rho_X = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3, \qquad \mu_i = E[X^i].$$

For T_r, T_m we have

$$\mu_1^3 = O(n^{-3/2}), \qquad \mu_2 = 1 + O(n^{-1}), \qquad \mu_3 = O(n^{-1/2}),$$

$$E\left[T_m^3\right] = E\left[T_0^3 s_0^3\right] + \frac{3}{\sqrt{n}}\left(T_0^2 T_1 s_0^3 + T_0^3 s_0^2 s_{11}\right) + O(n^{-3/2}),$$

$$E\left[T_r^3\right] = E\left[T_0^3 s_0^3\right] + \frac{3}{\sqrt{n}} E\left[T_0^2 T_1 s_0^3 + T_0^3 s_0^2 s_{11} + T_0^3 s_0^2 s_{12}\right] + O(n^{-3/2}),$$

and so

$$\begin{aligned}
\delta_3 &= \rho_{T_m} - \rho_{T_r} \\
&= \frac{3}{\sqrt{n}}\left\{E\left[T_0^3 s_0^2 s_{12}\right] - \left(E\left[T_m^2\right]E[T_m] - E\left[T_r^2\right]E[T_r]\right)\right\} + O(n^{-3/2}) \\
&= \frac{3}{\sqrt{n}}\left\{E\left[T_0^3 s_0^2 s_{12}\right] - E[d]\right\} + O(n^{-3/2}) = \frac{3}{\sqrt{n}} E\left[T_0^3 s_0^2 s_{12}\right] + O(n^{-3/2}).
\end{aligned}$$

Calculations in the Appendix show $E[T_0^3 s_0^2 s_{12}] = 0$ for each pair of competing statistics and hence $\delta_3 = O(n^{-3/2})$. One important consequence of this is that the signed root retains the property that its skewness is $O(n^{-3/2})$ under robust standardization [DiCiccio (1984)].

For two competing statistics with similar bias and skewness properties, a comparison of variances will provide insight into relative behaviour. In particular, the statistics with larger variance will typically fall into a common rejection region more often. One cannot expect this to always be the case, especially given that other distributional characteristics, like kurtosis, have an effect as well. However, unfavourable variance comparisons can alert us to inefficiencies.

A natural way to make variance comparisons is through a measure of asymptotic relative efficiency (ARE)

$$\Upsilon = \mathrm{ARE}(T_m, T_r) = \mathrm{Var}(T_r)/\mathrm{Var}(T_m).$$

Given model correctness, both s_m and s_r are consistent for the same quantity. Hence it becomes necessary to examine higher order terms.

We have

$$\begin{aligned}\mathrm{Var}[T_r] &= \mathrm{Var}[T_m + (T_r - T_m)] = \mathrm{Var}[T_m + d]\\ &= \mathrm{Var}[T_m] + \mathrm{Var}[d] + 2\,\mathrm{Cov}[T_m, d].\end{aligned}$$

Let

$$\mathrm{Var}[T_m] = 1 + \frac{\nu}{n} + O(n^{-2}).$$

Since $d = O_p(n^{-1/2})$, both $\mathrm{Var}[d]$ and $\mathrm{Cov}[T_m, d]$ are $O(n^{-1})$, and so

$$\begin{aligned}\Upsilon &= \{\mathrm{Var}[T_m] + \mathrm{Var}[d] + 2\,\mathrm{Cov}[T_m, d]\}/\mathrm{Var}[T_m]\\ &= \left\{1 + \frac{\nu}{n} + \mathrm{Var}[d] + 2\,\mathrm{Cov}[T_m, d] + O(n^{-2})\right\}\left\{1 + \frac{\nu}{n} + O(n^{-2})\right\}^{-1}\\ &= \left\{1 + \frac{\nu}{n} + \mathrm{Var}[d] + 2\,\mathrm{Cov}[T_m, d]\right\}\left\{1 - \frac{\nu}{n}\right\} + O(n^{-2})\\ &= 1 + \mathrm{Var}[d] + 2\,\mathrm{Cov}[T_m, d] + O(n^{-2})\\ &= \{1 + E[d^2] + 2E[T_m d]\} + O(n^{-2})\\ &= 1 + E\left[\frac{2s_{12}T_0^2 s_0}{n^{-1/2}} + n^{-1}\{s_{12}^2 T_0^2 + 2s_{22}T_0^2 s_0 + 4T_1 T_0 s_{12} s_0 + 2T_0^2 s_{11} s_{12}\}\right]\\ &\quad + O(n^{-2}).\end{aligned}$$

Calculations in the Appendix show that for each pair of competing statistics the asymptotic relative efficiency is

$$\begin{aligned}\Upsilon = 1 &- \frac{\Lambda}{nk_0^4}\{I_{rstu} + I_{rs,tu} + 3I_{r,s,tu} + 4I_{r,stu} + 2A_{rsi}I^{ij}A_{tuj} + I_{i,r,s}I^{ij}A_{tuj}\}\\ &- \frac{I^{\psi r}I^{\psi s}I^{tu}}{2nk_0^2}\{I_{rstu} + 6I_{rt,su} + 8I_{r,t,su} + 2I_{r,stu} + 2I_{t,rsu}\\ &\qquad + A_{rsi}I^{ij}A_{tuj} + 4I_{i,rt}I^{ij}A_{suj}\} + O(n^{-2}),\end{aligned}$$

where $\Psi = I^{\psi r}I^{\psi s}I^{\psi t}I^{\psi u}$ and $A_{rsi} = I_{r,si} + I_{s,ri} + I_{rsi}$. Note Υ possesses the same invariance properties as the robust profile likelihood.

4. Examples. In this section, we consider a variety of examples where the above expression for the asymptotic relative efficiency is evaluated and terms of order n^{-2} or smaller are ignored. In most cases, the robust methods suffer losses in efficiency that are severe for small sample sizes. Some simulation results are also presented where we focus on the signed root of the likelihood ratio statistic. For these, and others not reported, the robust methods tend to result in confidence intervals that under-cover. In the examples, the maximum likelihood estimate is either a common summary statistic, like a mean, variance or correlation coefficient, or it is a regression coefficient, where we assume the developments of Gould and Lawless (1988). That is, the maximum likelihood estimate is meaningful in the presence of a model misspecification.

Computer algebra methods are used for the evaluation of Υ. These methods involve operators for evaluating highly nested sums and the expected information quantities that compose the terms in a sum. The relevant code, written in *Mathematica* Version 2.0 [see Wolfram (1988)], is available at the ftp site *utstat.toronto.edu* in the file *ARE.tar*. The expansion for Υ is provided in a form that may be manipulated in *Mathematica*, and instructions for the use of the code are given along with an example. For detailed information related to these tools see Andrews and Stafford (1993), Stafford and Andrews (1993) and Stafford, Andrews and Wang (1994).

EXAMPLE 1 (Normal mean, variance unknown). For this example, data x_i, $i = 1, \dots, n$, are assumed to be from a univariate normal distribution with mean ψ and variance λ. For this case, the observed information is diagonal and so the sandwich estimator is simply $\{J^{\psi\psi}\}^2 J_{\psi,\,\psi}$. However, $J_{\psi,\,\psi} = -J_{\psi\psi} = \sum_{i=1}^n (x_i - \bar{x})^2/n$ and hence the robust methods reduce to the model-specific methods. That is, the usual methods of inference based on the maximum likelihood estimate, score or signed root are robust in this case. Here, the above expression equals 1 as it should.

EXAMPLE 2 (Normal regression). This example also appears in DiCiccio and Stern (1994b). We assume data (y_i, x_i), $i = 1, \dots, n$, derive from a regression model where the conditional expectation of y given x is a linear function of x, $x_i^t \beta$; errors are normally distributed with variance σ^2. The parameter of interest is σ and the resultant expression for the asymptotic relative efficiency is $1 + 9/n$. Note that this is not only independent of the parameters in this model, but also of the design. That is, this case is equivalent to Example 1 where the parameter of interest is the standard deviation rather than the mean.

For this example, we simulated data from a regression with two covariates, $n = 10$ and $\sigma = 1$. This value of σ was tested at 95% level for 5000 simulations. The usual signed root rejected this hypothesis at the rate 0.0594, while the signed root of the robust profile likelihood rejected this hypothesis at the rate 0.1084. Figure 1 is a plot of the quantiles for the normal distribution against the simulated values of the signed root of the likelihood ratio statistic and its robust version. The quantiles of the robust statistics are shifted a greater distance from the diagonal. The dotted line represents where the theory predicts the robust quantiles should lie. The line passes through the origin with slope determined by Υ. Here, there is close agreement between theory and simulation.

EXAMPLE 3 (Exponential regression). In this example, data are assumed to have exponential distribution with mean $\lambda^{-1}\exp(-\psi z_i)$, where the slope ψ is the parameter of interest. For this case, a component of the log-likelihood is

$$\psi z_i - \log(\lambda) - \lambda^{-1} y_i \exp(-\psi z_i).$$

Expressions derived are simplified by standardizing the covariate so $\sum z_i = 0$.

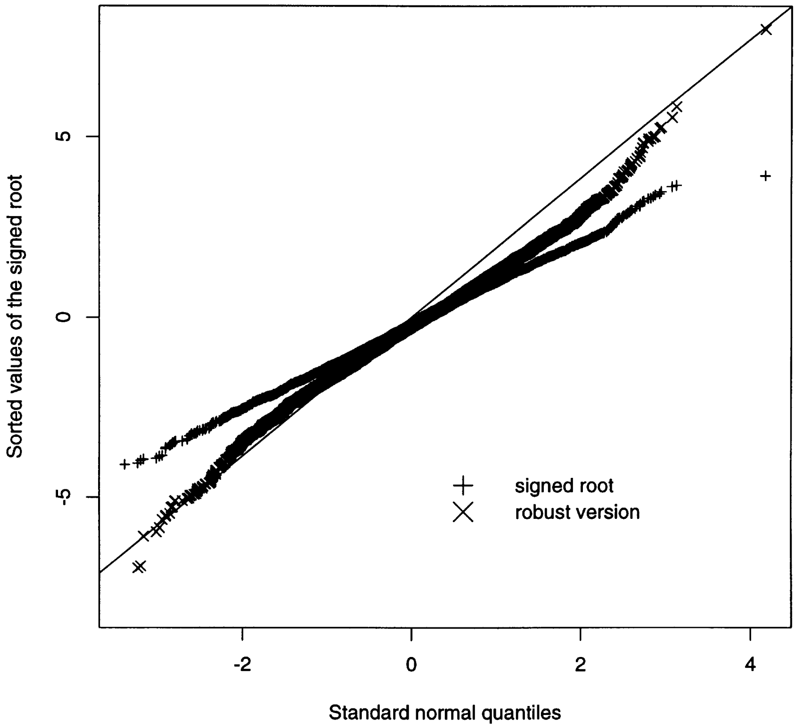

FIG. 1. *Normal regression: Q-Q plots for the simulated values of the signed square root of the profile likelihood ratio statistic and its robust counterpart. The line gives the location of the quantiles of the robust statistic as predicted by the theory.*

This also has the effect of orthogonalizing the parameters. Letting $\bar{z}_j = n^{-1}\Sigma z_i^j$, the asymptotic relative efficiency is

$$\Upsilon = 1 + \frac{17\bar{z}_4}{\bar{z}_2^2 n} - \frac{5\bar{z}_2^2}{\bar{z}_2^3 n} - \frac{1}{2n}.$$

EXAMPLE 4 (Correlation coefficient). For this example, we assume bivariate data $x_i^T = (x_{1i}, x_{2i})$, $i = 1, \dots, n$, derive from $N(\mu, \Sigma)$, where $\mu^T = (\mu_1, \mu_2)$ and

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{bmatrix}.$$

A component of the log-likelihood is

$$-\tfrac{1}{2}\log(1-\rho^2) - \log(\sigma) - \log(\tau) - \tfrac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu).$$

The calculations for this example are complicated by the relatively large number of parameters. For instance, the derivative $l_{i,\rho\rho\rho\rho}$ has 33 terms and the evaluation of the above expression for asymptotic relative efficiency involves inner products of arrays with dimension 4 or 6. This means that some inner products are nested sums with possibly 5^6 terms that are expected values of products of terms like $l_{i,\rho\rho\rho\rho}$. This is an extremely complicated calculation.

Lawley (1956) used properties specific to the normal distribution to simplify calculations in this case. We choose to explicitly evaluate all inner products. The use of computer algebra tools greatly simplifies such a calculation. The value of Υ is simply $1 + 6/n$.

EXAMPLE 5 [Normal regression (revisited)]. We again consider the case of a normal regression model where

$$y_i = \mu + \psi x_{1i} + \beta x_{2i} + \varepsilon_i; \qquad \varepsilon_i \rightsquigarrow N(0, \sigma^2), \qquad i = 1, \dots, n.$$

The parameter of interest is the coefficient ψ of the first covariate. The second covariate is assumed to contribute nuisance variation in the data. Calculations are again complicated for this example and again computer algebra is of valuable assistance. Letting $\hat{\mu}_{jk} = n^{-1}\Sigma x_{1i}^{j} x_{2i}^{k}$, the asymptotic relative efficiency is

$$\begin{aligned} 1 + \Big(& -4\hat{\mu}_{02}^2 \hat{\mu}_{11}^4 + 2\hat{\mu}_{04}\hat{\mu}_{11}^4 - 10\hat{\mu}_{02}\hat{\mu}_{11}^3\hat{\mu}_{13} + 17\hat{\mu}_{02}^2\hat{\mu}_{11}^2\hat{\mu}_{22} - 12\hat{\mu}_{02}^3\hat{\mu}_{11}\hat{\mu}_{31} \\ & + 8\hat{\mu}_{02}^3\hat{\mu}_{11}^2\hat{\mu}_{20} + \hat{\mu}_{02}\hat{\mu}_{04}\hat{\mu}_{11}^2\hat{\mu}_{20} - 2\hat{\mu}_{02}^2\mu_{11}\hat{\mu}_{13}\hat{\mu}_{20} + \hat{\mu}_{02}^3\hat{\mu}_{22}\hat{\mu}_{20} \\ & - 4\hat{\mu}_{02}^4\hat{\mu}_{20}^2 + 3\hat{\mu}_{02}^4\hat{\mu}_{40}\Big) \Big/ \Big(n\hat{\mu}_{02}^2\big(-\hat{\mu}_{11}^2 + \hat{\mu}_{02}\hat{\mu}_{20}\big)^2\Big). \end{aligned}$$

When the design is orthogonal, this expression simplifies considerably to

$$1 + \big(3\hat{\mu}_{40} + \hat{\mu}_{22} - 4\big)/n.$$

EXAMPLE 6 (Parameter orthogonality). In this last example, we test the invariance property of Υ by considering two examples where parameters may be orthogonalized [Cox and Reid (1987)]. Consider first, bivariate data (x_{1i}, x_{2i}), $i = 1, \dots, n$, from exponential distributions with means $(\psi\lambda)^{-1}$ and λ^{-1}, respectively. A component of the log-likelihood is

$$-\log(\psi) + 2\log(\lambda) - x_1/(\psi\lambda) - x_2/\lambda,$$

and the parameter of interest is ψ, the ratio of the two means. Under both this parameterization and the orthogonal parameterization given in Cox and Reid (1987), Υ was calculated to be $1 + 5/n$.

In the case of a Weibull distribution, where data x_i, $i = 1, \dots, n$, are assumed to have density

$$\psi/\lambda(x/\lambda)^{\psi-1}e^{-(x/\lambda)^{\psi}},$$

and interest lies in the index ψ, Υ was calculated to be $1 + 9.813/n$ for both parameterizations.

5. Final remarks. In conclusion, the use of a model-robust variance estimate for the signed square root, score or Wald statistic, while leaving bias and skewness characteristics relatively unchanged, can increase variability considerably. In the examples considered, robust methods are much less efficient than their model-based counterparts. This suggests caution should be used when employing a robust method. It also reflects the importance of model assumptions in improving the precision of inference.

Robust methods were derived from an adjusted profile likelihood due to McCullagh and Tibshirani (1990). An important aspect in the development was to approximate the integral $\tilde{l}_r(\psi) = \int^{\psi}\hat{v}(t)u_p(t)\,dt$ by $l_r(\psi) = \hat{v}(\hat{\psi})l_p(\psi)$ from which the standard results of Kent (1982) and Royall (1986) can be obtained. A question of interest is whether this is wise given $\tilde{l}_r(\psi)$ makes an adaptive adjustment over the entire range of ψ while $l_r(\psi)$ does not. A simple example is illuminating. Consider data $x_1, \dots, x_n$ from a $N(\psi, \sigma^2)$ distribution where model assumptions incorrectly specify that $\sigma = 1$. The log-likelihood function is $l(\psi) = l_p(\psi) = -\Sigma(x_i - \psi)^2/2$, $\hat{v}(\psi) = \Sigma(x_i - \psi)^2/n$ and $\hat{v}(\hat{\psi}) = \Sigma(x_i - \bar{x})^2/n$ and the profile log-likelihood under the true model is $l_{tp}(\psi) = -n\log\{\hat{v}(\psi)\}/2$. What is compelling about this example is that $\tilde{l}_r(\psi)$ is exactly $l_{tp}(\psi)$ while this is not the case for $l_r(\psi)$. So even though $l_r(\psi)$ is computationally convenient, a compromise is sometimes made. This is perhaps worth further investigation.

APPENDIX

Below are the detailed calculations involved in deriving the results of Section 3. Computer algebra tools such as those discussed in Stafford, Andrews and Wang (1994) were instrumental. Let $k_0 = \{I^{\psi\psi}\}^{1/2}$, $z^r = I^{rs}z_s$, $z^{\lambda_r} = I^{\lambda_r\lambda_s}z_{\lambda_s}$ and $z^{\psi} = I^{\psi s}z_s$. Expansions of the signed root, Wald and score statistic are well known [Lawley (1956); DiCiccio and Stern (1994b)] and from these we have

	T_0	T_1	c_1	c_2	c_3	c_4	c_5
$\sqrt{n}(\hat{\psi} - \psi)$	z^{ψ}	$I^{\psi r}z_{rs}z^s + \frac{1}{2}I^{\psi r}I_{rst}z^sz^t$	1	0	-3	1	3
$n^{-1/2}u(\psi)$	z^{ψ}/k_0^2	$(1/2k_0^2)\{I^{\psi r}z_{r\lambda_s}z^{\lambda_s} + \frac{1}{2}I^{\psi r}I_{r\lambda_s\lambda_t}z^{\lambda_s}z^{\lambda_t}\}$	-1	-4	1	0	1
$r(\psi)$	z^{ψ}/k_0	$(1/2k_0)\{I^{\psi r}z_{rs}z^s - I^{\psi r}z_{r\lambda_s}z^{\lambda_s} - \frac{1}{3}I^{\psi r}I_{rst}z^sz^t - \frac{1}{3}I^{\psi r}I_{rs\lambda_t}z^sz^{\lambda_t} - \frac{1}{3}I^{\psi r}I_{r\lambda_s\lambda_t}z^{\lambda_s}z^{\lambda_t}\}$	0	-2	-1	$\frac{1}{2}$	2

together with coefficients $c_1, \ldots, c_5$ defined for later use. To determine the coefficients for the various estimates, let

$$
\begin{aligned}
A_{rs} &= -\left(z_{rs} + I_{rst} z^t\right) \\
B_{rs} &= -\left(z_{rst} z^t + \tfrac{1}{2} I_{rstu} z^t z^u + I^{tv} I_{rst} z_{uv} z^u + \tfrac{1}{2} I^{tw} I_{rst} I_{uvw} z^u z^v\right), \\
C_{rs} &= z_{r,s} + (I_{r,st} + I_{s,rt}) z^t, \\
D_{rs} &= (z_{r,st} + z_{s,rt}) z^t + \tfrac{1}{2} z^t z^u (I_{s,rtu} + I_{r,stu} + 2 I_{rt,su}) \\
&\quad + \left(I^{tv} z_{uv} z^u + \tfrac{1}{2} I^{tw} I_{uvw} z^u z^v\right)(I_{r,st} + I_{s,rt}), \\
E^{rs} &= -I^{rt} A_{tu} I^{us}, \\
F^{rs} &= I^{rt} A_{tu} I^{uv} A_{vw} I^{ws} - I^{rt} B_{tu} I^{us}, \\
G_{\psi\psi} &= I^{\psi r} C_{rs} I^{s\psi} + 2 E^{\psi\psi}, \\
H_{\psi\psi} &= I^{\psi r} D_{rs} I^{s\psi} + 2 F^{\psi\psi} + 2 E^{\psi r} C_{rs} I^{s\psi} + E^{\psi r} I_{r,s} E^{s\psi}.
\end{aligned}
$$

Using these expressions, we have the expansions

$$
\begin{aligned}
-J_{rs} &= I_{r,s} + n^{-1/2} A_{rs} + n^{-1} B_{rs} + O_p(n^{-3/2}), \\
J_{r,s} &= I_{r,s} + n^{-1/2} C_{rs} + n^{-1} D_{rs} + O_p(n^{-3/2}), \\
J^{rs} &= I^{rs} + n^{-1/2} E^{rs} + n^{-1} F^{rs} + O_p(n^{-3/2}), \\
J^{\psi r} J_{r,s} J^{s\psi} &= I^{\psi\psi} + n^{-1/2} G_{\psi\psi} + n^{-1} H_{\psi\psi} + O_p(n^{-3/2}), \\
\{J^{\psi\psi}\}^{-1/2} &= k_0^{-1} - \frac{E^{\psi\psi}}{2\sqrt{n}\, k_0^3} - \frac{F^{\psi\psi}}{2nk_0^3} + \frac{3\{E^{\psi\psi}\}^2}{8nk_0^5}, \\
\left\{J^{\psi r} J_{r,s} J^{s\psi}\right\}^{-1/2} &= k_0^{-1} - \frac{G_{\psi\psi}}{2\sqrt{n}\, k_0^3} - \frac{H_{\psi\psi}}{2nk_0^3} + \frac{3G_{\psi\psi}^2}{8nk_0^5}, \\
\{J^{\psi\psi}\}^{1/2} &= k_0 + \frac{E^{\psi\psi}}{\sqrt{n}\, 2k_0} + \frac{F^{\psi\psi}}{2nk_0} - \frac{\{E^{\psi\psi}\}^2}{8nk_0^3}, \\
\left\{\frac{J^{\psi r} J_{r,s} J^{s\psi}}{[J^{\psi\psi}]^2}\right\}^{-1/2} &= k_0 - \frac{G_{\psi\psi} - 2E^{\psi\psi}}{2\sqrt{n}\, k_0} - \frac{H_{\psi\psi} - 2F^{\psi\psi}}{2nk_0} + \frac{3G_{\psi\psi}^2 - 4G_{\psi\psi} E^{\psi\psi}}{8nk_0^4}, \\
\left\{\frac{J^{\psi r} J_{r,s} J^{s\psi}}{J^{\psi\psi}}\right\}^{-1/2} &= 1 - \frac{G_{\psi\psi} - E^{\psi\psi}}{2\sqrt{n}\, k_0^2} - \frac{H_{\psi\psi} - F^{\psi\psi}}{2nk_0^2} \\
&\quad + \frac{3G_{\psi\psi}^2 - 2G_{\psi\psi} E^{\psi\psi} - \{E^{\psi\psi}\}^2}{8nk_0^4}.
\end{aligned}
$$

Let $\gamma = G_{\psi\psi} - E^{\psi\psi}$ and $\kappa = H_{\psi\psi} - F^{\psi\psi}$. From the above expansions for T, s_m and s_r and using the coefficients $c_1, \dots, c_5$, we have

$$T_0 s_{12} = -z^{\psi}\gamma/\left(2k_0^3\right), \qquad T_0^3 s_0^2 s_{12} = -\{z^{\psi}\}^3\gamma/\left(2k_0^5\right),$$
$$T_0^2 s_0 s_{12} = -\{z^{\psi}\}^2\gamma/\left(2k_0^4\right), \qquad T_0^2 s_{12}^2 = \{z^{\psi}\}^2\gamma^2/\left(4k_0^6\right),$$
$$T_0 T_1 s_0 s_{12} = T_1 z^{\psi}\gamma/\left(2k_0^4\right), \qquad T_0^2 s_{11} s_{12} = c_1\{z^{\psi}\}^2\gamma E^{\psi\psi}/\left(4k_0^6\right),$$
$$T_0^2 s_0 s_{22} = -\{z^{\psi}\}^2\kappa/\left(2k_0^4\right) + \{z^{\psi}\}^2\left(3G_{\psi\psi}^2 + c_2 G_{\psi\psi}E^{\psi\psi} + c_3\{E^{\psi\psi}\}^2\right)/\left(8k_0^6\right).$$

In particular,

$$2T_0^2 s_0 s_{22} + 2T_0^2 s_{11}s_{12} + T_0^2 s_{12}^2$$
$$= -\frac{\{z^{\psi}\}^2\kappa}{k_0^4} + \frac{\Lambda\{z^{\psi}\}^2}{k_0^6}(C_{rs} - A_{rs})(C_{tu} - c_5 A_{tu}).$$

For the calculations that follow, we make use of numerous identities,

$$\gamma'_{rs} = z_{r,s} + z_{r,s} + A_{rsi}z^i,$$
$$\gamma = I^{\psi r}(C_{rs} - A_{rs})I^{s\psi} = I^{\psi r}\gamma'_{rs}I^{s\psi}, \qquad I_{rsi} + I_{r,si} + I_{s,ri} + I_{i,rs} + I_{r,s,i} = 0,$$
$$E\left[I^{\psi r}I^{\psi s}\{z^{\psi}\}^2\gamma'_{rs}\right] = n^{-1/2}\Lambda\left\{I_{r,s,t,u} + I_{r,s,tu} + A_{rsi}I^{ij}I_{t,u,j}\right\},$$
$$E\left[\{z^{\psi}\}^2 z_{r,s}\gamma'_{tu}\right] = I^{\psi\psi}\left\{I_{r,s,t,u} + I_{r,s,tu} + I_{r,s,j}I^{ij}A_{tui}\right\} + O(n^{-1}),$$
$$E\left[\{z^{\psi}\}^2 z_{r,s}\gamma'_{tu}\right] = I^{\psi\psi}\left\{I_{r,s,tu} + I_{rs,tu} + I_{i,rs}I^{ij}A_{tuj}\right\} + O(n^{-1}),$$
$$E\left[z^{\psi}z^{i}z_{ti}\gamma'_{rs}\right] = I^{\psi u}\left\{I_{r,s,tu} + I_{rs,tu} + A_{tui}I^{ij}I_{j,tu}\right\} + O(n^{-1}),$$
$$E\left[z^{\psi}\gamma'_{rs}z_{tj}z^{j}\right] = I^{\psi u}\left\{I_{r,s,tu} + I_{rs,tu} + I_{j,tu}I^{ij}A_{rsi}\right\} + O(n^{-1}),$$
$$E\left[\{z^{\psi}\}^2 z^{i}\{2z_{r,si} + z_{rsi}\}\right] = (2I^{\psi i}I^{\psi j} + I^{\psi\psi}I^{ij})(2I_{i,r,js} + I_{i,jrs}) + O(n^{-1}),$$
$$E\left[\{z^{\psi}\}^2 z^{l}z_{jl}\right] = (2I^{\psi t}I^{\psi u} + I^{\psi\psi}I^{tu})I_{t,ju} + O(n^{-1}),$$
$$E\left[\{z^{\psi}\}^2 z^{k}z^{l}\right] = 2I^{\psi t}I^{\psi u} + I^{\psi\psi}I^{tu} + O(n^{-1}),$$

and the expected values $E[z^{\psi}\gamma'_{rs}]$, $E[\{z^{\psi}\}^3\gamma'_{rs}]$, $E[\{z^{\psi}\}^2 z^j\gamma'_{rs}]$, $E[z^{\psi}z^j z^k\gamma'_{rs}]$, $E[z^{\psi}\gamma'_{rs}z_{t\lambda_j}z^{\lambda_j}]$, $E[z^{\psi}\gamma'_{rs}I_{tj\lambda_k}z^j z^{\lambda_k}]$ and $E[z^{\psi}\gamma'_{rs}I_{t\lambda_j\lambda_k}z^{\lambda_j}z^{\lambda_k}]$ are all $O(n^{-1})$ or smaller. From these identities, we immediately have that $E[T_0 s_{12}]$ and $E[T_0^3 s_0 s_{12}]$ are $O(n^{-1})$ and hence $E[\delta_1] = E[\delta_3]$ are $O(n^{-3/2})$. Also, after lengthy calculations we have

$$E\big[T_0^2 s_0 s_{12}\big] = -\frac{\Lambda}{\sqrt{n}\,2k_0^4}\big(I_{r,s,t,u} + I_{r,s,tu} + A_{rsi}I^{ij}I_{t,u,j}\big),$$

$$E[T_0 T_1 s_0 s_{12}] = -\frac{c_4\Lambda}{2k_0^4}\big\{(I_{r,s,tu} + I_{rs,tu})I^{\psi u} + I_{j,tu}I^{ij}A_{rsi}\big\},$$

$$\begin{aligned}E\Big[\{z^{\psi}\}^2\kappa\Big] = {}& \Lambda\big\{I_{rstu} + 4I_{r,stu} + 4I_{r,s,tu} + 2I_{rs,tu} + A_{rsi}I^{ij}A_{tuj}\big\}\\ &+\frac{1}{2}I^{\psi r}I^{\psi s}I^{\psi\psi}I^{tu}\big\{I_{rstu} + 2I_{r,stu} + 2I_{t,rsu} + 8I_{r,t,su}\\ &\qquad +6I_{rt,su} + 4I_{j,rt}I^{ij}A_{usi} + A_{rsi}I^{ij}A_{tuj}\big\},\end{aligned}$$

$$\begin{aligned}&E\left[\frac{\Lambda}{k_0^6}(C_{rs} - A_{rs})(C_{tu} - c_5 A_{tu})\right]\\ &\quad= \frac{\Lambda}{k_0^6}I^{\psi\psi}\big\{I_{r,s,t,u} + (c_5+1)I_{r,s,tu} + c_5 I_{rs,tu} + (I_{r,s,i} + c_5 I_{i,rs})I^{ij}A_{tuj}\big\}.\end{aligned}$$

Combining all the relevant terms yields the same expression for the signed root, score and Wald statistics:

$$\begin{aligned}\Upsilon = 1 &- \frac{\Lambda}{nk_0^4}\big\{I_{rstu} + I_{rs,tu} + 3I_{r,s,tu} + 4I_{r,stu} + 2A_{rsi}I^{ij}A_{tuj} + I_{i,r,s}I^{ij}A_{tuj}\big\}\\ &-\frac{I^{\psi r}I^{\psi s}I^{tu}}{2nk_0^2}\big\{I_{rstu} + 6I_{rt,su} + 8I_{r,t,su} + 2I_{r,stu} + 2I_{t,rsu}\\ &\qquad +A_{rsi}I^{ij}A_{tuj} + 4I_{i,rt}I^{ij}A_{suj}\big\} + O(n^{-2}).\end{aligned}$$

Acknowledgments. I wish to thank Tom DiCiccio, Anthony Davison, Rob Tibshirani and the referees for very useful suggestions that aided the preparation of this paper.

REFERENCES

ANDREWS, D. F. and STAFFORD, J. E. (1993). Tools for the symbolic computation of asymptotic expansions. *J. Roy. Statist. Soc. Ser. B* **55** 613–627.

BARNDORFF-NIELSEN, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365.

BARNDORFF-NIELSEN, O. E. and SORENSON, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes. *Internat. Statist. Rev.* **62** 133–165.

COX, D. R. and REID, N. (1987). Parametric orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49** 1–39.

COX, D. R. and REID, N. (1993). A note on the calculation of adjusted profile likelihood. *J. Roy. Statist. Soc. Ser. B* **55** 467–471.

DICICCIO, T. J. (1984). On parameter transformations and interval estimation. *Biometrika* **71** 477–485.

DiCiccio, T. J. and Stern, S. E. (1994a). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods *J. Roy. Statist. Soc. Ser. B* **56** 397–408.

DiCiccio, T. J. and Stern, S. E. (1994b). Constructing approximately standard normal pivots from signed roots at adjusted likelihood ratio statistics. *Scand. J. Statist.* **21** 447–460.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65** 457–481.

Gould, A. and Lawless, J. F. (1988). Consistency and efficiency of regression coefficient estimates in location-scale models. *Biometrika* **75** 535–540.

Huber, P. J. (1967). The behaviour of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

Kent, T. J. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69** 19–27.

Lawley, D. N. (1956). A general method for approximating the distribution of likelihood ratio criteria. *Biometrika* **43** 295–303.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84** 1074–1078.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. Ser. B* **52** 325–344.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *Internat. Statist. Rev.* **54** 221–226.

Stafford, J. E. and Andrews, D. F. (1993). A symbolic algorithm for studying adjustments to the profile likelihood. *Biometrika* **80** 715–730.

Stafford, J. E., Andrews, D. F. and Wang, Y. (1994). Symbolic computation: A unified approach to studying likelihood. *Statistics and Computing* **4** 235–245.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **82** 1–25.

Wolfram, S. (1988). *Mathematica: A System for Doing Mathematics by Computer*. Addison-Wesley, Reading, MA.

Department of Statistical and
Actuarial Sciences
University of Western Ontario
Room 262 Western Science Center
London, Ontario
Canada N6A 5B7