AISM (2008) 60:61–83
DOI 10.1007/s10463-006-0085-1

Accurate confidence intervals in regression analyses of non-normal data

Robert J. Boik

Received: 25 February 2005 / Revised: 26 June 2006 / Published online: 18 October 2006

Abstract A linear model in which random errors are distributed independently and identically according to an arbitrary continuous distribution is assumed. Second- and third-order accurate confidence intervals for regression parameters are constructed from Charlier differential series expansions of approximately pivotal quantities around Student's t distribution. Simulation verifies that small sample performance of the intervals surpasses that of conventional asymptotic intervals and equals or surpasses that of bootstrap percentile-t and bootstrap percentile-$|t|$ intervals under mild to marked departure from normality.

Keywords Bootstrap · Charlier differential series · Cornish-Fisher transformation · Edgeworth expansion · Kurtosis · One-sample t · Skewness

1 Introduction

Non-normality of the parent population can greatly influence type I error rate of t tests as well as coverage of confidence intervals for means, especially if sample size is small (Pearson and Please, 1975, Bowman et al., 1977, Posten, 1979, Cressie, 1980). If sample size is large and departure from normality is not too severe, however, then the t test might still be used. Ractliffe (1968), Posten (1979), and Cressie et al. (1984) provided guidelines for this strategy. If the distribution has heavy tails or if outliers are present then a procedure based on trimmed means or on a robust estimating function (e.g., M or S) could be employed (Rousseeuw and Leroy, 1987, Staudte and Sheather, 1990, Wilcox,

R. J. Boik (✉)
Department of Mathematical Sciences, Montana State University,
Bozeman, MT 59717-2400, USA
e-mail: rjboik@math.montana.edu

1997, 1998). These robust procedures enable investigators to make accurate inferences about location parameters under widely varying conditions, but the location parameters are not necessarily means.

If interest truly is in means, and the tails of the distribution are not too heavy, then a compromise approach – something between the robust methods and the ordinary t test – might be reasonable. Johnson (1978) proposed a modified one-sample t test based on a Cornish-Fisher transformation (Cornish and Fisher, 1937). Johnson's test is not strongly affected by population skewness, provided that sample size is not too small. The test statistic, however, is not a monotonic function of t. Accordingly, the test cannot be inverted to produce a confidence interval with arbitrary confidence coefficient. Hall (1992a) proposed a monotonic cubic transformation to correct this deficiency.

In this article, the methods of Johnson (1978) and Hall (1992a) are extended to the linear models setting. Specifically, it is assumed that a linear regression function $\mathrm{E}(Y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ holds, where $\mathbf{x}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of regression coefficients. Random errors are assumed to be identically, but not necessarily normally, distributed. Section 2 summarizes the conventional asymptotic inference procedures and develops notation that is used throughout the article. Cumulants of $V = (\widehat{\psi} - \psi)/\,\mathrm{SE}(\widehat{\psi})$ are derived in Sect. 3, where $\psi = \mathbf{c}'\boldsymbol{\beta}$ is estimable, $\mathrm{SE}(\widehat{\psi})$ is an estimator of $\sqrt{\mathrm{Var}(\widehat{\psi})}$, and $\widehat{\psi}$ is a linear estimator of ψ. The cumulants are used in Sect. 4 to obtain Charlier differential series expansions of the distribution of V as well as monotonic approximations to generalized Cornish–Fisher transformations of V. The transformations are employed in Sect. 5 to construct second- and third-order accurate one- and two-sided confidence intervals for ψ. The intervals are illustrated in Sect. 6. Section 7 summarizes a simulation study that compares the performance of the proposed intervals to that of normal-theory and bootstrap intervals. A Matlab function for computing the proposed intervals as well as a supplement that contains proofs and numerical algorithms can be downloaded from <http://www.math.montana.edu/~rjboik/interests.html>.

2 Notation and background

2.1 Model and assumptions

The linear model under consideration is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Y}$ is an N-vector of responses, $\mathbf{X}$ is an $N \times p$ matrix of fixed explanatory variables, $\mathrm{rank}(\mathbf{X}) = r \leq p$, $\boldsymbol{\beta}$ is a p-vector of regression coefficients, and $\boldsymbol{\varepsilon}$ is an N-vector of random deviations. Estimability of $\mathbf{c}'\boldsymbol{\beta}$ requires that $\mathbf{c} \in \mathcal{R}(\mathbf{X}')$, where $\mathcal{R}(\mathbf{X}')$ is the vector space generated by the columns of $\mathbf{X}'$. The following assumptions are made about $\boldsymbol{\varepsilon}$:

$$\begin{aligned}&\text{(a) } \{\boldsymbol{\varepsilon}_i\}_{i=1}^N \text{ are iid with mean 0 and positive variance } \sigma^2 < \infty,\\ &\text{(b) } \mathrm{E}|\varepsilon_i|^8 < \infty, \text{ and}\\ &\text{(c) } \limsup_{||\boldsymbol{\xi}||\to\infty} |\phi_{\mathbf{W}}(\boldsymbol{\xi})| < 1, \text{ where } \mathbf{W} = (\varepsilon_i \ \ \varepsilon_i^2)',\end{aligned} \tag{2}$$

and $\phi_{\mathbf{W}}(\boldsymbol{\xi})$ is the characteristic function of $\mathbf{W}$. It is likely that (2b) can be weakened. Hall (1987) showed that the Edgeworth expansion of the one sample t with remainder $o(n^{-k/2})$ requires only $k+2$ rather than $2(k+2)$ finite moments. Hall's result, however, is applicable to studentized sums of iid random variables, whereas $(\widehat{\psi}-\psi)/\,\mathrm{SE}(\widehat{\psi})$ is a studentized sum of independent but not identically distributed random variables.

Under (2), the best linear unbiased estimator of ψ is $\widehat{\psi} = \mathbf{c}'\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$, and $(\mathbf{X}'\mathbf{X})^{-}$ is any generalized inverse of $\mathbf{X}'\mathbf{X}$. Properties of $\widehat{\psi}$ include the following:

$$\begin{aligned}&\sqrt{n}\left(\widehat{\psi}-\psi\right) = \mathbf{b}'\boldsymbol{\varepsilon}, \quad \mathrm{E}(\widehat{\psi}) = \psi \quad \text{and} \quad \mathrm{Var}(\widehat{\psi}) = n^{-1}\sigma_\psi^2, \text{ where}\\ &\mathbf{b} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{c}\sqrt{n}, \quad \sigma_\psi^2 = q_0\sigma^2, \quad q_0 = \mathbf{b}'\mathbf{b} = n\mathbf{c}'\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{c}, \text{ and } n = N-r.\end{aligned} \tag{3}$$

The conventional unbiased estimator of σ_ψ^2 is

$$\widehat{\sigma}_\psi^2 = q_0 S^2, \quad \text{where} \quad S^2 = n^{-1}\mathbf{Y}'\mathbf{A}\mathbf{Y}, \quad \text{and} \quad \mathbf{A} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'. \tag{4}$$

Scaling $\widehat{\psi}-\psi$ by $\mathrm{SE}(\widehat{\psi}) = \widehat{\sigma}_\psi/\sqrt{n}$ yields an approximately pivotal quantity,

$$V \stackrel{\text{def}}{=} \sqrt{n}(\widehat{\psi}-\psi)/\widehat{\sigma}_\psi, \tag{5}$$

which is denoted by V rather than T to avoid the implication that it has a t distribution.

Denote the ith column of $\mathbf{X}'$ by $\mathbf{x}_i$, and denote the smallest non-zero singular value of $\mathbf{X}$ by d_N. The following assumptions are made about $\mathbf{X}$:

$$\begin{aligned}&\text{(a) } \mathcal{R}(\mathbf{X}') \text{ is invariant } \forall\ N > r, \text{ where } r = \mathrm{rank}(\mathbf{X}) \text{ is constant,}\\ &\text{(b) } \lim_{N\to\infty} \min_i a_{ii} = 1, \text{ where } \mathbf{A} = \{a_{ij}\} \text{ is defined in (4),}\\ &\text{(c) } \mathbf{1}_N \in \mathcal{R}(\mathbf{X}), \text{ where } \mathbf{1}_N \text{ is an } N\text{-vector of ones, and}\\ &\text{(d) } M_N = O(n^h), \text{ where } M_N = \max_i ||\mathbf{x}_i|| \text{ and } h \in [0, 1/2),\\ &\text{(e) } \limsup_{N\to\infty} N^{-1}\sum_{i=1}^N ||\mathbf{x}_i||^5 < \infty, \text{ and}\\ &\text{(f) } \liminf_{N\to\infty} d_N^2/N > 0.\end{aligned} \tag{6}$$

Assumption (6a) ensures that the space of estimable functions of $\boldsymbol{\beta}$ does not depend on N and that $\mathrm{rank}(\mathbf{X})$ remains constant. Assumption (6b) ensures asymptotic normality of $\sqrt{n}(\widehat{\psi}-\psi)$ for any estimable function Huber

(1973). Assumption (6c) ensures that $\widehat{\psi}$ is location equivariant. Without loss of generality, the first column of $\mathbf{X}$ is taken to be $\mathbf{1}_N$ and, therefore, the first entry in $\boldsymbol{\beta}$, say β_0, is an intercept parameter. Assumptions (6d–f) are parts of assumptions C2–C3 in Yanagihara (2003) and ensure that the coefficients of the Hermite polynomials in the Edgeworth expansions are finite and that the jth cumulant of V in (5) has magnitude $O(n^{-(j-2)/2})$ for $j = 3, 4$.

The following quantities are needed in later sections:

$$\mathbf{a} = (a_{11}, \ldots, a_{NN})', \quad q_1 = \frac{\mathbf{a}'\mathbf{b}}{\sqrt{nq_0}}, \quad q_2 = \frac{\mathbf{a}'\mathbf{a}}{n}, \quad q_3 = \frac{1}{q_0}\sum_{i=1}^{N} a_{ii}b_i^2,$$
$$q_4 = \frac{\sqrt{n}}{3q_0^{3/2}}\sum_{i=1}^{N} b_i^3, \quad q_5 = \frac{n}{3q_0^2}\sum_{i=1}^{N} b_i^4, \quad q_6 = \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^3, \quad q_7 = \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^4,$$
$$q_8 = \frac{1}{q_6\sqrt{nq_0}}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^3 b_j, \quad q_9 = 3q_5 - q_2 - 2q_3, \quad \text{and} \quad q_{10} = 2q_3 - q_2 - q_5, \tag{7}$$

where a_{ij} is defined in (4). It is readily shown that $q_i = O(1)$ for $i \geq 0$.

Under (2) and (6), the central limit theorem together with Slutsky's theorem (Sen and Singer, 1993, Sects. 3.3–3.4) imply that $V \xrightarrow{\text{dist}} \mathrm{N}(0, 1)$ as $N \to \infty$, where V is defined in (5). More precisely, the cumulative distribution function (cdf) of V in (5) satisfies

$$\Phi_V(v) \stackrel{\text{def}}{=} P(V \leq v) = \Phi_{T_n}(v) + O\left(n^{-\frac{1}{2}}\right), \tag{8}$$

where $\Phi_{T_n}(v)$ is the cdf of T_n, a central t random variable with $df = n$. Equation (8) justifies the conventional upper, lower, and two-sided symmetric confidence intervals;

$$\left(-\infty, \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}} t_\alpha\right), \; \left(\widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}} t_{1-\alpha}, \infty\right) \quad \text{and}$$
$$\left(\widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}} t_{1-\alpha/2}, \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}} t_{\alpha/2}\right), \tag{9}$$

respectively, where t_α is the 100α percentile of the t distribution with n degrees of freedom.

2.2 Accuracy of interval estimators

Let $I = (L, U)$ be an interval estimator of ψ with nominal coverage $1 - \alpha$, where L and/or U are functions of $\mathbf{Y}$, and $\alpha \in (0, 1)$. Exact one-sided intervals satisfy $P(\psi \leq U) = 1 - \alpha$ or $P(L \leq \psi) = 1 - \alpha$. Exact two-sided equal-tailed intervals satisfy $P(L \le \psi) = P(\psi \le U) = 1 - \alpha/2$. A two-sided interval is symmetric if $L = \widehat{\psi} - C$ and $U = \widehat{\psi} + C$, where C is a non-negative function of $\mathbf{Y}$. If the distribution of V is skewed, then an exact symmetric interval cannot be equal-tailed. An interval, I, is said to have kth-order accurate coverage if $P(\psi \in I) = 1 - \alpha + O(n^{-k/2})$ uniformly in $\alpha \in (\epsilon, 1-\epsilon)$ (Hall, 1992b, Sect.3.5), where $\epsilon > 0$. The one-sided intervals in (9) have first-order accurate coverage and the two-sided symmetric interval in (9) has second-order accurate coverage.

The difference between the cdf of $Z_\psi = \sqrt{n}(\widehat{\psi} - \psi)/\sigma_\psi$ and the N(0, 1) cdf can be expressed either as $O(n^{-1/2})$ or as $O(N^{-1/2})$. The actual rate of convergence, however, varies depending on the degree to which $\mathbf{b}$ in (7) contains dominating large values. Under weaker assumptions about $\mathbf{X}$ than those in (6), Navidi (1989) showed that the error in the two term Edgeworth expansion for Z_ψ is $O(N_b^{-1})$ rather than $O(n^{-1})$, where

$$N_b = q_0 / \max_i b_i^2. \tag{10}$$

The quantity N_b can be thought of as the effective sample size and is bounded by $(\max_i \ell_i)^{-1} \le N_b \le N$, where $\ell_i = 1 - a_{ii}$ is the leverage of the ith observation. The lower bound is attained by equating $\mathbf{c}$ to the column of $\mathbf{X}'$ that has the largest leverage. The upper bound is obtained by equating $\mathbf{c}$ to $\bar{\mathbf{x}} = \mathbf{X}'\mathbf{1}_N N^{-1}$.

3 Moments and cumulants of V

An approximation to the moments of V can be obtained by using a stochastic Taylor series expansion of V^j around $S^2 = \sigma^2$. The expansion, up to $O_p(n^{-3/2})$, is the following:

$$V^j = \frac{n^{\frac{j}{2}}(\widehat{\psi} - \psi)^j}{\sigma_\psi^j \left(1 + \dfrac{Z_2}{\sqrt{n}}\right)^{j/2}} = Z_1^j \left[1 - \frac{j}{2\sqrt{n}} Z_2 + \frac{j(j+2)}{8n} Z_2^2 + O_p\left(n^{-\frac{3}{2}}\right)\right], \tag{11}$$

where $Z_1 = \sqrt{n}(\widehat{\psi} - \psi)/\sigma_\psi$ and $Z_2 = \sqrt{n}(S^2 - \sigma^2)/\sigma^2$. Theorem 1 gives the moments of V. The expressions were obtained by constructing an Edgeworth expansion of the joint distribution of (Z_1, Z_2) and then taking the expectation of (11) with respect to the Edgeworth expansion. A proof is given in the supplement. Corollary 1 gives the first four cumulants of V.

Theorem 1 *If* (2) *and* (6) *are satisfied and* $j \ge 0$ *is an integer, then*

$$\begin{aligned} \mathrm{E}(V^{2j}) = \frac{(2j)!}{2^j j!} \Bigg\{ 1 + \frac{j(j+1)}{n} &+ \frac{j\kappa_3^2}{n}\left[j(j+1)q_1^2 - 2j(j-1)q_1 q_4 + (j-1)(j-2)q_4^2\right] \\ &+ \frac{j\kappa_4}{2n}\left[(j+1)q_2 - 2jq_3 + (j-1)q_5\right]\Bigg\} + O\left(n^{-2}\right), \quad \text{and} \end{aligned}$$

$$\mathrm{E}(V^{2j+1}) = \frac{(2j+1)!\,\kappa_3}{j!\,2^j\sqrt{n}}\left[jq_4 - \left(j+\frac{1}{2}\right)q_1\right] + O\left(n^{-\frac{3}{2}}\right),$$

where κ_j is the jth *cumulant of ε_i/σ, and the remaining terms are defined in* (7).

Corollary 1 *The first four cumulants of V are the following:*

$$\rho_1(V) = \mathrm{E}(V) = \frac{\omega_1}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right), \quad \rho_2(V) = \mathrm{Var}(V) = 1 + \frac{\omega_2}{n} + O\left(n^{-2}\right),$$
$$\rho_3(V) = \frac{\omega_3}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right), \quad \text{and} \quad \rho_4(V) = \frac{\omega_4}{n} + O\left(n^{-2}\right),$$

where $\omega_1 = -q_1\kappa_3/2$, $\omega_2 = 2 + 7q_1^2\kappa_3^2/4 + \kappa_4(q_2 - q_3)$, $\omega_3 = 3(q_4 - q_1)\kappa_3$, and $\omega_4 = 6 + 18q_1(q_1 - q_4)\kappa_3^2 - 3q_{10}\kappa_4$. Furthermore, $\omega_i = O(1)$ for $i = 1, \ldots, 4$.

If $\mathbf{c} = \bar{\mathbf{x}}$, then $\mathbf{b} \propto \mathbf{1}_N$, V in (5) simplifies to the one-sample t statistic, the first four moments of V in Theorem 1 agree with those reported by Gayen (1949, Eq. 4.3) and the first four cumulants agree with those reported by Geary (1947, Eq. 2.18).

The first four cumulants of T_n, a central t random variable with $df = n$, are

$$\rho_1(T_n) = 0,\ \rho_2(T_n) = 1 + 2/n + O\left(n^{-2}\right),\ \rho_3(T_n) = 0,$$
$$\text{and } \rho_4(T_n) = 6/n + O\left(n^{-2}\right).$$

Note that $\rho_1(T_n) - \rho_1(V)$ and $\rho_3(T_n) - \rho_3(V)$ are $O(n^{-1/2})$, whereas $\rho_2(T_n) - \rho_2(V)$ and $\rho_4(T_n) - \rho_4(V)$ are only $O(n^{-1})$. Furthermore, the leading terms in $\rho_1(V)$ and $\rho_3(V)$ depend only on κ_3. Accordingly, $\kappa_3 \neq 0$ is the major issue when inference is based on V. Bias and/or skewness of V can be small, however, even if $|\kappa_3|$ is large. For example, if ψ is a contrast in an ANOVA model and data are balanced, then $q_1 = 0$ and $\rho_1(V)$ decreases to $O(n^{-3/2})$. If the coefficients of the contrast are symmetric around zero, then $q_4 = 0$ and $\rho_3(V)$ decreases to $O(n^{-3/2})$. This robustness of normal-theory ANOVA procedures when data are balanced is well known (Scheffé, 1959). If data are not balanced and $\kappa_3 \neq 0$, then $\rho_1(V)$ and $\rho_3(V)$ can vary widely depending on the coefficient vector $\mathbf{c}$.

4 Asymptotic expansions

Existing expansions of the one sample t (Chung, 1946, Geary, 1947, Gayen, 1949, Tiku, 1963) are applicable to V in (5) but only if $\mathbf{c} \propto \bar{x}$. In this section, Corollary 1 is used to construct expansions of the distributions of V and $|V|$ for arbitrary $\mathbf{c} \in \mathcal{R}(\mathbf{X}')$.

Denote the probability density functions (pdfs) of V and T_n by $\varphi_V(v)$ and $\varphi_{T_n}(v)$. Also, denote the cdfs of V, $|V|$, and T_n by $\Phi_V(v)$, $\Phi_{|V|}(|v|)$, and $\Phi_{T_n}(v)$. A Charlier differential series can be used to express φ_V in terms of φ_{T_n} (Wallace, 1958, Finney, 1963, Hill and Davis, 1968). Specifically, the pdf of V is recovered from $\phi_V(u)$ by inversion:

$$\varphi_V(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{iuv}}\phi_V(u)\,\mathrm{d}u = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{iuv}}\mathrm{e}^{K_V(u)-K_{T_n}(u)}\phi_{T_n}(u)\,\mathrm{d}u, \quad (12)$$

where $K_{\mathbf{W}}(u) \overset{\text{def}}{=} \ln[\phi_{\mathbf{W}}(u)]$ and $\phi_{\mathbf{W}}(u)$ is defined in (2). Define $H_j(v)$ as

$$H_j(v) \overset{\text{def}}{=} \left(\frac{(-1)^j}{\varphi_{T_n}(v)}\right)\frac{\partial^j\varphi_{T_n}(v)}{(\partial\, v)^j}, \qquad \text{for } j = 1, 2, \ldots.$$

With a remainder of $O(n^{-1})$, $H_j(v)$ is the jth Hermite polynomial. It follows from Corollary 1 that the difference between cumulant generating functions of V and T_n is

$$K_V(u) - K_{T_n}(u) = \frac{\omega_1 iu}{\sqrt{n}} + \frac{(\omega_2-2)(iu)^2}{2n} + \frac{\omega_3(iu)^3}{6\sqrt{n}} + \frac{(\omega_4-6)(iu)^4}{24n} + O\left(n^{-\frac{3}{2}}\right).$$

Term by term integration in (12) yields $\varphi_V(u)$. The expansion is given in Theorem 2.

Theorem 2 *If assumptions* (2) *and* (6) *are satisfied, then*

$$\varphi_V(v) = \varphi_{T_n}(v)\left[1 + \frac{\omega_1 H_1(v)}{\sqrt{n}} + \frac{g_1 H_2(v)}{2n} + \frac{\omega_3 H_3(v)}{6\sqrt{n}} + \frac{g_2 H_4(v)}{24n} + \frac{\omega_3^2 H_6(v)}{72n}\right] + O\left(n^{-\frac{3}{2}}\right),$$

$$\Phi_V(v) = \Phi_{T_n}(v) - \varphi_{T_n}(v)\left[\frac{\omega_1}{\sqrt{n}} + \frac{g_1 H_1(v)}{2n} + \frac{\omega_3 H_2(v)}{6\sqrt{n}} + \frac{g_2 H_3(v)}{24n} + \frac{\omega_3^2 H_5(v)}{72n}\right] + O\left(n^{-\frac{3}{2}}\right),$$

and

$$\Phi_{|V|}(|v|) = 2\Phi_{T_n}(|v|) - 1 - \frac{\varphi_{T_n}(|v|)}{n}\left[g_1 H_1(|v|) + \frac{g_2 H_3(|v|)}{12} + \frac{\omega_3^2 H_5(|v|)}{36}\right] + O\left(n^{-2}\right)$$

uniformly in v, *where* $g_1 = \omega_2 + \omega_1^2 - 2$, $g_2 = \omega_4 + 4\omega_1\omega_3 - 6$, *for* ω_i *in Corollary 1.*

The validity of Edgeworth expansions for functions of studentized regression coefficients under assumptions comparable to (2) and (6) was established by Qumsiyeh (1990, 1994) and Yanagihara (2003). To verify that the Charlier expansion in Theorem 2 is valid, Fisher's (1925) expansions for $\Phi_{T_n}(v)$ and $\varphi_{T_n}(v)$, namely,

$$\Phi_{T_n}(v) = \Phi_Z(v) - \frac{\varphi_Z(v)}{4n}(v^3 + v) + O\left(n^{-2}\right) \quad \text{and}$$
$$\varphi_{T_n}(v) = \varphi_Z(v)\left[1 + \frac{v^4 - 2v^2 - 1}{4n}\right] + O\left(n^{-2}\right), \tag{13}$$

can be employed, where $Z \sim \mathrm{N}(0,1)$. An analysis similar to that in Peiser (1949) shows that the expansions in (13) are accurate to $O(n^{-2})$ uniformly in v. Edgeworth expansions are obtained by substituting (13) for $\Phi_{T_n}(v)$ and $\varphi_{T_n}(v)$ in Theorem 2.

Generalized Cornish-Fisher expansions (Hill and Davis, 1968, Hall, 1983, Pace and Salvan, 1997, Sect. 10.6) based on the cdf expansions in Theorem 2 are summarized in Corollary 2.

Corollary 2 *The* 100α *percentiles of* T, V, $|T_n|$, *and* $|V|$ *are*

$$t_\alpha = \left(1 + \frac{g_3}{n}\right) v_\alpha - \frac{\omega_1}{\sqrt{n}} + \frac{\omega_3}{6\sqrt{n}}(1 - v_\alpha^2) + \frac{g_4}{n} v_\alpha^3 + O\left(n^{-\frac{3}{2}}\right),$$
$$v_\alpha = \left(1 - \frac{g_5}{n}\right) t_\alpha + \frac{\omega_1}{\sqrt{n}} - \frac{\omega_3}{6\sqrt{n}}(1 - t_\alpha^2) + \frac{g_6}{n} t_\alpha^3 + O\left(n^{-\frac{3}{2}}\right),$$
$$|t|_\alpha = \left(1 + \frac{g_7}{n}\right)|v|_\alpha + \frac{g_8}{n}|v|_\alpha^3 - \frac{\omega_3^2}{72n}|v|_\alpha^5 + O\left(n^{-2}\right), \quad \textit{and}$$
$$|v|_\alpha = \left(1 - \frac{g_7}{n}\right) t_{\frac{1+\alpha}{2}} - \frac{g_8}{n} t_{\frac{1+\alpha}{2}}^3 + \frac{\omega_3^2}{72n} t_{\frac{1+\alpha}{2}}^5 + O\left(n^{-2}\right)$$

uniformly in $\alpha \in (\epsilon, 1 - \epsilon)$, *where* $\epsilon > 0$, $8g_3 = (q_1^2 + 6q_1q_4 - 14q_4^2)\kappa_3^2 + q_9\kappa_4$, $4g_4 = (q_1 - q_4)(q_1 - 4q_4)\kappa_3^2 + q_{10}\kappa_4/2$, $8g_5 = (10q_4^2 - 2q_1q_4 - q_1^2)\kappa_3^2 - q_9\kappa_4$, $4g_6 = (q_1 - q_4)(q_1 + 2q_4)\kappa_3^2 - q_{10}\kappa_4/2$, $8g_7 = (q_1^2 + 6q_1q_4 - 15q_4^2)\kappa_3^2 + q_9\kappa_4$, $4g_8 = (q_1 - q_4)(q_1 - 5q_4)\kappa_3^2 + q_{10}\kappa_4/2$, *and* q_i *is defined in* (7).

If α is a uniform $(0, 1)$ random variable, then t_α and $|t|_\alpha$ in Corollary 2 are random variables with cdfs $\Phi_{T_n}(t)$ and $\Phi_{|T_n|}(t)$, respectively. Accordingly, V and $|V|$ can be transformed to random variables having the same distributions as T_n and $|T_n|$ with errors $O(n^{-3/2})$ and $O(n^{-2})$, respectively. These results are summarized in Corollary 3.

Corollary 3 *Define* T_n^* *and* M *as*

$$T_n^* \stackrel{\text{def}}{=} \left(1 + \frac{g_3}{n}\right) V - \frac{\omega_1}{\sqrt{n}} + \frac{\omega_3}{6\sqrt{n}}(1 - V^2) + \frac{g_4}{n} V^3 \textit{ and}$$
$$M \stackrel{\text{def}}{=} \left(1 + \frac{g_7}{n}\right)|V| + \frac{g_8}{n}|V|^3 - \frac{\omega_3^2}{72n}|V|^5,$$

where g_3–g_8 *are defined in Corollary* 2. *Then,* $P(T_n^* \leq t) = \Phi_{T_n}(t) + O\left(n^{-3/2}\right)$ *and* $P(M \leq |t|) = \Phi_{|T_n|}(|t|) + O\left(n^{-2}\right)$ *uniformly in* t *on compact intervals.*

5 Interval estimators

5.1 Estimators of κ_3 and κ_4

The proposed estimators of κ_3 and κ_4 are ratios of unbiased or nearly unbiased estimators and satisfy $\kappa_4 \geq -2$ and $\kappa_3^2 \leq \kappa_4 + 2$. Denote the vector of residuals as $\mathbf{e}$ and denote the ith residual as e_i. Note that $\mathbf{e} = \mathbf{Ay} = \mathbf{A}\boldsymbol{\varepsilon}$ and $e_i = \sum_{j=1}^{N} a_{ij}\varepsilon_j$, where $\mathbf{A}$ is defined in (4). Define $\mu_j \stackrel{\text{def}}{=} \mathrm{E}(\varepsilon_i^j)$. Using the same methods as in Boik (1998, Theorem 5), it can be shown that unbiased estimators of μ_3 and μ_4 are given by

$$\widehat{\mu}_3 = \frac{1}{nq_6}\sum_{i=1}^{N} \mathrm{e}_i^3 \quad \text{and} \quad \widehat{\mu}_4 = \frac{(n+2-3q_2)}{n\left[(n+2)q_7 - 3q_2^2\right]}\sum_{i=1}^{N} \mathrm{e}_i^4 + \frac{3n(q_7 - q_2)}{(n+2)q_7 - 3q_2^2} S^4 \tag{14}$$

respectively, where q_j is defined in (7). Expanding S^{2r} around $S^2 = \sigma^2$ and taking expectations yields $\mathrm{E}(S^{2r}) = \sigma^{2r}\left[1 + r(r-1)\,\mathrm{Var}(S^2)/(2\sigma^4)\right] + O\left(n^{-2}\right)$, where $\mathrm{Var}(S^2) = \sigma^4(q_2\kappa_4 + 2)/n$. It follows from Theorem 5 in Boik (1998) that

$$\widehat{\mathrm{Var}}(S^2) = \frac{1}{q_7(n+2) - 3q_2^2}\left[\frac{q_2}{n}\sum_{i=1}^{N} \mathrm{e}_i^4 - \left(3q_2^2 - 2q_7\right) S^4\right]$$

is unbiased for $\mathrm{Var}(S^2)$. Accordingly, an estimator of σ^{2r} with bias $O(n^{-2})$ is given by $\widehat{\sigma}^{2r} = S^{2r}/[1 + r(r-1)\widehat{\mathrm{Var}}(S^2)/(2S^4)]$ and the proposed estimators of κ_3, κ_4, and κ_3^2 are

$$\widehat{\kappa}_3 = \frac{\widehat{\mu}_3}{S^3}\left[1 + \frac{3\widehat{\mathrm{Var}}(S^2)}{8S^4}\right], \quad \widehat{\kappa}_4 = \max\left\{\frac{\widehat{\mu}_4}{S^4}\left[1 + \frac{\widehat{\mathrm{Var}}(S^2)}{S^4}\right] - 3, -2\right\}$$
$$\text{and} \quad \widehat{\kappa}_{32} = \min(\widehat{\kappa}_3^2, \widehat{\kappa}_4 + 2), \tag{15}$$

respectively. Estimators of ω_i and g_i, namely $\widehat{\omega}_i$ and $\widehat{g}_i$, can be obtained by substituting the above estimators for unknown cumulant functions in Corollary 1.

5.2 Second-order accurate intervals

5.2.1 Quadratic Cornish–Fisher transformation

Johnson (1978) proposed a modified one-sample t test that is less affected by population skewness than is the conventional t test. Johnson's procedure is readily generalized to the regression setting by employing Corollary 3. Define $\widehat{T}_{n,1}^*$ by

$$\widehat{T}^*_{n,1} \stackrel{\text{def}}{=} V - \frac{\widehat{\omega}_1}{\sqrt{n}} + \frac{\widehat{\omega}_3}{6\sqrt{n}}(1 - V^2), \tag{16}$$

where V is defined in (5). It follows from Corollary 3 that $P(\widehat{T}^*_{n,1} \leq t) = \Phi_{T_n}(t) + O(n^{-1})$. The inequality $\widehat{T}^*_{n,1} \leq t_{1-\alpha}$ can not always be inverted to obtain a confidence interval for ψ, however, because $\widehat{T}^*_{n,1}$ is not monotonic with respect to V.

5.2.2 Hall's modification of the quadratic transformation

Hall's (1992a) monotonic modification to the quadratic transformation is

$$\widehat{T}^*_{n,2} \stackrel{\text{def}}{=} V - \frac{\widehat{\omega}_1}{\sqrt{n}} + \frac{\widehat{\omega}_3}{6\sqrt{n}}(1 - V^2) + \frac{\widehat{\omega}_3^2}{108n}V^3. \tag{17}$$

Generalizations of (17) were given by Fujioka and Maesono (2000) and Yanagihara and Yuan (2005). Note that $\widehat{T}^*_{n,2} = \widehat{T}^*_{n,1} + O_p(n^{-1})$. Inverting the inequality $t_\alpha \leq \widehat{T}^*_{n,2}$ yields the endpoint of a one-sided upper $100(1-\alpha)\%$ confidence interval:

$$\begin{aligned} U &= \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}}\hat{v}_{n,2,\alpha} \quad \text{where} \\ \hat{v}_{n,2,\alpha} &= \frac{6\sqrt{n}}{\widehat{\omega}_3}\left\{1 - \left[1 + \frac{\widehat{\omega}_3}{2\sqrt{n}}\left(\frac{\widehat{\omega}_3}{6\sqrt{n}} - \frac{\widehat{\omega}_1}{\sqrt{n}} - t_\alpha\right)\right]^{\frac{1}{3}}\right\}. \end{aligned} \tag{18}$$

If $|\widehat{\omega}_3|$ is near zero, then $\hat{v}_{n,2,\alpha}$ can be replaced by

$$\begin{aligned} \hat{v}_{n,2,\alpha} &= t_\alpha + \frac{\widehat{\omega}_1}{\sqrt{n}} + \left(\frac{\widehat{\omega}_1}{\sqrt{n}} + t_\alpha - 1\right)\left(\frac{\widehat{\omega}_1}{\sqrt{n}} + t_\alpha + 1\right)\frac{\widehat{\omega}_3}{6\sqrt{n}} \\ &\quad + \left[5\left(\frac{\widehat{\omega}_1}{\sqrt{n}} + t_\alpha\right)^3 - 6\left(\frac{\widehat{\omega}_1}{\sqrt{n}} + t_\alpha\right)\right]\frac{\widehat{\omega}_3^2}{108n} + O\left(\frac{\widehat{\omega}_3^3}{n^{3/2}}\right). \end{aligned}$$

5.2.3 Box–Cox modifications of the quadratic transformation

Konishi (1991) proposed an alternative modification of the quadratic transformation. Konishi's method begins by applying a power transformation in the Box–Cox family (Box and Cox, 1964) to $\exp\{V/\sqrt{n}\}$. Specifically, a value of λ is chosen so that the distribution of $B(V) = \sqrt{n}(\exp\{\lambda V/\sqrt{n}\} - 1)/\lambda$ is symmetric to order $O(n^{-1/2})$. The second-order Taylor expansion of $B(V)$ around $V/\sqrt{n} = 0$ is $B(V) = V + \lambda V^2/(2\sqrt{n}) + O_p(n^{-1})$. If λ is chosen to be $\widehat{\lambda} = -\widehat{\omega}_3/3$, then $P[B(V) - \widehat{\omega}_1/\sqrt{n} + \widehat{\omega}_3/(6\sqrt{n}) \leq t_{1-\alpha}] = 1 - \alpha + O(n^{-1})$. Inverting yields the endpoint for a one-sided lower $100(1-\alpha)\%$ confidence interval for ψ:

$$L = \widehat{\psi} + \frac{3\widehat{\sigma}_\psi}{\widehat{\omega}_3} \ln\left[1 - \frac{\widehat{\omega}_3}{3\sqrt{n}}\left(t_{1-\alpha} + \frac{\widehat{\omega}_1}{\sqrt{n}} - \frac{\widehat{\omega}_3}{6\sqrt{n}}\right)\right].$$

The endpoint for a one-sided upper interval is obtained by replacing $t_{1-\alpha}$ by t_α. If sample size is small and/or $\widehat{\omega}_3$ is large, then the argument to the log function can be negative. In this case, the method fails to produce an interval. One can avoid using the log function by replacing it by its Taylor series expansion to obtain

$$L = \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}}\widehat{v}_{1-\alpha}, \quad \text{where } \widehat{v}_{1-\alpha} = t_{1-\alpha} + \frac{\widehat{\omega}_1}{\sqrt{n}} - \frac{\widehat{\omega}_3}{6\sqrt{n}}(1 - t_{1-\alpha}^2),$$

but $\widehat{v}_{1-\alpha}$ is not monotonic in α. Konishi's interval also can be obtained by applying the Box–Cox transformation to $\exp\{\widehat{\psi}\}$ rather than to $\exp\{V/\sqrt{n}\}$. This alternative method was employed by DiCiccio and Monti (2002) in the context of M-estimators.

5.2.4 An exponential modification of the quadratic transformation

Consider the transformation from V to $\widehat{T}^*_{n,3}$, where

$$\widehat{T}^*_{n,3} \overset{\text{def}}{=} V - \frac{\widehat{\omega}_1}{\sqrt{n}} + \frac{\widehat{\omega}_3}{6\sqrt{n}} - \exp\left\{-\frac{\widehat{d}_1}{2}V^2\right\}\frac{\widehat{\omega}_3}{6\sqrt{n}}V^2 = \widehat{T}^*_{n,1} + O_p(n^{-3/2}), \quad \text{and}$$
$$\widehat{d}_1 = \frac{\widehat{\omega}_3^2\,(31 - 7\sqrt{17})}{72n}\exp\left\{-\frac{1}{2}(5 - \sqrt{17})\right\} \tag{19}$$

is the smallest non-negative number such that $\widehat{T}^*_{n,3}$ is non-decreasing in V. Monotonicity of $\widehat{T}^*_{n,3}$ ensures that the inequalities $t_\alpha \le \widehat{T}^*_{n,2}$ and $\widehat{T}^*_{n,2} \le t_{1-\alpha}$ can be inverted to obtain second-order accurate one-sided confidence intervals for ψ with endpoints

$$L = \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}}\hat{v}_{n,3,1-\alpha} \quad \text{and} \quad U = \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}}\hat{v}_{n,3,\alpha}, \tag{20}$$

where $\hat{v}_{n,3,\alpha}$ is the solution to $\widehat{T}^*_{n,3} = t_\alpha$ for V. The solution can be computed using the modified Newton method described in the supplement.

5.3 Third-order accurate intervals

5.3.1 Third-order accurate two-sided symmetric intervals

It follows from Corollary 3 that $P(\widehat{M} \le m) = \Phi_{|T_n|}(m) + O(n^{-3/2})$, where $\widehat{M} = |V|(1+\widehat{g}_7/n) + |V|^3\widehat{g}_8/n - |V|^5\widehat{\omega}_3^2/(72n)$. Consider the modified transformation of $|V|$:

$$\widehat{M}^* = |V| \exp\left\{\frac{\widehat{g}_7}{n}\right\}^* + \exp\left\{-\frac{1}{2}V^2\widehat{d}_2\right\}\left(|V|^3\frac{\widehat{g}_8}{n} - |V|^5\frac{\widehat{\omega}_3^2}{72n}\right),$$

where $\widehat{d}_2$ is the smallest non-negative number such that $\widehat{M}^*$ is monotonic in $|V|$, and $\exp\{\widehat{g}_7/n\}^* = \max\left[\exp\{\widehat{g}_7/n\}, 1/(2n^2)\right]$. It can be shown that $\widehat{d}_2 = O_p(n^{-1/2})$ and, therefore, $\widehat{M}^* = \widehat{M} + O_p(n^{-3/2})$. The associated $100(1-\alpha)\%$ symmetric confidence interval with third-order accurate coverage is

$$\widehat{\psi} \pm \frac{\widehat{\sigma}_\psi}{\sqrt{n}}|\hat{v}|^*_{1-\alpha}, \tag{21}$$

where $|\hat{v}|^*_{1-\alpha}$ is the solution to $\widehat{M}^* = t_{1-\frac{\alpha}{2}}$ for $|V|$. The quantities $|\hat{v}|^*_{1-\alpha}$ and $\widehat{d}_2$ can be computed using the modified Newton methods that are described in the supplement.

Alternatively, it follows from Corollary 2 that $P(|V| \le |\hat{v}|_{1-\alpha}) = 1 - \alpha + O(n^{-3/2})$, where $|\hat{v}|_{1-\alpha} = |t|_{1-\alpha}(1 - \widehat{g}_7/n) - |t|^3_{1-\alpha}\widehat{g}_8/n + |t|^5_{1-\alpha}\widehat{\omega}_3^2/(72n)$, but $|\hat{v}|_{1-\alpha}$ need not be monotonic with respect to α. The proposed remedy is to replace $|\hat{v}|_{1-\alpha}$ by

$$|\hat{v}|^{**}_{1-\alpha} = t_{1-\frac{\alpha}{2}} \exp\left\{-\frac{\widehat{g}_7}{n} - t^2_{1-\frac{\alpha}{2}}\frac{\widehat{g}_8}{n} + t^4_{1-\frac{\alpha}{2}}\frac{\widehat{g}_9}{n}\right\}, \quad \text{where } \widehat{g}_9 = \frac{\widehat{\omega}_3^2}{72} + \frac{\widehat{g}_8^2}{4n},$$

which is monotonic in α. Furthermore, $P(|V| \le |\hat{v}|^{**}_{1-\alpha}) = 1 - \alpha + O(n^{-3/2})$ because $|\hat{v}|^{**}_{1-\alpha} = |\hat{v}|_{1-\alpha} + O_p(n^{-2})$. The associated third-order accurate symmetric interval is

$$\widehat{\psi} \pm \frac{\widehat{\sigma}_\psi}{\sqrt{n}}|\hat{v}|^{**}_{1-\alpha}. \tag{22}$$

5.3.2 Third-order accurate one-sided intervals

A transformation of $\widehat{T}^*_{n,1}$ in (16) whose cumulants differ from those of T_n by only $O(n^{-3/2})$ can be constructed, but the random quantity $V^j\sqrt{n}(\widehat{\kappa}_3 - \kappa_3)$ must be taken into account. An expression for $\mathrm{E}\left[V^j\sqrt{n}(\widehat{\kappa}_3 - \kappa_3)\right]$ is given in Theorem 3. A proof is sketched in the supplement. Cumulants of $\widehat{T}^*_{n,1}$ are summarized in Corollary 4. Associated Cornish-Fisher transformations are summarized in Corollary 5.

Theorem 3 *If* (2) *and* (6) *are satisfied, then*

$$\mathrm{E}\left[V^{2j}\sqrt{n}(\widehat{\kappa}_3-\kappa_3)\right]=O\left(n^{-\frac{1}{2}}\right)\ \textit{and}$$

$$\mathrm{E}\left[V^{2j+1}\sqrt{n}(\widehat{\kappa}_3-\kappa_3)\right]=\frac{(2j+1)!}{2^j j!}\left(\kappa_4 q_8-\frac{3q_1}{2}\kappa_3^2\right)+O\left(n^{-1}\right),$$

where j is any nonnegative integer, q_i is defined in (7)*, and V is defined in* (5).

Corollary 4 *Under* (2) *and* (6)*, the cumulants of $\widehat{T}^*_{n,1}$ in* (16) *are*

$$\mathrm{E}(\widehat{T}^*_{n,1})=O\left(n^{-\frac{3}{2}}\right),\quad \mathrm{Var}(\widehat{T}^*_{n,1})=1+\frac{\omega_2^*}{n}+O\left(n^{-\frac{3}{2}}\right),$$

$$\rho_3(\widehat{T}^*_{n,1})=O\left(n^{-\frac{3}{2}}\right),\ \textit{and}\ \rho_4(\widehat{T}^*_{n,1})=\frac{\omega_4^*}{n}+O\left(n^{-\frac{3}{2}}\right),$$

where $\omega_2^*=2-(25q_1^2-36q_1q_4+10q_4^2)\kappa_3^2/4+[q_2-q_3+q_8(3q_1-2q_4)]\kappa_4$, $\omega_4^*=6-24(q_1-q_4)^2\kappa_3^2+3[4q_8(q_1-q_4)-q_{10}]\kappa_4$, and q_i is defined in (7)*. To ensure that $1+\omega_2^*/n>0$ and $\omega_4^*/n\geq-2$ are satisfied, ω_2^* and ω_4^* can be replaced by $n\left[\exp(\omega_2^*/n)-1\right]$ and $\max(\omega_4^*,-2n)$, respectively. These substitutions do not change the order of accuracy.*

Corollary 5 *Denote the 100α percentile of $\widehat{T}^*_{n,1}$ by $t^*_{n,1,\alpha}$ and define*

$$\widehat{T}^*_{n,4}\stackrel{\text{def}}{=}\left(1+\frac{g_{10}}{n}\right)\widehat{T}^*_{n,1}+\frac{g_{11}}{n}\widehat{T}^{*3}_{n,1}\ \textit{and}\ \widehat{t}^*_{n,1,\alpha}\stackrel{\text{def}}{=}\left(1-\frac{g_{10}}{n}\right)t_\alpha-\frac{g_{11}}{n}t_\alpha^3,$$

where $g_{10}=(2+\omega_4^*-4\omega_2^*)/8$, $g_{11}=(6-\omega_4^*)/24$, and t_α is the 100α percentile T_n. Then, $P(\widehat{T}^*_{n,4}\leq t)=\Phi_{T_n}(t)+O(n^{-3/2})$ uniformly in t on compact intervals and $P(\widehat{T}^*_{n,1}\leq\widehat{t}^*_{n,1,\alpha})=\alpha+O(n^{-3/2})$ uniformly in $\alpha\in(\epsilon,1-\epsilon)$, where $\epsilon>0$. Furthermore, ω_2^* and ω_4^* may be replaced by $\widehat{\omega}_2^*$ and $\widehat{\omega}_4^*$ from (15) *without affecting the order of accuracy.*

The remaining issue is that the transformations V to $\widehat{T}^*_{n,1}$, $\widehat{T}^*_{n,1}$ to $\widehat{T}^*_{n,4}$, V to $\widehat{T}^*_{n,5}$, and t_α to $\widehat{t}^*_{n,1,\alpha}$ are not necessarily monotonic. Corollaries 6–8 describe several remedies in which polynomial functions are replaced by monotonic exponential functions.

Corollary 6 *Define $\widehat{T}^*_{n,5}$ and $\hat{v}_{n,5,\alpha}$ as*

$$\widehat{T}^*_{n,5}\stackrel{\text{def}}{=}\left(1+\frac{\widehat{g}_{10}}{n}\right)V+\frac{q_4}{2\sqrt{n}}\widehat{\kappa}_3+\frac{(q_1-q_4)}{2\sqrt{n}}\widehat{\kappa}_3V^2+\frac{\widehat{g}_{11}}{n}V^3,\ \textit{and}$$

$$\hat{v}_{n,5,\alpha}\stackrel{\text{def}}{=}\left(1+\frac{\widehat{g}_{12}}{n}\right)t_\alpha-\frac{q_4}{2\sqrt{n}}\widehat{\kappa}_3+\frac{(q_4-q_1)}{2\sqrt{n}}\widehat{\kappa}_3t_\alpha^2+\frac{\widehat{g}_{13}}{n}t_\alpha^3,$$

where $g_{12} = q_4(q_1-q_4)\kappa_3^2/2+(4\omega_2^*-\omega_4^*-2)/8$, $g_{13} = (q_4-q_1)^2\kappa_3^2/2+(\omega_4^*-6)/24$, *and* t_α *is the* 100α *percentile of the distribution of* T_n. *Then,* $\widehat{T}^*_{n,5} = \widehat{T}^*_{n,4} + O_p(n^{-3/2})$, $P(\widehat{T}^*_{n,5} \le t) = \Phi_{T_n}(t) + O(n^{-3/2})$ *uniformly in* t *on compact intervals, and* $P(V \le \hat{v}_{n,5,\alpha}) = \alpha + O(n^{-3/2})$ *uniformly in* $\alpha \in (\epsilon, 1-\epsilon)$, *where* $\epsilon > 0$.

Corollary 7 *Define* $\widehat{T}^*_{n,6}$ *and* $\widehat{v}_{n,6,\alpha}$ *as*

$$\widehat{T}^*_{n,6} \stackrel{\text{def}}{=} \frac{q_4}{2\sqrt{n}}\widehat{\kappa}_3 + \exp\left\{\frac{\widehat{g}_{10}}{n}\right\}V + \exp\left\{-\frac{\widehat{d}_3}{2}V^2\right\}\left(\frac{(q_1-q_4)}{2\sqrt{n}}\widehat{\kappa}_3 V^2 + \frac{\widehat{g}_{11}}{n}V^3\right) \text{ and}$$

$$\hat{v}_{n,6,\alpha} \stackrel{\text{def}}{=} -\frac{q_4}{2\sqrt{n}}\widehat{\kappa}_3 + \exp\left\{\frac{\widehat{g}_{12}}{n}\right\}t_\alpha + \exp\left\{-\frac{\widehat{d}_4}{2}t_\alpha^2\right\}\left(\frac{(q_4-q_1)}{2\sqrt{n}}\widehat{\kappa}_3 t_\alpha^2 + \frac{\widehat{g}_{13}}{n}t_\alpha^3\right),$$

where $\widehat{d}_3$ *and* $\widehat{d}_4$ *are chosen to be the smallest non-negative values for which the transformations are monotonic. Then,* $\widehat{T}^*_{n,6} = \widehat{T}^*_{n,5} + O_p(n^{-3/2})$, $\hat{v}_{n,6,\alpha} = \hat{v}_{n,5,\alpha} + O_p(n^{-3/2})$, $P(\widehat{T}^*_{n,6} \le t) = \Phi_{T_n}(t) + O(n^{-3/2})$ *uniformly in* t *on compact intervals, and* $P(V \le \hat{v}_{n,6,\alpha}) = \alpha + O(n^{-3/2})$ *uniformly in* $\alpha \in (\epsilon, 1-\epsilon)$, *where* $\epsilon > 0$. *Furthermore, if* $\widehat{\kappa}_3^2 \le 12\widehat{g}_{11}\exp\{\widehat{g}_{10}/n\}/(q_1-q_4)^2$, *then* $\widehat{d}_3 = 0$ *and if* $\widehat{\kappa}_3^2 \le 12\widehat{g}_{13}\exp\{\widehat{g}_{12}/n\}/(q_1-q_4)^2$, *then* $\widehat{d}_4 = 0$. *Otherwise,* $\widehat{d}_3$ *and/or* $\widehat{d}_4$ *can be computed using the modified Newton algorithm described in the supplement.*

Corollary 8 *Define* $\widehat{T}^*_{n,7}$ *and* $\widehat{t}^*_{n,3,\alpha}$ *as*

$$\widehat{T}^*_{n,7} \stackrel{\text{def}}{=} \exp\left\{\frac{\widehat{g}_{10}}{n}\right\}\widehat{T}^*_{n,3} + \exp\left\{-\frac{\widehat{d}_5}{2}\widehat{T}^{*2}_{n,3}\right\}\frac{\widehat{g}_{11}}{n}\widehat{T}^{*3}_{n,3} \quad \text{and}$$

$$\widehat{t}^*_{n,3,\alpha} \stackrel{\text{def}}{=} \exp\left\{-\frac{\widehat{g}_{10}}{n}\right\}t_\alpha - \exp\left\{-\frac{\widehat{d}_6}{2}t_\alpha^2\right\}\frac{\widehat{g}_{11}}{n}t_\alpha^3$$

where $\widehat{d}_5$ *and* $\widehat{d}_6$ *are the smallest non-negative numbers for which the transformations are monotonic. Then,* $\widehat{T}^*_{n,7} = \widehat{T}^*_{n,4} + O_p(n^{-3/2})$, $\widehat{t}^*_{n,3,\alpha} = \widehat{t}^*_{n,1,\alpha} + O_p(n^{-3/2})$, $P(\widehat{T}^*_{n,7} \le t) = \Phi_{T_n}(t) + O(n^{-3/2})$ *uniformly in* t *on compact intervals, and* $P(\widehat{T}^*_{n,3} \le \widehat{t}^*_{n,3,\alpha}) = \alpha + O(n^{-3/2})$ *uniformly in* $\alpha \in (\epsilon, 1-\epsilon)$, *where* $\epsilon > 0$. *Furthermore, the solutions for* $\widehat{d}_5$ *and* $\widehat{d}_6$ *are*

$$\widehat{d}_5 = \max\left[0, -\frac{2\widehat{g}_{11}}{n}\exp\left\{-\frac{2\widehat{g}_{10}+n}{2n}\right\}\right] \text{ and } \widehat{d}_6 = \max\left[0, \frac{2\widehat{g}_{11}}{n}\exp\left\{\frac{2\widehat{g}_{10}-n}{2n}\right\}\right].$$

Corollaries 7 and 8 provide a framework for constructing third-order accurate one-sided or equal-tailed two-sided confidence intervals. The endpoints for $100(1-\alpha)\%$ one-sided intervals can be written as

$$L = \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}}\widehat{v}_{1-\alpha} \quad \text{and} \quad U = \widehat{\psi} - \frac{\widehat{\sigma}_\psi}{\sqrt{n}}\widehat{v}_\alpha, \tag{23}$$

where the estimated percentiles $\widehat{v}_\alpha$ and $\widehat{v}_{1-\alpha}$ are method-specific. Corollary 9 summarizes four procedures for estimating the percentiles of V.

Corollary 9 *The* 100α *percentile of* V *can be estimated as follows:*

(a) $\widehat{v}_\alpha$ *is the solution to* $\widehat{T}^*_{n,6} = t_\alpha$ *for* V*, where* $\widehat{T}^*_{n,6}$ *is given in Corollary 7,*
(b) $\widehat{v}_\alpha$ *is* $\widehat{v}_{n,6,\alpha}$ *from Corollary 7,*
(c) $\widehat{v}_\alpha$ *is the solution to* $\widehat{T}^*_{n,3} = \widehat{t}^*_\alpha$ *for* V*, where* $\widehat{t}^*_\alpha$ *is the solution to* $\widehat{T}^*_{n,7} = t_\alpha$ *for* $\widehat{T}^*_{n,3}$*,* $\widehat{T}^*_{n,3}$ *is given in* (19)*, and* $\widehat{T}^*_{n,7}$ *is given in Corollary 8, and*
(d) $\widehat{v}_\alpha$ *is the solution to* $\widehat{T}^*_{n,3} = \widehat{t}^*_{n,3,\alpha}$ *for* V*, where* $\widehat{t}^*_{n,3,\alpha}$ *is given in Corollary 8.*

The algorithm in the supplement can be used to solve the exponential equations. The resulting confidence intervals in (23) *are third-order accurate.*

If the linear function $\psi = \mathbf{c}'\boldsymbol{\beta}$ does not involve the intercept, β_0, then Theorem 3 and its corollaries simplify. Partition the matrix of explanatory variables, $\mathbf{X}$, as $\mathbf{X} = (\mathbf{1}_N\ \mathbf{X}_2)$, where $\mathbf{X}_2$ has dimensions $N \times (p-1)$. If $c_1 = 0$, then $\mathbf{b}$ in (7) satisfies

$$\mathbf{b} = \mathbf{X}_{2\cdot 1}\left(\mathbf{X}'_{2\cdot 1}\mathbf{X}_{2\cdot 1}\right)^{-}\mathbf{c}\sqrt{n} \quad \text{and} \quad \mathbf{1}'_N\mathbf{b} = 0, \tag{24}$$

where $\mathbf{X}_{2\cdot 1} = (\mathbf{I}_N - \mathbf{P}_1)\mathbf{X}_2$ and $\mathbf{P}_1 = \mathbf{1}_N(1/N)\mathbf{1}'_N$. Note that $\mathbf{1}'_N\mathbf{b} = 0$ implies that q_1 and q_8 each have magnitude $O(n^{-1/2})$ rather than $O(1)$. Accordingly, the bias of V has magnitude $O(n^{-1})$ rather than $O(n^{-1/2})$. Furthermore, it follows from Theorem 3 that if $\mathbf{1}'_N\mathbf{b} = 0$, then $\mathrm{E}\left[V^j\sqrt{n}(\widehat{\omega}_3 - \omega_3)\right] = O(n^{-1/2})$ for any non-negative integer, j. This result is the basis of Hall's claim (Hall, 1989) that if $c_1 = 0$, then percentile-t bootstrap confidence intervals are third-order accurate. Under (24), the cumulants of $\widehat{T}^*_{n,1}$ that were reported in Corollary 4 simplify.

Corollary 10 *If* $c_1 = 0$ *then* $q_1 = O(n^{-1/2})$*,* $q_8 = O(n^{-1/2})$*, and the cumulants of* $\widehat{T}^*_{n,1}$ *in* (16) *are those in Corollary 4 in which* ω^*_2 *and* ω^*_4 *simplify to* $\omega^*_2 = 2 - 5q_4^2\kappa_3^2/2 + \kappa_4(q_2 - q_3)$ *and* $\omega^*_4 = 6 - 24q_4^2\kappa_3^2 - 3q_{10}\kappa_4$*. Furthermore,* $g_3 = g_{10} + O(n^{-1/2})$*,* $g_4 = g_{11} + O(n^{-1/2})$*,* $g_5 = g_{12} + O(n^{-1/2})$*, and* $g_6 = g_{13} + O(n^{-1/2})$*, where* g_3*–*g_6 *and* g_{10}*–*g_{13} *are defined in Corollaries 2, 5, and 6.*

6 Illustration

This section illustrates the proposed intervals using the Venables and Ripley (2002, pp. 234) data set. The response is log permeability obtained on $N = 12$ rocks. The explanatory variables are total area $\div 10{,}000$, total perimeter $\div 2{,}000$, and roundness. The explanatory variables were measured at four cross-sections within each rock and the geometric mean of the four measures was used to construct $\mathbf{X}$. Using (15), $\widehat{\kappa}_3 = 0.913$ and $\widehat{\kappa}_4 = -1.356$. Table 1 displays the coefficients of six linear functions together with effective sample size, $\widehat{\rho}_1(V)$, and $\widehat{\rho}_3(V)$. The coefficients in row one maximize $|\rho_1(V)|$ whereas the coefficients in

Table 1 Rock permeability study: bias, skewness, and effective sample size for selected linear functions

Function	Coefficients for ψ				Bias	Skewness	
	c_0	c_1	c_2	c_3	$\widehat{\rho}_1(V)$	$\widehat{\rho}_3(V)$	N_b
1	1	0.717	1.393	0.203	−0.133	−0.529	9.43
2	0	0.320	0.938	−0.132	−0.021	−0.592	1.54
3	1	0.	0.	0.	−0.027	−0.111	4.90
4	0	1.	0.	0.	−0.007	−0.011	3.19
5	0	0.	1.	0.	0.006	−0.126	2.50
6	0	0.	0.	1.	0.015	0.343	1.92

Table 2 Rock permeability: estimated percentiles of V for function 1

Method	Order	Reference	Percentile	
			$\widehat{v}_{0.05}$	$\widehat{v}_{0.95}$
1. Conventional t	1	Equation (9)	−1.860	1.860
2. Hall: $\widehat{T}^*_{n,2}$	2	Equation (18)	−2.363	1.583
3. Exponential: $\widehat{T}^*_{n,3}$	2	Equation (20)	−2.409	1.592
4. Exponential: $\widehat{T}^*_{n,6}$	3	Corollary 9a	−1.770	1.403
5. Exponential: $\widehat{v}^*_{n,6,\alpha}$	3	Corollary 9b	−1.837	1.205
6. Exponential: $\widehat{T}^*_{n,7}$	3	Corollary 9c	−1.927	1.352
7. Exponential: $\widehat{t}^*_{n,3,\alpha}$	3	Corollary 9d	−1.756	1.257
8. Bootstrap t	2		−1.969	1.788
			Percentile	
Method	Order	Reference	$-\lvert\widehat{v}\rvert_{0.95}$	$\lvert\widehat{v}\rvert_{0.95}$
9. Conventional $\lvert t\rvert$	2	Equation (9)	−2.306	2.306
10. Exponential: $\widehat{M}^*$	3	Equation (21)	−2.662	2.662
11. Exponential: $\lvert\widehat{v}\rvert^{**}_{1-\alpha}$	3	Equation (22)	−2.633	2.633
12. Bootstrap $\lvert t\rvert$	3		−2.351	2.351

row 2 maximize $|\rho_3(V)|$ subject to $c_0 = 0$. It is apparent that bias, skewness, and effective sample size can vary widely for different linear functions.

Table 2 displays estimates of $v_{0.05}$, $v_{0.95}$, and $\pm|v|_{0.95}$ that correspond to intervals for ψ, where $\mathbf{c}'$ is given in row one of Table 1. These estimated percentiles can be used to compute one-sided 95% intervals (methods 1–8), two-sided equal-tailed 90% intervals, (methods 1–8), and two-sided symmetric 95% intervals (methods 9–12). Methods 8 and 12 correspond to percentile-t and percentile-$|t|$ bootstrap methods, based on 100,000 bootstrap samples. To compute the estimated percentiles for method 6, one must first solve $\widehat{T}^*_{n,7} = t_{0.95}$ for $\widehat{T}^*_{n,3}$ as described in Corollary 9 part (c). The solution is $\widehat{T}^*_{n,3} = 1.5571$. To compute the estimated percentiles for method 7, one must first compute $\widehat{t}^*_{n,3,0.95}$ as described in Corollary 9 part (d). The solution is $\widehat{t}^*_{n,3,0.95} = 1.4413$. Table 2 reveals that even though skewness appears to be mild, the adjustments to one-sided intervals can be appreciable. The adjustments to two-sided symmetric intervals are less dramatic because the usual two-sided symmetric intervals already are second-order accurate.

7 Simulation

Realizations of random N-vectors $\{\boldsymbol{\varepsilon}_i\}_{i=1}^N$ were sampled from one of three zero-centered log-normal distributions, chosen so that $(\kappa_3, \kappa_4) = (1, 1.83)(3, 19.40)$, or $(5, 65.26)$ and $\text{Var}(\varepsilon_i) = 1$. Sample size was $N = 10$, 20, 30, 40, 60, or 100. The response was modeled as a function of a constant term plus four explanatory variables; i.e., $p = 5$. Without loss of generality, the vector of regression coefficients, $\boldsymbol{\beta}$, was set to zero. The $N \times p$ matrix of explanatory variables was constructed as $\mathbf{X} = (\mathbf{1}_k \otimes \mathbf{X}^*)$, where $\mathbf{X}^*$ is a 10×5 matrix whose ith row is $\mathbf{x}_i' = [(i-5.5)^0, \ldots, (i-5.5)^4]$, and $k = N/10$. The replication structure (blocks of 10) was chosen so that the effects of increasing sample size are not confounded with effects of varying the model matrix.

The vector of coefficients, $\mathbf{c}$, was chosen in one of four ways: (a) c_1 was set to 0 and the remaining coefficients were chosen to maximize $|\rho_3(V)|$, (b) c_1 was set to 1 and the remaining coefficients were chosen to maximize $|\text{E}(V)|$, (c) c_1 was set to 0 and the remaining coefficients were chosen to maximize $[\text{E}(V)]^2 + [\rho_3(V)]^2$ subject to $\kappa_4^*(b) = 2$, and (d) c_1 was set to 1 and the remaining coefficients were chosen to maximize $[\text{E}(V)]^2 + [\rho_3(V)]^2$ subject to $\kappa_4^*(b) = 2$, where $\kappa_4^*(b) = \sum_{i=1}^N b_i^4 / \left(\sum_{i=1}^N b_i^2\right)^2 - 3$, and $\mathbf{b}$ is defined in (7). The quantity $\kappa_4^*(b)$ is a proxy for effective sample size, N_b in (10). The effective sample sizes under condition (a) are near the minimal values and under condition (b) are near the maximal values. Conditions (c) and (d) yield intermediate effective sample sizes and likely are somewhat more representative of actual practice. Bias, skewness, and effective sample sizes are displayed in Table 3. These quantities depend only on $\mathcal{R}(\mathbf{X})$ and not on the specific structure of $\mathbf{X}$. Accordingly, multi-collinearity is not an issue. For each of the $3 \times 6 \times 4 = 240$ conditions, 5,000 data sets were generated and analyzed.

Figures 1 and 2 display the coverage probabilities of eight one-sided nominal 95% interval estimators under conditions (b) and (c). The results under conditions (a) and (d) are similar to those under (c) and are displayed in the supplement. Sub-plots in each figure display coverage for one method over the three log-normal distributions. Solid (dashed) line segments reflect coverage of one-sided lower (upper) intervals. The interval estimators are numbered as in Table 2. Bootstrap intervals were based on 1,000 bootstrap samples. Johnson's (1978) and Konishi's (1991) transformations were not evaluated because they fail to produce an interval if sample size is too small and skewness is too large.

Figures 1 and 2 reveal that second- and third-order accurate intervals (methods 2–8) are superior to conventional t intervals. Furthermore, under conditions (a), (c), and (d), method 6 third-order accurate intervals generally are

Table 3 Simulation study: bias, skewness, and effective sample size

		(a) Maximum $\lvert\rho_3(V)\rvert$, $c_1 = 0$				(b) Maximum $\lvert \mathrm{E}(V)\rvert$, $c_1 = 1$			
N	$\mathrm{Min}_{\mathbf{c}} N_b$	$\mathrm{E}(V)/\kappa_3$	$\rho_3(V)/\kappa_3$	N_b	$\kappa_4^*(b)$	$\mathrm{E}(V)/\kappa_3$	$\rho_3(V)/\kappa_3$	N_b	$\kappa_4^*(b)$
10	1.07	−0.052	−1.036	1.21	3.871	−0.173	−0.680	6.91	−1.702
20	2.13	−0.011	−0.590	2.40	4.027	−0.113	−0.448	16.72	−1.930
30	3.20	−0.006	−0.460	3.59	4.035	−0.092	−0.365	26.74	−1.972
40	4.27	−0.003	−0.390	4.78	4.039	−0.079	−0.316	36.76	−1.985
50	6.40	−0.002	−0.312	7.17	4.040	−0.064	−0.258	56.78	−1.994
100	10.67	−0.001	−0.239	11.96	4.042	−0.050	−0.200	96.80	−1.998
		(c) Maximum $\mathrm{E}^2(V) + \rho_3^2(V)$, $c_1 = 0$				(d) Maximum $\mathrm{E}^2(V) + \rho_3^2(V)$, $c_1 = 1$			
N	$\mathrm{Min}_{\mathbf{c}} N_b$	$\mathrm{E}(V)/\kappa_3$	$\rho_3(V)/\kappa_3$	N_b	$\kappa_4^*(b)$	$\mathrm{E}(V)/\kappa_3$	$\rho_3(V)/\kappa_3$	N_b	$\kappa_4^*(b)$
10	1.07	−0.064	−0.953	1.44	2.000	−0.081	−1.000	1.44	2.000
20	2.13	−0.015	−0.493	2.87	2.000	−0.032	−0.555	2.87	2.000
30	3.20	−0.007	−0.373	4.31	2.000	−0.022	−0.428	4.31	2.000
40	4.27	−0.005	−0.312	5.75	2.000	−0.017	−0.361	5.74	2.000
60	6.40	−0.002	−0.247	8.62	2.000	−0.013	−0.288	8.62	2.000
100	10.67	−0.001	−0.186	14.37	2.000	−0.009	−0.219	14.36	2.000

superior and never are inferior to other methods, including the bootstrap. The advantage of method 6 intervals over second-order accurate (methods 2 and 3) is less apparent under condition (b), where effective sample size increases most rapidly. In this case, intervals based on methods 2, 3, and 6 are similar and are superior to bootstrap t intervals. Coverage of nominal 95% two-sided symmetric intervals under conditions (b) and (c) is displayed in Fig. 3. Coverage under conditions (a) and (d) is similar and is displayed in the supplement. In all cases, method 11 intervals are superior to both conventional $|t|$ intervals and bootstrap-$|t|$ intervals.

8 Concluding comments

The simulation results reported in Sect. 7 verify that skewness and kurtosis corrections can yield interval estimators that are superior to conventional large sample intervals. Furthermore, the third-order accurate intervals based on methods 6 and 11 appear to perform as well or better than do the the remaining intervals, including those based on the bootstrap. Method 6 (Corollary 9c) is recommended for one-sided or equal-tailed two-sided confidence intervals. Method 11 in (22) is recommended for two-sided symmetric intervals. Simulations in which data were sampled from gamma or beta distributions also were performed. The results of these simulations are not displayed because they do not lead to different conclusions. In practice, it is unlikely that the linear functions of interest will happen to be those that maximize skewness, bias, or some function of skewness and bias. Accordingly, the coverage of the recommended third-order accurate intervals likely will be even closer to the nominal coverage than Figs. 1, 2, 3 would suggest.

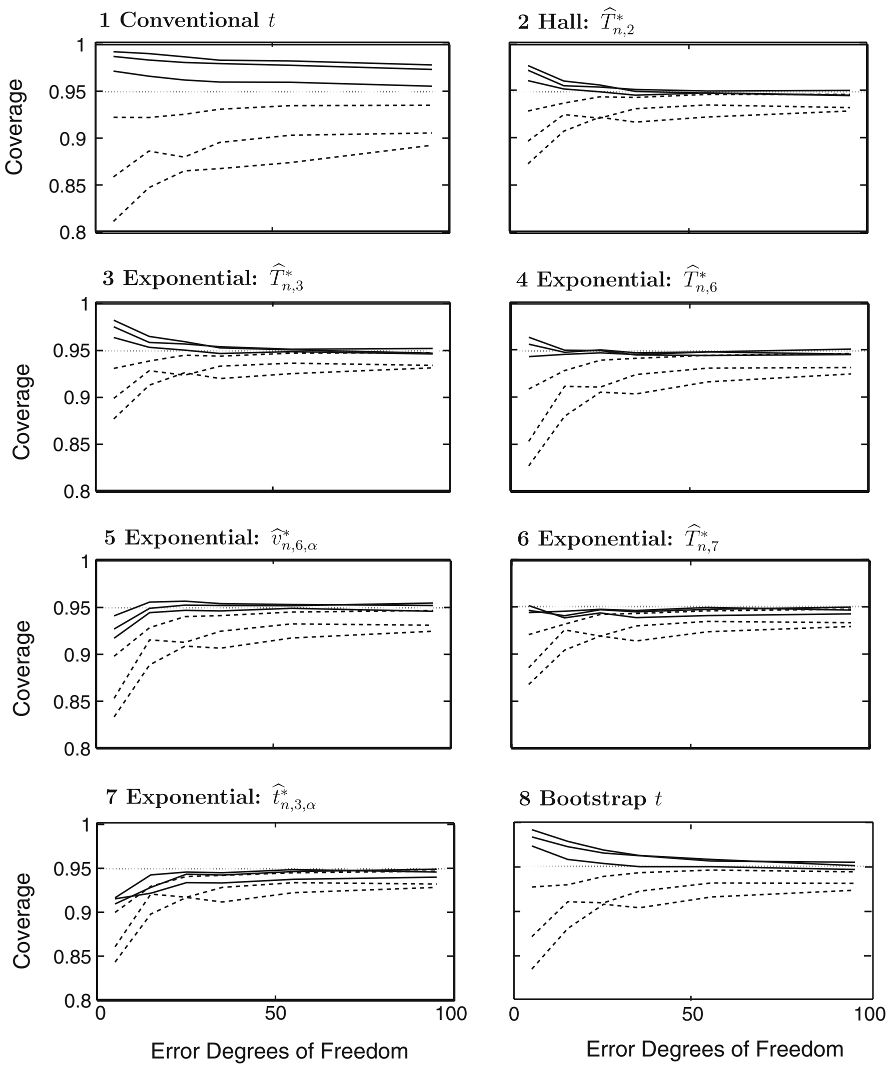

Fig. 1 Coverage of one-sided intervals under condition (**b**) in Table 3

An alternative approach to constructing accurate confidence intervals could begin by developing an accurate approximation to the distribution of the conventional F statistic for testing a family of linear functions under non-normal conditions. One such approximation was obtained by Yanagihara (2003). The modified omnibus test could be inverted to obtain modified simultaneous confidence intervals. One disadvantage of this alternative approach is that higher-order accurate simultaneous coverage is attained by making identical adjustments to each interval. As illustrated in Sect. 6, however, intervals for different linear functions within the same study require different adjustments

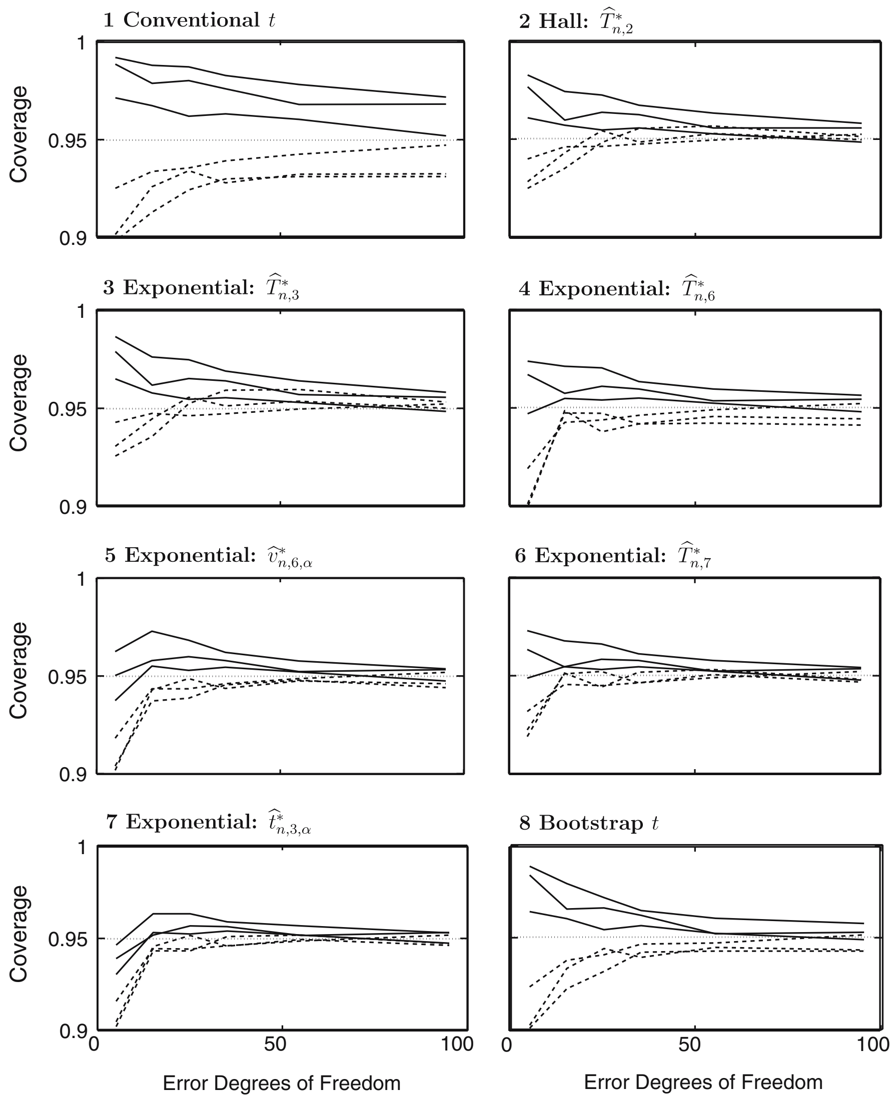

Fig. 2 Coverage of one-sided intervals under condition (**c**) in Table 3

to attain second- or third-order accurate coverage on an interval by interval basis.

The expansions in Sect. 4 are valid if the distribution of ε_i contains an absolutely continuous part. If the distribution of ε_i is discrete, then the expansions may still hold if the values of the explanatory variables do not cluster around too few points. See Kong and Levin (1996) for a verification of this result in logistic regression.

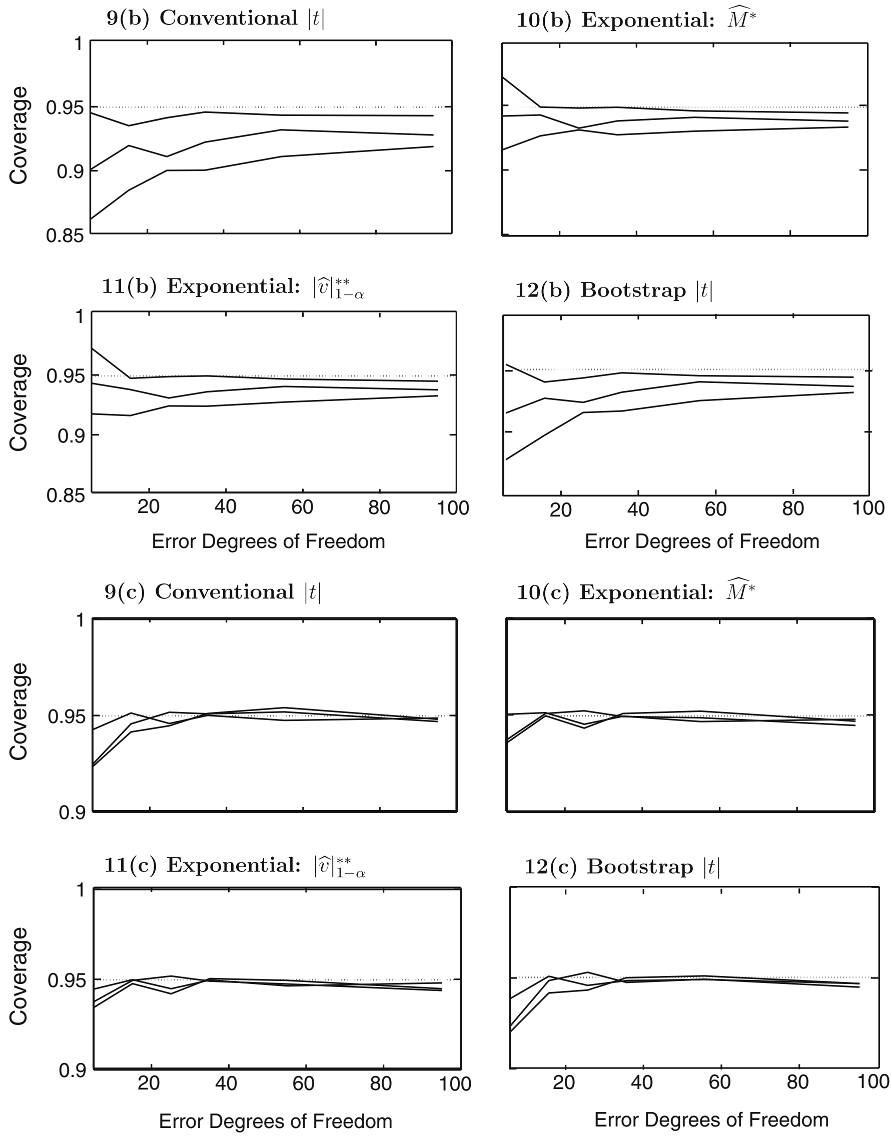

Fig. 3 Coverage of two-sided symmetric intervals under conditions (**b**) and (**c**) in Table 3

References

Boik, R. J. (1998). A local parameterization of orthogonal and semi-orthogonal matrices with applications. *Journal of Multivariate Analysis*, *67*, 244–276.

Box, G. E. P., Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*, 211–252.

Bowman, K. O., Beauchamp, J. J., Shenton, L. R. (1977). The distribution of the *t*-statistic under non-normality. *International Statistical Review*, *45*, 233–242.

Chung, K. L. (1946). The approximate distribution of Student's statistic. *The Annals of Mathematical Statistics*, *17*, 447–465.
Cornish, E. A., Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique*, *5*, 307–320.
Cressie, N. A. C. (1980). Relaxing the assumptions in the one sample *t*-test. *Australian Journal of Statistics*, *22*, 143–153.
Cressie, N. A. C., Sheffield, L. J., Whitford, H.J. (1984). Use of the one-sample *t*-test in the real world. *Journal of Chronic Diseases*, *37*, 107–114.
DiCiccio, T. J., Monti, A. C. (2002). Accurate confidence limits for scalar functions of vector *M*-estimands. *Biometrika 89*, 437–450.
Finney, D. J. (1963). Some properties of a distribution specified by its cumulants. *Technometrics, 5*, 63–69.
Fisher, R. A. (1925). Expansion of "Student's" integral in powers of n^{-1}. *Metron 5*, 109–120.
Fujioka, Y., Maesono, Y. (2000). Higher order normalizing transformations of asymptotic U-statistics for removing bias, skewness and kurtosis. *Journal of Statistical Planning and Inference 83*, 47–74.
Gayen, A. K. (1949). The distribution of Student's *t* in random samples of any size drawn from non-normal universes. *Biometrika* , *36*, 353–369.
Geary, R. C. (1947). Testing for normality. *Biometrika 34*, 209–242.
Hall, P. (1983). Inverting an Edgeworth expansion. *The Annals of Statistics 11*, 569–576.
Hall, P. (1987). Edgeworth expansion for Student's *t* statistic under minimal moment conditions. *The Annals of Probability 15*, 920–931.
Hall, P. (1989). Unusual properties of bootstrap confidence intervals in regression problems. *Probability Theory and Related Fields 81*, 247–273.
Hall, P. (1992a). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B 54*, 221–228.
Hall, P. (1992b). *The Bootstrap and Edgeworth Expansion*, Berlin Heidelberg New York: Springer.
Hill, G. W., Davis, A. W. (1968). Generalized expansions of Cornish-Fisher type. *The Annals of Mathematical Statistics 39*, 1264–1273.
Huber, P. (1973). Robust regression, asymptotics, conjectures and Monte Carlo. *The Annals of Statistics 1*, 799–821.
Johnson, N. J. (1978). Modified *t* tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association 73*, 536–544.
Kong, F., Levin, B. (1996). Edgeworth expansions for the conditional distributions in logistic regression models. *Journal of Statistical Planning and Inference 52*, 109–129.
Konishi, S. (1991). Normalizing transformations and bootstrap confidence intervals. *The Annals of Statistics 19*, 2209–2225.
Navidi, W. (1989). Edgeworth expansions for bootstrapping regression models. *The Annals of Statistics 4*, 1472–1478.
Pace, L., Salvan, A. (1997). Principles of Statistical Inference from a Neo-Fisherian Perspective. World Scientific, Singapore.
Pearson, E. S., Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika 62*, 223–241.
Peiser, A. M. (1949). Correction to "Asymptotic formulas for significance levels of certain distributions". *The Annals of Mathematical Statistics 20*, 128–129.
Posten, H. O. (1979). The robustness of the one-sample *t*-test over the Pearson system. *Journal of Statistical Computation and Simulation 9*, 133–149.
Qumsiyeh, M. B. (1990). Edgeworth expansions in regression models. *Journal of Multivariate Analysis 35*, 86–101.
Qumsiyeh, M. B. (1994). Bootstrapping and empirical Edgeworth expansions in multiple linear regression models. *Communications in Statistics – Theory and Methods 23*, 3227–3239.
Ractliffe, J. F. (1968). The effect on the *t* distribution of non-normality in the sampled population. *Applied Statistics*, *17*, 42–48. Corr. (1968), 17, 203.
Rousseeuw, P. J., Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.
Sen, P. K., Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. London: Chapman & Hall.

Staudte, R. G., Sheather, S. J. (1990). *Robust Estimation and Testing*. New York: Wiley.
Tiku, M. L. (1963). Approximation to Student's t distribution in terms of Hermite and Laguerre polynomials. *The Journal of the Indian Mathematical Society*, *27*, 91–102.
Venables, W. N., Ripley, B. D. (2002). *Modern Applied Statistics with S,* 4th edition, Berlin Heidelberg New York: Springer.
Wallace, D. L. (1958). Asymptotic approximations to distributions. *The Annals of Mathematical Statistics*, *29*, 635–654.
Wilcox, R. R. (1997). *Introduction to Robust Estimation and Hypothesis Testing*. San Diego: Academic Press.
Wilcox, R. R. (1998). The goals and strategies of robust methods. *British Journal of Mathematical and Statistical Psychology*, *51*, 1–39.
Yanagihara, H. (2003). Asymptotic expansion of the null distribution of test statistic for linear hypothesis in nonnormal linear model. *Journal of Multivariate Analysis*, *84*, 222–246.
Yanagihara, H., Yuan, K. H. (2005). Four improved statistics for contrasting means by correcting skewness and kurtosis. *British Journal of Mathematical and Statistical Psychology, 58*, 209–237.