# High-order asymptotic expansions for robust tests 

By FRANÇOIS-XAVIER DE ROSSI<br>Swiss Federal Statistical Office, 2010 Neuchâtel, Switzerland<br>francois-xavier.derossi@bfs.admin.ch<br>and RICCARDO GATTO<br>Institut für Mathematische Statistik und Versicherungslehre, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland<br>gatto@stat.unibe.ch


#### Abstract

Summary This paper provides high-order asymptotic expansions for the $M$-ratio and the signedroot $M$-ratio robust test statistics, which allow one to compute accurate approximations to significance or critical values using the Edgeworth approximation, the Bartlett correction, the variance correction or the saddlepoint approximation. Specific results are obtained for the linear regression model with the Huber $M$-estimator. A Monte Carlo study illustrates the numerical accuracy of these approximations, with respect to the usual first-order approximations.

Some key words: Bartlett correction; Edgeworth approximation; Likelihood ratio test; $M$-ratio test; Nuisance parameter; Regression model; Robust $p$-value and quantile; Saddlepoint approximation; Signed-root likelihood ratio test; Signed-root $M$-ratio test.


## 1. Introduction

The classical testing problem assumes that the data are a realisation of an underlying parametric model. Inference is usually based on the first-order asymptotic approximation, given by a central limit theorem, of the distribution of a test statistic. This approach has two major drawbacks: one concerns the robustness of the test and the other the accuracy of the asymptotic approximation. To limit the influence of the model hypotheses, robust tests consider the assumed model of an idealised approximation only, so that the results remain stable under small deviations from the model; see Markatou \& Ronchetti (1997) for a survey. Although there has been more intensive research in robust estimation than in robust testing, the need for robust test procedures is obvious, since it is not advisable to estimate robustly the parameters of a model and then use classical tests about these parameters. The significance probabilities, or $p$-values, are usually obtained by first-order approximations with a typical error of $O\left(n^{-\frac{1}{2}}\right)$, as $n \rightarrow \infty$. In small sample sizes, these approximations often become inaccurate, especially for the typically small tail probabilities required by testing problems, and higher-order asymptotic approximations are more reliable. In this paper we propose some high-order approximations for robust tests and show, by simulation studies, the numerical accuracy of the new approximations.

We focus on the likelihood ratio test. Suppose that the data $Z_{1}, \ldots, Z_{n}$ are independent and identically distributed random variables whose distribution belongs to a para-
metric model $\left\{F_{\theta}, \theta \in \Theta\right\}$, where $\theta=\left(\theta^{1}, \ldots, \theta^{p}\right)$, and suppose there is a null hypothesis $H_{0}: \theta=\theta_{0}$, where $\theta_{0}^{q+1}=\ldots=\theta_{0}^{p}=0$ and $\theta_{0}^{1}, \ldots, \theta_{0}^{q}$ are unspecified nuisance parameters. The likelihood ratio test statistic for this situation is given by

$$
\begin{equation*}
V_{q n}=2\left\{\sum_{l=1}^{n} \log f\left(Z_{l}, T_{n}\right)-\sum_{l=1}^{n} \log f\left(Z_{l}, T_{q n}\right)\right\}, \tag{1}
\end{equation*}
$$

where $T_{n}$ is the maximum likelihood estimator of $\theta, T_{q n}$ is the maximum likelihood estimator constrained by $\theta^{q+1}=\ldots=\theta^{p}=0$, and $f(x, \theta)=d F_{\theta}(x) / d x$. For the case without nuisance parameters, $q=0$ and $T_{q n}=\theta_{0}$. If we denote by $\chi_{p-q}^{2}$ a chi-squared random variable with $(p-q)$ degrees of freedom, the limiting distribution under $H_{0}$ of $V_{q n}$ is given by $\operatorname{pr}\left(V_{q n} \leqslant v\right)=\operatorname{pr}\left(\chi_{p-q}^{2} \leqslant v\right)+O\left(n^{-1}\right)$. The latter error term can be diminished by applying the Bartlett correction (Bartlett, 1937; Lawley, 1956) outlined as follows. The Bartlett factor $b$ obtained by the expansion $E\left(V_{q n}\right)=(p-q)(1+b / n)+O\left(n^{-2}\right)$ determines the necessary amount of shifting of the chi-squared distribution to give

$$
\begin{equation*}
\operatorname{pr}\left(V_{q n} \leqslant v\right)=\operatorname{pr}\left(\chi_{p-q}^{2} \leqslant \frac{v}{1+b n^{-1}}\right)+O\left(n^{-2}\right) . \tag{2}
\end{equation*}
$$

Note that, with nuisance parameters, $b=b\left(\theta_{0}\right)$ can be replaced by $\tilde{b}=b\left(\widetilde{T}_{n}\right)$, where $\widetilde{T}_{n}$ is an estimator of $\theta_{0}$ which satisfies $\widetilde{T}_{n}-\theta_{0}=O_{p}\left(n^{-\frac{1}{2}}\right)$. This does not increase the order of error in (2). A natural choice satisfying this condition is $\widetilde{T}_{n}=T_{q n}$. For the case where $q=p-1$, that is $\theta_{0}^{p}=0$ and $\theta_{0}^{1}, \ldots, \theta_{0}^{p-1}$ unspecified, the definition of the signed-root likelihood ratio test as

$$
\begin{equation*}
R_{n}=\operatorname{sgn}\left\{T_{n}^{p}-\theta_{0}^{p}\right\} V_{p-1, n}^{1 / 2} \tag{3}
\end{equation*}
$$

is motivated by its standard normal limiting distribution. The error of the limiting normal approximation can be brought from $O\left(n^{-1 / 2}\right)$ down to $O\left(n^{-3 / 2}\right)$ by the following bias and variance corrections (DiCiccio, 1984). The bias correction factor $a\left(\widetilde{T}_{n}\right)$ is obtained by the expansion $E\left(R_{n}\right)=n^{-1 / 2} a\left(\widetilde{T}_{n}\right)+O_{p}\left(n^{-3 / 2}\right)$ and the variance correction factor $c\left(\widetilde{T}_{n}\right)$ is obtained by the expansion $\operatorname{var}\left\{R_{n}-n^{-1 / 2} a\left(\widetilde{T}_{n}\right)\right\}=1+n^{-1} c\left(\widetilde{T}_{n}\right)+O_{p}\left(n^{-2}\right)$. It then follows that $\widetilde{R}_{n}=\left\{R_{n}-n^{-\frac{1}{2}} a\left(\widetilde{T}_{n}\right)\right\}\left\{1+n^{-1} c\left(\widetilde{T}_{n}\right)\right\}^{-\frac{1}{2}}$ is distributed as a standard normal random variable, up to $O\left(n^{-3 / 2}\right)$. Alternative modifications of (3) have been widely explored (Barndorff-Nielsen, 1991; DiCiccio \& Martin, 1993).

The likelihood ratio test statistic (1) can be modified by replacing the logarithmic density by a convex function defining a robust $M$-estimator. To be precise, by considering now $T_{n}$ as the solution of

$$
\begin{equation*}
\max _{\theta} \sum_{l=1}^{n} \varrho\left(Z_{l}, \theta\right), \tag{4}
\end{equation*}
$$

where $-\varrho: \mathbb{R}^{k} \times \Theta \rightarrow \mathbb{R}$ is a convex function in $\theta$, we can define the $M$-ratio test by

$$
\begin{equation*}
W_{q n}=2\left\{\sum_{l=1}^{n} \varrho\left(Z_{l}, T_{n}\right)-\sum_{l=1}^{n} \varrho\left(Z_{l}, T_{q n}\right)\right\}, \tag{5}
\end{equation*}
$$

where $T_{q n}$ is the $H_{0}$-restricted $M$-estimator. This leads to tests that are robust in validity and efficiency: the level and the power remain stable under small and arbitrary departures from the null and alternative hypotheses. Heritier \& Ronchetti (1994) derived the asymptotic distribution of (5), which is, under $H_{0}$, that of a linear combination of $(p-1) \chi_{1}^{2}$ random variables, and they discussed some robustness properties. In the next section, we
propose high-order approximations to this distribution, which enable us to compute $p$-values with smaller errors.

Section 2 provides asymptotic expansions for the $M$-ratio test in the general parametric model and for the regression model, both without nuisance parameters. For the case with nuisance parameters, expansions for the $M$-ratio and the signed-root $M$-ratio statistics are given for the linear regression model. All these expansions allow us to approximate the necessary cumulants for the Bartlett and variance corrections, and for the Edgeworth, Cornish-Fisher and saddlepoint approximations. Section 3 gives numerical simulations and comparisons that show the effectiveness of the proposed high-order approximations. Proofs and regularity conditions are reported in the Appendices.

## 2. Derivation of the expansions

The next propositions provide high-order expansions for the $M$-ratio and its signedroot statistics. These expansions are not direct generalisations of their classical version (McCullagh, 1987) because Bartlett identities, such as

$$
E\{(\partial / \partial t) \log f(Z, t)\}^{2}=E\left\{\left(\partial^{2} / \partial t^{2}\right) \log f(Z, t)\right\},
$$

do not hold generally with $M$-estimators.
Indeed, we show that Bartlett identities are not necessary: the particular form of the influence function of $M$-estimators (7) is sufficient to obtain all simplifications which will lead to the computational expansions (9) in Proposition 1 and (16) in Proposition 2. The results are expressed with tensor notation (McCullagh, 1987). The summation convention is used that, whenever identical indices appear in both superscript and subscript of an expression, term-wise summation over the indexed elements is meant. For matrices, transferring both indices from lower to upper levels, or conversely, indicates the elements of the inverse matrix. Also, unless explicitly mentioned, indices $r, s, t, u, v$ and $w$ range from 1 to $p$.

We first recall an auxiliary expansion of the $r$ th component $\Delta^{r}$ of $\Delta=n^{\frac{1}{2}}\left(T_{n}-\theta_{0}\right)$. We denote the derivatives of $\varrho\left(Z_{1}, \theta\right)$, defined in (4), by

$$
\begin{equation*}
\Psi_{r}\left(Z_{1}, \theta\right)=\frac{\partial}{\partial \theta^{r}} \varrho\left(Z_{1}, \theta\right), \quad \Psi_{r s}\left(Z_{1}, \theta\right)=\frac{\partial}{\partial \theta^{r}} \Psi_{s}\left(Z_{1}, \theta\right), \quad \Psi_{r s t}\left(Z_{1}, \theta\right)=\frac{\partial}{\partial \theta^{r}} \Psi_{s t}\left(Z_{1}, \theta\right), \tag{6}
\end{equation*}
$$

the corresponding cumulants by $\kappa_{r s}=E\left\{\Psi_{r s}\left(Z_{1}, \theta_{0}\right)\right\}, \kappa_{r s t}=E\left\{\Psi_{r s t}\left(Z_{1}, \theta_{0}\right)\right\}$, the matrix inverse of ( $\kappa_{r s}$ ) by ( $\kappa^{r s}$ ), and we define the random variables

$$
\begin{gathered}
Y_{r}=n^{-\frac{1}{2}} \sum_{l=1}^{n} \Psi_{r}\left(Z_{l}, \theta_{0}\right), \quad Y_{r s}=n^{-\frac{1}{2}} \sum_{l=1}^{n} \Psi_{r s}\left(Z_{l}, \theta_{0}\right)-n^{\frac{1}{2}} \kappa_{r s}, \\
Y_{r s t}=n^{-\frac{1}{2}} \sum_{l=1}^{n} \Psi_{r s t}\left(Z_{l}, \theta_{0}\right)-n^{\frac{1}{2}} \kappa_{r s t} .
\end{gathered}
$$

The regularity conditions required for the following results are given in Appendix 1, and some proofs appear in Appendix 2.

Lemma 1. When Conditions (A) in Appendix 1 hold for $\theta=\theta_{0} \in \Theta$, the following expansions are valid as $n \rightarrow \infty$ :
(i) $\Delta^{r}=B^{r}+o_{p}(1)$, where

$$
\begin{equation*}
B^{r}=\left(-\kappa^{r s}\right) Y_{s} ; \tag{7}
\end{equation*}
$$

(ii) if Condition (C1) also holds, then $\Delta^{r}=B^{r}+n^{-1 / 2} C^{r}+O_{p}\left(n^{-1}\right)$, where

$$
\begin{align*}
C^{r} & =\left(-\kappa^{r t}\right) B^{u} Y_{t u}+\frac{1}{2}\left(-\kappa^{r u}\right) B^{v} B^{w} \kappa_{u v w} \\
& =\left(-\kappa^{r t}\right)\left(-\kappa^{s u}\right) Y_{s} Y_{t u}+\frac{1}{2}\left(-\kappa^{r i}\right)\left(-\kappa^{s j}\right)\left(-\kappa^{t k}\right) \kappa_{i j k} Y_{S} Y_{t} \tag{8}
\end{align*}
$$

Note that $B^{r}$ is currently called the influence function. This result was derived by Fan $\&$ Field (1995) by expanding the $p$ equations defining the $M$-estimator and by applying the implicit function theorem, and by Gatto \& Ronchetti (1996) as a particular case of the von Mises functional expansion. Lemma 1 is necessary for establishing the expansions for the $M$-ratio statistic and its expectation. For the general parametric model without nuisance parameters, these expansions are given by Proposition 1, which generalises to $M$-estimators the results in formulae (26) and (27) of McCullagh \& Cox (1986). We define the cumulants

$$
\begin{gathered}
\kappa_{r, s}=E\left\{\Psi_{r}\left(Z_{1}, \theta_{0}\right) \Psi_{s}\left(Z_{1}, \theta_{0}\right)\right\}, \quad \kappa_{r, s, t}=E\left\{\Psi_{r}\left(Z_{1}, \theta_{0}\right) \Psi_{s}\left(Z_{1}, \theta_{0}\right) \Psi_{t}\left(Z_{1}, \theta_{0}\right)\right\}, \\
\kappa_{r, s, t u}=E\left\{\Psi_{r}\left(Z_{1}, \theta_{0}\right) \Psi_{s}\left(Z_{1}, \theta_{0}\right) \Psi_{t u}\left(Z_{1}, \theta_{0}\right)\right\}-E\left\{\Psi_{r}\left(Z_{1}, \theta_{0}\right) \Psi_{s}\left(Z_{1}, \theta_{0}\right)\right\} E\left\{\Psi_{t u}\left(Z_{1}, \theta_{0}\right)\right\}, \\
\kappa_{t, u v}=E\left\{\Psi_{t}\left(Z_{1}, \theta_{0}\right) \Psi_{u v}\left(Z_{1}, \theta_{0}\right)\right\}, \quad \kappa_{t, u v w}=E\left\{\Psi_{t}\left(Z_{1}, \theta_{0}\right) \Psi_{u v w}\left(Z_{1}, \theta_{0}\right)\right\}, \\
\kappa_{t u, v w}=E\left\{\Psi_{t u}\left(Z_{1}, \theta_{0}\right) \Psi_{v w}\left(Z_{1}, \theta_{0}\right)\right\}-E\left\{\Psi_{t u}\left(Z_{1}, \theta_{0}\right)\right\} E\left\{\Psi_{v w}\left(Z_{1}, \theta_{0}\right)\right\} .
\end{gathered}
$$

Proposition 1 (General parametric model). When Conditions (A) hold for $\theta=\theta_{0} \in \Theta$, the following expansions are valid as $n \rightarrow \infty$.
(i) If Condition (C1) holds,

$$
\begin{align*}
W_{0 n}= & \left(-\kappa^{r s}\right) Y_{r} Y_{s}+n^{-\frac{1}{2}} \frac{1}{3} \kappa^{r s t} Y_{r} Y_{s} Y_{t}+n^{-\frac{1}{2}}\left(-\kappa^{r t}\right)\left(-\kappa^{s u}\right) Y_{r} Y_{s} Y_{t u} \\
& +n^{-1}\left\{\frac{1}{12} \kappa^{r s t u}+\frac{1}{4} \kappa^{r s v} \kappa^{t u w}\left(-\kappa_{v w}\right)\right\} Y_{r} Y_{s} Y_{t} Y_{u}+n^{-1} \kappa^{r t u}\left(-\kappa^{s v}\right) Y_{r} Y_{s} Y_{t} Y_{u v} \\
& +n^{-1} \frac{1}{3}\left(-\kappa^{r u}\right)\left(-\kappa^{s v}\right)\left(-\kappa^{t w}\right) Y_{r} Y_{s} Y_{t} Y_{u v w} \\
& +n^{-1}\left(-\kappa^{r u}\right)\left(-\kappa^{s v}\right)\left(-\kappa^{t w}\right) Y_{r} Y_{s} Y_{t u} Y_{v w}+O_{p}\left(n^{-\frac{3}{2}}\right) \tag{9}
\end{align*}
$$

where

$$
\kappa^{r s t}=\left(-\kappa^{r i}\right)\left(-\kappa^{s j}\right)\left(-\kappa^{t k}\right) \kappa_{i j k}, \quad \kappa^{r s t u}=\left(-\kappa^{r i}\right)\left(-\kappa^{s j}\right)\left(-\kappa^{t k}\right)\left(-\kappa^{u m}\right) \kappa_{i j k m}
$$

(ii) If Condition (C2) holds,

$$
\begin{align*}
E\left(W_{0 n}\right)= & \left(-\kappa^{r s}\right) \kappa_{r, s}+n^{-1}\left\{\frac{1}{3} \kappa^{r s t} \kappa_{r, s, t}+\left(-\kappa^{r t}\right)\left(-\kappa^{s u}\right) \kappa_{r, s, t u}\right. \\
& +\frac{1}{4} \kappa^{r s t u} \kappa_{r, s} \kappa_{t, u}+\frac{1}{4} \kappa^{r s v} \kappa^{t u w}\left(-\kappa_{v w}\right)\left(\kappa_{r, s} \kappa_{t, u}+2 \kappa_{r, t} \kappa_{s, u}\right) \\
& +\kappa^{r t u}\left(-\kappa^{s v}\right)\left(2 \kappa_{r, s} \kappa_{t, u v}+\kappa_{r, t} \kappa_{s, u v}\right)+\left(-\kappa^{r u}\right)\left(-\kappa^{s v}\right)\left(-\kappa^{t w}\right) \kappa_{r, s} \kappa_{t, u v w} \\
& \left.+\left(-\kappa^{r u}\right)\left(-\kappa^{s v}\right)\left(-\kappa^{t w}\right)\left(\kappa_{r, s} \kappa_{t u, v w}+\kappa_{r, t u} \kappa_{s, v w}+\kappa_{r, v w} \kappa_{s, t u}\right)\right\} \\
& +O\left(n^{-2}\right) . \tag{10}
\end{align*}
$$

The proof of Proposition 1 is given in Appendix 2. As a special case, consider the regression model given by

$$
\begin{equation*}
\eta_{l}=g\left(\xi_{l 1}, \ldots, \xi_{l p}, \theta\right)+\varepsilon_{l} \quad(l=1, \ldots, n) \tag{11}
\end{equation*}
$$

where $Z_{l}=\left(\eta_{l}, \xi_{l 1}, \ldots, \xi_{l p}\right)$ are $n$ independent and identically distributed random vectors in $\mathbb{R} \times \mathbb{R}^{p}$ representing the data, $\varepsilon_{l}$ are independent and identically distributed errors of $\left(\xi_{l 1}, \ldots, \xi_{l p}\right)$, with $E\left(\varepsilon_{1}\right)=0$ and $\operatorname{var}\left(\varepsilon_{1}\right)=\sigma^{2}$, and $g$ satisfies Condition (B1). A Huber type $M$-estimator is defined as the solution $T_{n}$ of

$$
\begin{equation*}
\min _{\theta \in \Theta} \sum_{l=1}^{n} \rho\left(\frac{\eta_{l}-g\left(\xi_{l 1}, \ldots, \xi_{l p}, \theta\right)}{\sigma}\right) \tag{12}
\end{equation*}
$$

where $\rho: \mathbb{R} \rightarrow \mathbb{R}^{+}$is a convex function satisfying Condition (B2). The $M$-ratio test statistic for this model, without nuisance parameters, is given by

$$
\begin{equation*}
W_{0 n}=2\left\{\sum_{l=1}^{n} \rho\left(\frac{e_{l}}{\sigma}\right)-\sum_{l=1}^{n} \rho\left(\frac{\varepsilon_{l}}{\sigma}\right)\right\} \tag{13}
\end{equation*}
$$

where $e_{l}=\eta_{l}-g\left(\xi_{l 1}, \ldots, \xi_{l p}, T_{n}\right)$ are the residuals. An equivalent definition of $T_{n}$ is given by $\sum_{l=1}^{n} \Psi\left(Z_{l}, T_{n}\right)=0$, where the generalised score function $\Psi$ has the form

$$
\Psi_{r}\left(Z_{l}, \theta_{0}\right)=\psi\left(\varepsilon_{l}\right) g_{r}\left(\xi_{l 1}, \ldots, \xi_{l p}, \theta_{0}\right)
$$

with $\psi=\rho^{\prime}$ and $g_{r}=g_{r}\left(\xi_{1}, \ldots, \xi_{p}, \theta_{0}\right)=\partial g\left(\xi_{1}, \ldots, \xi_{p}, \theta_{0}\right) / \partial \theta^{r}$. Note that the regression model with fixed regressors corresponds to the particular case of the above random regressors model, in which the empirical distribution of the regressors is used instead of the distribution of the random regressors. In considering the fixed regressors as a special case in our theorems, we replace a fixed distribution by another depending on $n$. This is possible as long as the fixed regressors can be thought of as the result of a random draw from a fixed distribution, so that the sequence of the empirical distributions would converge to a fixed distribution. Otherwise, our results hold for random regressors only.

In what follows, we propose specific results for the linear model $g\left(\xi_{1}, \ldots, \xi_{p}, \theta\right)=\theta^{r} \xi_{r}$. This model is of practical importance and we will see that it leads to simple expressions. We define the moments $\zeta_{r}=E\left(\xi_{r}\right), \zeta_{r s}=E\left(\xi_{r} \xi_{s}\right)$, and so on.

Corollary 1 (Linear regression model). When Conditions $(\mathrm{A})$ and $(\mathrm{B})$ hold for $\theta=$ $\theta_{0} \in \Theta$, then the following expansion for the $M$-ratio test statistic in a linear regression model is valid as $n \rightarrow \infty$ :

$$
\begin{aligned}
W_{0 n}= & \left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\} \zeta^{r s} Y_{r} Y_{s}+n^{-\frac{1}{2}}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2} \zeta^{r t \varphi s u} Y_{r} Y_{s} Y_{t u} \\
& +n^{-1} \frac{1}{12}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{4} E\left(-\psi^{\prime \prime \prime}\right) \zeta^{r i \zeta s j} \zeta^{t k \zeta u m} \zeta_{i j k m} Y_{r} Y_{s} Y_{t} Y_{u} \\
& +n^{-1} \frac{1}{3}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{3} \zeta^{r u \zeta s v \zeta t w} Y_{r} Y_{s} Y_{t} Y_{u v w}+n^{-1}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{3} \zeta^{r u \zeta s v \zeta t w} Y_{r} Y_{s} Y_{t u} Y_{v w} \\
& +O_{p}\left(n^{-3 / 2}\right)
\end{aligned}
$$

where

$$
Y_{r}=n^{-\frac{1}{2}} \sum_{l=1}^{n} \psi\left(\varepsilon_{l}\right) \xi_{l r}, \quad Y_{r s}=n^{\frac{1}{2}} E\left(\psi^{\prime}\right) \zeta_{r s}-n^{-\frac{1}{2}} \sum_{l=1}^{n} \psi^{\prime}\left(\varepsilon_{l}\right) \xi_{l r} \xi_{l s}, \quad Y_{r s t}=n^{-\frac{1}{2}} \sum_{l=1}^{n} \psi^{\prime \prime}\left(\varepsilon_{l}\right) \xi_{l r} \xi_{l s} \xi_{l t}
$$

With nuisance parameters, the expansion of the $M$-ratio statistic for the linear regression model is given by Proposition 2, and by Corollary 2 for the Huber $M$-estimator given by (21). We denote by $e_{q l}=\eta_{l}-T_{q n}^{r} \xi_{r l}$ the residual of the reduced model, where $T_{q n}$ is the solution of (12) with $\theta^{q+1}=\ldots=\theta^{p}=0$. The $M$-ratio statistic, also called the $\tau$-test (Hampel et al., 1986, pp. 345-6), is hence given by

$$
W_{q n}=2\left\{\sum_{l=1}^{n} \rho\left(\frac{\varepsilon_{l}}{\sigma}\right)-\sum_{l=1}^{n} \rho\left(\frac{e_{q l}}{\sigma}\right)\right\} .
$$

Lemma 2 is useful for the expansion of $W_{q n}(0 \leqslant q \leqslant p-1)$.
Lemma 2. Let $A$ be a real $p \times p$ symmetric positive definite matrix. Then the matrix $\Omega$, defined by $\Omega^{r s}=\zeta^{r s}-\zeta^{\alpha \beta} \delta_{\alpha}^{r} \delta_{\beta}^{s}$, is positive semidefinite of $\operatorname{rank}(p-q)$, and there exists a nonsingular $p \times p$ matrix $V$ such that $V^{\mathrm{T}} A^{-1} V=I$ and $V^{\mathrm{T}} \Omega V=\Lambda$, where $\Lambda=$ $\operatorname{diag}\left(0, \ldots, 0, \lambda_{q+1}, \ldots, \lambda_{p}\right)$ with $\lambda_{q+1}=\ldots=\lambda_{p}=1$, and where $\delta_{\alpha}^{r}$ equals 1 if $\alpha=r$, and 0 otherwise; Greek letters indices range from 1 to $q$.

This result was used by E. Ronchetti in his 1982 Ph.D. thesis at the Swiss Federal Institute of Technology, Zurich, in a similar context. An explicit form of $V$ can be obtained by considering the Choleski decomposition $A^{-1}=L L^{\mathrm{T}}$, where $L$ is lower triangular, by defining $H=L^{-1} \Omega\left(L^{-1}\right)^{\mathrm{T}}$, and by using the singular value decomposition $H=\Gamma \Lambda \Gamma^{\mathrm{T}}$ along with the fact that $\Gamma^{\mathrm{T}} \Gamma=I$, to obtain $V=\left(L^{-1}\right)^{\mathrm{T}} \Gamma$.
Proposition 2 (Linear regression model). Assume that Conditions (A) and (B) hold for $\theta=\theta_{0} \in \boldsymbol{\Theta}$. Then the following expansions are valid as $n \rightarrow \infty$.
(i) If Condition (C1) holds, then

$$
\begin{equation*}
W_{q n}=\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}\left(\zeta^{r s}-\zeta^{\alpha \beta} \delta_{\alpha}^{r} \delta_{\beta}^{s}\right) S_{r} S_{s}+O_{p}\left(n^{-3 / 2}\right), \tag{14}
\end{equation*}
$$

where

$$
\begin{align*}
S_{r}= & Y_{r}+n^{-\frac{1}{2}} \frac{1}{2}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\} \zeta^{s t} Y_{r s} Y_{t}+n^{-1} \frac{1}{24}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{3} E\left(-\psi^{\prime \prime \prime}\right) \zeta_{r x} \varphi^{x s t u} Y_{s} Y_{t} Y_{u} \\
& +n^{-1} \frac{1}{6}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2} \zeta^{s u} \zeta^{t v} Y_{r s t} Y_{u} Y_{v}+n^{-1} \frac{3}{8}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2} \zeta^{s t \zeta} \zeta^{u v} Y_{r s} Y_{t u} Y_{v} . \tag{15}
\end{align*}
$$

(ii) If Condition (C1) holds, then

$$
\begin{equation*}
W_{q n}=\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\} \sum_{r=q+1}^{p} U_{r}^{2}+O_{p}\left(n^{-3 / 2}\right), \tag{16}
\end{equation*}
$$

where $U_{r}$ is defined by

$$
\begin{equation*}
U_{r}=\left\{\frac{1}{E\left(\psi^{2}\right)}\right\}^{\frac{1}{2}} v^{j s} \delta_{j r} S_{s}, \tag{17}
\end{equation*}
$$

in which $v^{j s}$ denotes the $(j, s)$ th element of $V^{-1}, V$ being defined by a particular application of Lemma 2 to the matrix $A=\left(\zeta_{r s}\right)$, and where $\delta_{j r}$ equals 1 , if $j=r$, and 0 otherwise.
(iii) If Condition (C1) holds, then the first four cumulants of $U_{r}$ are given by

$$
\begin{gather*}
\omega_{r}=O\left(n^{-3 / 2}\right), \quad \omega_{r, r}=1+n^{-1} c_{\psi} h_{r}+O\left(n^{-2}\right),  \tag{18}\\
\omega_{r, r, r}=O\left(n^{-3 / 2}\right), \\
\omega_{r, r, r, r}=n^{-1} e_{\psi} m_{r}+O\left(n^{-2}\right),
\end{gather*}
$$

where $c_{\psi}, e_{\psi}, h_{r}$ and $m_{r}$ are given by

$$
\begin{gather*}
c_{\psi}=\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}\left\{\frac{1}{E\left(\psi^{2}\right)}\right\}\left[\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}\left\{E\left(\psi^{\prime} \psi^{\prime}\right)+E\left(\psi^{\prime \prime} \psi\right)\right\}-E\left(\psi^{\prime} \psi^{2}\right)-\frac{1}{4}\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{2} E\left(\psi^{\prime \prime \prime}\right)\right]  \tag{19}\\
e_{\psi}=\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2}\left[E\left(\psi^{4}\right)+\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{2}\left\{3 E\left(\psi^{\prime} \psi^{\prime}\right)+4 E\left(\psi^{\prime \prime} \psi\right)\right\}\right. \\
\left.-6\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\} E\left(\psi^{\prime} \psi^{2}\right)-\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{3} E\left(\psi^{\prime \prime \prime}\right)\right]  \tag{20}\\
h_{r}=v^{x i} v^{v j} \delta_{x r} \delta_{y r} \zeta^{z w} \zeta_{i j z w}, \quad m_{r}=v^{x i} v^{v j} v^{z k} v^{w m} \delta_{x r} \delta_{y r} \delta_{z r} \delta_{w r} \zeta_{i j k m} .
\end{gather*}
$$

The proof of Proposition 2 is given in Appendix 2. For the Huber $M$-estimator defined by

$$
\rho_{c}(\omega)= \begin{cases}\frac{1}{2} \omega^{2}, & \text { for }|\omega| \leqslant c,  \tag{21}\\ c|\omega|-\frac{1}{2} c^{2}, & \text { otherwise },\end{cases}
$$

with $c>0$, Proposition 2 applies with the specific values given by Corollary 2.
Corollary 2 (Huber $M$-estimator). For the Huber $M$-estimator, the terms (19) and (20) in Proposition 2 are given by

$$
\begin{gathered}
c_{\psi_{c}}=\left\{\frac{1}{E\left(\psi_{c}^{\prime}\right)}\right\}\left\{\frac{1}{E\left(\psi_{c}^{2}\right)}\right\}\left[\left\{\frac{E\left(\psi_{c}^{2}\right)}{E\left(\psi_{c}^{\prime}\right)}\right\} E\left(\psi_{c}^{\prime}\right)-E\left(\psi_{c}^{\prime} \psi_{c}^{2}\right)\right], \\
e_{\psi_{c}}=\left\{\frac{1}{E\left(\psi_{c}^{2}\right)}\right\}^{2}\left[E\left(\psi_{c}^{4}\right)+3\left\{\frac{E\left(\psi_{c}^{2}\right)}{E\left(\psi_{c}^{\prime}\right)}\right\}^{2} E\left(\psi_{c}^{\prime}\right)-6\left\{\frac{E\left(\psi_{c}^{2}\right)}{E\left(\psi_{c}^{\prime}\right)}\right\} E\left(\psi_{c}^{\prime} \psi_{c}^{2}\right)\right],
\end{gathered}
$$

respectively, where $\psi_{c}(\omega)=\max \{-c, \min (\omega, c)\}$.
The proof of Corollary 2 is given in Appendix 2. The previous expansions can be used in the following formulae for approximating significance or critical values.

For a scalar parameter of interest, that is $q=p-1$, under the conditions of Proposition 2,

$$
R_{n}=\operatorname{sgn}\left(T_{n}^{p}\right) W_{p-1, n}^{\frac{1}{2}}=J_{\hat{\psi}}^{\frac{1}{2}} U_{p}+O_{p}\left(n^{-3 / 2}\right),
$$

where $J_{\psi}=E\left(\psi^{2}\right) / E\left(\psi^{\prime}\right)$.
If we write $\phi=\Phi^{\prime}$, the second-order Edgeworth expansion of $\operatorname{pr}\left(R_{n} \leqslant r\right)$ is given by

$$
\operatorname{pr}\left(U_{p} \leqslant x\right)=\Phi(x)-n^{-1} \phi(x)\left\{\frac{1}{2} c_{\psi} h_{p} x+\frac{1}{24} e_{\psi} m_{p}\left(x^{3}-3 x\right)\right\}+O\left(n^{-3 / 2}\right),
$$

at $x=r J_{\psi}^{-1 / 2}$. The Cornish-Fisher expansion provides an approximation to the $\beta$ th quantile of $R_{n}$; it is given by

$$
r(\beta)=J_{\psi \psi}^{\frac{1}{\psi}} z_{\beta}+n^{-1} J_{\psi}^{\frac{1}{2}}\left\{\frac{1}{2} c_{\psi} h_{p} z_{\beta}+\frac{1}{24} e_{\psi} m_{p}\left(z_{\beta}^{3}-3 z_{\beta}\right)\right\}+O\left(n^{-3 / 2}\right),
$$

where $z_{\beta}$ is the $\beta$ th quantile of the standard normal. Note that the $n^{-1 / 2}$ terms do not appear because of symmetry assumptions in $\psi$; see Condition (B2). In this situation, the
validation of the Edgeworth expansion, for a statistic for which we can obtain expansions to its cumulants with sufficiently small errors, is established by Bhattacharya \& Ghosh (1978).

Alternatively, the saddlepoint approximation has a bounded relative error and leads therefore to very accurate approximations of small tail probabilities; see Field \& Tingley (1997) for a survey. Define

$$
\begin{equation*}
H_{n}(\alpha)=\frac{1}{2}\left(n+c_{\psi} h_{p}\right) \alpha^{2}+\frac{1}{24} n^{2} e_{\psi} m_{p} \alpha^{4} . \tag{22}
\end{equation*}
$$

The saddlepoint approximation of $\operatorname{pr}\left(R_{n} \leqslant r\right)$ is derived by applying the method of Easton \& Ronchetti (1986), and it is given by

$$
\operatorname{pr}\left(U_{p} \leqslant x\right)=\left\{\Phi\left(r_{x}\right)-\phi\left(r_{x}\right)\left(\frac{1}{s_{x}}-\frac{1}{r_{x}}\right)\right\}\left\{1+O\left(n^{-1}\right)\right\},
$$

at $x=r J_{\psi}^{-1 / 2}$, where

$$
s_{x}=\alpha_{x}\left\{n H_{n}^{\prime \prime}\left(\alpha_{x}\right)\right\}^{\frac{1}{2}}, \quad r_{x}=\operatorname{sgn}\left(\alpha_{x}\right)\left[2 n\left\{\alpha_{x} x-H_{n}\left(\alpha_{x}\right)\right\}\right]^{\frac{1}{2}},
$$

and $\alpha_{x}$ is the saddlepoint defined as the solution of the equation

$$
\begin{equation*}
H_{n}^{\prime}(\alpha)=x . \tag{24}
\end{equation*}
$$

Here $H_{n}$ is constructed from the fourth-order Taylor expansion of the logarithm of the characteristic function of $U_{p}$, and is based on the first four approximated cumulants given by (18). Hence this saddlepoint approximation does not require the existence of the cumulant generating function, as does the standard saddlepoint approximation of the sample mean. The following one-step algorithm proposed by Wang (1995) provides approximated quantiles with the same asymptotic error as we would obtain by inverting the saddlepoint probabilities. The steps are the following. We define $r_{x}^{*}=r_{x}+\log \left(s_{x} / r_{x}\right) / r_{x}$. We find the normal approximation to the $\beta$ th quantile $x^{(0)}=z_{\beta}\left(1+n^{-1} c_{\psi} h_{p}\right)^{\frac{1}{2}}$, and determine the starting value

$$
x^{(1)}=x^{(0)}+\frac{z_{\beta}^{2}-\left(r_{x^{*}(0)}^{*}\right)^{2}}{2 n x_{x^{(0)}}}
$$

and the increment

$$
\Delta\left(x^{(1)}\right)=\frac{z_{\beta}^{2}-\left(r_{x_{1}^{(1)}}^{*(1)}\right)^{2}}{2 n \alpha_{x^{(1)}}^{(1)}} .
$$

The one-step approximation to the $\beta$ th quantile of $R_{n}$ is $J_{\psi}^{1 / 2} x^{(2)}$, where $x^{(2)}=x^{(1)}+\Delta\left(x^{(1)}\right)$. It can be shown that, if $J_{\psi}^{1 / 2} x(\beta)$ denotes the exact $\beta$ th quantile, then $x^{(2)}=$ $x(\beta)\left\{1+O\left(n^{-3 / 2}\right)\right\}$. It was pointed out by Wang (1992) that (24) could have multiple solutions and to avoid this problem he proposed the modification of $H_{n}(\alpha)$ of the form

$$
\tilde{H}_{n}(\alpha ; a)=\frac{1}{2}\left(n+c_{\psi} h_{p}\right) \alpha^{2}+\frac{1}{24} n^{2} e_{\psi} m_{p} \alpha^{4} \gamma_{n}(\alpha ; a),
$$

where

$$
\gamma_{n}(\alpha ; a)=\exp \left\{-\left(n+c_{\psi} h_{p}\right) a^{2} \alpha^{2} / 2\right\}, \quad a=\max \left[\frac{1}{2}, \inf \left\{b \mid \tilde{H}_{n}^{\prime \prime}(\alpha ; b)>0\right\}\right] .
$$

Replacing $H_{n}(\alpha)$ by $\tilde{H}_{n}(\alpha ; a)$ in (23) and (24) guarantees a unique solution of (24) and does not change the order of the approximations.

Another type of approximation is obtained by the variance-corrected statistic, now given by $\widetilde{R}_{n}=R_{n}\left[J_{\psi}\left\{1+\left(n^{-1} c_{\psi} h_{p}\right)\right\}\right]^{-\frac{1}{2}}$, which leads to the approximation

$$
\operatorname{pr}\left(R_{n} \leqslant r\right) \bumpeq \Phi\left(r\left[J_{\psi}\left\{1+\left(n^{-1} c_{\psi} h_{p}\right)\right\}\right]^{-\frac{1}{2}}\right) .
$$

It must however be stressed that the asymptotic error here is $O\left(n^{-1}\right)$ and not $O\left(n^{-3 / 2}\right)$, as in the classical case. The reason for this is that Lawley (1956) made use of the Bartlett identities to prove that the third and fourth cumulants of $R_{n}$ are $O\left(n^{-3 / 2}\right)$ and $O\left(n^{-2}\right)$, respectively, while without these identities they are $O\left(n^{-1 / 2}\right)$ and $O\left(n^{-1}\right)$.

For several parameters of interest, we can define a Bartlett correction for the $M$-ratio statistic. By Proposition 2 we can define the Bartlett factor $b$ from the expansion

$$
E\left(\sum_{r=q+1}^{p} U_{r}^{2}\right)=(p-q)\left(1+\frac{b}{n}\right)+O\left(n^{-2}\right)
$$

which leads to

$$
b=\frac{1}{p-q} c_{\psi} \sum_{r=q+1}^{p} h_{r}
$$

Thus, the Bartlett-corrected approximation of the distribution of the $M$-ratio is

$$
\operatorname{pr}\left(W_{q n} \leqslant w\right)=\operatorname{pr}\left\{\chi_{p-q}^{2} \leqslant \frac{w}{J_{\psi}(1+b / n)}\right\}+O\left(n^{-1}\right)
$$

This result is obtained by following the same lines that give the classical result (2). It should however be noted that, for the same reasons mentioned in connection with the variance-corrected statistic, the error of the above expression would be $O\left(n^{-2}\right)$ in the classical situation.

## 3. Monte Carlo results

This section presents a numerical study of the Bartlett correction for the $M$-ratio statistic, and of the Edgeworth approximation and the variance correction with the signedroot $M$-ratio statistic. We consider the linear regression model and the Huber $M$-estimator. The errors $\varepsilon_{l}$ of the model are generated from the standard normal, the Student $t_{3}$, the contaminated normal $0 \cdot 9 \cdot \mathcal{N}(0,1)+0 \cdot 1 \cdot \mathcal{N}\left(0,10^{2}\right)$ and the Cauchy distributions, for which the 'tail weights' are $1 \cdot 00,1 \cdot 72,4 \cdot 93$ and $9 \cdot 22$, respectively; we measure the tail weight of a distribution function $F$ by

$$
\operatorname{TW}(F)=\frac{\left\{F^{-1}(0.99)-F^{-1}(0.5)\right\} /\left\{\Phi^{-1}(0.99)-\Phi^{-1}(0.5)\right\}}{\left\{F^{-1}(0.75)-F^{-1}(0.5)\right\} /\left\{\Phi^{-1}(0.75)-\Phi^{-1}(0.5)\right\}}
$$

All moments involving the error distribution are computed with respect to the normal distribution since in practice we do not know the true distribution. We consider $p=4$, $\xi_{l 1}=1$ and $\xi_{l 2}, \xi_{l 3}$ and $\xi_{l 4}$ as fixed regressors drawn from the $\operatorname{Un}(0,1)$ distribution. We used sample size $n=40,20,16,12$ or 8 and set $p-q$, the number of components of $\theta$ to be tested, equal to 1 , for $n=40,12$ or 8 , equal to 2 , for $n=20$, and equal to 1 or 2 for $n=16$. A standard computation shows that the number of draws used, 50000 , leads to sufficiently small standard errors for the Monte Carlo estimators of $p$-values, so that we can consider the Monte Carlo $p$-value as exact. The scale parameter $\sigma$ in (12) is estimated
simultaneously with $\theta$, by solving

$$
\frac{1}{n-p} \sum_{l=1}^{n} \psi_{c}^{2}\left(\frac{\eta_{l}-\theta^{1} \xi_{l 1}-\ldots-\theta^{p} \xi_{l p}}{\sigma}\right)-E\left\{\psi_{c}^{2}\left(\varepsilon_{1}\right)\right\}=0,
$$

where $\varepsilon_{1}$ is standard normal and $c=1.345$ (Huber, 1981, p. 177). For some of the scenarios, Fig. 1 shows the relative errors, as $\{($ approximated - simulated $) /$ simulated $\}$, for $p$-values ranging from 1 to 10 percent for a scalar null hypothesis. The solid line represents the

(e) Normal, $n=12$

(f) $t_{3}, n=12$

(i) Normal, $n=8$


(g) $\mathrm{CN}, n=12$

(k) $\mathrm{CN}, n=8$

(h) Cauchy, $n=12$

(1) Cauchy, $n=8$


Fig. 1. Relative errors versus $p$-values for a scalar null hypothesis. Errors are generated from standard normal, Student $t_{3}$, contaminated normal (CN) and Cauchy distributions, for 40, 12 and 8 observations. Solid lines, Bartlett correction; dotted lines, first-order approximation.

Bartlett correction and the dotted line represents the first-order approximation. The Bartlett correction generally improves on the first-order approximation, although it becomes inaccurate for extremely small $p$-values. With $n=40$, the asymptotic approximation still remains generally far from the simulated $p$-values, even at the underlying model. In contrast, the Bartlett correction provides very low errors. Figure 2 shows good approximations obtained with the Bartlett correction, for $p$-values equal to $5 \%$. The shaded bars represent the Bartlett correction and the white bars the first-order approximation. Regarding the robustness of the approximations, Fig. 3 shows small errors with the Bartlett correction at $p$-values equal to $5 \%$. The solid lines represent the Bartlett correction and the dotted lines the first-order approximation. The Bartlett correction deteriorates as $\mathrm{Tw}(F)$ increases, becoming inaccurate under the Cauchy distribution, which is extremely heavy-tailed. A comparison of Figs 2 and 3 suggests that the number of parameters $(p-q)$ to be tested should also be taken into account in the determination of the effective sample size.


Fig. 2. Absolute errors under various distributions at $p$-values equal to $5 \%$. Shaded bars, Bartlett correction; white bars, first-order approximation.

For the signed-root $M$-ratio statistic, we show comparisons between the first-order, the Edgeworth and the variance-corrected $p$-values. For $n=16$, Table 1 shows that both variance-corrected and Edgeworth approximations are accurate. Both improve significantly on the first-order approximation, as does the Bartlett correction.

In conclusion, provided that we are not approximating extremely small $p$-values or that a quantity like $n / p-(p-q-1)$ is not smaller than 3 , the use of a high-order approximation leads to substantially more accurate numerical results, for various underlying distributions of the regression errors, except for heavy-tailed ones like the Cauchy distribution.

## 4. Discussion

The previous section illustrates the advantage, in terms of numerical accuracy, of using our high-order approximations with robust test statistics. As in the classical setting, approximations with small asymptotic errors lead to small numerical errors. Since it is well established that robust methods lead to stable inference, with respect to model deviations (Markatou \& Ronchetti, 1997), the addition of high-order approximations does guarantee very reliable inference.


Fig. 3. Absolute errors versus $\operatorname{Tw}(F)$ at $p$-values equal to $5 \%$. Solid lines, Bartlett correction; dotted lines, first-order approximation.

Table 1. $p$-values $\times 100$ under four distributions, for $n=16$

| Normal |  |  |  | $t_{3}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $p$-value | FO | ED | vc | $p$-value | FO | ED | vc |
| $3 \cdot 0$ | $1 \cdot 6$ | $2 \cdot 9$ | $2 \cdot 8$ | $3 \cdot 0$ | $1 \cdot 9$ | $3 \cdot 2$ | $3 \cdot 1$ |
| $5 \cdot 0$ | 3.6 | 5.6 | $5 \cdot 5$ | $5 \cdot 0$ | 3.5 | $5 \cdot 4$ | $5 \cdot 3$ |
| $7 \cdot 0$ | $5 \cdot 3$ | 7.7 | $7 \cdot 6$ | $7 \cdot 0$ | 5.6 | $8 \cdot 1$ | 8.0 |
| 10.0 | 7.7 | 10.5 | $10 \cdot 4$ | 10.0 | 8.8 | $11 \cdot 8$ | 11.8 |
| Contaminated normal |  |  |  | Cauchy |  |  |  |
| $p$-value | FO | ED | vc | $p$-value | FO | ED | vc |
| $3 \cdot 0$ | $1 \cdot 6$ | $2 \cdot 9$ | $2 \cdot 7$ | $3 \cdot 0$ | $2 \cdot 2$ | $3 \cdot 8$ | $3 \cdot 6$ |
| $5 \cdot 0$ | $3 \cdot 1$ | 5.0 | $4 \cdot 8$ | $5 \cdot 0$ | 3.6 | 5.6 | 5. |
| $7 \cdot 0$ | $4 \cdot 6$ | 7.0 | 6.9 | $7 \cdot 0$ | $5 \cdot 1$ | 7.5 | $7 \cdot 4$ |
| 10.0 | $7 \cdot 3$ | 10.0 | $10 \cdot 0$ | 10.0 | $7 \cdot 9$ | $10 \cdot 8$ | 10 |

Approximations: fo, first-order; Ed, Edgeworth; vc, variance correction.

S-Plus routines for computing the $p$-values of the $M$-ratio and its signed-root test statistics are described in F.-X. de Rossi's 1996 Ph.D. thesis from the University of Geneva. We have focused on the linear regression model, but computer algebra (Andrews \& Stafford, 1993) would allow generalisation of the expansions to more complicated situations, such as generalised linear models or nonlinear models, for which the amount of algebraic manipulation, although simple in nature, becomes daunting.

## Acknowledgement

The authors are grateful to E. Ronchetti for many discussions and suggestions, and to T. DiCiccio, C. Field, J.-P. Imhof, P. Stalder, the editor and an anonymous referee for many helpful comments which improved the quality of this paper. The research of F.-X. de Rossi was supported by the University of Geneva.

## Appendix 1

Regularity conditions
The sample space $R \subset \mathbb{R}^{k}$ is an open convex set and $\mathscr{G}$ denotes the metric space of probability distributions on $R$ with distance $d$. The parameter space $\Theta \subset \mathbb{R}^{p}$ is an open convex set.

Conditions for the existence of $M$-estimators
Condition (A1a). The function $\varrho: R \times \Theta \rightarrow \mathbb{R}$ is differentiable up to the fifth order with respect to $\theta \in \Theta$.

Condition (A1b). The function $-\varrho$ is strictly convex with respect to $\theta$.
Condition (A1c). Given $\theta \in \Theta$, there exists a compact neighbourhood $D$ of $\theta$ such that the sign of $\Psi_{r}(z, \tau)$ is not constant on $D$, for all $z \in R$.

## Integrability

Condition (A2). Given $\theta \in \Theta$, there exists a compact neighbourhood $D$ of $\theta$ such that, for all $\tau \in D$,

$$
\left\|d_{\Psi}(z, \tau)\right\| \leqslant h(z)
$$

where $d_{\Psi}: R \times \Theta \rightarrow \mathbb{R}$ is any product of functions (6), and where $h: R \rightarrow \mathbb{R}$ is some function that is integrable with respect to all $G \in B\left(F_{\theta}, \varepsilon\right)$, for some $\varepsilon>0, B\left(F_{\theta}, \varepsilon\right)$ denoting an open ball and $\|\cdot\|$ the Euclidean norm.

Fisher consistency
Condition (A3). We require that $\int_{R} \Psi_{r}(z, \theta) d F_{\theta}(z)=0$, for all $\theta \in \Theta$.
Continuity at $F_{\theta}$
Condition (A4). For all $\delta>0$, there exists $\varepsilon>0$ such that, for all $G \in B\left(F_{\theta}, \varepsilon\right)$,

$$
\sup _{\tau \in D}\left\|\int_{R} d_{\Psi}(z, \tau) d G(z)-\int_{R} d_{\Psi}(z, \tau) d F_{\theta}(z)\right\|<\delta
$$

for any $d_{\Psi}$.
Fréchet differentiability at $F_{\theta}$
Condition (A5a). The matrix whose $(r, s)$ element is $\int_{R} \Psi_{r s}(z, \theta) d F_{\theta}(z)$ is nonsingular.
Condition (A5b). There exists a neighbourhood $\mathscr{N}_{F_{\theta}}$ of $F_{\theta}$ such that, for all $G \in \mathscr{N}_{F_{\theta}}$,

$$
\int_{R} \Psi_{r}(z, \tau) d\left(G-F_{\theta}\right)(z)=O\left\{d\left(G, F_{\theta}\right)\right\} .
$$

Central limit theorem condition
Condition (A6). We require $d\left(F_{n}, F_{\theta}\right)=O_{p}\left(n^{-1 / 2}\right)$ for the empirical distribution $F_{n}$ of a sample $Z_{1}, \ldots, Z_{n}$ with individual distribution $F_{\theta}$.

Regression model
Condition (B1). The function $g: \mathbb{R}^{p} \times \Theta \rightarrow \mathbb{R}$ is four-fold differentiable with respect to $\theta$.
Condition (B2). The function $\rho: \mathbb{R} \rightarrow \mathbb{R}^{+}$is an even function which satisfies

$$
\varrho\left(\eta, \xi_{1}, \ldots, \xi_{p}, \theta\right)=-\rho\left\{\eta-g\left(\xi_{1}, \ldots, \xi_{p}, \theta\right)\right\} .
$$

## High-order Fréchet differentiability

Condition (C1). Given $\theta \in \Theta$, the $M$-functional $T_{\Psi}: \mathscr{G} \rightarrow \mathbb{R}$ which solves $\int_{R} \Psi_{r}(z, \tau) d G(z)=0$ is Fréchet differentiable up to the third order at $F_{\theta}$.

Condition (C2). Given $\theta \in \Theta$, the $M$-functional $T_{\Psi}$ is Fréchet differentiable up to the fourth order at $F_{\theta}$.

Condition (A1a) is a smoothness condition for an elementary treatment of Taylor expansions involving $\varrho$. Conditions (A1b) and (A1c) ensure uniqueness and existence of the $M$-estimator. Uniqueness and continuity are proved under weaker conditions by Clarke (1983) and even by relaxing differentiability assumptions (Clarke, 1986). Condition (A2) ensures integrability for the moments of interest, and allows the interchange of differentiation and integration. Condition (A3) is the Fisher consistency for the $M$-functional $T_{\Psi}$, and Condition (A4) is a continuity condition that allows us to establish the validity of formal expansions for the $M$-ratio statistic. It can be derived under some conditions on the integrand; see Clarke (1983) and F.-X. de Rossi's thesis. Conditions (A5a) and (A5b) are conditions for Fréchet differentiability, and Condition (A6) is a condition for the central limit theorem, holding for the Kolmogorov distance through the Dvoretzky, Kiefer and Wolfowitz inequality (Serfling, 1980, pp. 59-61). Conditions (A2)-(A5b) include those of Clarke (1983), who demonstrated that they imply the Fréchet differentiability of the $M$-functional $T_{\Psi}$. Since other tractable conditions implying high-order Fréchet differentiability are still an open question, Conditions (C1) and (C2) assume the required differentiability for our propositions. Fréchet differentiability together with Condition (A6) implies the asymptotic normality of an $M$-estimator in a neighbourhood of the parametric model; see Bednarski (1993).

## Appendix 2

## Proofs

Proof of Proposition 1. By Taylor expansion,

$$
\begin{aligned}
\frac{1}{2} W_{0 n}= & \Delta^{r} Y_{r}+\frac{1}{2} \Delta^{r} \Delta^{s} \kappa_{r s}+n^{-\frac{1}{2}}\left(\frac{1}{2} \Delta^{r} \Delta^{s} Y_{r s}+\frac{1}{6} \Delta^{r} \Delta^{s} \Delta^{t} \kappa_{r s t}\right) \\
& +n^{-1}\left(\frac{1}{6} \Delta^{r} \Delta^{s} \Delta^{t} Y_{r s t}+\frac{1}{24} \Delta^{r} \Delta^{s} \Delta^{t} \Delta^{u} \kappa_{r s t u}\right)+O_{p}\left(n^{-3 / 2}\right)
\end{aligned}
$$

where $\Delta^{r}=B^{r}+n^{-1 / 2} C^{r}+n^{-1} D^{r}+O_{p}\left(n^{-3 / 2}\right)$, and $B^{r}$ and $C^{r}$ are given by Lemma 1 . By use of (7), an explicit expression for $D^{r}$ is not necessary, and we can obtain

$$
\begin{aligned}
\frac{1}{2} W_{0 n}= & \frac{1}{2} B^{r} Y_{r}+n^{-\frac{1}{2}}\left(\frac{1}{2} B^{r} B^{s} Y_{r s}+\frac{1}{6} B^{r} B^{s} B^{t} \kappa_{r s t}\right) \\
& +n^{-1}\left\{\frac{1}{2}\left(C^{r} C^{s} \kappa_{r s}+2 B^{r} C^{s} Y_{r s}+B^{r} B^{s} C^{t} \kappa_{r s t}\right)+\frac{1}{6} B^{r} B^{s} B^{t} Y_{r s t}+\frac{1}{24} B^{r} B^{s} B^{t} B^{u} \kappa_{r s t u}\right\}+O_{p}\left(n^{-3 / 2}\right) .
\end{aligned}
$$

The use of (8) leads to

$$
\begin{aligned}
W_{0 n}=B^{r} Y_{r}+ & n^{-\frac{1}{2}}\left(\frac{1}{3} B^{r} B^{s} B^{t} \kappa_{r s t}+B^{r} B^{s} Y_{r s}\right) \\
+ & n^{-1}\left\{\frac{1}{12} B^{r} B^{s} B^{t} B^{u} \kappa_{r s t u}+\frac{1}{4}\left(-\kappa^{s u}\right) B^{v} B^{w} B^{r} B^{t} \kappa_{u v w} \kappa_{r s t}\right. \\
& \left.+\left(-\kappa^{s u}\right) B^{v} B^{r} B^{t} Y_{u v} \kappa_{r s t}+\frac{1}{3} B^{r} B^{s} B^{t} Y_{r s t}+\left(-\kappa^{s u}\right) B^{v} B^{r} Y_{u v} Y_{r s}\right\}+O_{p}\left(n^{-3 / 2}\right),
\end{aligned}
$$

and replacing $B^{r}$ by (7) gives (9). By standard transformations from moments to cumulants (McCullagh, 1987, § 2.3), (10) is proved.

Proof of Proposition 2. Write $W_{q n}=W_{0 n}=W_{\text {oqn }}$, where $W_{0 q n}$ is defined by replacing $e_{l}$ by $e_{q l}$ in (13). By (9), we can obtain

$$
W_{0 n}=\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\} \zeta^{r s} S_{r} S_{s}+O_{p}\left(n^{-3 / 2}\right)
$$

By a similar computation for $W_{o q n}$, we can obtain (14) and (16). We denote by $\omega_{r}^{*}, \omega_{r, s}^{*}, \ldots$ the cumulants of $S_{r}$. From (15) we have, for instance,

$$
\begin{aligned}
\omega_{r, i}^{*}= & n^{-\frac{1}{2}} \frac{1}{2}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}\left(\zeta^{j k} \kappa_{r,(i j, k)}^{*}+\zeta^{s t} \kappa_{i,(r s, t)}^{*}\right)+n^{-1} \frac{1}{4}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2} \zeta^{j k} \zeta^{s t} \kappa_{(i, k),(r s, t)}^{*} \\
& +n^{-1} \frac{1}{24}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{3} E\left(-\psi^{\prime \prime \prime}\right)\left(\zeta_{i y} \zeta^{v j k m} \kappa_{r,(j, k, m)}^{*}+\zeta_{r x} \zeta^{x s t u} \kappa_{i,(s, t, u)}^{*}\right) \\
& +n^{-1} \frac{1}{6}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2}\left(\zeta^{j m} \zeta^{k n} \kappa_{r,(i j k, m, n)}^{*}+\zeta^{s u} \zeta^{t v} \kappa_{i,(r s t, u, v)}^{*}\right) \\
& +n^{-1} \frac{3}{8}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}^{2}\left(\zeta^{j k} \zeta^{m n} \kappa_{r,(i j, k m, n)}^{*}+\zeta^{s t} \zeta^{u v} \kappa_{i,(r s, t u, v)}^{*}\right)+O_{p}\left(n^{-3 / 2}\right)
\end{aligned}
$$

where $\kappa_{r,(i j, k)}^{*}=\operatorname{cov}\left(Y_{r}, Y_{i j} Y_{k}\right), \ldots$ are the cumulants associated with the $Y_{r}, Y_{r s}$ and $Y_{r s t}$. By applying transformations from generalised to simple cumulants (McCullagh, 1987, Ch. 4), we may obtain

$$
\begin{gathered}
\omega_{r}^{*}=O\left(n^{-3 / 2}\right), \\
\begin{array}{c}
\omega_{r, i}^{*}=E\left(\psi^{2}\right) \zeta_{r i}+n^{-1}\left\{\frac{1}{E\left(\psi^{\prime}\right)}\right\}\left[E\left(-\psi^{\prime} \psi^{2}\right)+\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\} E\left(\psi^{\prime} \psi^{\prime}\right)+\frac{1}{4}\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{2} E\left(-\psi^{\prime \prime \prime}\right)\right. \\
\\
\left.+\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\} E\left(\psi^{\prime \prime} \psi\right)\right] \zeta^{s t} \\
=\zeta_{r i s t}+O\left(n^{-2}\right), \\
\omega_{r, i, j, k}^{*}=n^{-1}\left[E\left(\psi^{4}\right)+6\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\} E\left(-\psi^{\prime} \psi^{2}\right)+3\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{2} E\left(\psi^{\prime} \psi^{\prime}\right)\right. \\
\\
\left.\quad+\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{3} E\left(-\psi^{\prime \prime \prime}\right)+4\left\{\frac{E\left(\psi^{2}\right)}{E\left(\psi^{\prime}\right)}\right\}^{2} E\left(\psi^{\prime \prime} \psi\right)\right] \zeta_{r, i, j, k}+O\left(n^{-2}\right), \\
\omega_{r, i, j, k, m}^{*}=O\left(n^{-3 / 2}\right), \quad \omega_{r, i, j, k, m, n}^{*}=O\left(n^{-3 / 2}\right), \quad \omega_{r, s, i, j, k, m, n}^{*}=O\left(n^{-3 / 2}\right),
\end{array}
\end{gathered}
$$

and so on. Since by (17) the $U_{r}$ are linear combinations of the $S_{r}$, standard transformations of cumulants lead to (18).

Proof of Corollary 2. We need to give only the reasons for which Corollary 2 remains valid even though some of the hypotheses of Proposition 2 regarding the differentiability of $\psi_{c}$ are not satisfied. Here, we define $\psi_{c}$ implicitly in the following way, using the theory of mathematical distributions. We consider a sequence of differentiable functions $\rho_{m}: \mathbb{R} \rightarrow \mathbb{R}^{+}$with $\psi_{m}=\rho_{m}^{\prime}$ and $\rho_{c}=\lim _{m \rightarrow \infty} \rho_{m}$, such that the real-valued functionals of $F, \gamma\left(F ; \psi_{m}\right)=\int \psi_{m} d F$ and $\lim _{m \rightarrow \infty} \gamma\left(F ; \psi_{m}\right)$, exist for every $F$ in some functional neighbourhood. Then we define $\psi_{c}$, the derivative of $\rho_{c}$, implicitly by $\gamma\left(F ; \psi_{c}\right)=$ $\lim _{m \rightarrow \infty} \gamma\left(F ; \psi_{m}\right)$. We can select $\psi_{m}$ as an odd function and, since the standard normal density is
an even function, by integration by parts we can show for instance that

$$
E\left(\psi_{m}^{\prime \prime \prime}\right)=\int_{-\infty}^{\infty} \psi_{m}^{\prime \prime \prime}(r) \phi(r) d r=-\int_{-\infty}^{\infty} \psi_{m}(r) \phi^{\prime \prime \prime}(r) d r=0
$$

With $m \rightarrow \infty, E\left(\psi_{c}^{\prime \prime \prime}\right)=0$, which justifies Corollary 2.

## References

Andrews, D. F. \& Stafford, J. E. (1993). Tools for the symbolic computation of asymptotic expansions. J. R. Statist. Soc. B 55, 613-27.

Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. Biometrika 78, 557-63.
Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proc. R. Soc. Lond. A 160, 268-82.
Bednarski, T. (1993). Fréchet differentiability of statistical functionals and implications to robust statistics. In New Directions in Statistical Data Analysis and Robustness, Ed. S. Morgenthaler, E. Ronchetti and W. A. Stahel, pp. 26-34. Berlin: Birkhäuser.
Bhattacharya, R. N. \& Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. Ann. Statist. 6, 434-51.
Clarke, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. Ann. Statist. 11, 1196-205.
Clarke, B. R. (1986). Non-smooth analysis and Fréchet differentiability of $M$-functionals. Prob. Theory Rel. Fields 73, 197-209.
DiCiccio, T. J. (1984). On parameter transformation and interval estimation. Biometrika 71, 477-85.
DiCiccio, T. J. \& Martin, M. A. (1993). Simple modifications for signed roots of likelihood ratio statistics. J. R. Statist. Soc. B 55, 305-16.

Easton, G. S. \& Ronchetti, E. (1986). General saddlepoint approximations with applications to $L$ statistics. J. Am. Statist. Assoc. 81, 420-30.

Fan, R. Y. K. \& Field, C. A. (1995). Approximations for marginal densities of M-estimators. Can. J. Statist. 23, 185-97.
Field, C. A. \& Tingley, M. A. (1997). Small sample asymptotics: applications in robustness. In Handbook of Statistics, 15, Ed. G. S. Maddala and C. R. Rao, pp. 513-36. Amsterdam: North Holland.
Gatto, R. \& Ronchetti, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. J. Am. Statist. Assoc. 91, 666-73.
Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. \& Stahel, W. A. (1986). Robust Statistics. The Approach Based on Influence Functions. New York: Wiley and Sons.
Heritier, S. \& Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. J. Am. Statist. Assoc. 89, 897-904.
Huber, P. J. (1981). Robust Statistics. New York: Wiley and Sons.
Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. Biometrika 43, 295-303.
Markatou, M. \& Ronchetti, E. (1997). Robust inference: the approach based on influence functions. In Handbook of Statistics, 15, Ed. G. S. Maddala and C. R. Rao, pp. 49-75. Amsterdam: North Holland.
McCullagh, P. (1987). Tensor Methods in Statistics. London: Chapman and Hall.
McCullagh, P. \& Cox, D. R. (1986). Invariants and likelihood ratio statistics. Ann. Statist. 14, 1419-30.
Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. New York: Wiley and Sons. Wang, S. (1992). General saddlepoint approximations in the bootstrap. Statist. Prob. Lett. 13, 61-6.
Wang, S. (1995). One-step saddlepoint approximations for quantiles. Comp. Statist. Data Anal. 20, 65-74.
[Received January 1999. Revised March 2001]

