Higher Order Properties of Block Empirical Likelihood methods*

S. N. Lahiri
North Carolina State University

OUTLINE

- The Block Empirical Likelihood method
- Existing results on its higher order properties
- A new result on the block EL
- Tapered Block EL
- Higher order properties of the TBEL

*Joint work with Arindam Chatterjee and Dan Nordman.

The Empirical Likelihood Method - iid r.v.s

Let $X_1, \ldots, X_n$ be a collection of iid random variables with mean μ. For simplicity, we first describe the EL method for the mean parameter μ in the iid case.

- Assign probabilities $\{p_i\}_{i=1}^n$ to the observations $\{X_i : i = 1, \ldots, n\}$ and define the *EL function* for μ as

$$L_n(\mu) = \sup \left\{ \prod_{i=1}^n p_i : p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu \right\}$$

- Without the "mean constraint", the product $\prod_{i=1}^n p_i$ is maximized when $p_1 = \ldots = p_n = n^{-1}$.
- Thus, we define the *EL Ratio for the mean* at μ as

$$R_n^{\mathrm{IID}}(\mu) \ = \ \frac{L_n(\mu)}{n^{-n}}.$$

The Empirical Likelihood Method - iid r.v.s

- An EL confidence interval (CI) for μ is given by

$$\{\mu : R_n(\mu) \geq A\},$$

where $A > 0$ is chosen to obtain a desired confidence level [cf. Owen (1988, 1990), Hall and La Scala (1990)].

- Owen (1988, 1990) showed that under mild conditions,

$$\Lambda_n \equiv -2 \log R_n(\mu_0) \to^d \chi^2(1) \quad \text{as} \quad n \to \infty$$

where μ_0 is the true mean.

- Note that
  - the EL ratio statistic is asymptotically pivotal,
  - no explicit studentization is needed.

Hence EL CIs can be constructed more easily than CIs based on a t-statistic (i.e., studentization).

The Empirical Likelihood Method - Rates

- DiCCicio, Hall and Romano (1991) ([DHR]) showed that for the "smooth function model",

$$\sup_{x\in(0,\infty)} \Big| P\Big(\Lambda_n \le x\Big) - P\Big(\chi^2(1) \le x\Big)\Big| = O(n^{-1}).$$

- [DHR] also showed that the statistic Λ_n admits Bartlett correction: i.e.,

$$\sup_{x\in(0,\infty)} \Big| P\Big(\Lambda_n^{\mathrm{BART}} \le x\Big) - P\Big(\chi^2(1) \le x\Big)\Big| = O(n^{-2}).$$

  where $\Lambda_n^{Bart} = \Lambda_n(1 + n^{-1}\hat{\phi}_n)$ for some suitable ϕ.
- The form of ϕ is rather complicated. For the mean parameter,

$$\phi = 2^{-1}\mu_4\mu_2^{-2} - 3^{-1}\mu_3^2\mu_2^{-3}$$

  where $\mu_j = E(X_1 - \mu)^j$, $j \ge 2$.
- Question:What happens when the X_i's are dependent?

The Block Empirical Likelihood Method (Kitamura, 1997)

For simplicity, we consider the (Maximum Overlapping) Block Empirical Likelihood for the mean parameter $E(X_1) = \mu$.

- Let $1 \leq \ell \leq n$, $\mathcal{B}(i, \ell) = (X_i, \ldots, X_{i+\ell-1})$, $1 \leq i \leq N \equiv n - \ell + 1$. Let

$$\bar{X}_{i,\ell} = \ell^{-1} \sum_{j=i}^{i+\ell-1} X_j$$

  denote the sample mean of the ℓ variables in $\mathcal{B}(i, \ell)$, $1 \leq i \leq N$.
- Assign probabilities $\{p_i\}_{i=1}^N$ to each block sample mean $\{\bar{X}_{i,\ell}\}_{i=1}^N$ and define the blockwise EL function for μ as

$$L_n^{[\mathrm{K}]}(\mu) = \sup \left\{ \prod_{i=1}^N p_i : p_i > 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \bar{X}_{i,\ell} = \mu \right\}$$

The Block Empirical Likelihood Method [contd.]

- The *block empirical likelihood ratio for the mean* μ is then given by

$$R_n^{[K]}(\mu) = \frac{L_n^{[K]}(\mu)}{N^N}$$

and a confidence interval for μ is $\{\mu : R_n(\mu) \geq A\}$, where $A > 0$.

- Kitamura(1997) proved the following version of Wilk's Theorem:

Theorem (Kitamura,1997)

Under some suitable regularity conditions, for $1 \ll \ell \ll n^{1/2}$,

$$\Lambda_n^{[K]} \equiv -2a_n \log R_n^{[K]}(\mu_0) \to^d \quad \chi^2(1) \quad \textit{as} \quad n \to \infty$$

where $a_n = \ell^{-1}$ *and where* μ_0 *is the true mean.*

Block Empirical Likelihood

- Note that the BEL does NOT have the automatic scaling as in the iid EL. But it is asymptotically pivotal.
- The factor a_n in Kitamura's Theorem adjusts for the strong dependence among the neighboring ℓ blocks.

- QUESTION 1: What is the rate of convergence in Kitamura's Theorem?
- QUESTION 2: Is the Block Empirical Likelihood Ratio Statistic Bartlett correctable?
- It is clear that the rate depends on the block size ℓ.
- QUESTION 3: What is the optimal block size?

Block Empirical Likelihood - Rates

- Kitamura (1997) claims (cf. (4.5a), page 2095) that with $\ell \sim Cn^{1/3}$,

$$\sup_{z \in (0,\infty)} \left| P\Big(\Lambda_n^{[\mathrm{K}]} \le z\Big) - P\Big(\chi^2(1) \le z\Big) \right| = O(n^{-2/3}).$$

- He also claims that, with $\ell \sim Cn^{1/3}$, the BEL ratio statistic is Bartlett correctable and (cf. (4.5b), page 2095)

$$\sup_{z \in (0,\infty)} \left| P\Big(\Lambda_n^{[\mathrm{K}]}[1 - N^{-1}\varphi_n] \le z\Big) - P\Big(\chi^2(1) \le z\Big) \right| = O(n^{-5/6}),$$

where φ_n is defined in analogy to the iid case.

A new result

- Let us consider the first of the two rate results. Validity of this is addressed in the following theorem:

Theorem (Chatterjee, Lahiri and Nordman, 2016)

Suppose that the regularity conditions for valid Edgeworth expansions for the partial sums (cf. Götze and Hipp(1983), Lahiri (2007, 2009, 2010)) hold. If, in addition, $\sum_{k=1}^{\infty} k Cov(X_1, , x_{1+k}) \neq 0$*, then*

$$\liminf_{n\to\infty} \quad n^{1/2}\left[\inf_{1\le\ell\le n}\sup_{z\in(0,\infty)}\left|P\left(\Lambda_n^{[K]}\le z\right)-P\left(\chi^2(1)\le z\right)\right|\right]\in(0,\infty).$$

- Thus, under the conditions of the theorem, Kitamura's result can not be true, as the rate $O(n^{-2/3})$ is not achievable for any block length.

Remarks

- Indeed, the optimal block length that achieves the best possible rate in the theorem is $\ell \sim Cn^{1/2}$, which is much larger than the optimal block size for variance estimation (viz. $\ell \sim Cn^{1/3}$, as used by Kitamura (1997)).
- This implies that the BEL CIs for μ has an error in coverage probablity of order $O(n^{-1/2})$, under the best possible choice of the block size.
- In contrast, a two-sided CI for μ based on Normal critical points can achieve an error rate of $O(n^{-1}[\log n]^C)$ for some $C \in (0, \infty)$ under exponential strong mixing.
- It can be shown that the Bartlett Correction result is also false.

Remarks

- Question: Why are the rates in Kitamura (1997)'s results WRONG?
- Although the stochastic approximation in the BEL has a structure similar to the independent case (barring the scaling factor a_n), the form of the Edgeworth expansion (EE) under dependence is different from that in the iid case.
- For iid r.v.s, under some standard regularity conditions,

$$P(\Lambda_n \leq x) = P(W \leq x) + [n^{-1}p_1(x) + n^{-2}p_2(x)]\phi(x) + o(n^{-2})$$

uniformly in $x \in (0, \infty)$, where $W \sim \chi^2(1)$.

- However, in the dependent case, the EE for the BEL ratio statistic is a superposition of three distinct series

$$\begin{aligned} P(\Lambda_n^{[K]} \leq x) &= P(W \leq x) + \Big[\ell^{-1}q_1(x) + b_n^{-1}q_2(x) \\ &\quad + n^{-1}q_3(x)\Big]\phi(x) + o(n^{-1} + \ell^{-1}) \end{aligned}$$

uniformly in $x \in (0, \infty)$, where $b_n = n/\ell$.

Related Block Empirical Likelihood Methods

- We consider two variants of the Maximum Overlapping Block EL - one simpler than the overlapping BEL, and the other more complex!!
- The simplest is perhaps the NON-overlapping BEL, based on the blocks

$$\mathcal{B}_i^{[\mathrm{NO}]} = (X_{(i-1)\ell+1}, \ldots, X_{i\ell}),\ \ i = 1, \ldots, b$$

  where $b = \lfloor n/\ell \rfloor$.
- Define the Non-overlapping BEL function for μ as

$$L_n^{[\mathrm{NO}]}(\mu) = \sup \left\{ \prod_{i=1}^{b} p_i : p_i > 0, \sum_{i=1}^{b} p_i = 1, \sum_{i=1}^{b} p_i \bar{X}_{i,\ell} = \mu \right\}$$

The Nonoverlapping Block Empirical Likelihood Method

- Define the *block empirical likelihood ratio for the mean* μ as before:

$$R_n^{[\mathrm{NO}]}(\mu) \;=\; \frac{L_n^{[\mathrm{NO}]}(\mu)}{b^b}.$$

- Then, the following version of Wilk's Theorem holds for the Non-overlapping BEL:

Theorem (Kitamura,1997)

Under some suitable regularity conditions, for $1 \ll \ell \ll n^{1/2}$,

$$\Lambda_n^{[NO]} \equiv -2\log R_n^{[K]}(\mu_0) \to^d \quad \chi^2(1) \quad as \quad n \to \infty.$$

where μ_0 *is the true mean.*

- Thus, compared to the overlapping case, the scaling $a_n = \ell^{-1}$ is not required!!!

Tapered Block Empirical Likelihood

- Nordaman (2007) proposed a version of the BEL, called the Tapered Block Empirical Likelihood (TBEL), where the blockwise sample means are replaced by a weighted average or tapered sum.
- It is an adaptation of an idea of Paparoditis and Politis (2001) on block bootstrap to the Block EL context.
- Let $1 \leq \ell \leq n$ and let $\mathcal{B}(i, \ell) = (X_i, \ldots, X_{i+\ell-1})$, $1 \leq i \leq N$, as in the BEL. Let $w_{1n}, \ldots, w_{\ell n} \in \boldsymbol{R}$ be nonrandom weights with

$$\sum_{k=1}^{\ell} w_{kn} \neq 0.$$

Define the weighted or tapered block sums

$$\mathcal{Y}_{i,l} = \frac{\sum_{k=1}^{\ell} w_{kn} X_{i+k-1}}{\sum_{k=1}^{\ell} w_{kn}}, \quad i = 1, \ldots, N.$$

TBEL-contd.

- The TBEL function for μ is now defined as

$$L_n^{[\mathrm{TB}]}(\mu) = \sup\left\{\prod_{i=1}^{N} p_i : p_i > 0, \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \mathcal{Y}_{i,\ell} = \mu\right\}.$$

- The TBEL ratio statistic is now given by $R_n^{[\mathrm{TB}]}(\mu) = N^N L_n^{[\mathrm{TB}]}(\mu)$.
- For studying the properties of $R_n^{[\mathrm{TB}]}(\mu)$, we restrict attention to weights

$$w_{kn} = w([k - .5]/n)$$

where $w : [0,1] \to \boldsymbol{R}$ is symmetric about $u = 1/2$ and positive in a neighborhood of $u = 1/2$.

- For example, $w(u) \equiv 1$ for $u \in [0,1]$ gives the BEL.

A result on TBEL

- Define $\Lambda_n^{[\mathrm{TB}]}(\mu) = -2c_n \log R_n^{[\mathrm{TB}]}(\mu)$ where $c_n = \frac{\sum_{k=1}^{\ell} w_{kn}^2}{[\sum_{k=1}^{\ell} w_{kn}]^2}$.
- Then, we have the following:

Theorem (Chatterjee, Lahiri and Nordman (2016))

Suppose that the self convolution $w * w$ is twice continuously differentiable in a neighborhood of $u = 0$ and $(w * w)''(0)$ is nonzero. Further, suppose that the regularity conditions similar to those in the last theorem hold. Then, for $\ell \sim Cn^{1/3}$,

$$\sup_{z \in (0,\infty)} \left| P\Big(\Lambda_n^{[TB]}(\mu_0) \le z\Big) - P\Big(\chi^2(1) \le z\Big) \right| = O(n^{-2/3}).$$

- Note the optimal block size for variance estimation by Tapered block bootstrap is $\ell \sim Cn^{1/5}$.

Higher order accuracy

- We consider two approaches to achieving higher order accurate results:
  - Edgeworth expansion (EE) based calibration
  - Block Bootstrap based calibration

Higher order accuracy based on EE

- We can use higher order terms from the EE to achieve a smaller error of coverage probability.
- Let $\chi^2(1;\alpha)$ denote the α quantile of $\chi^2(1)$ distribution and let $\hat{q}_n(\alpha)$ (respectively, $\hat{r}_n(\alpha)$) be a suitable estimator of the co-efficient of the $n^{-2/3}$-term (the n^{-1}-term) in the EE for $\Lambda_n^{\mathrm{TB}}(\mu_0)$ at $\chi^2(1;\alpha)$.
- Define the EE corrected TBEL CI for μ by

$$\mathcal{I}_n(\alpha) = \{\mu : \Lambda_n^{\mathrm{TB}}(\mu) \leq \hat{t}_n(\alpha)\}$$

  where $\hat{t}_n(\alpha) = \chi^2(1;\alpha) - n^{-2/3}\hat{q}_n(\alpha) - n^{-1}\hat{r}_n(\alpha)\}$.
- Then, under regularity conditions as in the last theorem,

$$P(\mu \in \mathcal{I}_n(\alpha)) = \alpha + O(n^{-7/6}[\log n]^C).$$

- This beats the 2-sided Normal CIs, but not as good as the Bartlett corrected Empirical Likelihood CIs in the iid case.

Higher order accuracy : Block Bootstrap Calibration

- Generate a MBB sample $X_1^*, \ldots, X_n^*$ based on resampling the overlapping blocks of size ℓ.
- Apply the Non-Overlapping version of the TBEL to the resampled values, to get Λ_n^*.
- Define the BB corrected TBEL CI for μ by

$$\mathcal{J}_n(\alpha) = \{\mu : \Lambda_n^{\text{TB-nl}}(\mu) \leq \tilde{t}_n(\alpha)\}$$

  where $\tilde{t}_n(\alpha)$ the α quantile of the conditional distribution of Λ_n^* given $X_1, \ldots, X_n$.
- Then, under some regularity conditions (similar to the last theorem), with $\ell \sim Cn^{1/3}$,

$$P(\mu \in \mathcal{J}_n(\alpha)) = \alpha + O(n^{-1}).$$

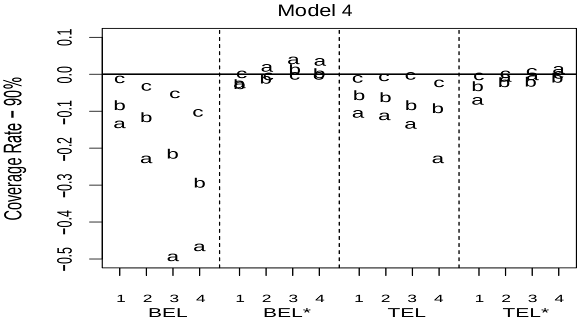
Some simulation results

- We consider 2 models:
  - $X_t = 0.4X_{t-1} + 0.2X_{t-2} + 0.1X_{t-3} + \epsilon_t + 0.2\epsilon_{t-1} + 0.3\epsilon_{t-2} + 0.2\epsilon_{t-3}$, $\epsilon_1 \sim \chi^2(1) - 1$.
  - $X_t = 0.6\sin(X_{t-1}) + 0.1\epsilon_t, \quad \epsilon_1 \sim \mathrm{N}(0,1)$.
- BEL and TBEL methods are based on usual chi-square calibrations as well as versions of BEL/TBEL, denoted as BEL* and TBEL*, based on block bootstrap calibrations.
- We consider block lengths $b = Cn^{1/2}$ for BEL/BEL* and $b = Cn^{1/3}$ for TBEL/TBEL* where $C = 1, 2, 3, 4$.
- These block lengths are of optimal order for each method, respectively.

Simulation results

Model 4		Chi-square Calibration				Bootstrap Calibration			
n	Type	1	2	3	4	1	2	3	4
80	BEL	78.4	68.9	42.3	45.0	87.1	91.7	93.7	93.2
	TBEL	81.2	80.6	78.2	68.9	84.8	89.0	89.2	91.2
200	BEL	83.6	80.3	70.5	62.4	87.3	88.6	91.4	90.3
	TBEL	86.3	85.7	83.6	82.7	87.8	88.8	89.0	90.0
1000	BEL	88.6	86.6	84.6	81.5	90.0	89.4	89.6	89.5
	TBEL	88.8	89.1	89.6	88.5	89.4	89.8	90.4	89.7
Model 7		Chi-square Calibration				Bootstrap Calibration			
n	Type	1	2	3	4	1	2	3	4
80	BEL	81.0	70.8	44.2	44.4	88.8	90.8	92.4	93.0
	TBEL	84.9	82.6	80.2	71.2	87.1	90.1	90.4	91.5
200	BEL	84.7	78.6	70.0	61.0	87.5	88.3	90.9	90.6
	TBEL	87.6	86.2	85.0	82.4	88.7	88.9	89.6	90.2
1000	BEL	86.7	85.6	83.2	79.5	88.2	89.0	88.9	88.9
	TBEL	89.6	88.7	88.8	87.5	89.9	89.0	89.5	88.6

Simulation results



Some simulation results

- TBEL CIs broadly have better coverage rates than BEL CIs, over a range of block scaling $b = Cn^a$ of optimal order ($a = 1/2$ for BEL/ $a = 1/3$ for TBEL).
- Bootstrap calibrations improve the coverage accuracy of both BEL/TBEL methods, especially for small n.
- The improvement is particularly substantial for BEL across differing blocks b.
- The bootstrap version TBEL* tends to exhibit the best coverage accuracy over a range of block lengths b and sample sizes n.

Thank you!!