Ann Inst Stat Math (2014) 66:961–981
DOI 10.1007/s10463-013-0434-9

Extensions of saddlepoint-based bootstrap inference

Robert L. Paige · A. Alexandre Trindade · R. Indika P. Wickramasinghe

Received: 24 October 2012 / Revised: 29 August 2013 / Published online: 6 November 2013

Abstract We propose two substantive extensions to the saddlepoint-based bootstrap (SPBB) methodology, whereby inference in parametric models is made through a monotone quadratic estimating equation (QEE). These are motivated through the first-order moving average model, where SPBB application is complicated by the fact that the usual estimators, method of moments (MOME), least squares, and maximum likelihood (MLE), all have mixed distributions and tend to be roots of high-order polynomials that violate the monotonicity requirement. A unifying perspective is provided by demonstrating that these estimators can all be cast as roots of appropriate QEEs. The first extension consists of two double saddlepoint-based Monte Carlo algorithms for approximating the Jacobian term appearing in the approximated density function of estimators derived from a non-monotone QEE. The second extension considers inference under QEEs from exponential power families. The methods are demonstrated for the MLE under a Gaussian distribution, and the MOME under a joint Laplace distribution for the process.

R. L. Paige
Department of Mathematics and Statistics, Missouri University of Science and Technology,
Rolla, MO 65409, USA
e-mail: paigero@mst.edu

A. A. Trindade (✉)
Department of Mathematics and Statistics, Texas Tech University,
Lubbock, TX 79409, USA
e-mail: alex.trindade@ttu.edu

R. I. P. Wickramasinghe
Department of Mathematical Sciences, Eastern New Mexico University,
Portales, NM 88130, USA
e-mail: indika.wickramasinghe@enmu.edu

Keywords Saddlepoint approximation · Estimating equation · Mixed distribution · Moving average model · MA(1) · Elliptically contoured distribution · Exponential power distribution

1 Introduction

Many statistical applications involve inference primarily on a (scalar) parameter θ in the presence of a finite dimensional nuisance parameter $\boldsymbol{\lambda}$. Paige et al. (2009) considered a class of problems in which the estimator $\hat{\theta}$ of θ is a solution of

$$\Psi(\theta; \theta_0, \boldsymbol{\lambda}_0) \equiv \Psi_0(\theta) = \mathbf{y}^{\mathsf{T}} A_\theta \mathbf{y} = 0, \tag{1}$$

where A_θ is a conformable symmetric matrix whose entries are functions of θ, and the random vector $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}_{\theta_0,\boldsymbol{\lambda}_0}, \Sigma_{\theta_0,\boldsymbol{\lambda}_0})$ is normally distributed with mean $\boldsymbol{\mu}_0 \equiv \boldsymbol{\mu}_{\theta_0,\boldsymbol{\lambda}_0}$ and covariance matrix $\Sigma_0 \equiv \Sigma_{\theta_0,\boldsymbol{\lambda}_0}$, where we make explicit the dependence of these quantities on the true or hypothesized values of $\theta = \theta_0$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$. We call $\Psi_0(\theta)$ a *quadratic estimating equation (QEE)*. The multivariate normality immediately furnishes a closed-form expression for the moment generating function (MGF) of the QEE. If the QEE is monotone in θ, then it is possible to relate the cumulative distribution function (CDF) or probability density function (PDF) of $\hat{\theta}$ to that of $\Psi_0(\theta)$. The nuisance parameter(s) $\boldsymbol{\lambda}$ is dealt with by substituting a conditional maximum likelihood estimator (MLE) $\hat{\boldsymbol{\lambda}}_\theta$. Using saddlepoint approximations, it is then possible to accurately approximate the distribution of the estimator of interest. Confidence intervals for θ can be produced by inverting (or pivoting) this distribution.

Paige et al. (2009) developed the *SaddlePoint-Based Bootstrap* (SPBB) methodology in a coherent manner. Their choice of name reflects the fact that the technique is identical to a parametric bootstrap, but with (slow) Monte Carlo simulation replaced by (fast) saddlepoint approximation. The essential steps of this approach are:

(i) formulation of an estimating equation for the estimator of interest;
(ii) substitution of conditional MLEs for any nuisance parameters, resulting in a monotone (in θ) profile estimating equation in the form of a QEE having the estimator as its unique root;
(iii) a distributional assumption for the data vector $\mathbf{y}$ which ensures the QEE has a closed-form expression for the MGF;
(iv) inversion of the MGF of the QEE to produce accurate saddlepoint approximations to the CDF or PDF of the estimator; and
(v) pivoting of this CDF to produce a confidence interval (C.I.) for the parameter of interest.

A saddlepoint approximation for the CDF of $\hat{\theta}$ at any specified value t in its support can now be obtained from the formula of Lugannani and Rice (1980)

$$\hat{F}_{\hat{\theta}}(t; \theta_0, \boldsymbol{\lambda}_0) = \hat{F}_{\Psi(t)}\,(0; \theta_0, \boldsymbol{\lambda}_0) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})\left[\hat{w}^{-1} - \hat{u}^{-1}\right], & \text{if } \mathbb{E}[\Psi(t)] \neq 0, \\ \frac{1}{2} + K'''_{\Psi(t)}(0; \theta_0, \boldsymbol{\lambda}_0)\left[72\pi K''_{\Psi(t)}(0; \theta_0, \boldsymbol{\lambda}_0)^3\right]^{-1/2}, & \text{if } \mathbb{E}[\Psi(t)] = 0, \end{cases}$$

where $\hat{w} = \operatorname{sgn}(\hat{s})[-2K_{\Psi(t)}(\hat{s}; \theta_0, \boldsymbol{\lambda}_0)]^{1/2}$, $\hat{u} = \hat{s}[K''_{\Psi(t)}(\hat{s}; \theta_0, \boldsymbol{\lambda}_0)]^{1/2}$, and $\Phi(\cdot)$ and $\phi(\cdot)$ denote, respectively, the CDF and PDF of a standard normal random variable. A corresponding formula for the PDF follows from Daniels (1983). Dropping the subscript on $\Psi(\cdot)$ for notational expediency, these formulas are functions of the cumulant generating function (CGF) of $\Psi(t)$, denoted by $K_{\Psi(t)}(s; \theta_0, \boldsymbol{\lambda}_0)$, and its first few derivatives with respect to both s and t. The most computationally expensive step involves finding $\hat{s}$ at each point t, by solving the (nonlinear) *saddlepoint equation*

$$\frac{\partial}{\partial s} K_{\Psi(t)}(s; \theta_0, \boldsymbol{\lambda}_0)\bigg|_{s=\hat{s}} \equiv K'_{\Psi(t)}\left(\hat{s}; \theta_0, \boldsymbol{\lambda}_0\right) = 0.$$

The result, upon substitution of the conditional MLE $\hat{\boldsymbol{\lambda}}_{\theta_0}$ for the nuisance parameter $\boldsymbol{\lambda}_0$, is the CDF approximation

$$F_{\hat{\theta}}(t; \theta_0, \boldsymbol{\lambda}_0) \approx \hat{F}_{\hat{\theta}}\left(t; \theta_0, \hat{\boldsymbol{\lambda}}_{\theta_0}\right) = \hat{F}_{\Psi(t)}\left(0; \theta_0, \hat{\boldsymbol{\lambda}}_{\theta_0}\right),$$

which is third-order accurate over sets of bounded central tendency. With $\hat{\theta}_{\text{obs}}$ the estimated θ obtained from the observed sample, this leads immediately to an automatic approximate percentile confidence set construction method, where the lower and upper bounds (θ_L, θ_U) of the desired $(1-\alpha)100\%$ (equi-tailed) C.I. for θ_0 are determined by solving

$$\hat{F}_{\hat{\theta}}\left(\hat{\theta}_{\text{obs}}; \theta_L, \hat{\boldsymbol{\lambda}}_{\theta_L}\right) = 1 - \alpha/2 \quad \text{and} \quad \hat{F}_{\hat{\theta}}\left(\hat{\theta}_{\text{obs}}; \theta_U, \hat{\boldsymbol{\lambda}}_{\theta_U}\right) = \alpha/2,$$

or equivalently,

$$\hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}\left(0; \theta_L, \hat{\boldsymbol{\lambda}}_{\theta_L}\right) = 1 - \alpha/2 \quad \text{and} \quad \hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}\left(0; \theta_U, \hat{\boldsymbol{\lambda}}_{\theta_U}\right) = \alpha/2. \tag{2}$$

Paige et al. (2009) showed that the resulting SPBB C.I. is 2nd-order accurate in the sense of Hall (1988), so that the coverage probability of a nominal $(1-\alpha)100\%$ C.I. is $1-\alpha+O(n^{-1})$. The profiling out of $\boldsymbol{\lambda}$ limits the general applicability of SPBB, but when it is possible it provides a device for eliminating nuisance parameters in the absence of a pivotal quantity. Extensive examples include AR(1) time series models (Paige and Trindade 2008); nonlinear parameters in conditionally linear nonlinear regression and ratios of regression parameter problems (Paige et al. 2009; Paige and Fernando 2008); smoothing parameters in penalized spline models with independent errors (Paige and Trindade 2013) and with correlated errors (Paige and Trindade 2010).

All indications are that SPBB yields a C.I. with length and coverage probability that compares very favorably with those obtained from competing methods, many of which have 2nd or 3rd order accuracy, and some of which are exact. As Young (2009) concludes, the parametric bootstrap with the conditional MLE substituted for the nuisance parameter (what he calls the "constrained bootstrap") generally yields the easiest route to $O(n^{-3/2})$ inference. Additionally, SPBB enjoys faster computational speeds. Paige and Trindade (2013) compared the performance of bootstrap, exact, and

SPBB methods for inference on the smoothing parameter in penalized spline models under a variety of optimality criteria, such as those commonly denoted by ML, REML, GCV, and AIC. A key insight in Paige and Trindade (2013) and Paige and Trindade (2010) was a unification of all these criteria by viewing the estimator as the root of an appropriate QEE. In benchmark comparisons, Paige and Trindade (2013) find SPBB to be an order of magnitude faster than exact methods devised for ML and REML, and two orders of magnitude faster than the bootstrap. In summary, not only does SPBB compete well with exact methods (e.g. ML and REML), but is also the only computationally feasible alternative where no other methods exist (e.g. GCV and AIC). Furthermore, in all instances where a comparison can be made, SPBB appears to deliver a performance that is nearly exact.

Although SPBB has proven to be very successful, the most serious restrictions occur in steps (ii) and (iii). In a general situation, the profiled estimating equation may not be a QEE, and even if it is, may not be one in normal random variables. In fact, as we shall show, all that is required is that the underlying estimating equation $\Psi(\theta)$ has a tractable MGF, or indeed, that it can be approximated with one. Also, even though the monotonicity requirement can be relaxed to just assuming that there exists an interval where this happens and within which with high probability the MLE lies (assured by the appropriate regularity conditions guaranteeing consistency and asymptotic normality), this tends to happen only for larger sample sizes. Since the real utility of SPBB is its good small-sample performance (inherited from the accuracy of the saddlepoint approximation), it would be of interest to find a solution that would not compromise this aspect of the methodology.

The purpose of this paper is to, therefore, extend SPBB in these directions. Although we strive for general results, these will be illustrated with respect to the first-order moving average model or MA(1) in time series (e.g. Brockwell and Davis 1991)

$$X_t = \theta_0 Z_{t-1} + Z_t, \quad Z_t \sim \text{iid}\ (0, \sigma^2), \tag{3}$$

where $\{Z_t\}$ is sequence of zero-mean independent and identically distributed (iid) random variables with variance σ^2. Seemingly a simple model, for n observations the (Gaussian) MLE for the (true) coefficient θ_0 can be expressed as the root of a QEE, but is a polynomial of degree (approximately) $2n$ in θ, as is shown in Section 2. The small-sample properties of the MLE were discussed by Cryer and Ledolter (1981), from which it emerges that it is a mixture of a continuous density on $(-1, 1)$ and point masses at ± 1 (this is also true of simpler estimators like method of moments and least squares). Basic issues in the asymptotic regimes were only recently settled by Davis and Dunsmuir (1996). From this perspective, the *unit-root* case of $\theta_0 = -1$ deserves special attention as it has implications concerning, for example, over-differencing of a series. Thus, devising unit-root tests, and investigating the asymptotics of unit-root and near unit-root settings, has been the subject of some activity (Tanaka 1990; Davis et al. 1995; Davis and Dunsmuir 1996; Davis and Song 2011).

Application of SPBB in this setting is, therefore, challenging. For estimators like method of moments which is endowed with a monotone QEE, the existing methodology works well. For estimators with a non-monotone QEE like the MLE, the results can be poor for small sample sizes. This is due to the Daniels (1983) approach for constructing the Jacobian being heavily dependent on the monotonicity. As a solution we propose in Sect. 3 that the Jacobian be computed via the method of Skovgaard (1990). Since the latter relies on the computation of an intractable conditional expectation, we propose an algorithm to saddlepoint approximate the underlying conditional density. Implementation then proceeds by either numerically integrating this density, or via a Monte Carlo scheme such as importance sampling.

Section 4 considers SPBB generalizations to estimating equations under distributions with tractable QEE MGF promise such as skew-normal and elliptically contoured families. A member of the latter we focus on is the *exponential power* (EP) distribution introduced by Gomez et al. (1998). An expression is developed for the MGF of an EP QEE that is "closed-form" to within a one-dimensional integral. To be amenable to SPBB, however, the expression must furnish tractable derivatives for the logarithm of the MGF. We show how these can be obtained by Laplace approximating the integral after truncation of the characteristic polynomial expansion for a key determinant term appearing in the integrand. The method is shown to work well under monotone QEEs.

2 SPBB for the Gaussian MA(1)

Consider the QEE in (1) with $\hat{\theta}$ as the target estimator. As outlined in Paige et al. (2009), under monotonicity of the QEE, $\hat{\theta}$ is its unique root, and one then has that either $\hat{\theta} \leq t \Leftrightarrow \Psi(t) \leq 0$ or $\hat{\theta} \leq t \Leftrightarrow \Psi(t) \geq 0$. If we assume the former (monotone decreasing) case without loss of generality, this leads to the device $P(\hat{\theta} \leq t) = P(\Psi(t) \leq 0)$, and so the CDF of $\hat{\theta}$ at t, denoted by $F_{\hat{\theta}}(t)$, can be expressed in terms of the CDF of $\Psi(t)$ at 0, denoted by $F_{\Psi(t)}(0)$. Furthermore, if $\mathbf{y}$ is multivariate normal, one has a closed-form expression for the MGF of $\Psi(t)$, and this permits one to accurately (saddlepoint) approximate $F_{\hat{\theta}}(t)$ via $\hat{F}_{\Psi(t)}(0)$. Pivoting this approximated CDF then allows for the construction of a test and C.I. In this section, we apply the method to explore the small-sample properties of some common estimators of the MA(1) model.

Let I_n denote the $(n \times n)$ identity matrix, and J_n the $(n \times n)$ symmetric matrix with 1 on the first-order off-diagonals (and zero otherwise). For a vector of observations $\boldsymbol{x} = [x_1, \ldots, x_n]^\mathsf{T}$ from model (3), we note from Cryer and Ledolter (1981) that the Gaussian profile log-likelihood for θ is proportional to $\ell(\theta) \propto -\log(\boldsymbol{x}^\mathsf{T} \Omega_\theta^{-1} |\Omega_\theta|^{1/n} \boldsymbol{x})$, where

$$\Omega_\theta = \theta J_n + (1 + \theta^2) I_n. \tag{4}$$

Thus, maximization of $\ell(\theta)$ to obtain the MLE $\hat{\theta}_{\mathrm{ML}}$ is equivalent to determining an appropriate root of the QEE $\Psi_{\mathrm{ML}}(\theta) = \boldsymbol{x}^\mathsf{T}[\partial(\Omega_\theta^{-1}|\Omega_\theta|^{1/n})/\partial\theta]\boldsymbol{x} \equiv \boldsymbol{x}^\mathsf{T} A_\theta^{(\mathrm{ML})} \boldsymbol{x}$. Cryer and Ledolter (1981) also provide an explicit expression for the entries of the matrix $\Omega_\theta^{-1}|\Omega_\theta|^{1/n}$, and note that $\ell(\theta) = \ell(1/\theta)$, from which it follows that ± 1 is always a critical value of $\ell(\theta)$ and hence there is a positive probability that the MLE occurs at the invertibility boundary. As such, the estimator is a mixture of a continuous density on $(-1, 1)$ and point masses at ± 1. Values of $\theta = \pm 1$ result in a non-invertible MA(1), and the probabilities of such occurring for the MLE were further investigated by Anderson and Takemura (1986).

Other common estimators are (unconditional) least squares, conditional least squares (CLSE), and method of moments (MOME). Their respective densities exhibit the same mixture distribution structure as the MLE, and all are expressible also as the root of a QEE. The unconditional least squares estimator is essentially identical to the MLE, and we will not discuss it further. The MOME is obtained by equating the sample autocovariances of the process to their model counterparts, resulting in the estimator (Brockwell and Davis 1991, § 8.5)

$$\hat{\theta}_{\text{MoM}} = \mathbb{1}_{[\hat{\rho}_1 > 1/2]} + \frac{1 - \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1} \mathbb{1}_{[|\hat{\rho}_1| \leq 1/2]} - 1\mathbb{1}_{[\hat{\rho}_1 < -1/2]},$$

where $\hat{\rho}_1 = (\boldsymbol{x}^\top J_n \boldsymbol{x})/(\boldsymbol{x}^\top 2I_n \boldsymbol{x})$ is the sample autocorrelation function (acf) at lag 1. It can be obtained as the unique root of the monotone QEE $\Psi_{\text{MoM}}(\theta) = \boldsymbol{x}^\top A_\theta^{(\text{MoM})} \boldsymbol{x}$, where the form of $A_\theta^{(\text{MoM})}$ is given in Proposition 1 below. The CLSE is defined as the minimizer of the least-squares expression $S(\theta) = \sum_{t=1}^n z_t^2$ (e.g. Cryer and Chan 2008, §7.2; Shumway and Stoffer 2011, §3.6). Appealing to the invertible representation of the MA(1) process $Z_t = \sum_{j=0}^\infty (-\theta)^j X_{t-j}$, and substituting into $S(\theta)$ gives the criterion

$$\begin{aligned} S(\theta) = x_1^2 + (x_2 - \theta x_1)^2 + \left(x_3 - \theta x_2 + \theta^2 x_1\right)^2 \\ + \cdots + \left(x_n - \theta x_{n-1} + ... + (-\theta)^{n-1} x_1\right)^2, \end{aligned}$$

which is obtained by setting non-observed values of X_t equal to (their mean value of) zero. As shown by Plosser and Schwertz (1977), this is equivalent to minimizing the expression $\boldsymbol{x}^\top \Omega_\theta^{-1} \boldsymbol{x}$, and thus the resulting estimator, $\hat{\theta}_{\text{LS}}$, is an appropriate root of the QEE $\Psi_{\text{LS}}(\theta) = \boldsymbol{x}^\top A_\theta^{(\text{LS})} \boldsymbol{x}$. We summarize these findings and give the form of the respective QEEs in the following proposition.

Proposition 1 *With I_n, J_n, and Ω_θ as in (4), the estimators $\hat{\theta}_{MoM}$, $\hat{\theta}_{LS}$, and $\hat{\theta}_{ML}$ described above are obtained equivalently as appropriate roots of the QEE $\Psi(\theta) = \boldsymbol{x}^\top A_\theta \boldsymbol{x}$, where the form of the symmetric matrix A_θ, $|\theta| < 1$, is given in each respective case as follows.*

MOME: $A_\theta^{(MoM)} = (1 + \theta^2) J_n - 2\theta I_n$, *and the resulting QEE* $\Psi_{MoM}(\theta) = \boldsymbol{x}^\top A_\theta^{(MoM)} \boldsymbol{x}$ *is monotone in θ over the interval $|\theta| < 1$.*
CLSE: $A_\theta^{(LS)} = \Omega_\theta^{-1}[J_n + 2\theta I_n]\Omega_\theta^{-1}$.
MLE: $A_\theta^{(ML)} = \frac{2\theta[1-(n+1)\theta^{2n} + n\theta^{2n+2}]}{(1-\theta^2)^2} \Omega_\theta^{-1} - \frac{n(1-\theta^{2n+2})}{1-\theta^2} \Omega_\theta^{-1}[J_n + 2\theta I_n]\Omega_\theta^{-1}$.

Proof For $|\hat{\rho}_1| \leq 1/2$, the MOME is obtained by equating model and sample acf's at lag 1

$$\rho_1 = \frac{\theta}{1+\theta^2} \overset{\text{set}}{=} \frac{\boldsymbol{x}^\top J_n \boldsymbol{x}}{\boldsymbol{x}^\top 2 I_n \boldsymbol{x}} = \hat{\rho}_1 \quad \Longrightarrow \quad \Psi_{\text{MoM}}(\theta) \equiv \boldsymbol{x}^\top [(1+\theta^2) J_n - 2\theta I_n] \boldsymbol{x} = 0.$$

To see why $\Psi_{\text{MoM}}(\theta)$, being a quadratic in θ, is monotone over $|\theta| < 1$, it suffices to show that its turning point $\tilde{\theta}$ must occur outside the interval $(-1, 1)$. To this end, differentiating and setting equal to zero yields

$$\frac{\partial \Psi_{\text{MoM}}(\theta)}{\partial \theta} = \boldsymbol{x}^{\mathsf{T}}[2\theta J_n - 2I_n]\boldsymbol{x} \overset{\text{set}}{=} 0 \quad \Longrightarrow \quad \tilde{\theta} = \frac{\boldsymbol{x}^{\mathsf{T}} J_n \boldsymbol{x}}{\boldsymbol{x}^{\mathsf{T}} I_n \boldsymbol{x}} = \frac{1}{2\hat{\rho}_1}.$$

Now, since $|\hat{\rho}_1| \leq 1/2$, we have $|2\hat{\rho}_1| \leq 1$ which implies $|\tilde{\theta}| \geq 1$, and the result follows. The result for CLSE follows straightforwardly by differentiating the QEE $\boldsymbol{x}^{\mathsf{T}}\Omega_\theta^{-1}\boldsymbol{x}$. For the MLE case, see Wickramasinghe (2012). For implementation purposes, note that Plosser and Schwertz (1977) gave the form of the (i, j)th element of Ω_θ^{-1} as

$$\Omega_\theta^{-1}(i, j) = (-\theta)^{j-i}\frac{(1-\theta^{2i})(1-\theta^{2(n+1-j)})}{(1-\theta^2)(1-\theta^{2(n+1)})}, \quad i \leq j \text{ and } |\theta| < 1. \qquad \square$$

All estimators are consistent and asymptotically normal, i.e. $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, v_\theta)$, where the asymptotic variance is given by $v_\theta = 1 - \theta^2$ for the CLSE and MLE, and $v_\theta = (1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8)/(1 - \theta^2)^2$ for the MOME (Brockwell and Davis 1991, §8.5). Probabilities of the MLE occurring at the non-invertible boundary values ± 1 were computed by Cryer and Ledolter (1981) for the relatively simple case of $n = 2$. For larger n, this is a difficult task due to the lack of an explicit criterion for determining when the MLE and CLSE occur on the boundary. For the MOME, however, the criterion is simply $|\hat{\rho}_1| > 1/2$, and SPBB can, therefore, be routinely used since the fact that $\hat{\rho}_1$ can be expressed as a ratio of quadratic forms yields the following equivalence in terms of QEEs,

$$\begin{aligned} P(\hat{\theta}_{\text{MoM}} = -1) = P\left(\boldsymbol{x}^{\mathsf{T}}[J_n + I_n]\boldsymbol{x} < 0\right) \quad \text{and} \quad P(\hat{\theta}_{\text{MoM}} = +1) \\ = P\left(\boldsymbol{x}^{\mathsf{T}}[J_n - I_n]\boldsymbol{x} > 0\right). \end{aligned}$$

Table 1 shows some values of these saddlepoint approximated MOME probabilities for sample sizes $n = \{10, 20\}$, and $\theta_0 = \{0.4, 0.8\}$. Also included for later reference are Monte Carlo approximations to the corresponding probabilities for the MLE.

Table 1 Approximate probabilities for the MOME (MoM) and MLE (ML) to occur at the non-invertible boundary values of ± 1 for samples of size n from a Gaussian MA(1) model with coefficient θ_0 and $\sigma^2 = 1$

n	θ_0	$P(\hat{\theta}_{\text{MoM}} = -1)$	$P(\hat{\theta}_{\text{MoM}} = +1)$	$P(\hat{\theta}_{\text{ML}} = -1)$	$P(\hat{\theta}_{\text{ML}} = +1)$
10	0.4	0.00000	0.00006	0.017	0.188
10	0.8	0.00000	0.00008	0.007	0.498
20	0.4	0.00000	0.00003	0.001	0.058
20	0.8	0.00000	0.00009	0.000	0.364

The MOME and MLE values were obtained via saddlepoint and Monte Carlo approximation, respectively

Remark 1 In all cases, and for the remainder of this paper, it suffices to consider an MA(1) with $\sigma^2 = 1$. The reason for this is that QEE-based inference is invariant to scale parameters. If $\hat{\theta}$ is a root of $\boldsymbol{x}^\mathsf{T} A_\theta \boldsymbol{x}$ with σ a one-dimensional scale parameter for $\boldsymbol{x}$, then $\boldsymbol{z} = \boldsymbol{x}/\sigma$ is scale-free. Thus, we obtain

$$0 = \boldsymbol{x}^\mathsf{T} A_{\hat{\theta}} \boldsymbol{x} = \sigma^2 (\boldsymbol{x}/\sigma)^\mathsf{T} A_{\hat{\theta}} (\boldsymbol{x}/\sigma) = \sigma^2 \boldsymbol{z}^\mathsf{T} A_{\hat{\theta}} \boldsymbol{z},$$

so that $\hat{\theta}$ is also a root of an equivalent QEE with $\sigma = 1$.

Figure 1 displays SPBB-approximated PDF's for the 3 estimators of θ in Proposition 1 (MOME, CLSE, MLE), for samples of size $n = \{10, 20\}$ and $\theta_0 = \{0.4, 0.8\}$. The AN curve is the PDF of a $N(\theta, (1-\theta^2)/n)$, which corresponds to the asymptotically normal distribution for the CLSE and MLE. In this regard, we notice a substantial deterioration in the performance of the AN approximation for low sample sizes and θ_0 close to the invertibility boundary. As is well-known the MOME exhibits substantial bias; its only real utility being to initialize optimization algorithms for nonlinear estimators like MLE and CLSE.

Theoretical justification for the use of saddlepoint CDF approximations in mixed distributions is provided by Lund et al. (1999). Although justification of the same for the PDF is not yet a settled issue, the PDF plots in Fig. 1 help shed light on the matter. To assess their accuracy, a common tool is the calculation of *percent relative errors* (PREs). Let $\hat{F}_{\text{sim}}(t)$ and $\hat{F}_{\text{sad}}(t)$ denote, respectively, the true and saddlepoint approximated CDF at t, where the true CDF is estimated empirically (based on 10^6 simulations). The PRE at t is then defined as follows:

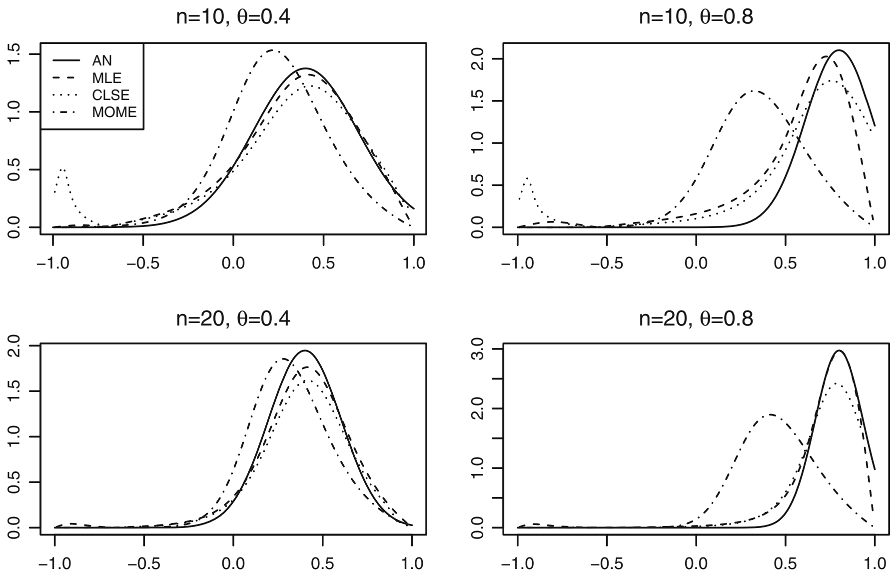

Fig. 1 Saddlepoint approximated densities of the 3 estimators of θ in Proposition 1 (MOME, CLSE, MLE) for the moving average coefficient θ in samples of size n from an MA(1) model. The AN curve corresponds to the asymptotically normal PDF for the CLSE and MLE

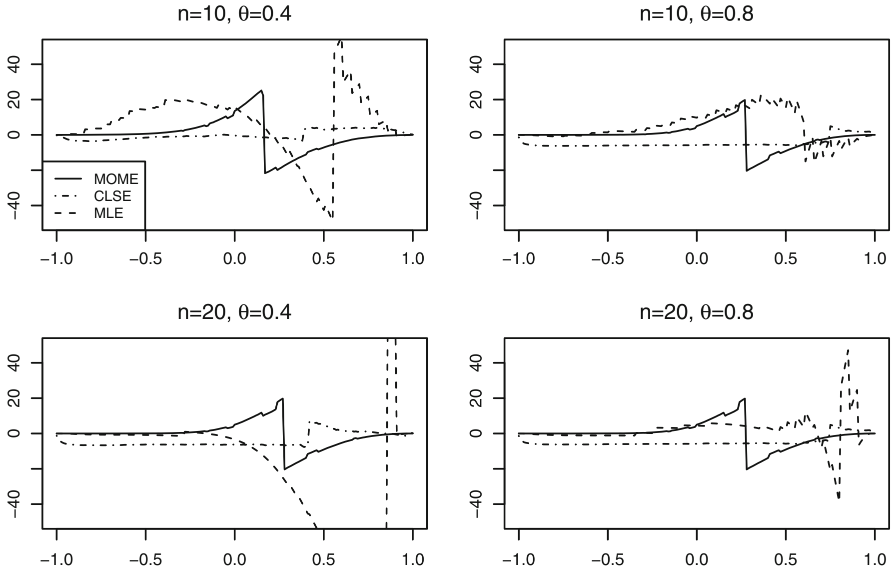

Fig. 2 Percent relative errors (PREs) for the saddlepoint approximated densities of the 3 estimators of θ in Proposition 1 (MOME, CLSE, MLE) for the moving average coefficient θ_0 in samples of size n from an MA(1) model

$$\text{PRE}(t) = \begin{cases} \frac{\hat{F}_{\text{sad}}(t) - \hat{F}_{\text{sim}}(t)}{\hat{F}_{\text{sim}}(t)} 100, & \text{if } \hat{F}_{\text{sim}}(t) \le 0.5, \\ \frac{\hat{F}_{\text{sim}}(t) - \hat{F}_{\text{sad}}(t)}{1 - \hat{F}_{\text{sim}}(t)} 100, & \text{if } \hat{F}_{\text{sim}}(t) > 0.5. \end{cases}$$

Figure 2 displays PREs for the SPBB-approximated PDF's of the 3 estimators (MOME, CLSE, MLE) in Fig. 1. As expected, the PREs for MOME are fairly low ($\pm 20\,\%$). Those for MLE can be quite high, but what is surprising is that the PREs for CLSE are the lowest of all ($\pm 10\,\%$).

Let us now consider the construction of SPBB confidence bounds for θ based on the MOME $\hat{\theta}$. For a vector of observations $\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathsf{T}}$ from model (3), let $\hat{\theta}_{\text{obs}}$ denote the observed MOME value. Suppose that $\hat{\theta}$ has a point mass at -1 of size p_1 and a point mass at $+1$ of size p_2 (and is continuous otherwise). Then, $p_1 \le F_{\hat{\theta}}(\hat{\theta}_{\text{obs}}|\theta_0) \le 1 - p_2$, for $-1 < \hat{\theta}_{\text{obs}} < +1$. It follows that $F_{\hat{\theta}}(\hat{\theta}_{\text{obs}}|\theta_0)$ has a mixed distribution; it takes on the value -1 with probability p_1, the value 1 with probability p_2, and is uniform on $(-1, 1)$ with probability $1 - p_1 - p_2$.

If we are interested in a nominal 95 % C.I., we will have *coverage* whenever $0.025 \le F_{\hat{\theta}}(\hat{\theta}_{\text{obs}}|\theta_0) \le 0.975$. Also note that we will have *underage* when $F_{\hat{\theta}}(\hat{\theta}_{\text{obs}}|\theta_0) < 0.025$, this corresponds to the situation where the C.I. lies completely below θ_0. Similarly, *overage* occurs when $F_{\hat{\theta}}(\hat{\theta}_{\text{obs}}|\theta_0) > 0.975$, and in this situation the C.I. lies completely above θ_0. As an example, generating a 95 % lower confidence bound $(\theta_L, 1)$ would involve solving the equation $F_{\hat{\theta}}(\hat{\theta}_{\text{obs}}|\theta_L) = \alpha$, with $\alpha = 0.05$ if $F_{\hat{\theta}}(\cdot|\theta_0)$ is monotone increasing in θ_0, and $\alpha = 0.95$ if $F_{\hat{\theta}}(\cdot|\theta_0)$ is monotone decreasing in θ_0. In the context of

Table 2 Empirical lengths and coverage probabilities for nominal 95 % lower, upper, and two-sided (interval) confidence bounds for parameter θ_0 produced via MOME, according to the methods: SPBB, nonparametric bootstrap (Boot), and asymptotic normal (AN)

95 % Confidence	Settings		Coverage probability			Average length		
Bound type	n	θ_0	SPBB	Boot	AN	SPBB	Boot	AN
Lower	10	0.4	1.000	1.000	1.000	1.435	1.802	0.875
	10	0.8	0.997	1.000	1.000	1.899	1.774	1.442
	20	0.4	1.000	1.000	1.000	1.110	1.830	0.764
	20	0.8	1.000	1.000	1.000	0.950	1.790	1.681
Upper	10	0.4	0.900	0.452	1.000	1.945	1.749	1.675
	10	0.8	0.903	0.315	0.246	1.998	1.989	1.906
	20	0.4	0.919	0.451	1.000	1.894	1.709	1.531
	20	0.8	0.916	0.311	0.313	1.998	1.985	1.876
Interval	10	0.4	0.940	0.432	0.997	1.484	1.438	0.561
	10	0.8	0.948	0.358	0.259	1.336	1.653	1.300
	20	0.4	0.953	0.717	1.000	1.095	1.560	0.334
	20	0.8	0.960	0.524	0.693	1.005	1.692	1.616

Results are based on 1000 realizations, each of sample size n, simulated from a Gaussian MA(1) model with $\sigma^2 = 1$

SPBB, and with $\hat{F}_{\Psi(\theta)}$ denoting the saddlepoint approximation to the CDF of $\Psi_{\text{MoM}}(\theta)$ given in Proposition 1, this 95% lower confidence bound would be produced by solving

$$\hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}(0 \mid \theta_L) = \begin{cases} 0.05, & \text{if } F_{\Psi(\hat{\theta}_{\text{obs}})}(0|\theta_0) \text{ monotone increasing in } \theta_0, \\ 0.95, & \text{if } F_{\Psi(\hat{\theta}_{\text{obs}})}(0|\theta_0) \text{ monotone decreasing in } \theta_0. \end{cases} \tag{5}$$

Table 2 shows the result of some simulations carried out to compare the length and coverage probability of nominal 95 % MOME confidence bounds for the sample sizes and θ_0 settings as in Table 1. Three methods are compared: SPBB, nonparametric bootstrap (Boot), and the asymptotic normal approximation (AN). The "Lower" bounds correspond to (5); the "Upper" to the same but with the roles of the monotonicity switched; and "Interval" to an equal probability split C.I. obtained according to (2) with $\alpha = 0.05$. Each set of results is based on 10^3 simulations.

As expected, the bootstrap method performs poorly, and the asymptotic method only marginally better, both missing the nominal coverage in almost all cases. It is perhaps surprising that SPBB does so well, especially in the two-sided confidence bounds. This is a vindication for the parametric bootstrap, of which SPBB is an approximation; but of course it is accomplished orders of magnitude faster. Results for the interval lengths are somewhat mixed, but it is practically pointless to compare lengths when the coverages are far off nominal.

3 Extensions to non-monotone estimating equations

Monotonicity of the QEE in (1) is one of the key points of SPBB, permitting the implementation of step (iv) as explained in the Sect. 1. Without the monotonicity, one no longer has the simple equivalence of events that enables saddlepoint approximation of the CDF of $\hat{\theta}$ through that of $\Psi(t)$. This is also a key assumption in the Daniels (1983) saddlepoint approximation for the PDF of $\hat{\theta}$ in the context of QEEs,

$$\hat{f}_{\hat{\theta}}(t) = \hat{f}_{\Psi(t)}(0) J_D(t), \quad \hat{f}_{\Psi(t)}(0) = \frac{M_{\Psi(t)}\left(\hat{s}\right)}{\sqrt{2\pi K''_{\Psi(t)}\left(\hat{s}\right)}}, \quad J_D(t) = -\frac{1}{\hat{s}}\frac{\partial K_{\Psi(t)}(\hat{s})}{\partial t}, \tag{6}$$

where $M_{\Psi(t)}(s)$ and $K_{\Psi(t)}(s)$ denote, respectively, the MGF and CGF of $\Psi(t)$ evaluated at s, and $\hat{s}$ solves the saddlepoint equation $K'_{\Psi(t)}(\hat{s}) = 0$, with $K'(s) \equiv \partial K(s)/\partial s$. Butler (2007, §12.2.1) provides an illuminating proof of this result, and gives an expression for $\hat{f}_{\hat{\theta}}$ where the Jacobian term is in the form of a conditional expectation

$$\hat{f}_{\hat{\theta}}(t) = \hat{f}_{\Psi(t)}(0) J_S(t), \quad J_S(t) = \mathbb{E}\left[|\dot{\Psi}(t)| \mid \Psi(t) = 0\right], \quad \dot{\Psi}(t) \equiv \partial\Psi(t)/\partial t, \tag{7}$$

which is due to Skovgaard (1990) and Spady (1991). Although more computationally challenging, the PDF approximation in (7) does not require monotonicity of $\Psi(t)$ in t.

The fact that $J_S(t)$ constitutes an intractable conditional expectation would seemingly rule out an application of (7). However, we propose the following approach that combines a *double-saddlepoint approximation* (a device for approximating a conditional density) with a Monte Carlo scheme. Defining $U_t \equiv \Psi(t)$ and $V_t \equiv \dot{\Psi}(t)$ to ease notation, the basic strategy is as follows:

$$\begin{aligned} J_S(t) &= \mathbb{E}[|V_t||U_t = 0] = \int |v| f_{V_t|U_t}(v|u=0)\,\mathrm{d}v \approx \int |v| \hat{f}_{V_t|U_t}(v|u=0)\mathrm{d}v \\ &\approx \frac{1}{m}\sum_{i=1}^{m} |z_i|, \end{aligned} \tag{8}$$

where $v_1, \ldots, v_m$ is a random draw from $\hat{f}_{V_t|U_t}(v|u=0)$, the (double-saddlepoint) approximation to the PDF $f_{V_t|U_t}$, which, adapted from Butler (2007, §4.2) for the case $u = 0$, is given by

$$\hat{f}_{V_t|U_t}(v|u=0) = \left[\frac{|K''_{U_t,V_t}\left(\hat{s}_1, \hat{s}_2\right)|}{K''_{U_t}\left(\hat{s}_0\right)}\right]^{-\frac{1}{2}} \frac{\mathrm{e}^{-\hat{\omega}^2/2}}{\sqrt{2\pi}}. \tag{9}$$

To define its ingredients, we start with the joint CGF of (U_t, V_t),

$$K_{U_t,V_t}(s_1, s_2) = \log[\mathbb{E}\exp\{s_1 U_t + s_2 V_t\}] = -\frac{1}{2}\log\left|I_n - 2\sigma^2\Omega_\theta(s_1 A_t + s_2 \dot{A}_t)\right|,$$

from which we obtain $K_{U_t}(s_0) = K_{U_t,V_t}(s_0, 0)$ as the marginal CGF of U_t. If we let $K'_{U_t,V_t}(s_1, s_2)$ denote the gradient vector and $K''_{U_t,V_t}(s_1, s_2)$ the Hessian matrix of the joint CGF,

$$K'_{U_t,V_t}(s_1,s_2)=\begin{bmatrix}\frac{\partial K_{U_t,V_t}(s_1,s_2)}{\partial s_1}\\ \frac{\partial K_{U_t,V_t}(s_1,s_2)}{\partial s_2}\end{bmatrix},\quad K''_{U_t,V_t}(s_1,s_2)=\begin{bmatrix}\frac{\partial^2 K_{U_t,V_t}(s_1,s_2)}{\partial s_1^2} & \frac{\partial^2 K_{U_t,V_t}(s_1,s_2)}{\partial s_1 \partial s_2}\\ \frac{\partial^2 K_{U_t,V_t}(s_1,s_2)}{\partial s_1 \partial s_2} & \frac{\partial^2 K_{U_t,V_t}(s_1,s_2)}{\partial s_2^2}\end{bmatrix},$$

then the two-dimensional saddlepoint $(\hat{s}_1, \hat{s}_2)$ solves $K'_{U_t,V_t}(\hat{s}_1, \hat{s}_2) = [0, v]^{\mathsf{T}}$. Similarly, $K'_{U_t}(s_0) = \partial K_{U_t}(s_0)/\partial s_0$, $K''_{U_t}(s_0) = \partial^2 K_{U_t}(s_0)/\partial s_0^2$, and the saddlepoint $\hat{s}_0$ solves the equation $K'_{U_t}(\hat{s}_0) = 0$. Finally,

$$\hat{\omega} = \operatorname{sgn}(\hat{s}_2)\sqrt{2\left[K_{U_t}(\hat{s}_0) - K_{U_t,V_t}(\hat{s}_1, \hat{s}_2) + \hat{s}_2 v\right]}.$$

Details of the computation, including expressions for the derivatives of the CGF's (obtained by elementary matrix calculus), can be found in Wickramasinghe (2012). To avoid computing normalizing constants for the double-saddlepoint approximated PDF's, a Monte Carlo scheme like importance sampling can be used. With $g(z)$ an appropriate instrumental density, the proposed implementation to compute $\hat{f}_{\hat{\theta}}(t)$ via (7) would then proceed according to Algorithm 1 as follows.

Algorithm 1 For a sufficiently large integer m, instrumental density $g(z)$, and a grid of values $t \in [-1, 1]$, do:

- draw an iid sample $z_1, \ldots, z_m$ from $g(z)$;
- for $i = 1, \ldots, m$, obtain $\hat{f}_{V_t|U_t}(z_i|0) \equiv f(z_i)$ from (9);
- form the importance sampling approximation to $J_S(t)$ as

$$\hat{J}_S(t) = \frac{\sum_{i=1}^m |z_i| f(z_i)/g(z_i)}{\sum_{i=1}^m f(z_i)/g(z_i)};$$

- obtain $\hat{f}_{\Psi(t)}(0)$ from (6), and set $\hat{f}_{\hat{\theta}}(t) = \hat{f}_{\Psi(t)}(0)\hat{J}_S(t)$.

For efficiency reasons, the instrumental distribution g should have heavier tails than the target f (Robert and Casella 2004), and a t distribution with 3 degrees of freedom seems to be a reasonable default choice. Note that a numerical integration method like Gauss quadrature could be used to approximate $J_S(t)$ instead, but the importance sampling approach has the advantage of not requiring the computation of the normalizing constant in the saddlepoint approximated $\hat{f}_{\hat{\theta}}(t)$, which, therefore, lessens the computational burden.

In fact, making use of the double-saddlepoint approximation to the CDF, the following alternate algorithm that relies on the probability integral transform could be used instead. The trade-off between these two is that Algorithm 2 does not require the selection of an instrumental density with its associated efficiency issues, but it does require greater programming effort to determine the saddlepoints $\hat{s}_0$, $\hat{s}_1$, and $\hat{s}_2$, which have to be found multiple times in the inherent CDF inversion operation of the second step.

Algorithm 2 For a sufficiently large integer m, and a grid of values $t \in [-1, 1]$, do:

- draw an ordered iid sample $z_1 \leq \cdots \leq z_m$ from a uniform distribution on $(0, 1)$;
- for $i = 1, \ldots, m$, find the value y_i that solves $\hat{F}_{V_t|U_t}(y_i|0) = z_i$, the double-saddlepoint approximation to the CDF of $V_t|U_t$ (Butler 2007, §4.2);
- form the empirical approximation to $J_S(t)$ as

$$\hat{J}_S(t) = \frac{1}{m}\sum_{i=1}^{m} |y_i|;$$

- obtain $\hat{f}_{\Psi(t)}(0)$ from (6), and set $\hat{f}_{\hat{\theta}}(t) = \hat{f}_{\Psi(t)}(0)\hat{J}_S(t)$.

Figure 3 shows saddlepoint approximations to the PDF's of the MLE of θ_0 for the MA(1) model settings considered in Fig. 1. The approximations are based on the two Jacobians: Daniels from equation (6), and Skovgaard from equation (7) obtained via Algorithm 1. For comparison with the "truth", the empirical PDF is represented by the histograms (based on 10^6 simulated replicates), and is displayed with the exclusion of the point masses at $\theta = \pm 1$. For meaningful comparisons, all the PDF's are, therefore, normalized to integrate to unity.

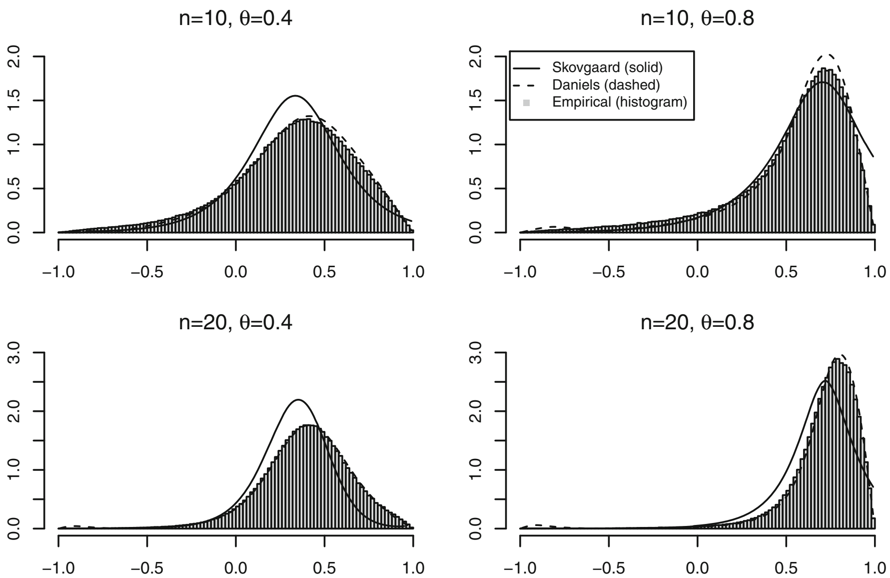

Fig. 3 Saddlepoint approximated densities of the MLEs of θ in Proposition 1 for the moving average coefficient $\theta_0 = 0.4$ (left panels) and $\theta_0 = 0.8$ (right panels), in samples of size $n = 10$ (top panels) and $n = 20$ (bottom panels) from an MA(1) model, using two different Jacobians: Skovgaard (solid) and Daniels (dashed). The empirical PDF displayed in the histograms is based on 10^6 simulated replicates

From this we can see that the Daniels method completely misses the point masses. The fact that the MLE QEE is monotone with high probability here is responsible for the excellent quality of the Daniels approximation in the open interval $(-1, 1)$. The Skovgaard method on the other hand seems to pick up the point masses, as the PDF shows a tendency to curve upward close to the upper end of the support at 1, and thus gives a much more realistic picture of the true state of affairs. This tendency of saddlepoint approximations to exhibit smooth jumps at point masses was already noted by Lund et al. (1999). As a final note, a C.I. could be obtained by numerical integration of the approximate PDF.

4 Extensions to estimating equations with tractable MGFs

Consider the QEE $\Psi_0(\theta)$ defined in (1). The multivariate normality immediately furnishes a closed-form expression for its MGF as

$$M_{\Psi_0(\theta)}(s; \boldsymbol{\mu}_0, \Sigma_0) = \exp\left\{s\boldsymbol{\mu}_0^{\mathsf{T}}(I_n - 2s\Sigma_0 A_\theta)^{-1} A_\theta \boldsymbol{\mu}_0\right\} |I_n - 2s\Sigma_0 A_\theta|^{-1/2}. \tag{10}$$

In this section, we consider extensions of SPBB which still result in a tractable or "near-tractable" MGF for the underlying estimating equation, and thus permit the implementation of step (iii) outlined in the Introduction.

4.1 Estimating equations of linear quadratic form type

A straightforward generalization of (10) occurs if the estimating equation $\Psi(\cdot)$ contains linear and constant terms,

$$\Psi(\theta) = c_\theta + \mathbf{b}_\theta^{\mathsf{T}}\mathbf{y} + \mathbf{y}^{\mathsf{T}} A_\theta \mathbf{y} = 0, \tag{11}$$

where c_θ and $\mathbf{b}_\theta$ are, respectively, scalar and vector-valued constants (possibly depending on θ). If $\mathbf{y}$ is Gaussian as above, and defining $\Lambda_\theta^{-1} \equiv \Sigma_0^{-1} - 2sA_\theta$, the resulting MGF becomes

$$\begin{aligned} M_{\Psi(\theta)}(s; \boldsymbol{\mu}_0, \Sigma_0) &= M_{\Psi_0(\theta)}(s; \boldsymbol{\mu}_0, \Sigma_0) \\ &\times \exp\left\{(c_\theta + \mathbf{b}_\theta^{\mathsf{T}}\boldsymbol{\mu}_0)s + \left(2\mathbf{b}_\theta^{\mathsf{T}}\Lambda_\theta A_\theta \boldsymbol{\mu}_0 + \mathbf{b}_\theta^{\mathsf{T}}\Lambda_\theta \mathbf{b}_\theta/2\right)s^2\right\}. \end{aligned}$$

An application of this is if the underlying estimating equation $g(\mathbf{y}; \theta)$ is a function that is itself not a QEE, but can be approximated as such via a Taylor series expansion to second-order terms, so that $g(\mathbf{y}; \theta) \approx \Psi(\theta)$. An example is presented by Feuerverger and Wong (2000), where it is desired to approximate the distribution of a scalar-valued function of a Gaussian random vector of returns. They show how the tractable MGF of (11) leads directly to a saddlepoint approximation for the desired distribution. This kind of scenario is subsumed as a special case of the simple SPBB approach, which can handle instances when only the estimating equation giving rise to the estimator of interest is known.

4.2 Skew-normal and elliptically contoured families

Generalizations of the Gaussian distribution with tractable QEE MGF promise include the extensively studied elliptically contoured and skew-normal families. The latter is a relatively recent innovation; see for instance Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001), Gupta and Huang (2002), Gupta et al. (2004), Huang and Chen (2006), and Wang et al. (2009). Some of these references provide closed-form expressions for the MGF of a quadratic form, the starting point for SPBB implementation.

The elliptically contoured family is extensively documented in Fang et al. (1990), Fang and Anderson (1990), and Fang and Zhang (1990). The n-dimensional random vector $\mathbf{y} \sim EC_n(\boldsymbol{\mu}, \Sigma, \phi)$ is *elliptically contoured* with location and dispersion parameters $\boldsymbol{\mu}$ and Σ, if its characteristic function is of the form $\Xi(\boldsymbol{s}) = \mathrm{e}^{i\boldsymbol{s}^\top\boldsymbol{\mu}}\phi(\boldsymbol{s}^\top\Sigma\boldsymbol{s})$ for some function $\phi(u)$, $u \in \mathbb{R}$. An equivalent characterization is through the "generator" function $h(\cdot)$ of the PDF, in which case we write $\mathbf{y} \sim EC_n(\boldsymbol{\mu}, \Sigma, h)$, with PDF

$$f(\mathbf{y}) = |\Sigma|^{-1/2} c_n h(z) \equiv |\Sigma|^{-1/2} g(z), \quad \text{with } z = (\mathbf{y}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}) \in \mathbb{R}^+, \tag{12}$$

where c_n is a known normalizing constant.

Provost and Cheong (2002, Lemma 2) presented an expression for the PDF of $\mathbf{y} \sim EC_n(\boldsymbol{\mu}, \Sigma, \phi)$, as

$$f(\mathbf{y}) = \int_0^\infty w(t)\,\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma/t)\,\mathrm{d}t, \tag{13}$$

where $\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma/t)$ denotes the PDF of an n-dimensional normal with mean $\boldsymbol{\mu}$ and covariance matrix Σ/t, and $w(t)$ is a "weighting" function that is defined through the inverse Laplace transform of $f(t)$, with $t = (\mathbf{y}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})/2 \equiv z/2$, denoted by $\mathcal{L}_f^{-1}(t)$. More explicitly, we have

$$w(t) = (2\pi)^{n/2}|\Sigma|^{1/2}t^{-n/2}\mathcal{L}_f^{-1}(t), \quad \text{with} \quad \int_0^\infty w(t)\,\mathrm{d}t = 1, \tag{14}$$

which results in the following relationship between the PDF at z and $w(t)$,

$$f(z) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\int_0^\infty t^{n/2}\mathrm{e}^{-tz/2}w(t)\,\mathrm{d}t.$$

With this construct, and if the weighting function can be analytically obtained, it is possible to derive an integral representation for the MGF of a QEE in elliptically contoured random variables. Since $w(\cdot)$ integrates to 1 over $\mathbb{R}^+$, the resulting MGF can be regarded as an infinite mixture of normal MGF's.

Proposition 2 (MGF of Elliptical QEE) *Let* $\mathbf{y} \sim EC_n(\boldsymbol{\mu}_0, \Sigma_0, \phi)$ *be an* n*-dimensional elliptically contoured random vector with PDF* $f(t)$, $t = (\mathbf{y}-\boldsymbol{\mu}_0)^\top\Sigma_0^{-1}(\mathbf{y}-\boldsymbol{\mu}_0)/2$, *and weighting function* $w(t)$ *as defined by* (14). *Define the QEE*

$\Psi_0^{EC}(\theta) \equiv \mathbf{y}^\mathsf{T} A_\theta \mathbf{y}$, *where* $\boldsymbol{\mu}_0$ *and* Σ_0 *are functions of the parameter of interest* θ_0 *and nuisance parameter vector* $\boldsymbol{\lambda}_0$. *Then, provided the integral converges in a neighborhood of* $s = 0$, *the MGF of* $\Psi_0^{EC}(\theta)$ *is given by*

$$M_{\Psi_0^{EC}(\theta)}(s; \boldsymbol{\mu}_0, \Sigma_0) = \int_0^\infty w(t)\, M_{\Psi_0(\theta)}(s; \boldsymbol{\mu}_0, \Sigma_0/t)\, \mathrm{d}t, \tag{15}$$

where $M_{\Psi_0(\theta)}(s; \boldsymbol{\mu}_0, \Sigma_0/t)$ *is the expression in* (10) *with* Σ_0/t *replacing* Σ_0.

Proof This result follows straightforwardly by writing out the n-dimensional integral that defines $M_{\Psi_0^{EC}(\theta)}(\cdot)$, substituting expression (13) in for the PDF, interchanging the order of integration in the resulting double integral, and noticing that the inner integral defines the MGF of $\Psi_0(\theta)$ with respect to a normal with mean $\boldsymbol{\mu}_0$ and covariance matrix Σ_0/t. $\square$

Remark 2 Note that inverse Laplace transform $\mathcal{L}_f^{-1}(t)$ is by definition a positive integrable function. When it is continuous and does not have infinite jumps, such as one would see in a Dirac delta function, then it could be viewed as the unnormalized PDF of some random variable. Therefore, the saddlepoint PDF could in principle be used to approximate $\mathcal{L}_f^{-1}(t)$. The issue of the unknown normalization constant would then be handled using the condition that the approximate weight function integrates to one, as in (14).

Some noteworthy special cases of the elliptically contoured family with finite moments of all orders are presented in Table 3. The weighting function for the normal is the Dirac delta centered at 1. Two versions of multivariate Laplace are presented. The Laplace-A, proposed by Kotz et al. (2001, Chap. 5), has an explicit form for the characteristic function but not the PDF, which makes the calculation of the weighting function intractable as far as we can determine (but could be approximated as in Remark 2). The Laplace-B listed here is a scaled version of the “bilateral exponential” in Table 1 of Provost and Cheong (2002). Both the Gaussian and Laplace-B are special cases of the multivariate *exponential power* (EP) sub-family proposed by Gomez et al. (1998), with shape parameters $\beta = 1$ and $\beta = 1/2$, respectively. The EP is a multivariate generalization of the univariate exponential power distribution[1], where $\beta > 0$ controls the thickness of the tails: for $\beta < 1$, the tails are heavier than the normal, and $\beta \to \infty$ results in a uniform. Gomez et al. (1998) originally termed this *power exponential*, but later realized the inadvertent switching of the names (Gomez et al. 2002).

If $\mathbf{y} \sim EP_n(\boldsymbol{\mu}, \Sigma, \beta)$ denotes an n-dimensional EP random vector, the values $\boldsymbol{\mu}$, Σ, and β play the role of location, scale, and shape parameters, with PDF given by

[1] Also variously called *Subbotin*, *Generalized Error Distribution* (Mineo and Ruggieri 2005), and *Generalized Normal Distribution* (Nadarajah 2005), with slight differences in the parametrizations.

$$f(z) = \frac{n\Gamma(n/2)}{\pi^{n/2}\Gamma(1+\frac{n}{2\beta})2^{1+\frac{n}{2\beta}}} |\Sigma|^{-1/2} \exp\{-z^{\beta}/2\}, \quad z = (\mathbf{y}-\boldsymbol{\mu})^{\top}\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}).$$

The mean and variance are related to the location and scale parameters by the relations

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu}, \quad \text{and} \quad \mathrm{Var}(\mathbf{y}) = \frac{2^{1/\beta}\Gamma\left(\frac{n+2}{2\beta}\right)}{n\Gamma\left(\frac{n}{2\beta}\right)}\Sigma. \tag{16}$$

It is obvious that $\mathbf{y} \sim EP_n(\boldsymbol{\mu}, \Sigma, \beta)$ implies $\mathbf{y} \sim EC_n(\boldsymbol{\mu}, \Sigma, h)$, with generator function $h(z) = \exp\{-z^{\beta}/2\}$. For more on the EP and some results for the MA(1) under EP noise and EP likelihood, see Barnard et al. (2013).

4.3 SPBB computations for non-gaussian QEE's

SPBB computations with multivariate skew-normal distributions will be relatively straightforward since the relevant MGF's are given in closed-form. In contrast, computations with elliptically contoured data will be more involved since the calculation of the MGF for the QEE in Proposition 2 requires (one-dimensional) integration. However, this is not necessarily an impediment to implementation of SPBB, since it does not require numerical inversion of the MGF and the ill-conditioned problems it brings. One possible solution is to utilize a *Laplace approximation* e.g., Butler (2007).

We apply this idea to the MA(1) model (with $\sigma^2 = 1$) when the data follow a Laplace-B distribution, $\boldsymbol{x} \sim EP_n\left(\boldsymbol{\mu}_0 = \mathbf{0}, \Sigma_0 = (4n+4)^{-1}\Omega_{\theta_0}, \beta = 1/2\right)$. From Proposition 2, and with $w(t)$ as given in the last row of in Table 3, we then have the following expression for the MGF of $\Psi(\theta) = \boldsymbol{x}^{\top}A_{\theta}\boldsymbol{x}$,

$$M_{\Psi(\theta)}(s) = \int_0^{\infty} w(t)\,|I_n - (2s/t)\Sigma_0 A_{\theta}|^{-\frac{1}{2}}\,\mathrm{d}t \equiv c_n \int_0^{\infty} \mathrm{e}^{-q(t)}\mathrm{d}t,$$
$$c_n = \frac{\Gamma(n/2)}{\Gamma(n)2^{(5+n)/2}\sqrt{\pi}}, \tag{17}$$

Table 3 Some elements of the elliptically contoured family

Distribution	$\phi(u)$	$g(z)$	$w(t)$
Gaussian	$\mathrm{e}^{-u/2}$	$(2\pi)^{-n/2}\mathrm{e}^{-z/2}$	$\delta(t-1)$
Laplace-A	$(1+u/2)^{-1}$	$2(2\pi)^{-n/2}(z/2)^{(2-n)/4}K_{1-n/2}(\sqrt{2z})$	Intractable
Laplace-B	Not Explicit	$\frac{(2\pi)^{-n/2}\Gamma(n/2)}{\Gamma(n)2^{1+n/2}}\mathrm{e}^{-\sqrt{z}/2}$	$\frac{\Gamma(n/2)\mathrm{e}^{-1/(8t)}t^{-(n+3)/2}}{\Gamma(n)2^{(5+n)/2}\sqrt{\pi}}$

The Laplace-A is that proposed by Kotz et al. (2001, Chap. 5) with an explicit characteristic function, but whose PDF involves a modified Bessel function of the 2nd kind. Both the Gaussian and Laplace-B are elements of the EP sub-family (Gomez et al. 1998), with shape parameters $\beta = 1$ and $\beta = 1/2$, respectively where $q(t) = 1/(8t) + (3/2)\log(t) + (1/2)\log[p(t)]$ and $p(t) = |tI_n - 2s\Sigma_0 A_\theta|$. We can now appeal to a Laplace approximation for the integral,

$$\int_0^\infty \mathrm{e}^{-q(t)}\mathrm{d}t \approx \frac{\sqrt{2\pi}\mathrm{e}^{-q(\hat{t})}}{\sqrt{q''(\hat{t})}},$$

where the interior point $\hat{t} \equiv \hat{t}(s, \theta, \theta_0)$ is the global minimum of $q(t)$ over $(0, \infty)$.

Progress in terms of SPBB from here on requires explicit expressions for the first and second derivatives of $M_{\Psi(\theta)}(s)$ with respect to s and θ. This in turns means that an explicit expression for $\hat{t}$ is needed. Noting that $p(t)$ is the *characteristic polynomial* of the matrix $2s\Sigma_0 A_\theta$, we can expand and then truncate it as follows,

$$\begin{aligned} p(t) &= t^n - \mathrm{tr}(2s\Sigma_0 A_\theta)t^{n-1} + \text{intermediate terms} + (-1)^n|2s\Sigma_0 A_\theta| \\ &= \tilde{p}(t) + O(s^2), \qquad \text{where} \quad \tilde{p}(t) = t^n - \mathrm{tr}(2s\Sigma_0 A_\theta)t^{n-1}. \end{aligned}$$

The justification for truncating $p(t) \approx \tilde{p}(t)$ is that higher order terms are $O(s^2)$, for $|s| \to 0$. Since the behavior of $M_{\Psi(\theta)}(s)$ in the neighborhood of $s = 0$ is the key component in the saddlepoint equation, the linear term should play the principal role in the resulting approximations. Substitution of $\tilde{p}(t)$ in the expression for $q(t)$, calling the result $\tilde{q}(t)$, and differentiation in t leads to

$$\tilde{q}'(t) = -\frac{1}{8t^2} + \frac{n+2}{2t} + \frac{1}{2[t - \mathrm{tr}(2s\Sigma_0 A_\theta)]}.$$

Equating this expression to zero to find the minimum is now seen to result in a quadratic equation in t, and thus a closed-form solution for $\hat{t}$ is obtained. Note that convergence of (17) is assured for small enough $|s|$, since at $s = 0$ we have

$$M_{\Psi(\theta)}(0) = \int_0^\infty t^{-(n+3)/2}\mathrm{e}^{-1/(8t)}\mathrm{d}t = \frac{1}{c_n}.$$

All that remains in implementing SPBB are explicit expressions for the first and second derivatives of $M_{\Psi(\theta)}(s)$ with respect to s and θ, but these can now be routinely obtained by careful application of the chain rule (Wickramasinghe 2012). Figure 4 shows the results for MOME. The histograms represent the exact PDF, and are obtained via simulation (based on 10^4 replications). One possible problem with the stratagem of truncating $p(t)$ is that the saddlepoint approximation may be less accurate in the tails. However, the plots in Fig. 4 show that even for small sample sizes like 5 and 10 this does not seem to be an issue.

5 Summary

Two substantive extensions to the SPBB methodology of Paige et al. (2009) have been proposed, whereby inference for a scalar parameter of interest is made through its underlying QEE. The first tackled the issue of non-monotone QEEs, by providing

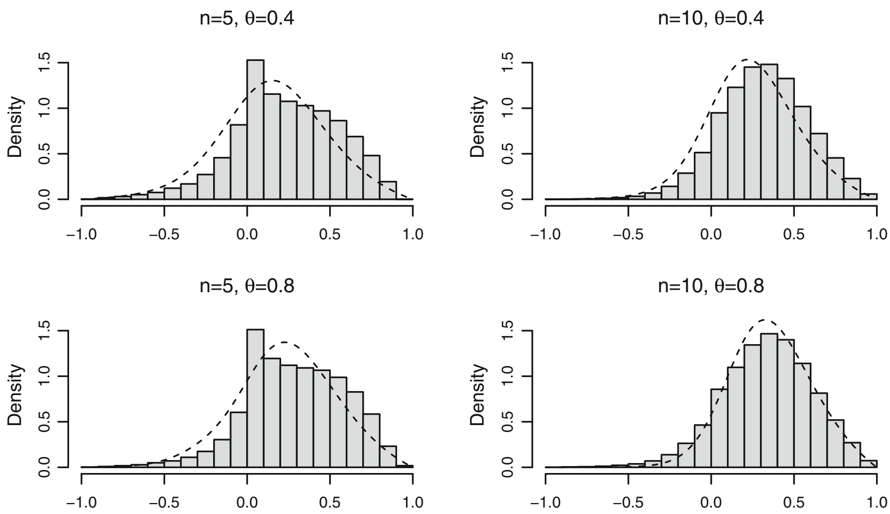

Fig. 4 Simulated (*histograms*) and SPBB approximated (*dashed lines*) PDF's for the MOME of the coefficient θ_0 of an MA(1) model ($\sigma^2 = 1$) under a Laplace-B distribution for the data

an alternative expression for the Jacobian term in the saddlepoint approximation to the density of the estimator. Application to common estimators in the MA(1) model, such as least squares and maximum likelihood, revealed that the existing approach works well when the QEE is monotone with high probability. When this fails, the Skovgaard Jacobian can alternatively be employed, which also seems to detect the existence of point masses in estimators with mixed distributions. Monte Carlo-based algorithms were proposed to deal with the intractable conditional expectation arising in its computation.

The second extension considered saddlepoint approximations to the density of estimators when the QEE is non-Gaussian. In this case, one has to search for an alternative suitable distributional structure that leads to a tractable expression for the MGF of the QEE. Two possibilities are the skew-normal (Gupta and Huang 2002) and exponential power (Gomez et al. 1998) families, both being flexible enough to include the normal as a special case. The multivariate Laplace version of the exponential power was illustrated in the MA(1) context, by developing an expression for the QEE MGF that is closed-form to within a one-dimensional integral. We showed how tractable derivatives for the logarithm of this MGF can then be obtained by Laplace approximating the integral after truncation of the characteristic polynomial expansion for the problematic determinant term in the integrand. This leads to a reasonably accurate approximation for the desired densities.

In future work, it would be interesting to address the issue of C.I. construction under these two extensions. With monotonicity holding, the CDF of the estimator for a multivariate Laplace QEE can be obtained analogously to the PDF, and SPBB inference is thus implemented with little additional effort. Without monotonicity, one lacks the straightforward mapping of events that makes SPBB possible, and our best recommendation would be numerical integration of the approximated density.

Acknowledgments The paper has greatly benefited from the comments of two anonymous referees.

References

Anderson, T. W., Takemura, A. (1986). Why do noninvertible estimated moving averages occur? *Journal of Time Series Analysis*, *7*, 235–254.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, *B61*, 579–602.

Azzalini, A., Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726.

Barnard, R. W., Trindade, A. A., Wickramasinghe, R. I. P. (2013). *Some results for the first order moving average model under exponential power distributions*. Preprint.

Branco, M. D., Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, *79*, 99–113.

Brockwell, P. J., Davis, R. A. (1991). *Time series: theory and methods* (2nd ed.). New York: Springer.

Butler, R. W. (2007). *Saddlepoint approximations with applications*. New York: Cambridge University Press.

Cryer, J. D., Chan, K.-S. (2008). *Time series analysis: with applications in R* (2nd ed.). New York: Springer.

Cryer, J. D., Ledolter, J. (1981). Small-Sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, *68*, 691–694.

Daniels, H. E. (1983). Saddlepoint approximations for estimating equations. *Biometrika*, *70*, 89–96.

Davis, R. A., Dunsmuir, W. T. M. (1996). Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. *Econometric Theory*, *12*, 1–29.

Davis, R. A., Song, L. (2011). Unit roots in moving averages beyond first order. *The Annals of Statistics*, *39*, 3062–3091.

Davis, R. A., Chen, M., Dunsmuir, W. T. M. (1995). Inference for MA(1) processes with a root on or near the unit circle. *Probability and Mathematical Statistics*, *15*, 227–242.

Fang, K. T., Anderson, T. W. (1990). *Statistical inference in elliptically contoured and related distributions*. New York: Allerton Press.

Fang, K. T., Zhang, Y. T. (1990). *Generalized multivariate analysis*. Berlin: Springer-Verlag.

Fang, K., Kotz, S., Ng, K. W. (1990). *Symmetric multivariate and related distributions, Monographs on Statistics and Applied Probability*. New York: Chapman and Hall/CRC.

Feuerverger, A., Wong, A. (2000). Computation of Value-at-Risk for nonlinear portfolios. *Journal of Risk*, *3*, 37–55.

Gomez, E., Gomez-Villegas, M. A., Marin, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics*, *A27*, 589–600.

Gomez, E., Gomez-Villegas, M. A., Marin, J. M. (2002). Continuous elliptical and exponential power linear dynamic models. *Journal of Multivariate Analysis*, *83*, 22–36.

Gupta, A. K., Huang, W. J. (2002). Quadratic forms in skew normal variates. *Journal of Mathematical Analysis and Applications*, *273*, 558–564.

Gupta, A. K., Gonzalez-Farias, G., Dominguez-Molina, J. A. (2004). A Multivariate skew normal distribution. *Journal of Multivariate Analysis*, *89*, 181–190.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *The Annals of Statistics*, *16*, 927–985.

Huang, W.-J., Chen, Y.-H. (2006). Quadratic forms of multivariate skew normal-symmetric distributions. *Statistics and Probability Letters*, *9*, 871–879.

Kotz, S., Kozubowski, T., Podgórski, K. (2001). *The laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*. Boston: Birkhäuser.

Lugannani, R., Rice, S. O. (1980). Saddlepoint approximations for the distribution of sums of independent of random variables. *Advances in Applied Probability*, *12*, 475–490.

Lund, R. B., Butler, R. W., Paige, R. L. (1999). Prediction of shot noise. *Journal of Applied Probability*, *36*, 374–388.

Mineo, A. M., Ruggieri, M. (2005). A software tool for the exponential power distribution: the normalp package. *Journal of Statistical Software*, *12*(4), 1–23.

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, *32*, 685–694.

Paige, R. L., Fernando, P. (2008). Robust inference in conditionally linear nonlinear regression models. *Scandinavian Journal of Statistics*, *35*, 158–168.

Paige, R. L., Trindade, A. A. (2008). Practical small sample inference for single lag subset autoregressive models. *Journal of Statistical Planning and Inference*, *138*, 1934–1949.
Paige, R. L., Trindade, A. A. (2010). The Hodrick-Prescott Filter: a special case of penalized spline smoothing. *Electronic Journal of Statistics*, *4*, 856–874.
Paige, R. L., Trindade, A. A. (2013). Fast and accurate inference for the smoothing parameter in semiparametric models. *Australian and New Zealand Journal of Statistics*, *55*, 25–41.
Paige, R. L., Trindade, A. A., Fernando, P. H. (2009). Saddlepoint-Based bootstrap inference for quadratic estimating equations. *Scandinavian Journal of Statistics*, *36*, 98–111.
Plosser, C. I., Schwertz, G. W. (1977). Estimation of a non-invertible moving average process: the case for overdifferencing. *Journal of Econometrics*, *6*, 199–224.
Provost, S. B., Cheong, Y.-H. (2002). The distribution of Hermitian quadratic forms in elliptically contoured random vectors. *Journal of Statistical Planning and Inference*, *102*, 303–316.
Robert, C. P., Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer.
Shumway, R. H., Stoffer, D. S. (2011). *Time series analysis and its applications: with R examples* (3rd ed.). New York: Springer.
Skovgaard, I. M. (1990). On the density of minimum contrast estimators. *The Annals of Statistics*, *18*, 779–789.
Spady, R. H. (1991). Saddlepoint approximations for regression models. *Biometrika*, *78*, 879–889.
Tanaka, K. (1990). Testing for a moving average unit root. *Econometric Theory*, *6*, 433–444.
Wang, T., Li, B., Gupta, A. K. (2009). Distribution of quadratic forms under skew normal settings. *Journal of Multivariate Analysis*, *100*, 533–545.
Wickramasinghe, R. I. P. (2012). *Extensions of saddlepoint-based bootstrap inference with application to the moving average model of order one, PhD Dissertation*. Texas Tech University, Lubbock.
Young, G. A. (2009). Routes to higher-order accuracy in parametric inference. *Australian and New Zealand Journal of Statistics*, *51*, 115–126.