Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2016)

On the coverage bound problem of empirical likelihood methods for time series

Xianyang Zhang

University of Missouri—Columbia, USA

and Xiaofeng Shao

University of Illinois at Urbana-Champaign, Champaign, USA

[Received January 2014. Final revision February 2015]

Summary. The upper bounds on the coverage probabilities of the confidence regions based on blockwise empirical likelihood and non-standard expansive empirical likelihood methods for time series data are investigated via studying the probability of violating the convex hull constraint. The large sample bounds are derived on the basis of the pivotal limit of the blockwise empirical log-likelihood ratio obtained under fixed b asymptotics, which has recently been shown to provide a more accurate approximation to the finite sample distribution than the conventional χ^2-approximation. Our theoretical and numerical findings suggest that both the finite sample and the large sample upper bounds for coverage probabilities are strictly less than 1 and the blockwise empirical likelihood confidence region can exhibit serious undercoverage when the dimension of moment conditions is moderate or large, the time series dependence is positively strong or the block size is large relative to the sample size. A similar finite sample coverage problem occurs for non-standard expansive empirical likelihood. To alleviate the coverage bound problem, we propose to penalize both empirical likelihood methods by relaxing the convex hull constraint. Numerical simulations and data illustrations demonstrate the effectiveness of our proposed remedies in terms of delivering confidence sets with more accurate coverage. Some technical details and additional simulation results are included in on-line supplemental material.

Keywords: Convex hull constraint; Coverage probability; Fixed b asymptotics; Heteroscedasticity–auto-correlation robustness; Moment condition

1. Introduction

Empirical likelihood (EL) (Owen, 1988, 1990) is a non-parametric methodology for deriving estimates and confidence sets for unknown parameters, which shares some of the desirable properties of parametric likelihood (see DiCiccio *et al.* (1991) and Chen and Cui (2006)). Because of its effectiveness and flexibility, it has advanced in many branches in statistics, such as regression models, time series and censored data; see Owen (2001) for a nice treatment of the subject.

The EL-based confidence sets inherit some good features from their parametric likelihood counterparts, but there is a finite sample upper bound for the coverage of the EL ratio confidence region (see Owen (2001), page 209, and Tsao (2004)) due to the convex hull constraint, which may limit its applicability and make it less appealing. For example, the EL confidence region for the mean of a random sample is nested within the convex hull of the data and its coverage level is necessarily smaller than that of the convex hull itself. The upper bound can be much smaller than

Address for correspondence: Xianyang Zhang, Department of Statistics, University of Missouri—Columbia, Columbia, MO 65211, USA.
E-mail: zhangxiany@missouri.edu

 1369–7412/16/78000

nominal coverage level $1-\alpha$ in the small sample and multi-dimensional situations. Following the terminology in Tsao and Wu (2013), the finite sample coverage bound problem is due to the mismatch between the domain of the EL and the parameter space, so it is also called a *mismatch problem*. There have been a few recent proposals to alleviate or resolve the mismatch problem; see, for example adjusted EL (Chen *et al.*, 2008; Emerson and Owen, 2009; Liu and Chen, 2010; Chen and Huang, 2012), penalized EL (Bartolucci, 2007; Lahiri and Mukhopadhyay, 2012) and the domain expansion approach (Tsao and Wu, 2013, 2014). However, all these works deal with independent estimation equations, and their direct applicability to the important time series case is not clear.

In this paper, our interest concerns the coverage bound problems for EL methods tailored to stationary and weakly dependent time series. Although many variants have been proposed to extend EL to the time series setting (see Nordman and Lahiri (2014) for a recent review), it seems that no investigation has been conducted regarding the coverage bound problem, which is expected to exist but its effect in the time series setting is unknown. We focus on two EL methods: blockwise EL (BEL), proposed by Kitamura (1997), and non-standard expansive BEL (EBEL), recently proposed by Nordman *et al.* (2013). BEL applies EL to the blockwise-averaged moment conditions to accommodate the dependence in time series non-parametrically and it has some useful properties of EL, such as Wilks's theorem. In Kitamura (1997), the limiting χ^2-distribution for the empirical log-likelihood ratio (up to a multiplicative constant) was shown under traditional small b asymptotics, in which b, the fraction of block size relative to sample size, goes to 0 as the sample size $n\to\infty$. Adopting fixed b asymptotics (Kiefer and Vogelsang, 2005), in which $b\in(0,1)$ is held fixed as $n\to\infty$, Zhang and Shao (2014) derived the pivotal limit of the empirical log-likelihood ratio at the true parameter value and used that as the basis for confidence region construction. The pivotal limit depends on b and the simulations show that the fixed-b-based confidence set has more accurate coverage than the small b counterpart, indicating that the approximation by the fixed b pivotal limit is more accurate than the small b counterpart (i.e. χ^2).

Since this paper is related to our previous work in Zhang and Shao (2014), it pays to highlight the difference. The focus of this paper is rather different from that of Zhang and Shao (2014), and we investigate the coverage upper bound problem of the block-based EL methods for time series. The technique that we use to derive the large sample bound, which depends on b, is completely different from that involved in the derivation of the fixed b limit of the EL ratio statistic in Zhang and Shao (2014). The main contribution of the current paper is

(a) to identify the coverage bound problem for block-based EL methods in time series settings and to study the factors (e.g. sample size, block size, joint distribution of time series and the form of moment conditions) that determine its magnitude (the large sample bound that we derive under fixed b asymptotics provides an approximation to its finite sample counterpart and the approximation is accurate for large n and
(b) to propose penalized BEL and EBEL methods as remedies of the coverage bound problem, and to show their effectiveness through theory and simulations.

Let $1-\beta_n$ denote the probability that the convex hull of the moment conditions at the true parameter value contains the origin as an interior point and it is a natural upper bound on the coverage probability of the BEL ratio confidence region (with any finite critical values) regardless of its confidence level. In Tsao (2004), a finite sample upper bound was derived for independent estimation equations and the EL method. Tsao's technique is tailored to the independent case and seems not applicable to time series data. The calculation of the finite sample bound in the dependent and BEL case is challenging since it depends on the sample size, block size, dimension and form of moment conditions as well as the joint distribution of time series. To shed some light on the coverage bound $1-\beta_n$, we approximate $1-\beta_n$ by its large sample counterpart $1-\beta$, where β is shown to be the probability that the pivotal limit (under fixed b asymptotics) equals ∞. We further provide an analytical formula for β as a function of b in the case $k=1$, and we derive an upper bound for $1-\beta$ in the case $k>1$, where k denotes the dimension of moment conditions. Interestingly, we discover that $\beta=\beta(b)>0$ for any $b>0$ and β can be close to 1 for fixed $b \in (0,1)$ if the dimension of moment conditions k is moderately large. Compared with Tsao (2004) and Kitamura (1997), the large sample bound problem (i.e. $\beta>0$) is a unique feature that is associated with BEL under fixed b asymptotics and it does not occur under the traditional small b asymptotic approximation or for independent estimation equations. It is also worth pointing out that the large sample bound is always 1 under small b asymptotics regardless of the choice of block size, and it provides an inaccurate approximation of the finite sample bound and could lead to an overoptimistic and thus misleading inference. In corroboration with our theoretical results, our simulations show that the finite sample coverage bound can deviate substantially from 1 when

(a) the block size is large relative to the sample size (i.e. b is large),
(b) the dimension of moment conditions is moderate or high and
(c) the time series dependence is positively strong.

In any one of these cases, constructing a confidence set of a conventional nominal level (say, 95% or 99%) is likely to lead to undercoverage. Thus our finding represents a cautionary note on the recent (theoretical) extension of BEL in the high dimensional setting (see Chang *et al.* (2015)), where the dimension of moment conditions can also grow to ∞ as the sample size n grows to ∞.

EBEL uses a sequence of nested blocks with growing sizes so no choice of block size is involved, and the empirical log-likelihood ratio at the true parameter value converges to a pivotal but non-standard limit. Unlike BEL, there is no large sample bound problem for EBEL as the probability that the pivotal limit of EBEL equals ∞ is 0. However, the finite sample bound can be far below the nominal level as shown in our simulations and results in a severe undercoverage. To alleviate the finite sample undercoverage problem that is caused by the convex hull constraint, we propose to penalize BEL and EBEL by dropping the convex hull constraint. Penalized EL (PEL) was first introduced by Bartolucci (2007) for the inference of the mean of independent and identically distributed (IID) data, and our generalization to the time series context requires a non-trivial modification. In particular, we introduce a new normalization matrix that takes the dependence into account and we derive the limit of log-EL ratio at the true value under fixed b asymptotics. Our numerical results in the on-line supplementary material suggest that fixed b asymptotics not only provide better approximation for the original BEL (see Zhang and Shao (2014)) but also tends to provide a better finite sample approximation for its penalized counterpart. Our new PEL ratio test statistic can be viewed as an intermediate between the empirical log-likelihood ratio test statistic and the self-normalized score test statistic (see expression (21) in Section 4.1) with the tuning parameter in the penalization term determining the amount of relaxation of the convex hull constraint. Our numerical results show the effectiveness of the two penalization-based EL methods in terms of delivering more accurate confidence sets for a range of tuning parameters. It is worth noting that the undercoverage problem that is associated with BEL methods may be alleviated by applying a block bootstrap approximation, as pointed out by a referee. However, the block bootstrap does not completely solve the coverage bound problem. In particular, when the finite sample bound is below the nominal level, the undercoverage is bound to occur for any finite critical values, including bootstrap based. By contrast, the finite sample bounds for our penalized BEL and EBEL are always 1, so they are free of the coverage bound problem; see remark 2 in Section 3.1 for more discussions.

A word on notation: let $D[0,1]$ be the space of functions on $[0,1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (Billingsley, 1999). Weak convergence in $D[0,1]$ or more generally in the $\mathbb{R}^q$-valued function space $D^q[0,1]$ is denoted by '$\Rightarrow$', where $q \in \mathbb{N}$. Convergence in probability and convergence in distribution are denoted by '$\rightarrow^{\mathrm{p}}$' and '$\rightarrow^{\mathrm{d}}$' respectively. Let $\lfloor a \rfloor$ be the integer part of $a \in \mathbb{R}$. The notation $N(v, \Sigma)$ is used to denote the multivariate normal distribution with mean v and covariance Σ. Technical details and some simulation results are gathered in the on-line supplementary material. The data sets and R code that are used for this paper can be found at the second author's personal web page `https://publish.illinois.edu/xshao/publications-full-list/`.

2. Blockwise empirical likelihood and expansive blockwise empirical likelihood

Suppose that we are interested in the inference of a p-dimensional parameter vector θ, which is identified by a set of moment conditions. Denote by θ_0 the true parameter of θ which is an interior point of a compact parameter space $\Theta \subseteq \mathbb{R}^p$. Let $\{z_t\}_{t=1}^n$ be n observations from an $\mathbb{R}^l$-valued stationary time series and assume that the moment conditions

$$\mathbb{E}[f(z_t, \theta_0)] = 0, \qquad t = 1, 2, \ldots, n, \tag{1}$$

hold, where $f(z_t, \theta): \mathbb{R}^l \times \Theta \rightarrow \mathbb{R}^k$ is a map which is differentiable with respect to θ and

$$\operatorname{rank}\{E[\partial f(z_t, \theta_0)/\partial \theta']\} = p$$

with $k \geqslant p$. To deal with time series data, we consider the fully overlapping smoothed moment condition (Kitamura, 1997) which is given by

$$f_{tn}(\theta) = \frac{1}{m} \sum_{j=t}^{t+m-1} f(z_j, \theta)$$

with $t = 1, 2, \ldots, n - m + 1$ and $m = \lfloor nb \rfloor$ for $b \in (0, 1)$. The overlapping data blocking scheme aims to preserve the underlying dependence between neighbouring time observations. Consider the profile empirical log-likelihood function based on the fully overlapping smoothed moment conditions,

$$\mathcal{L}_n(\theta) = \sup\left\{ \sum_{t=1}^N \log(\pi_t) : \pi_t \geqslant 0, \ \sum_{t=1}^N \pi_t = 1, \ \sum_{t=1}^N \pi_t f_{tn}(\theta) = 0 \right\}, \qquad N := n - m + 1. \tag{2}$$

Standard Lagrange multiplier arguments imply that the maximum is attained when

$$\pi_t = \frac{1}{N\{1 + \lambda' f_{tn}(\theta)\}}, \qquad \sum_{t=1}^N \frac{f_{tn}(\theta)}{1 + \lambda' f_{tn}(\theta)} = 0,$$

where λ is the Lagrange multiplier. By duality, the empirical log-likelihood ratio function (up to a multiplicative constant) is given by

$$\operatorname{elr}(\theta) = \frac{2}{nb} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^N \log\{1 + \lambda' f_{tn}(\theta)\}, \qquad \theta \in \Theta. \tag{3}$$

Under traditional small b asymptotics, i.e. $nb^2 + 1/(nb) \rightarrow 0$ as $n \rightarrow \infty$, and suitable weak dependence assumptions (Kitamura (1997); also see theorem 1 of Nordman and Lahiri (2014)), it can be shown that

$$\mathrm{elr}(\theta_0) \xrightarrow{\mathrm{d}} \chi^2_k. \tag{4}$$

As pointed out by Nordman *et al.* (2013), the coverage accuracy of BEL can depend crucially on the block length $m = \lfloor nb \rfloor$ and appropriate choices can vary with respect to the joint distribution of the series. To capture the choice of block length in the asymptotics, Zhang and Shao (2014) adopted the fixed b approach that was proposed by Kiefer and Vogelsang (2005) in the context of heteroscedasticity–auto-correlation robust testing and derived the non-standard limit of $\mathrm{elr}(\theta_0)$. To proceed, we make the following assumption which can be verified under suitable moment and weak dependence assumptions on $f(z_j,\theta_0)$ (see for example Phillips (1987)).

Assumption 1. Assume that $\Sigma_{j=1}^{\lfloor nr \rfloor} f(z_j,\theta_0)/\sqrt{n} \Rightarrow \Lambda W_k(r)$ for $r \in [0,1]$, where $\Lambda\Lambda' = \Omega = \Sigma_{j=-\infty}^{\infty} \Gamma_j$ with $\Gamma_j = \mathbb{E}[f(z_{t+j},\theta_0)\, f(z_t,\theta_0)']$ and $W_k(r)$ is a k-dimensional vector of independent standard Brownian motions.

Under assumption 1, Zhang and Shao (2014) showed that, when $n \to \infty$ and b is held fixed,

$$\mathrm{elr}(\theta_0) \xrightarrow{\mathrm{d}} U_{\mathrm{el},k}(b) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^{1-b} \log[1 + \lambda'\{W_k(r+b) - W_k(r)\}]\,\mathrm{d}r, \tag{5}$$

where we define $\log(x) = -\infty$ for $x \leqslant 0$. The asymptotic distribution $U_{\mathrm{el},k}(b)$ is non-standard yet pivotal for a given b, and its critical values can be obtained via simulation or the bootstrap. Given $b \in (0,1)$, a $100(1-\alpha)\%$ confidence region for the parameter θ_0 is then given by

$$\mathrm{CI}(1-\alpha; b) = \left\{ \theta \in \Theta : \frac{\mathrm{elr}(\theta)}{1-b} \leqslant u_{\mathrm{el},k}(b; 1-\alpha) \right\}, \tag{6}$$

where $u_{\mathrm{el},k}(b;1-\alpha)$ denotes the $100(1-\alpha)\%$ quantile of the distribution $P\{U_{\mathrm{el},k}(b)/(1-b) \leqslant x\}$. It was demonstrated in Zhang and Shao (2014) that the confidence region based on the fixed b approximation has more accurate coverage than the traditional counterpart. Our analysis in the next section reveals an interesting coverage upper bound problem associated with the fixed b approach in the BEL framework. This result provides some insight on the use of fixed-b-based critical values as suggested in Zhang and Shao (2014). It also sheds some light on the finite sample coverage bound problem that can occur as long as the BEL ratio statistic is used to construct the confidence region. Moreover, we propose a penalized version of the fixed-b-based BEL, which improves the finite sample performance of the method in Zhang and Shao (2014).

To avoid the choice of block length and also to improve the finite sample coverage, Nordman *et al.* (2013) proposed a new version of BEL which uses a non-standard data blocking rule. To describe their approach, we let

$$\tilde{f}_{tn}(\theta) = \frac{\omega(t/n)}{n} \sum_{j=1}^{t} f(z_j,\theta)$$

for $t = 1, 2, \ldots, n$, where $\omega(\cdot) : [0,1] \to [0,\infty)$ denotes a non-negative weight function. The block collection, which constitutes a type of forward scan in the block subsampling language of McElroy and Politis (2007), contains a data block of every possible length for a given sample size n. It is worth noting that this non-standard data blocking rule bears some resemblance to recursive estimation in the self-normalization approach of Shao (2010). Following Nordman *et al.* (2013), we define the EBEL ratio function as

$$\widetilde{\mathrm{elr}}(\theta) = \frac{1}{n} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^{n} \log\{1 + \lambda' \tilde{f}_{tn}(\theta)\}. \tag{7}$$

For the smooth function model, Nordman *et al.* (2013) showed that

$$\widetilde{\mathrm{elr}}(\theta_0) \xrightarrow{\mathrm{d}} U_{\mathrm{ebel},k}(\omega) = \max_{\lambda\in\mathbb{R}^k} \int_0^1 \log\{1+\lambda'\omega(r)\, W_k(r)\}\mathrm{d}r. \tag{8}$$

The numerical studies in Nordman *et al.* (2013) indicate that EBEL generally exhibits comparable (or in some cases even better) coverage accuracy than BEL with χ^2-approximation and suitable block size. Though the fixed-b-based BEL and EBEL provide an improvement over traditional χ^2-based BEL, our study in the next section reveals that both fixed-b-based BEL and EBEL can suffer seriously from the coverage upper bound problem in finite samples. To the best of our knowledge, this is the first time that the coverage upper bound problem has been revealed for EL methods in time series.

3. Bounds on the coverage probabilities

3.1. Large sample bounds

In the framework of BEL, asymptotic theory is typically established under small b asymptotics, where the large sample bound problem does not occur as the empirical log-likelihood ratio statistic converges to a χ^2-limit. However, in finite samples, the coverage upper bound $1-\beta_n$ can deviate significantly from 1. To shed some light on the finite sample coverage bound, we derive a limiting upper bound on the coverage probabilities of the BEL ratio confidence region based on the fixed b limiting distribution given in expression (5). The fixed b method that is adopted here reflects the coverage upper bound problem in the asymptotics, whereas the original BEL under the small b asymptotics is somewhat 'overoptimistic' as the corresponding upper bound in the limit is always 1 regardless of what the finite sample bound is. Define $D_k(r;b)=W_k(r+b)-W_k(r)$ and $\mathcal{A}=\mathcal{A}_b=\{\lambda\in\mathbb{R}^k : \min_{r\in(0,1-b)}\{1+\lambda' D_k(r;b)\}\geqslant 0\}$. Let

$$t_k(r;b)=\frac{D_k(r;b)}{|D_k(r;b)|}\mathbf{I}\{|D_k(r;b)|>0\}$$

be the direction of $D_k(r;b)$ on the $(k-1)$-dimensional sphere $\mathcal{S}^{k-1}$, where $|\cdot|$ denotes the Euclidean norm and $\mathbf{I}\{\cdot\}$ denotes the indicator function. We first present the following lemma regarding the unboundedness of $\mathcal{A}$.

Lemma 1. Define the convex hull $\mathcal{H}(D_k)=\{\Sigma_{j=1}^s \alpha_j D_k(r_j;b) : s\in\mathbb{N}, \alpha_j\geqslant 0, \Sigma_{j=1}^s\alpha_j=1, r_j\in(0,1-b)\}$. Then the set $\mathcal{A}$ is unbounded if and only if the origin is not an interior point of $\mathcal{H}(D_k)$.

From the proof of lemma 1 (which is given in the on-line supplementary material) and the fact that the components of $D_k(r;b)$ are linearly independent (with probability 1), we know that $\{\mathcal{A}$ is unbounded$\}$ implies that $\{U_{\mathrm{el},k}(b)=\infty\}$. However, when $U_{\mathrm{el},k}(b)=\infty$, it is easy to see that $\mathcal{A}$ cannot be bounded. Therefore we have $P(\mathcal{A}$ is unbounded$)=P\{U_{\mathrm{el},k}(b)=\infty\}$. Let $\mathcal{H}_n(\theta_0;b)=\{\Sigma_{t=1}^N \alpha_t f_{tn}(\theta_0) : \alpha_t\geqslant 0, \Sigma_{t=1}^N\alpha_t=1\}$ and denote by $\mathcal{H}_n^{\mathrm{o}}(\theta_0;b)$ the interior of $\mathcal{H}_n(\theta_0;b)$. By lemma 1 and strong approximation, we have, for large n, $P\{$the origin is not contained in $\mathcal{H}_n^{\mathrm{o}}(\theta_0;b)\}\approx P(\mathcal{A}$ is unbounded$)=P\{U_{\mathrm{el},k}(b)=\infty\}$.

It was conjectured in Zhang and Shao (2014) that $P(\mathcal{A}$ is unbounded$)>0$, which implies that $P(U_{\mathrm{el},k}(b)=\infty)>0$. In what follows, we give an affirmative answer to this conjecture and provide an explicit formula for the probability $P(\mathcal{A}$ is unbounded$)$ when $k=1$. For $k=1$, we must have $\{\mathcal{A}$ is unbounded$\}=\{D_1(r;b)\geqslant 0, \forall\, r\in(0,1-b]\}\cup\{D_1(r;b)\leqslant 0, \forall\, r\in(0,1-b]\}$. By the symmetry of a Wiener process, we have $P(\mathcal{A}$ is unbounded$)=2P\{D_1(r;b)\geqslant 0, \forall\, r\in(0,1-b]\}$. For $\beta>0$, we let $\phi_\beta(\cdot)=\phi(\cdot/\surd\beta)/\surd\beta$ with $\phi(x)=\{1/(\surd 2\pi)\}\exp(-x^2/2)$ being the standard normal density. For two vectors $x=(x_1,x_2,\ldots,x_L)'$ and $y=(y_1,y_2,\ldots,y_L)'$ of real numbers with $L\in\mathbb{N}$, define the matrix $\mathcal{Q}_{\beta,L}(x,y)=(\phi_\beta(x_i-y_j))_{i,j=1}^{L}$. Let $q_{\beta,L}(x,y)$ be the determinant of $\mathcal{Q}_{\beta,L}(x,y)$. For a vector $x=(x_1,x_2,\ldots,x_L)'$, denote by $x_{s_1:s_2}=(x_{s_1},x_{s_1+1},\ldots,x_{s_2})'$ the subvector of x for $1\leqslant s_1\leqslant s_2\leqslant L$. Using similar arguments to those in Shepp (1971) (also see Karlin and Mcgregor (1959)), we prove the following result.

Theorem 1. If $L=1/b$ is a positive integer, we have

$$P\{D_1(r;b)\geqslant 0,\forall r\in(0,1-b]) = \int_{0=x_1<x_2<x_3<\ldots<x_{L+1}} q_{1,L}(x_{1:L},x_{2:(L+1)})\,\mathrm{d}x_2\,\mathrm{d}x_3\ldots\mathrm{d}x_{L+1}, \tag{9}$$

where $x=(x_1,\ldots,x_{L+1})'$. If $bL+\tau=1$ with L being a positive integer and $0<\tau<b$, we have

$$P\{D_1(r;b)\geqslant 0,\forall r\in(0,1-b]\} = \int\ldots\int_S q_{\xi,L+1}(x,y)\,q_{1-\xi,L}(x_{2:(L+1)},y_{1:L})\,\mathrm{d}y_1\,\mathrm{d}x_2\,\mathrm{d}y_2 \ldots\mathrm{d}x_{L+1}\mathrm{d}y_{L+1}, \qquad 0<\xi=\tau/b<1, \tag{10}$$

where $x=(x_1,\ldots,x_{L+1})'$ with $x_1=0$, $y=(y_1,\ldots,y_{L+1})'$ and the integral is over the set $S:=\{(y_1,x_2,y_2,\ldots,x_{L+1},y_{L+1})\in\mathbb{R}^{2L+1}:0<x_2<\ldots<x_{L+1},y_1<y_2<\ldots<y_{L+1}\}$.

Theorem 1 provides an exact formula for the probability $P(\mathcal{A}$ is unbounded) when $k=1$. The probability can be manually calculated when L is small. In particular, if $b=\frac{1}{2}$ (i.e. $L=2$), we have

$$\begin{aligned}P(\mathcal{A}\text{ is unbounded}) &= 2\int_{0<x_2<x_3}\{\phi(-x_2)\,\phi(x_2-x_3)-\phi(-x_3)\,\phi(0)\}\,\mathrm{d}x_2\,\mathrm{d}x_3\\ &= 2\left\{\Phi^2(0)-\phi(0)\int_{-\infty}^{0}\Phi(x)\,\mathrm{d}x\right\}\\ &= 0.18169,\end{aligned}$$

where $\Phi(\cdot)$ denotes the distribution function of the standard normal random variable. When $b=\frac{1}{3}$ (i.e. $L=3$), direct calculation yields that

$$\begin{aligned}P(\mathcal{A}\text{ is unbounded}) &= 2\left\{\Phi^3(0)+\frac{\phi^2(0)}{4}+\int_{0<x_2<x_3}\phi(-x_3)\,\phi(x_3-x_2)\,\Phi(x_2-x_3)\,\mathrm{d}x_2\,\mathrm{d}x_3\right\}\\ &\quad -2\left\{\int_{0<x_2<x_3}\phi^2(x_3-x_2)\,\Phi(-x_3)\,\mathrm{d}x_2\,\mathrm{d}x_3+\phi^2(0)\,\Phi(0)\right.\\ &\quad \left.+\int_{0<x_2<x_3}\phi(-x_2)\,\phi(0)\,\Phi(x_2-x_3)\,\mathrm{d}x_2\,\mathrm{d}x_3\right\}\\ &= 2\left\{\frac{1}{8}+\frac{\phi^2(0)}{4}+\int_{-\infty}^{0}\phi(u)\,\Phi^2(u)\,\mathrm{d}u+\frac{\phi^2(0)}{2\surd 2}\right.\\ &\quad \left.-\frac{1}{\surd(4\pi)}\int_0^\infty\Phi(-x_3)\,\Phi(\surd 2x_3)\,\mathrm{d}x_3-\phi^2(0)\right\}\\ &= 0.03635.\end{aligned}$$

The calculation for larger L is still possible but is more involved. An alternative way is to approximate the probabilities in expression (9) and (10) by using Monte Carlo simulation; see Table 1 and Fig. 1. Utilizing the result in theorem 1, we can derive a (conservative) upper bound on $P(\mathcal{A}$ is bounded) (i.e. $1-\beta$) in the multi-dimensional case. For $k>1$, we let $D_k^{(j)}(r;b)$ be the

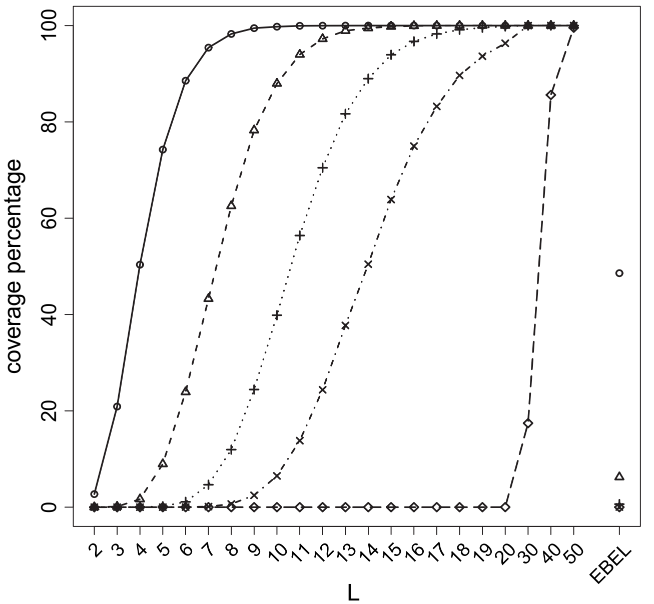

Fig. 1. Bounds on the coverage probabilities for the BEL and EBEL (the data are generated from a multivariate standard normal distribution with $n=5000$ and the number of Monte Carlo replications is 10000): $\bigcirc$, $k=5$; $\triangle$, $k=10$; $+$, $k=15$; $\times$, $k=20$; $\diamond$, $k=50$

jth element of $D_k(r;b)$ and $\mathcal{V}_j=\{D_k^{(j)}(r;b)\geqslant 0, \forall\, r\in(0,1-b]\}\cup\{D_k^{(j)}(r;b)\leqslant 0, \forall\, r\in(0,1-b]\}$ with $1\leqslant j\leqslant k$. By the independence of the components of $D_k(r;b)$, it is easy to derive that

$$\begin{aligned} P(\mathcal{A}\text{ is unbounded})\geqslant & P(\cup_{j=1}^k \mathcal{V}_j)=1-P(\cap_{j=1}^k \mathcal{V}_j^{\mathrm{c}})=1-P^k(\mathcal{V}_1^{\mathrm{c}}) \\ =&1-[1-2\,P\{D_k^{(j)}(r;b)\geqslant 0, \forall\, r\in(0,1-b]\}]^k. \end{aligned}$$

Therefore, we obtain the following result.

Proposition 1. When $L=1/b$ is a positive integer, we have

$$P\{U_{\mathrm{el},k}(b)<\infty)\leqslant\left\{1-2\int_{0=x_1<x_2<x_3<\ldots<x_{L+1}} q_{1,L}(x_{1:L},x_{2:(L+1)})\mathrm{d}x_2\,\mathrm{d}x_3\ldots\mathrm{d}x_{L+1}\right\}^k, \qquad (11)$$

where $x=(x_1,\ldots,x_{L+1})'$. When $bL+\tau=1$ with L being a positive integer and $0<\tau<b$, we have

$$\begin{aligned} &P\{U_{\mathrm{el},k}(b)<\infty\} \\ &\quad\leqslant\left\{1-2\int\ldots\int_S q_{\xi,L+1}(x,y)\,q_{1-\xi,L}(x_{2:(L+1)},y_{1:L})\,\mathrm{d}y_1\,\mathrm{d}x_2\,\mathrm{d}y_2\ldots\mathrm{d}x_{L+1}\,\mathrm{d}y_{L+1}\right\}^k, \\ &\qquad\qquad 0<\xi=\tau/b<1, \qquad (12) \end{aligned}$$

where $x=(x_1,\ldots,x_{L+1})'$ with $x_1=0$, $y=(y_1,\ldots,y_{L+1})'$ and the integral is over the set $S:=\{(y_1,x_2,y_2,\ldots,x_{L+1},y_{L+1})\in\mathbb{R}^{2L+1}:0<x_2<\ldots<x_{L+1}, y_1<y_2<\ldots<y_{L+1}\}$. When $k=1$, the inequality becomes equality in expressions (11) and (12).

If the (asymptotic) critical value based on the fixed b pivotal limit $U_{\mathrm{el},k}(b)$ is used to construct a $100(1-\alpha)\%$ confidence region, then the following several cases can occur.

Table 1. Bounds on the coverage probabilities for BEL†

n	ρ	k	*Bounds (%) for the following values of* $L=1/b$:										
			2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*15*	*20*
50	0.0	1	73.61	93.61	98.07	99.30	99.83	99.93	99.98	99.99	99.99	100.00	100.00
	0.0	2	37.22	75.83	90.47	95.71	98.80	99.52	99.82	99.97	99.97	100.00	100.00
	0.2	1	71.20	92.26	97.36	98.92	99.70	99.87	99.94	99.99	99.99	100.00	100.00
	0.2	2	33.45	71.65	87.90	93.80	97.89	99.02	99.68	99.91	99.91	100.00	100.00
	0.5	1	66.47	89.26	95.61	97.83	99.21	99.57	99.80	99.93	99.93	100.00	100.00
	0.5	2	26.73	63.07	81.76	89.19	95.31	97.20	98.60	99.43	99.43	99.98	100.00
	0.8	1	56.83	80.69	89.45	92.89	95.82	96.93	97.91	98.63	98.63	99.62	99.84
	0.8	2	15.98	44.42	62.81	72.30	81.15	85.25	89.09	92.53	92.53	97.24	98.84
	−0.5	1	80.16	96.44	99.24	99.80	99.98	99.99	100.00	100.00	100.00	100.00	100.00
	−0.5	2	48.66	85.17	95.78	98.61	99.84	99.90	100.00	100.00	100.00	100.00	100.00
100	0.0	1	76.36	94.13	98.34	99.57	99.90	99.97	99.99	99.99	99.99	100.00	100.00
	0.0	2	40.72	76.22	91.04	96.94	99.13	99.76	99.90	99.97	99.99	100.00	100.00
	0.2	1	74.59	93.17	97.86	99.40	99.85	99.95	99.98	99.99	99.99	100.00	100.00
	0.2	2	37.51	73.02	89.00	95.91	98.68	99.47	99.80	99.90	99.97	100.00	100.00
	0.5	1	71.16	91.01	96.77	98.92	99.65	99.84	99.95	99.98	99.99	100.00	100.00
	0.5	2	32.39	67.18	84.64	93.24	97.33	98.54	99.47	99.69	99.85	100.00	100.00
	0.8	1	63.71	85.58	93.20	96.71	98.53	99.08	99.49	99.64	99.76	99.98	99.99
	0.8	2	22.53	53.78	72.08	83.59	91.28	94.11	96.37	97.40	98.21	99.78	99.90
	−0.5	1	80.78	96.27	99.21	99.82	99.99	99.99	100.00	100.00	100.00	100.00	100.00
	−0.5	2	48.77	83.68	95.10	98.70	99.67	99.99	100.00	100.00	100.00	100.00	100.00
500	0.0	1	79.87	95.51	98.96	99.79	99.96	99.98	100.00	100.00	100.00	100.00	100.00
	0.0	2	45.60	80.19	93.63	98.12	99.57	99.87	99.95	99.99	100.00	100.00	100.00
	0.2	1	79.16	95.11	98.83	99.74	99.95	99.98	99.99	100.00	100.00	100.00	100.00
	0.2	2	44.10	79.22	92.90	97.75	99.44	99.83	99.94	99.98	99.99	100.00	100.00
	0.5	1	77.50	94.37	98.52	99.63	99.92	99.97	99.98	100.00	100.00	100.00	100.00
	0.5	2	41.33	76.71	91.50	96.83	99.10	99.68	99.86	99.96	99.98	100.00	100.00
	0.8	1	73.91	92.43	97.65	99.25	99.75	99.90	99.95	99.98	99.99	100.00	100.00
	0.8	2	35.56	70.69	87.36	94.58	97.81	99.13	99.58	99.84	99.92	100.00	100.00
	−0.5	1	81.92	96.36	99.30	99.86	99.97	100.00	100.00	100.00	100.00	100.00	100.00
	−0.5	2	49.76	83.39	95.42	98.87	99.74	99.91	99.97	100.00	100.00	100.00	100.00
1000	0.0	1	80.06	95.67	99.05	99.78	99.96	99.99	100.00	100.00	100.00	100.00	100.00
	0.0	2	45.25	80.59	94.01	98.16	99.54	99.87	99.97	99.99	100.00	100.00	100.00
	0.2	1	79.49	95.43	98.97	99.77	99.95	99.99	100.00	100.00	100.00	100.00	100.00
	0.2	2	44.09	79.60	93.50	98.02	99.39	99.84	99.97	99.98	100.00	100.00	100.00
	0.5	1	78.40	94.94	98.79	99.72	99.94	99.99	100.00	100.00	100.00	100.00	100.00
	0.5	2	42.06	77.83	92.50	97.55	99.23	99.73	99.89	99.95	100.00	100.00	100.00
	0.8	1	76.01	93.68	98.27	99.48	99.86	99.95	100.00	100.00	100.00	100.00	100.00
	0.8	2	37.93	73.21	89.65	96.21	98.57	99.49	99.76	99.90	99.97	100.00	100.00
	−0.5	1	81.64	96.30	99.24	99.85	99.98	100.00	100.00	100.00	100.00	100.00	100.00
	−0.5	2	48.21	82.96	95.11	98.76	99.74	99.92	99.98	100.00	100.00	100.00	100.00
∞	0.0	1	81.70	96.26	99.23	99.85	99.97	99.99	100.00	100.00	100.00	100.00	100.00
	0.0	2	48.58	82.93	95.04	98.72	99.70	99.91	99.99	100.00	100.00	100.00	100.00

†The number of Monte Carlo replications is 50000 for $k=1$ (10000 for $k=2$). For the last row $n=\infty$, we approximate the probability $P(\mathcal{A}$ is bounded) by simulating independent Wiener processes, where the Wiener process is approximated by a normalized partial sum of 50000 for $k=1$ (10000 for $k=2$) IID standard normal random variables and the number of replications is 100000 for $k=1$ (50000 for $k=2$).

(a) $P\{U_{\mathrm{el},k}(b)<\infty\}=1-\beta\leqslant 1-\alpha$; then the fixed-$b$-based critical value is ∞. In this case, it is impossible to construct a meaningful confidence region as $\{\theta\in\Theta|\mathrm{elr}(\theta)\leqslant\infty\}=\Theta$. In the case $k=1$, the value of β is known but, in the case $k=2$ or higher, only an upper bound for $1-\beta$ is provided in proposition 1. Thus, if the upper bound is no greater than $1-\alpha$, then we cannot construct a sensible confidence region based on fixed b critical values.

(b) $P\{U_{\mathrm{el},k}(b) < \infty\} = 1 - \beta > 1 - \alpha$; then the fixed-$b$-based critical value is finite. The $100(1-\alpha)\%$ quantile of the distribution of $U_{\mathrm{el},k}(b)/(1-b)$ (i.e. $u_{\mathrm{el},k}(b; 1-\alpha)$) is the $100\gamma\%$ quantile of the conditional distribution $P\{U_{\mathrm{el},k}(b)/(1-b) \leqslant x | U_{\mathrm{el},k}(b) < \infty\}$, where $\gamma = (1-\alpha)/(1-\beta)$. In the simulation experiment of Zhang and Shao (2014), the $100(1-\alpha)\%$ quantile of the conditional distribution was used as the critical value. Note that the largest b considered in Zhang and Shao (2014) was 0.2, which corresponds to $\beta \approx 1 - 0.9985 = 0.0015$ when $k=1$ and $\beta \approx 1 - 0.9872 = 0.0128$ when $k=2$, as seen from Table 1. This suggests that the critical values that were used in Zhang and Shao (2014) are wrong, but not by much.

(c) In the event that $u_{\mathrm{el},k}(b; 1-\alpha)$ is finite, which occurs in case (b) above or when the χ^2-based critical values are used,

$$P\{\theta_0 \in \mathrm{CI}(1-\alpha; b)\} \leqslant P\{\text{the origin is contained in } \mathcal{H}_n^{\mathrm{o}}(\theta_0; b)\} = 1 - \beta_n,$$

which is a finite sample bound. The quantity β_n depends on joint distributions of time series, the form of f, the block size and the sample size, so it is in general difficult to calculate. We present some numerical results on β_n in Section 3.2 below. If $1-\beta_n \leqslant 1-\alpha$, then the confidence region is bound to undercover and the amount of undercoverage becomes severe when β_n is further from 0.

Proposition 1 shows that, for any fixed $b \in (0,1)$, the bound decays exponentially to zero as the dimension k grows. This result suggests that caution needs to be taken in the recent extension of the BEL to the high dimensional setting (see Chang *et al.* (2015)), where the dimension of moment condition k can grow with respect to sample size n. In Chang *et al.* (2015), small b asymptotics were adopted, and no discussion on such a coverage bound issue (either finite sample or large sample) seems provided. It would be interesting to extend the fixed b asymptotic approach to BEL in the high dimensional setting and we leave it for future investigation.

The large sample bound on the coverage probabilities depends crucially on how the smoothed moment conditions are constructed. By lemma 1 of Nordman *et al.* (2013), we know that, for EBEL, the set $\mathcal{A}_\omega = \{\lambda \in \mathbb{R}^k : \min_{r\in(0,1)}\{1 + \lambda'\omega(r)\, W_k(r)\} \geqslant 0\}$ is bounded with probability 1,

Table 2. Bounds on the coverage probabilities for EBEL†

ρ	k	*Bounds (%) for the following values of n:*				
		50	*100*	*500*	*1000*	*5000*
0.0	1	84.02	89.51	94.64	96.18	98.30
0.0	2	52.45	62.66	78.48	84.22	92.09
0.2	1	82.20	87.93	93.99	95.74	98.12
0.2	2	48.95	59.47	76.50	82.55	91.51
0.5	1	77.57	84.31	92.35	94.57	97.72
0.5	2	41.24	52.96	72.31	79.09	89.84
0.8	1	65.99	75.78	88.32	91.91	96.52
0.8	2	26.56	39.06	62.10	71.32	85.62
−0.5	1	87.42	91.75	95.89	96.83	98.70
−0.5	2	60.24	69.35	82.52	87.16	93.87

†The number of Monte Carlo replications is 10000. The bounds on the coverage probabilities for EBEL do not depend on the choice of the weight function $\omega(\cdot)$.

which implies that $P\{U_{\mathrm{ebel},k}(\omega)<\infty\}=1$. Thus, for large samples, no upper bound problem occurs for EBEL. However, the numerical results in Table 2 show that the finite sample bounds on the coverage probabilities of the EBEL ratio confidence regions can be significantly lower than 1, which indicates that the convergence of the EBEL ratio statistic $\widetilde{\mathrm{elr}}(\theta_0)$ to its limit $U_{\mathrm{ebel},k}(\omega)$ is in fact slow and substantial undercoverage can be associated with EBEL-based confidence regions in any one of the following three cases:

(a) the dependence is positively strong;
(b) the sample size n is small;
(c) k is moderate, say $k\geqslant 3$.

Remark 1. The convex hull constraint is related to the underlying distance measure between $\pi=(\pi_1,\ldots,\pi_N)$ and $(1/N,\ldots,1/N)$ in EL. If we consider alternative non-parametric likelihood such as the Euclidean likelihood or, more generally, members of the Cressie–Read power divergence family of discrepancies, then the origin is allowed to be outside the convex hull of the smoothed moment conditions as long as the weights are allowed to be negative. No coverage upper bound problem occurs for these alternative non-parametric likelihoods but, since EL has a certain optimality property (Kitamura, 2006; Kitamura *et al.*, 2013), it is still a worthwhile effort to seek remedies of the coverage bound problem based on EL.

Remark 2. An alternative way to calibrate the sampling distribution is the block bootstrap. To illustrate the idea, consider the linear model, i.e. $y_t=x_t'\theta+u_t$, where x_t and θ are $(l-1)$-dimensional vectors ($p=l-1$ in this case), and the stationary time series $\{x_t\}$ and $\{u_t\}$ are uncorrelated. Define $f_t(\theta)=x_t(y_t-x_t'\theta)$ and $z_t=(x_t',y_t)'$. Assume that $n=d_nl_n$, where d_n denotes the block size in the bootstrap and l_n is the number of blocks. Let $M_1,\ldots,M_{l_n}$ be independent and identically distributed (IID) uniform random variables on $\{0,\ldots,n-d_n\}$ and let $z^*_{(j-1)d_n+i}=z_{M_j+i}$ with $1\leqslant j\leqslant l_n$ and $1\leqslant i\leqslant d_n$. Let $f^*_{tn}(\hat{\theta})$ be the smoothed moment condition based on the bootstrap sample $\{z_i^*\}$, where $\hat{\theta}$ is the ordinary least squares estimator based on the original sample. The naive bootstrap version of $\mathrm{elr}(\theta_0)$ is given by

$$\mathrm{elr}^*(\hat{\theta})=\frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{N}\log(1+\lambda'[f^*_{tn}(\hat{\theta})-\mathbb{E}^*\{f^*_{tn}(\hat{\theta})\}]),$$

where $\mathbb{E}^*$ denotes the expectation conditional on $\{z_i\}_{i=1}^n$. Alternatively, $\hat{\theta}$ may be replaced by $\mathbb{E}^*(\Sigma_{t=1}^n x_t^* x_t^{*\prime})^{-1}\mathbb{E}^*(\Sigma_{t=1}^n x_t^* y_t^*)$. The bootstrap critical value obtained from this procedure is expected to provide a better finite sample approximation compared with the χ^2-calibration. The intuition is that, besides the time series dependence (which is captured by the blocking strategy in the block bootstrap) and the effect of sample size, the choice of block size $\lfloor bn\rfloor$ is also reflected in the bootstrap statistic through the construction of f^*_{tn}. This is essentially the rationale behind the fixed b approach. On the basis of the arguments in Gonçalves and Vogelsang (2011), it is expected that the bootstrap test statistic $\mathrm{elr}^*(\hat{\theta})$ has the same limiting distribution (conditionally on the data) as that of $\mathrm{elr}(\theta^0)$ derived under fixed b asymptotics. Therefore, the bootstrap calibration indeed has a deep connection with the fixed b approach. Investigation along this direction is very interesting and will be pursued in the future.

However, a remedy to the coverage bound problem seems necessary when the (finite sample) bound is less than the nominal level, which could happen in the case of high dimension, small sample size, strong dependence or large b. In this case, a naive application of the above bootstrap method to calibrate may not work as there is an intrinsic coverage upper bound for whatever critical values (including bootstrap based). This motivates us to develop the penalized approach in the next section, which is free of the coverage bound problem.

3.2. Finite sample results on coverage bounds

To evaluate the upper bounds on the coverage probabilities for BEL and EBEL, we simulate time series from auto-regressive (AR(1)) models with AR(1) coefficient $\rho = -0.5, 0, 0.2, 0.5, 0.8$, and IID standard normal errors. The sample size $n = 50, 100, 500, 1000, 5000, \infty$. We approximate the probability $P(\mathcal{A}$ is bounded) by simulating independent Wiener processes, where the Wiener process is approximated by the normalized partial sum of 50000 IID standard normal random variables and the number of Monte Carlo replications is 100000. When $k > 1$, we simulate vector AR (VAR(1)) processes with the coefficient matrix $A_1 = \rho I_k$ for $\rho = -0.5, 0, 0.2, 0.5, 0.8$, and standard multivariate normal errors. Table 1 summarizes the upper bounds on the coverage probabilities for BEL with $b = 1/L$ for $L = 2, 3, \ldots, 10, 15, 20,$ and Table 2 provides the finite sample upper bounds on the coverage probabilities for EBEL. For BEL, it is seen from Table 1 that the upper bound on the coverage probability decreases as the block size increases and the positive dependence strengthens. The bound in the multi-dimensional case is lower than its counterpart in the univariate case, which is consistent with our theoretical finding. It is interesting that negative dependence (corresponding to $\rho = -0.5$) tends to bring the upper bound higher. In practice, if the dependence is expected to be positively strong, a large block size is preferable. However, our result indicates that the corresponding upper bound on the coverage probabilities will be lower for larger block size. It is also worth noting that the upper bounds on the coverage probabilities generally increase as the sample size grows and the result in proposition 1 provides conservative bounds on $1 - \beta$ when $k = 2$. For EBEL, though its large sample bound is 1, its finite sample bound can be significantly lower than 1 as seen from Table 2. To assess the effect of the dimensionality k further, we present the coverage upper bounds for $k = 5, 10, 15, 20, 50$ and $L = 2, 3, \ldots, 20, 30, 40, 50$ in Fig. 1, where data are generated from a multivariate standard normal distribution with sample size $n = 5000$. We observe that

(a) as k grows, a smaller b (or larger L) is required to deliver meaningful finite sample upper bounds (say, larger than nominal level) and
(b) the coverage upper bound for EBEL can be close to zero for $k = 15$ or larger.

We expect that the bound can grow worse when we increase the positive dependence in the observations. On the basis of the numerical results for this specific setting, we suggest that special attention is paid to the potential coverage bound problem for the following cases:

(i) the nominal level is close to 1 (such as 99%);
(ii) the dimension of moment conditions k is moderate or high;
(iii) the (positive) dependence is strong;
(iv) b is large.

4. Penalized blockwise empirical likelihood and expansive blockwise empirical likelihood

The convex hull constraint violation underlying the mismatch is well known in the EL literature (see Owen (1990, 2001)). Various methods have been proposed to bypass this constraint, such as PEL (Bartolucci, 2007; Lahiri and Mukhopadhyay, 2012), adjusted EL (Chen *et al.*, 2008; Emerson and Owen, 2009; Liu and Chen, 2010; Chen and Huang, 2012) and extended EL (Tsao and Wu, 2013, 2014). Motivated by the theoretical findings as well as the finite sample results in Section 3.2, we propose a remedy based on penalization to circumvent the coverage bound problem which leads to improved coverage accuracy under fixed b asymptotics.

4.1. Penalized blockwise empirical likelihood

To overcome the convex hull constraint violation problem, Bartolucci (2007) dropped the convex hull constraint in the formulation of EL for the mean of a random sample and defined the likelihood by penalizing the unconstrained EL by using the Mahalanobis distance. Recently, Lahiri and Mukhopadhyay (2012) introduced a modified version of Bartolucci's (2007) PEL in the mean case. Under the assumption that the observations are IID and the components of each observation are dependent, Lahiri and Mukhopadhyay (2012) derived asymptotic distributions of the PEL ratio statistic in the high dimensional setting. Other variants of PEL where a penalty function is added to the standard EL were considered by Otsu (2007) for efficient estimation in semiparametric models and Tang and Leng (2010) for consistent parameter estimation and variable selection in linear models. Otsu (2007) and Tang and Leng (2010) either penalized high dimensional parameters or roughness of unknown non-parametric functions, and their PELs still suffer from the same convex hull constraint violation problem as standard EL does. In what follows, we shall consider a penalized version of the BEL ratio test statistic in the moment condition models, which allows weak dependence within the moment conditions and may be computed even when the origin does not belong to the convex hull of the smoothed moment conditions. Compared with existing penalization methods in the literature, our method is different in three aspects. First, our method is designed for dependent data where existing methods are applicable only to independent moment conditions. Second, our theoretical result is established under fixed b asymptotics, which are expected to provide a better approximation to the finite sample distribution. And we suggest the use of the fixed-b-based critical values that capture the choice of tuning parameters (also see the simulations in the on-line supplementary material). Third, our formulation produces a new class of statistic between the empirical log-likelihood ratio statistic and the self-normalized score statistic which is of interest in its own right. To illustrate the idea, we first consider the case $k=p$, i.e. the moment condition is exactly identified (see remark 3 for the general overidentified case). Define the simplex $\mathfrak{F}_N=\{\pi=(\pi_1,\ldots,\pi_N):\pi_t\geqslant 0,\Sigma_{t=1}^N\pi_t=1\}$ and the quadratic distance measure $\delta_n(\mu):=\delta_{\Psi_n}(\mu)=\mu'\Psi_n^{-1}\mu$ for $\mu\in\mathbb{R}^k$, where $\Psi_n\in\mathbb{R}^{k\times k}$ is an invertible normalization matrix. Let $\mu_\pi(\theta)=\Sigma_{t=1}^N\pi_t f_{tn}(\theta)$ with $\pi=(\pi_1,\ldots,\pi_N)\in\mathfrak{F}_N$. We consider penalized BEL (PBEL) as follows:

$$\mathcal{L}_{\text{pbel},n}(\theta)=\max_{\pi\in\mathfrak{F}_N}\prod_{t=1}^N\pi_t\exp\left[-\frac{n\tau}{2}\delta_n\{\mu_\pi(\theta)\}\right]. \tag{13}$$

The PBEL ratio test statistic is then defined as

$$\text{elr}_{\text{pbel}}(\theta)=-\frac{2}{nb}\log\{N^N\mathcal{L}_{\text{pbel},n}(\theta)\}=\min_{\pi\in\mathfrak{F}_N}\left[-\frac{2}{nb}\sum_{t=1}^N\log(N\pi_t)+\frac{\tau}{b}\delta_n\{\mu_\pi(\theta)\}\right].$$

Under the constraint that $\mu=\Sigma_{t=1}^N\pi_t f_{tn}(\theta)$, it is not difficult to derive that

$$\pi_t=\frac{1}{N[1+\lambda'\{f_{tn}(\theta)-\mu\}]},\qquad\sum_{t=1}^N\frac{f_{tn}(\theta)-\mu}{1+\lambda'\{f_{tn}(\theta)-\mu\}}=0,$$

by using standard Lagrange multiplier arguments. Denote by $\mathcal{H}_n(\theta;b)=\{\Sigma_{t=1}^N\pi_t f_{tn}(\theta):\pi\in\mathfrak{F}_N\}$. Thus we deduce that

$$\text{elr}_{\text{pbel}}(\theta)=\min_{\mu\in\mathcal{H}_n(\theta;b)}\left(\frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N\log[1+\lambda'\{f_{tn}(\theta)-\mu\}]+\frac{\tau}{b}\delta_n(\mu)\right), \tag{14}$$

where μ is minimized to balance the empirical log-likelihood ratio and the penalty term.

Proposition 2. If the space that is spanned by $\{f_{tn}(\theta)\}_{t=1}^{N}$ is of k dimension, we have

$$\mathrm{elr}_{\mathrm{pbel}}(\theta)=\min_{\mu\in\mathbb{R}^k}\left(\frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{N}\log[1+\lambda'\{f_{tn}(\theta)-\mu\}]+\frac{\tau}{b}\delta_n(\mu)\right). \tag{15}$$

The condition that the space that is spanned by $\{f_{tn}(\theta)\}_{t=1}^{N}$ is of k dimension is fairly mild because k is fixed and N grows with n. Note that the minimizer μ^* of equation (15) is necessarily contained in $\mathcal{H}_n(\theta;b)$, which implies that the origin of $\mathbb{R}^k$ is contained in the convex hull of $\{f_{tn}(\theta)-\mu^*\}_{t=1}^{N}$. In addition, since the empirical log-likelihood ratio and the penalty term in equation (15) are both convex functions of μ, it is not difficult to obtain μ^* in practice. Let $\tau=c^*n$ with c^* being a non-negative constant which controls the magnitude of the penalty term, and suppose that $\Psi_n^{-1}\to^{\mathrm{d}}(\Lambda\Phi_k\Lambda')^{-1}$ as $n\to\infty$, where $\Phi_k\in\mathbb{R}^{k\times k}$ is a pivotal limit. For example, if we let $Q(\cdot,\cdot):[0,1]^2\to\mathbb{R}$ be a positive semidefinite kernel, then one possible choice of the normalization matrix Ψ_n is given by

$$\Psi_n(\hat{\theta}_n)=\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{n}Q\left(\frac{t}{n},\frac{j}{n}\right)f(z_t,\hat{\theta}_n)f'(z_j,\hat{\theta}_n), \tag{16}$$

where $\hat{\theta}_n$ is a preliminary estimator obtained by solving the equation $\Sigma_{j=1}^{n}f(z_j,\theta)=0$. In practice, one can choose $Q(r,s)=\kappa(r-s)$ with $\kappa(\cdot)$ being the kernels that are used in the heteroscedasticity and auto-correlation consistent estimation, such as the Bartlett kernel or the quadratic spectral kernel. Under appropriate conditions (see, for example, Kiefer and Vogelsang (2005) and Sun (2013)), it can be shown that

$$\Psi_n(\hat{\theta}_n)\xrightarrow{\mathrm{d}}\Lambda\int_0^1\int_0^1 Q(r,s)\,\mathrm{d}B_k(r)\,\mathrm{d}B_k'(s)\Lambda':=\Lambda\Phi_k\Lambda', \tag{17}$$

where $\Phi_k=\int_0^1\int_0^1 Q(r,s)\,\mathrm{d}B_k(r)\,\mathrm{d}B_k'(s)$ with $B_k(r)=W_k(r)-r\,W_k(1)$. Therefore, under assumption 1, we have

$$\begin{aligned}\mathrm{elr}_{\mathrm{pbel}}(\theta_0)&\xrightarrow{\mathrm{d}}U_{\mathrm{pbel},k}(b)\\&=\min_{\mu\in\mathcal{H}(b)}\left(\frac{2}{b}\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log\left[1+\lambda'\left\{\Lambda\frac{D_k(r;b)}{b}-\mu\right\}\right]\mathrm{d}r+\frac{c^*}{b}\mu'(\Lambda\Phi_k\Lambda')^{-1}\mu\right),\end{aligned} \tag{18}$$

where $\mathcal{H}(b)$ denotes the convex hull of $\{\Lambda\,D_k(r;b)/b:r\in(0,1-b)\}$. When μ is outside the convex hull of $\{\Lambda\,D_k(r;b)/b:r\in(0,1-b)\}$, the separating hyperplane theorem (see for example section 11 of Rockafellar (1970)) implies that $\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log[1+\lambda'\{\Lambda\,D_k(r;b)/b-\mu\}]\mathrm{d}r=\infty$. Thus we have the simplified expression

$$\begin{aligned}U_{\mathrm{pbel},k}(b)&=\min_{\mu\in\mathbb{R}^k}\left(\frac{2}{b}\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log\left[1+\lambda'\left\{\Lambda\frac{D_k(r;b)}{b}-\mu\right\}\right]\mathrm{d}r+\frac{c^*}{b}\mu'(\Lambda\Phi_k\Lambda')^{-1}\mu\right)\\&=\min_{\tilde{\mu}\in\mathbb{R}^k}\left(\frac{2}{b}\max_{\tilde{\lambda}\in\mathbb{R}^k}\int_0^{1-b}\log\left[1+\tilde{\lambda}'\left\{\frac{D_k(r;b)}{b}-\tilde{\mu}\right\}\right]\mathrm{d}r+\frac{c^*}{b}\tilde{\mu}'\Phi_k^{-1}\tilde{\mu}\right),\end{aligned} \tag{19}$$

where $\tilde{\lambda}=\Lambda'\lambda$ and $\tilde{\mu}=\Lambda^{-1}\mu$. Note that the limiting distribution $U_{\mathrm{pbel},k}(b)$ is pivotal and its critical values can be simulated by approximating the Brownian motion with the standardized or normalized partial sum of IID standard normal random variables. As to the pivotal limit $U_{\mathrm{pbel},k}(b)$, we have the following result.

Proposition 3. For $b\in(0,1)$ and $c^*>0$, $P\{U_{\mathrm{pbel},k}(b)<\infty\}=1$.

Thus, compared with BEL, PBEL is well defined and does not suffer from the convex hull violation problem in both large sample and finite sample cases, though it involves the choice of additional tuning parameters such as c^* and Ψ_n.

When $c^*=\infty$, we have $\mu^*=0$ and the PBEL ratio statistic reduces to the BEL ratio statistic. In contrast,

$$\text{elr}_{\text{pbel}}(\theta)=c^* \min_{\mu\in\mathbb{R}^k}\left(\frac{2}{c^*nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{N}\log[1+\lambda'\{f_{tn}(\theta)-\mu\}]+\frac{n}{b}\delta_n(\mu)\right), \tag{20}$$

and

$$\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{N}\log\left[1+\lambda'\left\{f_{tn}(\theta)-\frac{1}{N}\sum_{t=1}^{N}f_{tn}(\theta)\right\}\right]=0.$$

Thus, for small c^*, the minimizer μ^* should be close to $\Sigma_{t=1}^{N}f_{tn}(\theta)/N$. In this case, the penalty term dominates and the PBEL ratio statistic evaluated at the true parameter value behaves like the self-normalized score statistic which is defined as

$$\mathcal{S}_n(\theta_0)=n\delta_n\left(\sum_{t=1}^{N}\frac{f_{tn}(\theta_0)}{N}\right)=n\left(\sum_{t=1}^{N}\frac{f_{tn}(\theta_0)}{N}\right)'\Psi_n^{-1}(\hat{\theta}_n)\left(\sum_{t=1}^{N}\frac{f_{tn}(\theta_0)}{N}\right). \tag{21}$$

We call $\mathcal{S}_n(\theta_0)$ the self-normalized score statistics as $f_{tn}(\theta)$ plays the role of the score in likelihood-based inference and the self-normalizer $\Psi_n(\hat{\theta}_n)$ is an inconsistent estimator of the asymptotic variance matrix Ω in the spirit of the self-normalized approach of Shao (2010). Therefore, on the basis of the quadratic distance measure, the penalized BEL ratio statistic can be viewed as a combination of the BEL ratio statistic and the self-normalized score statistic.

Remark 3. When the moment condition is overidentified (i.e. $k>p$), we shall consider the normalization matrix $\Psi_n=\Psi_n(\hat{\theta}_n)$ with $\hat{\theta}_n$ being a preliminary estimator such as the one-step generalized method-of-moments estimator with the weighting matrix $W_n\rightarrow^{\text{p}}W_0$, where W_0 is a $k\times k$ positive definite matrix. To illustrate the idea, define

$$G_t(\theta)=\frac{1}{n}\sum_{j=1}^{t}\frac{\partial f(z_j,\theta)}{\partial\theta'}$$

and $G_0=\mathbb{E}\{G_n(\theta_0)\}$. Let $\hat{u}_j=\{G_n'(\hat{\theta}_n)W_nG_n(\hat{\theta}_n)\}^{-1}G_n'(\hat{\theta}_n)W_nf(z_j,\hat{\theta}_n)$. Consider the normalization matrix

$$\Psi_n(\hat{\theta}_n)=\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{n}Q\left(\frac{t}{n},\frac{j}{n}\right)\hat{u}_t\hat{u}_j'.$$

Under suitable conditions (see Kiefer and Vogelsang (2005)), it can be deduced that

$$\Psi_n(\hat{\theta}_n)\rightarrow^{\text{d}}\Delta\int_0^1\int_0^1 Q(r,s)\,\mathrm{d}B_p(r)\,\mathrm{d}B_p(s)\Delta',$$

where $\Delta\in\mathbb{R}^{p\times p}$ is an invertible matrix such that $\Delta\Delta'=(G_0'W_0G_0)^{-1}G_0'W_0\Omega W_0G_0(G_0'W_0G_0)^{-1}$. In this case, the PBEL ratio test statistic can be defined as

$$\text{elr}_{\text{pbel}}(\theta)=\min_{\mu\in\mathbb{R}^p}\left(\frac{2}{nb}\max_{\lambda\in\mathbb{R}^p}\sum_{t=1}^{N}\log[1+\lambda'\{g_{tn}(\theta)-\mu\}]+\frac{\tau}{b}\mu'\Psi_n^{-1}(\hat{\theta}_n)\mu\right),$$

where $g_{tn}(\theta)=\{G_n'(\hat{\theta}_n)W_nG_n(\hat{\theta}_n)\}^{-1}G_n'(\hat{\theta}_n)W_n f_{tn}(\theta)$ is the transformed smooth moment condition. Following the arguments above, it can be shown that $\text{elr}_{\text{pbel}}(\theta_0)$ admits the same pivotal limit,

$$\mathrm{elr}_{\mathrm{pbel}}(\theta_0) \xrightarrow{\mathrm{d}} U_{\mathrm{pbel},p}(b) = \min_{\tilde{\mu}\in\mathbb{R}^p} \left\{ \frac{2}{b} \max_{\tilde{\lambda}\in\mathbb{R}^p} \int_0^{1-b} \log\left[1+\tilde{\lambda}'\left\{\frac{D_p(r;b)}{b} - \tilde{\mu}\right\}\right]\mathrm{d}r + \frac{c^*}{b}\tilde{\mu}'\Phi_p^{-1}\tilde{\mu}\right\}. \tag{22}$$

4.2. Penalized expansive blockwise empirical likelihood

As demonstrated in Section 3.2, EBEL suffers seriously from the convex hull violation problem in finite samples. To deal with the convex hull condition, we introduce penalized EBEL (PEBEL) which is shown to provide significant finite sample improvement in Section 5. We describe the idea for exactly identified moment condition models. The results below can be extended to more general cases following the discussion in remark 3. Recall that

$$\tilde{f}_{tn}(\theta) = \frac{\omega(t/n)}{n} \sum_{j=1}^{t} f(z_j, \theta)$$

for $t = 1, 2, \ldots, n$. We consider the PEBEL ratio test statistic which is defined as

$$\mathrm{elr}_{\mathrm{pebel}}(\theta) = -\frac{1}{n}\log\{n^n \mathcal{L}_{\mathrm{pebel},n}(\theta)\} = \min_{\pi\in\mathfrak{F}_n}\left[-\frac{1}{n}\sum_{t=1}^{n}\log(n\pi_t) + \tau\,\delta_n\{\tilde{\mu}_\pi(\theta)\}\right], \qquad \tau = c^* n,$$

where

$$\mathcal{L}_{\mathrm{pebel},n}(\theta) = \max_{\pi\in\mathfrak{F}_n} \prod_{t=1}^{n} \pi_t \exp[-n\tau\,\delta_n\{\tilde{\mu}_\pi(\theta)\}], \tag{23}$$

and $\tilde{\mu}_\pi(\theta) = \Sigma_{t=1}^n \pi_t \tilde{f}_{tn}(\theta)$ with $\pi = (\pi_1, \ldots, \pi_n) \in \mathfrak{F}_n$. Following similar derivations in the proof of proposition 3, we deduce that

$$\begin{aligned}\mathrm{elr}_{\mathrm{pebel}}(\theta) &= \min_{\mu\in\tilde{\mathcal{H}}_n(\theta)}\left(\frac{1}{n}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{n}\log[1+\lambda'\{\tilde{f}_{tn}(\theta)-\mu\}] + \tau\,\delta_n(\mu)\right) \\ &= \min_{\mu\in\mathbb{R}^k}\left(\frac{1}{n}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{n}\log[1+\lambda'\{\tilde{f}_{tn}(\theta)-\mu\}] + \tau\,\delta_n(\mu)\right),\end{aligned} \tag{24}$$

where $\tilde{\mathcal{H}}_n(\theta)$ denotes the convex hull of $\{\tilde{f}_{tn}(\theta)\}_{t=1}^n$. Under suitable assumptions (see Nordman *et al.* (2013)), it can be shown that

$$\mathrm{elr}_{\mathrm{pebel}}(\theta_0) \xrightarrow{\mathrm{d}} \min_{\tilde{\mu}\in\mathbb{R}^k}\left(\max_{\tilde{\lambda}\in\mathbb{R}^k}\int_0^1 \log[1+\tilde{\lambda}'\{\omega(r)\,W_k(r)-\tilde{\mu}\}]\mathrm{d}r + c^*\tilde{\mu}'\Phi_k^{-1}\tilde{\mu}\right). \tag{25}$$

Note that PEBEL is free of b, but again it requires the choice of a tuning parameter c^*. For large c^*, we have $\mu^* \approx 0$ and $\delta_n(\mu^*) \approx 0$ with μ^* being the minimizer in equation (24). Thus PEBEL behaves like EBEL when c^* is large. Following the discussion in Section 4.1, as c^* grows close to 0, μ^* approaches $\Sigma_{t=1}^n \tilde{f}_{tn}(\theta)/n$ which satisfies that

$$\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{n}\log\left[1+\lambda'\left\{\tilde{f}_{tn}(\theta) - \sum_{t=1}^{n}\frac{\tilde{f}_{tn}(\theta)}{n}\right\}\right] = 0.$$

Thus, for small c^*, the behaviour of the PEBEL ratio statistic evaluated at the true parameter value is closely related to the self-normalized score statistic given by

$$\tilde{\mathcal{S}}_n(\theta_0) = n\delta_n\left(\sum_{t=1}^{n}\frac{\tilde{f}_{tn}(\theta_0)}{n}\right) = n\left(\sum_{t=1}^{n}\frac{\tilde{f}_{tn}(\theta_0)}{n}\right)'\Psi_n^{-1}(\hat{\theta}_n)\left(\sum_{t=1}^{n}\frac{\tilde{f}_{tn}(\theta_0)}{n}\right). \tag{26}$$

Remark 4. To resolve the coverage upper bound problem, we may consider adjusted versions of BEL and EBEL, which retain the formulation of BEL and EBEL but add one or two pseudo-observations to the sample (see Chen *et al.* (2007) and Emerson and Owen (2009)). However, a direct extension to the current setting may not work because of temporal dependence in moment conditions. A possible strategy is to add a small fraction of artificial data points instead of one or two pseudo-observations, and to derive the limiting distributions under fixed b asymptotics. This approach also requires the choice of additional tuning parameters such as the fraction of points being added. The extended EL method (Tsao and Wu, 2013, 2014) is a nice alternative to the original EL, and it has been shown to enjoy the Bartlett correctability in the IID data case. Nevertheless, an extension to the time series setting seems very non-trivial. We shall investigate these alternative solutions in future research.

5. Numerical results

In this section, we conduct simulation studies to evaluate the finite sample performance of the penalization methods that were proposed in Section 4. We shall focus on the confidence region for the mean of univariate or multivariate time series. In the univariate case, we consider the AR(1) process $z_t = \rho z_{t-1} + \varepsilon_t$ with $\rho = -0.5, 0.2, 0.5, 0.8$, and the moving average MA(1) process $z_t = \theta \epsilon_{t-1} + \epsilon_t$ with $\theta = -0.5, 0.2, 0.5, 0.95$, where $\{\varepsilon_t\}$ and $\{\epsilon_t\}$ are two sequences of IID standard normal errors. In the multi-dimensional case (i.e. $k > 1$), we generate multivariate time series with each component being independent AR(1) or MA(1) processes. The sample sizes considered are $n = 100$ and $n = 400$. In the on-line supplementary material, we present additional simulation results for time series regression models, where the results are qualitatively similar to those for the mean.

5.1. Penalized blockwise empirical likelihood

To implement PBEL, we consider the self-normalization matrix Ψ_n (Shao, 2010), which is defined as

$$\Psi_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(1 - \left|\frac{i-j}{n}\right|\right) (z_i - \bar{z}_n)(z_j - \bar{z}_n)', \tag{27}$$

where $\hat{\theta}_n = \bar{z}_n = \Sigma_{j=1}^{n} z_j / n$. The tuning parameter c^* is chosen between 0.01 and 2. As pointed out in Section 4.1, the limiting distribution of PBEL under the fixed b asymptotics is pivotal and it can be approximated numerically. Table 1 in the on-line supplementary material summarizes the simulated critical values for the limiting distributions of BEL and PBEL. Selected simulation results are presented in Figs 2 and 3. In the univariate case, the performances of PBEL with $c^* = 2$ and BEL are generally comparable in terms of the coverage probability and interval width. PBEL with $c^* = 0.01$ delivers more accurate coverage compared with the two alternatives especially when the positive dependence is strong, although the corresponding interval width is slightly wider for relatively small b. This finding is presumably because the finite sample bounds for BEL do not deviate much from 1 for $k = 1$ and not quite large b (see Table 1). The simulation results for the MA models are quantitatively similar and thus are not presented here for brevity. In the case $k = 2$, PBEL tends to provide better coverage uniformly over b compared with BEL (when the dependence is positive). The improvement becomes more significant as the block size grows. Also PBEL with $c^* = 0.01$ delivers the most accurate coverage in most cases. Unreported numerical results show that, for $k = 1, 2$ and c^* between 0.01 and 2, the performance of PBEL is generally between the two cases reported here. To assess the effect of dimensionality, we

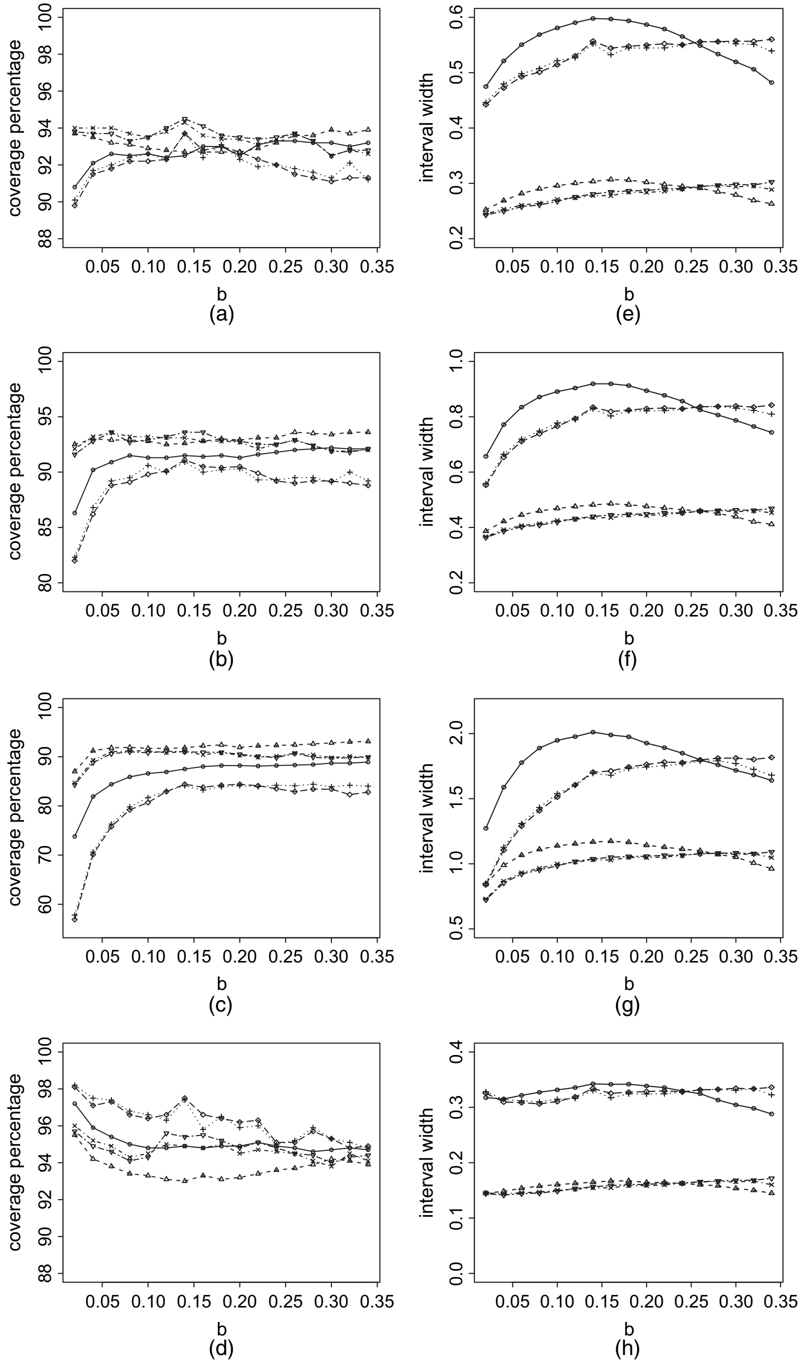

Fig. 2. (a)–(d) Coverage probabilities and (e)–(h) interval widths for the mean delivered by PBEL with $Q(r,s)=(1-|r-s|)\mathbf{I}(|r-s|\leqslant 1)$ and BEL, where $k=1$ (the nominal level is 95% and the number of Monte Carlo replications is 1000) (○, PBEL, fixed b asymptotics, $c^*=0.01$, $n=100$; △, PBEL, fixed b asymptotics, $c^*=0.01$, $n=400$; +, PBEL, fixed b asymptotics, $c^*=2$, $n=100$; ×, PBEL, fixed b asymptotics, $c^*=2$, $n=400$; ◇, BEL, fixed b asymptotics, $n=100$; ∇, BEL, fixed b asymptotics, $n=400$) : (a), (e) AR(1), $\rho=0.2$; (b), (f) AR(1), $\rho=0.5$; (c), (g) AR(1), $\rho=0.8$; (d), (h) AR(1), $\rho=-0.5$

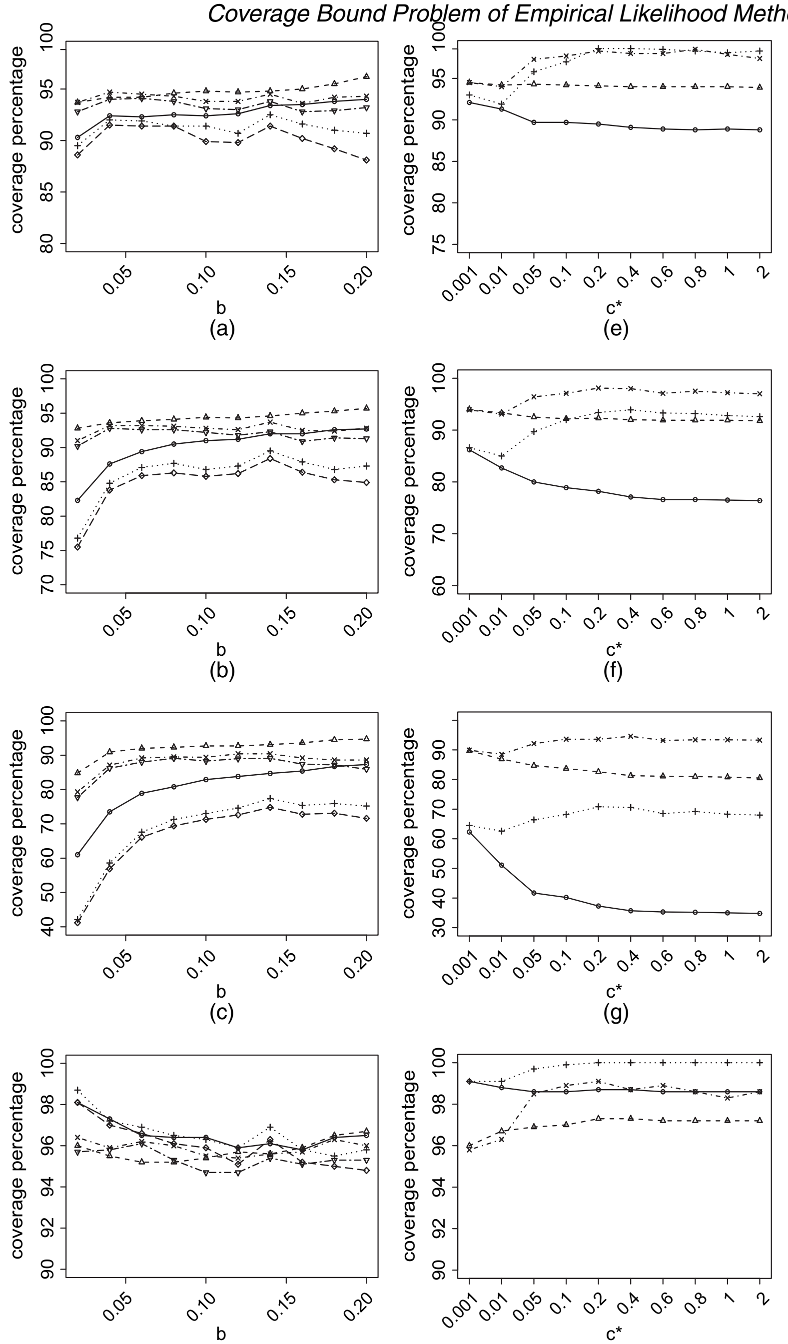

Fig. 3. Coverage probabilities for the mean delivered by PBEL with $Q(r,s)=(1-|r-s|)\mathbf{I}(|r-s|\leqslant 1)$, and BEL, where (a)–(d) $k = 2$ ($\bigcirc$, PBEL, fixed b asymptotics, $c^* = 0.01, n = 100$; Δ, PBEL, fixed b asymptotics, $c^* = 0.01, n = 400$; $+$, PBEL, fixed b asymptotics, $c^* = 2, n = 100$; $\times$, PBEL, fixed b asymptotics, $c^* = 2, n = 400$; $\Diamond$, BEL, fixed b asymptotics, $n = 100$; ∇, BEL, fixed b asymptotics, $n = 400$) and (e)–(h) $k = 5$ ($\bigcirc$, PBEL, fixed b asymptotics, $b^* = 0.05, n = 100$; Δ, PBEL, fixed b asymptotics, $c^* = 0.05, n = 400$); $+$, PBEL, fixed b asymptotics, $b = 0.1, n = 100$; $\times$, PBEL, fixed b asymptotics, $b = 0.1, n = 400$) (the nominal level is 95% and the number of Monte Carlo replications is 1000): (a), (e) AR(1), $\rho = 0.2$; (b), (f) AR(1), $\rho = 0.5$; (c), (g) AR(1), $\rho = 0.8$; (d), (h) AR(1), $\rho = -0.5$

Table 3. Coverage probabilities for the mean delivered by BEL†

n	b	*Coverage probabilities (%) for the following values of* ρ:			
		0.2	*0.5*	*0.8*	*−0.5*
100	0.05	88.5 (97.8)	76.1 (86.7)	34.3 (24.1)	98.6 (99.9)
100	0.10	84.7 (37.3)	74.6 (17.6)	42.4 (1.4)	97.5 (80.9)
400	0.05	93.8 (99.9)	91.8 (99.6)	80.3 (92.6)	97.2 (100.0)
400	0.10	93.0 (68.0)	90.2 (55.8)	78.7 (26.1)	96.5 (87.2)

†The data are generated from the VAR(1) process with the coefficient matrix being ρI_k for $k=5$ or $k=10$. The numbers in parentheses are the coverage upper bounds for the case $k=10$.

present the coverage probabilities for PBEL with $b=0.05$ and $b=0.1$, and various c^* when $k=5$ (see Figs 3(e)–3(h)). Along with Table 3, which also shows the coverage bound for the case $k=10$, we see that PBEL with suitable c^* offers improvement over the unpenalized counterpart. The coverage upper bound problem clearly shows up for BEL especially when the dependence is strong and dimension k is large. We also note that the choice of c^* (that delivers the most accurate coverage) is delicate in this case as it depends on b, the sample size n and the underlying dependence structure. Overall, the finite sample performance of PBEL is satisfactory in terms of delivering better coverage (especially when the bound is substantially below 1) compared with BEL under fixed b asymptotics.

5.2. *Penalized expansive blockwise empirical likelihood*

We implement PEBEL with Ψ_n being the self-normalization matrix in equation (27) and various choices of c^*. The simulated critical values for PEBEL are summarized in Table 2 in the on-line supplementary material. We present the coverage probabilities and interval widths for unweighted EBEL and PEBEL (i.e. $w(t)=1$) in Figs 4 and 5. Compared with EBEL, PEBEL significantly improves coverage probabilities in all the cases considered here. Figs 4(e)–4(h) suggest that PEBEL can also deliver smaller interval widths for the range of c^* being considered. In the univariate case, the choice of small c^* seems to provide both better coverage and shorter interval width. In the case of $k=2$, a relatively large c^* tends to provide good coverage as well, and the performance of PEBEL is not affected much by the choice of c^*. It is worth noting that, when $k=5$, the performance of PEBEL deteriorates for $c^*=2$, which, along with the above findings in the cases $k=1$ and $k=2$, suggests that the optimal c^* that delivers the most accurate coverage and the sensitivity of the coverage with respect to c^* can very much depend on the underlying dimensionality k. To sum up, the numerical results demonstrate the usefulness of PEBEL as it provides significant improvement over EBEL provided that c^* is suitably chosen.

In view of Figs 5(e)–5(h), there seems to be an optimal c^* in terms of delivering the most accurate coverage when $\rho=0.5$ and $\rho=0.8$. Below we present a simple block-bootstrap-based method for choosing the tuning parameter c^*. Suppose that $n=d_n l_n$ where $d_n, l_n \in \mathbb{Z}$. Conditionally on the sample $\{z_l\}_{l=1}^n$, we let $M_1, \ldots, M_{l_n}$ be IID uniform random variables on $\{0, \ldots, l_n-1\}$ and define $z^*_{(j-1)d_n+i} = z_{M_j d_n+i}$ with $1 \leqslant j \leqslant l_n$ and $1 \leqslant i \leqslant d_n$. In other words, $\{z^*_t\}_{t=1}^n$ is a non-overlapping block bootstrap sample with block size d_n. For each c^*, we can compute the times that the sample mean $\bar{z}_n$ is contained in the confidence region constructed on the basis of the bootstrap sample $\{z^*_l\}_{l=1}^n$ and then compute the empirical coverage probabilities based on B

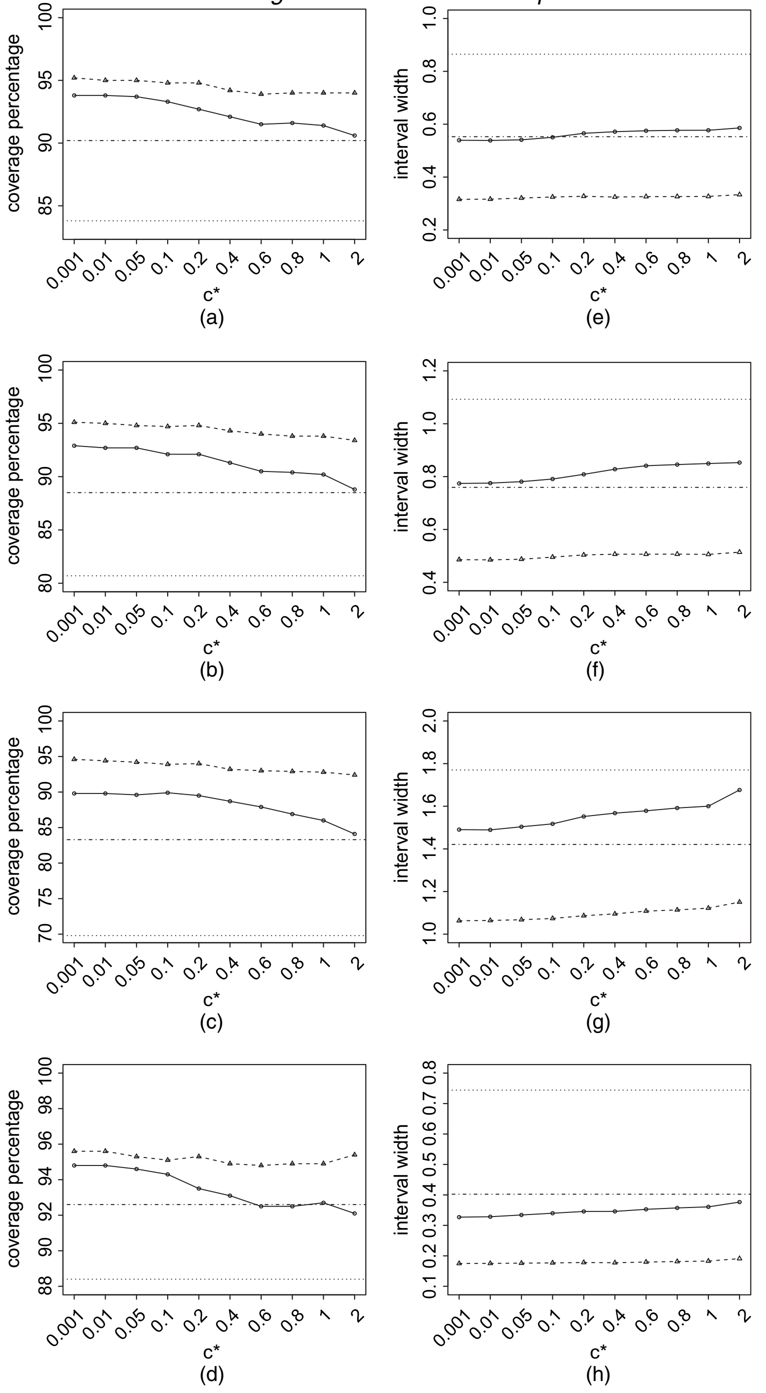

Fig. 4. (a)–(d) Coverage probabilities and (e)–(h) interval widths for the mean delivered by PEBEL with various c^* and $Q(r,s) = (1-|r-s|)\mathbf{I}(|r-s| \leqslant 1)$ and EBEL, where $k = 1$ (the nominal level is 95% and the number of Monte Carlo replications is 1000) (○, PEBEL, fixed b asymptotics, $n = 100$; Δ, PEBEL, fixed b asymptotics, $n = 400$; · · · · · ·, EBEL, fixed b asymptotics, $n = 100$; -·-·-·-, EBEL, fixed b asymptotics, $n = 400$): (a), (e) AR(1), $\rho = 0.2$; (b), (f) AR(1), $\rho = 0.5$; (c), (g) AR(1), $\rho = 0.8$; (d), (h) AR(1) $\rho = -0.5$

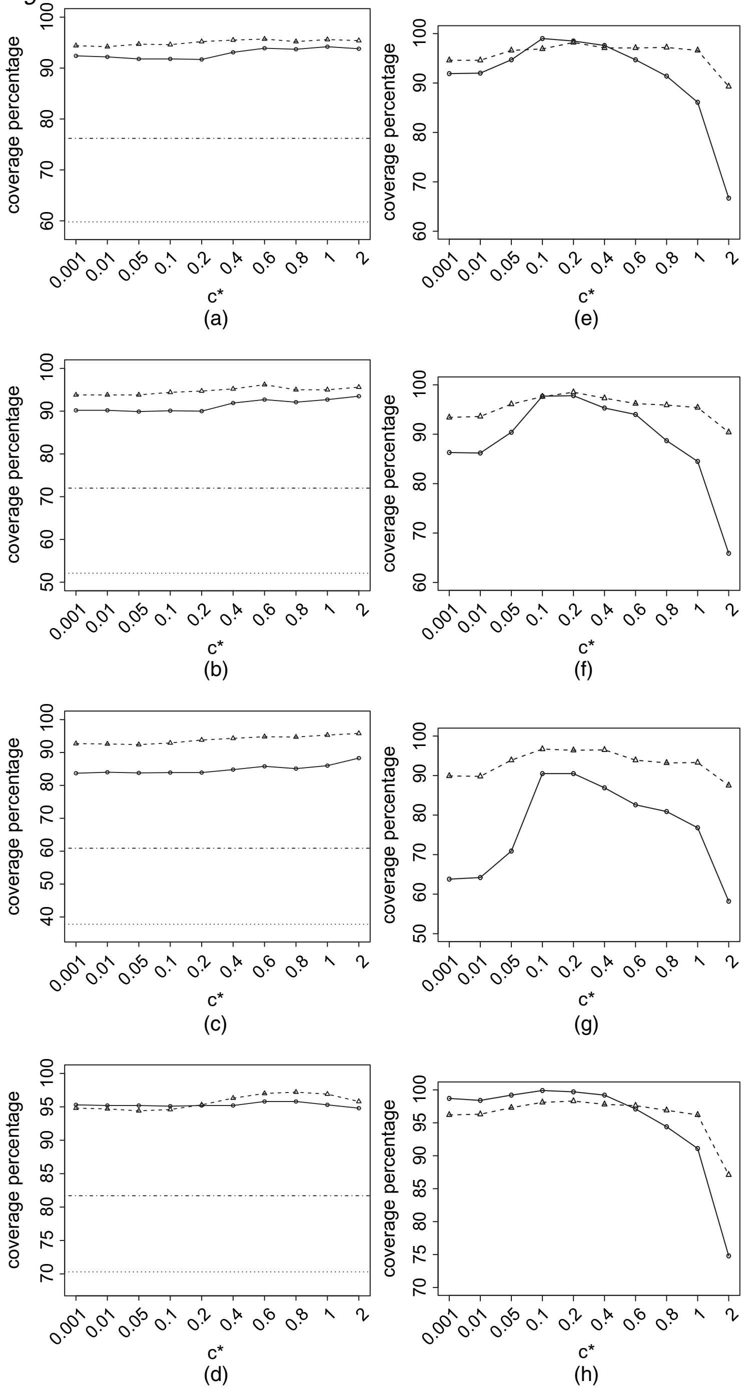

Fig. 5. Coverage probabilities for the mean delivered by PEBEL with various $c*$ and $Q(r,s) = (1-|r-s|)\mathbf{I}(|r-s| \leqslant 1)$ and EBEL, where (a)–(d) $k=2$ and (e)–(h) $k=5$ (the nominal level is 95% and the number of Monte Carlo replications is 1000) ($\bigcirc$, PEBEL, fixed b asymptotics, $n=100$; Δ, PEBEL, fixed b asymptotics, $n=400$; $\cdots\cdots$, EBEL, fixed$b$ asymptotics, $n=100$; -·-·-·-, EBEL, fixed b asymptotics, $n=400$): (a), (e) AR(1), $\rho=0.2$; (b), (f) AR(1), $\rho=0.5$; (c), (g) AR(1), $\rho=0.8$; (d), (h) AR(1), $\rho=-0.5$

bootstrap samples. This is based on the notion that $\bar{z}_n$ is the true mean for the bootstrap sample conditionally on the data and the c^* that delivers the most accurate coverage for bootstrap samples is an estimate of the optimal c^* for the original series. Specifically, we consider the case $n=100$ and $d_n=5$, and set $B=100$ and the number of Monte Carlo replications to be 100 to see whether this scheme works well. For the VAR(1) model with coefficient matrix $0.5I_5$, the coverage probability based on the above tuning parameter selection procedure is 98% and the most frequently chosen c^* is 0.4 (33%). When the coefficient matrix is $0.8I_5$, the corresponding coverage probability is 90% and the most frequently chosen c^* is 0.1 (51%), which is identical to the empirically optimal c^* as seen from Fig. 5(g). Hence the method of choosing the tuning parameter c^* seems to perform quite well. We shall leave a more detailed examination of this bootstrap-based tuning parameter selection method to a separate work.

6. Records of hemispheric temperatures

To illustrate the finite sample performance further, we apply the penalization methods (PBEL and PEBEL) and their unpenalized counterparts to the so-called hemisphere temperature anomaly time series HadCRUT3v that is available from the Climate Research Unit (UK). The data, consisting of adjusted monthly temperature averages from 1850 to 2010, combine the land and marine gridded temperature anomalies, after correcting for non-climatic (e.g. instrumental) errors and adjusting the variance (see for example Rayner *et al.* (2006) and Jones *et al.* (2011) and references therein for more details about the data set). Following Kim *et al.* (2013), we consider the annual average anomalies for months December–January–February and June–July–August over the years 1850–2009 in both northern and southern hemispheres; the December–February values are means of average temperature anomalies of December of the

Table 4. Confidence intervals CI and coverage probabilities for the hemispheric temperatures records†

Method	*m*	c^*	*Results for northern hemisphere*			*Results for southern hemisphere*		
			CI	CP_4 *(%)*	CP_8 *(%)*	*CI*	CP_4 *(%)*	CP_8 *(%)*
PBEL	4	0.05	[1.012, 1.336]	90.6	82.2	[0.862, 0.949]	93.5	91.3
	4	0.1	[0.961, 1.431]	98.5	95.6	[0.833, 0.977]	99.0	99.4
	4	0.2	[0.983, 1.404]	98.0	94.5	[0.837, 0.973]	98.7	99.1
	4	1	[1.028, 1.330]	93.2	86.0	[0.854, 0.957]	95.5	95.1
	4	2	[1.031, 1.326]	93.1	85.5	[0.854, 0.956]	95.1	94.8
BEL	4	—	[1.033, 1.323]	92.7	85.0	[0.855, 0.955]	95.0	94.8
PBEL	8	0.05	[0.983, 1.417]	91.5	84.8	[0.858, 0.948]	93.7	92.5
	8	0.1	[1.022, 1.377]	90.3	83.0	[0.860, 0.946]	92.1	89.2
	8	0.2	[0.974, 1.481]	98.1	95.1	[0.838, 0.976]	98.6	99.1
	8	1	[1.029, 1.383]	93.9	87.5	[0.853, 0.956]	95.5	94.6
	8	2	[1.034, 1.374]	93.0	86.2	[0.855, 0.955]	95.2	94.3
BEL	8	—	[1.039, 1.365]	92.3	84.9	[0.856, 0.953]	94.7	93.6
PEBEL	—	0.05	[0.738, 1.825]	93.7	90.5	[0.868, 0.952]	95.2	95.1
	—	0.1	[0.742, 1.830]	93.6	90.6	[0.869, 0.955]	95.2	95.3
	—	0.2	[0.746, 1.832]	93.6	90.8	[0.871, 0.962]	95.0	95.5
	—	1	[0.800, 1.786]	92.8	90.5	[0.876, 0.988]	94.6	94.8
	—	2	[0.831, 1.760]	92.3	89.7	[0.877, 1.002]	94.6	94.6
EBEL	—	—	[1.059, 1.538]	87.0	84.4	[0.880, 1.133]	89.6	93.3

†The columns CP_4 and CP_8 correspond to the coverage probabilities based on the bootstrap samples with block size $d_n=4$ and $d_n=8$ respectively.

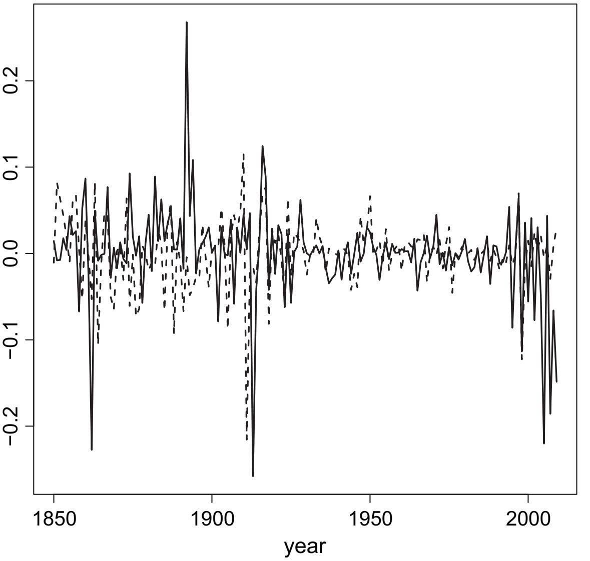

Fig. 6. Plot of $f(Z_t,\hat{\theta}) = X_t(Y_t - X_t\hat{\theta})$ with $Z_t = (X_t, Y_t)$: ———, northern hemisphere; – – –, southern hemisphere

current year and January and February of the next year. We consider fitting a simple linear regression model

$$Y_t = X_t\theta + \epsilon_t, \qquad t = 1, \ldots, 160,$$

for predicting the December–February temperature anomalies $\{Y_t\}$ from the June–August anomalies $\{X_t\}$. Define the estimating equation $f(Z_t,\theta) = X_t(Y_t - X_t\theta)$ with $Z_t = (X_t, Y_t)'$. If the model is correctly specified and $\mathbb{E}(X_t\epsilon_t) = 0$, then $\mathbb{E}\{f(Z_t,\theta_0)\} = 0$ with θ_0 being the true parameter. We apply the penalization methods and their unpenalized versions to compute 95% confidence intervals for θ_0 (Table 4). Since θ_0 is unknown to us, it makes a fair comparison of various EL methods difficult as we do not really know whether the constructed confidence interval covers θ_0 or not. For this, we propose to apply the EL methods to non-overlapping bootstrap samples which mimic the dependence structure of the original time series, and to make a fair comparison. In particular, let $n = d_n l_n$ where $n = 160$ and $d_n, l_n \in \mathbb{Z}$. Let $M_1, \ldots, M_{l_n}$ be IID uniform random variables on $\{0, \ldots, l_n - 1\}$ and let $(X^*_{(j-1)d_n+i}, Y^*_{(j-1)d_n+i}) = (X_{M_j d_n+i}, Y_{M_j d_n+i})$ with $1 \leqslant j \leqslant l_n$ and $1 \leqslant i \leqslant d_n$. It is not difficult to verify that $\mathbb{E}^*\{\Sigma_{t=1}^n X_t^*(Y_t^* - X_t^*\hat{\theta})\} = 0$, where $\hat{\theta} = \Sigma_{t=1}^{160} X_t Y_t / \Sigma_{t=1}^{160} X_t^2$ is the ordinary least square estimator and $\mathbb{E}^*$ denotes the expectation conditional on the sample $\{X_t, Y_t\}_{t=1}^{160}$. Thus, for the bootstrap sample, the true θ is $\hat{\theta}$ conditional on the data and we can compute the empirical coverage probabilities for $\hat{\theta}$ based on 1000 bootstrap samples, where the block size d_n is chosen to be 4 or 8. It is seen from Table 4 that, for the northern hemisphere, undercoverage occurs for BEL, whereas PBEL with suitable choice of c^* can deliver better coverage. In such cases, the corresponding interval widths delivered by PBEL based on the original data are wider. For the southern hemisphere, BEL provides quite accurate coverage and PBEL with $c^* = 1, 2$ is comparable with BEL in terms of the coverage accuracy based on the bootstrap samples and the confidence intervals based on the original data. In view of Table 4, PEBEL provides better coverage compared with the unpenalized version for all cases that are considered here. For the northern hemispheric temperature anomalies, the PEBEL-based confidence intervals are wider whereas, for the southern anomalies, PEBEL delivers shorter interval widths. On the basis of 1000 bootstrap samples, we can compute the percentages of convex hull violation for EBEL. For the northern hemisphere, the upper bounds are 90.1% and 88.2% for $d_n = 4$ and $d_n = 8$ respectively; for the southern hemisphere, the upper bounds are 93.2% and 96.5%, showing a serious deficiency of the EBEL method. It is worth pointing out that the penalized methods generally deliver wider interval widths for the northern hemispheric data (in particular, PEBEL seems quite conservative in this case). This finding might be due to the following facts. First, the June–August temperature anomalies appear to be worse predictors for the December–February anomalies in the northern hemisphere (with adjusted R^2 0.6234) than in the southern hemisphere (with adjusted R^2 0.8768). Second, the plot of $f(Z_t, \hat{\theta})$ in the northern hemisphere (Fig. 6) tends to exhibit certain non-stationarity features (e.g. in the second-order property), which may pose difficulty in constructing a confidence interval for θ.

7. Conclusion

In this paper, we studied the upper bounds on the coverage probabilities of the BEL- and EBEL-based confidence regions via theory and simulations. Our theoretical results, which are derived for the pivotal limit of the BEL ratio obtained under fixed b asymptotics, suggest that the large sample coverage upper bound for BEL is strictly less than 1 for any $b \in (0, 1)$. This result is in sharp contrast with those corresponding to EL for independent moment conditions, where the large sample bound is always equal to 1 because of the χ^2-limit. By numerical simulations, we discover that the finite sample coverage bounds for both BEL and EBEL can be far below the nominal level in the cases when

(a) the dimension of moment condition k is moderate or high,
(b) the dependence of moment conditions is positively strong or
(c) b is large for BEL.

The deterioration in the coverage for the EBEL-based confidence region with respect to k is especially noticeable. These phenomena appear to be discovered for the first time for these two important EL methods in the time series context, which will hopefully lead to a new research direction on EL methods for dependent data.

To overcome the convex hull constraint and related undercoverage problem, we introduce the penalization-based BEL and EBEL methods, which drop the convex hull constraint and penalize the original EL by using the quadratic distance measure, and derive their limiting distributions under fixed b asymptotics. Interestingly, the penalization generates a new class of statistics which lies in between the empirical log-likelihood ratio statistic and the self-normalized score statistic through the choice of a tuning parameter c^*. Our simulation studies show that the penalization-based methods can outperform their unpenalized counterparts in terms of coverage accuracy especially when the coverage bound is below the nominal level. In addition, we propose a method of choosing the tuning parameter and demonstrate its effectiveness through a simulation example. It is worth mentioning that our techniques (i.e. fixed b asymptotics and penalization) are expected to be extendable to other variants of EL methods for time series or spatial data, such as tapered blockwise EL (Nordman, 2009) and spatial EL (Nordman and Caragea, 2008). We shall leave these for future investigation.

Acknowledgements

The authors thank the Joint Editor, the Associate Editor and two reviewers for their constructive comments, which substantially improved the paper. Shao's research is supported in part by National Science Foundation grant DMS-11-04545.

References

Bartolucci, F. (2007) A penalized version of the empirical likelihood ratio for the population mean. *Statist. Probab. Lett.*, **77**, 104–110.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd edn. New York: Wiley.

Chang, J., Chen, S. and Chen, X. (2015) High dimensional empirical likelihood for generalized estimating equations with dependent data. *J. Econmetr.*, **185**, 283–304.

Chen, S. X. and Cui, H. J. (2006) On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, **93**, 215–220.

Chen, J. and Huang, Y. (2012) Finite-sample properties of the adjusted empirical likelihood. *J. Nonparam. Statist.*, **25**, 147–159.

Chen, J., Variyath, A. M. and Abraham, B. (2008) Adjusted empirical likelihood and its properties. *J. Computnl Graph. Statist.*, **17**, 426–443.

Diciccio, T. J., Hall, P. and Romano, J. P. (1991) Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053–1061.

Emerson, S. C. and Owen, A. B. (2009) Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.*, **3**, 1161–1192.

Gonçalves, S. and Vogelsang, T. J. (2011) Block bootstrap HAC robust tests: the sophistication of the naive bootstrap. *Econmetr. Theor.*, **27**, 745–791.

Jones, P. D., Parker, D. E., Osborn, T. J. and Briffa, K. R. (2011) Global and hemispheric temperature anomalies-land and marine instrumental records. In *Trends: a Compendium of Data on Global Change*. Oak Ridge: Oak Ridge National Laboratory.

Karlin, S. and Mcgregor, J. (1959) Coincidence probabilities. *Pacif. J. Math.*, **9**, 1141–1164.

Kiefer, N. M. and Vogelsang, T. J. (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econmetr. Theor.*, **21**, 1130–1164.

Kim, Y. M., Lahiri, S. N. and Nordman, D. J. (2013) A progressive block empirical likelihood method for time series. *J. Am. Statist. Ass.*, **108**, 1506–1516.

Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084–2102.

Kitamura, Y. (2006) Empirical likelihood methods in econometrics: theory and practice. *Cowles Foundation Discussion Paper 1569*. Department of Economics, Yale University, New Haven.

Kitamura, Y., Santos, A. and Shaikh, A. (2013) On the asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, **80**, 413–423.

Lahiri, S. N. and Mukhopadhyay, S. (2012) A penalized empirical likelihood method in high dimensions. *Ann. Statist.*, **40**, 2511–2540.

Liu, Y. and Chen, J. (2010) Adjusted empirical likelihood with high-order precision. *Ann. Statist.*, **38**, 1341–1362.

McElroy, T. and Politis, D. N. (2007) Computer-intensive rate estimation, diverging statistics, and scanning. *Ann. Statist.*, **35**, 1827–1848.

Nordman, D. J. (2009) Tapered empirical likelihood for time series data in time and frequency domains. *Biometrika*, **96**, 119–132.

Nordman, D. J., Bunzel, H. and Lahiri, S. N. (2013) A non-standard empirical likelihood for time series. *Ann. Statist.*, **4**, 3050–3073.

Nordman, D. J. and Caragea, P. C. (2008) Point and interval estimation of variogram models using spatial empirical likelihood. *J. Am. Statist. Ass.*, **103**, 350–361.

Nordman, D. J. and Lahiri, S. N. (2014) A review of empirical likelihood methods for time series. *J. Statist. Planng Inf.*, **155**, 1–18.

Otsu, T. (2007) Penalized empirical likelihood estimation of semiparametric models. *J. Multiv. Anal.*, **98**, 1923–1954.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. B. (1990) Empirical likelihood confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. B. (2001) *Empirical Likelihood*. London: Chapman and Hall.

Phillips, P. C. B. (1987) Time series regression with unit roots. *Econometrica*, **55**, 277–301.

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T. and Tett, S. F. B. (2006) Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *J. Clim.*, **19**, 446–469.

Rockafellar, T. R. (1970) *Convex Analysis*. Princeton: Princeton University Press.

Shao, X. (2010) A self-normalized approach to confidence interval construction in time series. *J. R. Statist. Soc.* B, **72**, 343–366.

Shepp, L. A. (1971) First passage time for a particular Gaussian process. *Ann. Math. Statist.*, **42**, 946–951.

Sun, Y. (2013) Fixed-smoothing asymptotics in a two-step GMM framework. *Working Paper*. Department of Economics, University of California at San Diego, San Diego.

Tang, C. Y. and Leng, C. (2010) Penalized high-dimensional empirical likelihood. *Biometrika*, **97**, 905–919.

Tsao, M. (2004) Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.*, **32**, 1215–1221.

Tsao, M. and Wu, F. (2013) Empirical likelihood on the full parameter space. *Ann. Statist.*, **41**, 2176–2196.

Tsao, M. and Wu, F. (2014) Extended empirical likelihood for estimating equations. *Biometrika*, **101**, 703–710.

Zhang, X. and Shao, X. (2014) Fixed-*b* asymptotics for blockwise empirical likelihood. *Statist. Sin.*, **24**, 1179–1194.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplement to “On the coverage bound problem of empirical likelihood methods for time series”’.