Journal of Statistical Planning and Inference 155 (2014) 1–18

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Review

A review of empirical likelihood methods for time series[☆]

Daniel J. Nordman[a,*], Soumendra N. Lahiri[b]

[a] *Department of Statistics, Iowa State University, United States*

[b] *Department of Statistics, North Carolina State University, United States*

ARTICLE INFO

Article history:
Received 18 December 2012
Received in revised form 24 September 2013
Accepted 3 October 2013
Available online 11 October 2013

MSC:
62G09
62F40
62M10

Keywords:
Blocking
Frequency domain
GARCH
High dimensional data
Kullback–Leibler distance
Long range dependence
Mahalanobis distance
Markov chains
Spatial data
Tapering
Whittle estimation

ABSTRACT

We summarize advances in empirical likelihood (EL) for time series data. The EL formulation for independent data is briefly presented, which can apply for inference in special time series problems, reproducing the Wilks phenomenon of chi-square limits for log-ratio statistics. For more general inference with time series, versions of time domain block-based EL, and its generalizations based on divergence measures, are described along with their distributional properties; some approaches are intended for mixing time processes and others are tailored to time series with a Markovian structure. We also present frequency domain EL methods based on the periodogram. Finally, EL for long-range dependent processes is reviewed as well as recent advantages in EL for high dimensional problems. Some illustrative numerical examples are given along with a summary of open research issues for EL with dependent data.

Contents

☆ Research partially supported by NSF grant numbers DMS-1007703 and DMS-0906588 and NSA grant number H98230-11-1-0130.

* Corresponding author. Tel.: +1-515-294-7777.
E-mail address: dnordman@iastate.edu (D.J. Nordman).

http://dx.doi.org/10.1016/j.jspi.2013.10.001
0378-3758/

1. Introduction

For independent, identically distributed (iid) data, Owen (1988, 1990) introduced Empirical Likelihood (EL) as a general nonparametric methodology for creating likelihood-type inference without specifying a joint distributional model for the data, as typical with parametric likelihood. The main idea is to formulate a non-parametric likelihood (or EL) function for assessing the plausibility of values of a given population parameter. The resulting EL function is built by a process of probability profiling of data and leads to likelihood-ratio statistics for constructing tests and confidence regions, which have some analogous properties to their fully parametric likelihood counterparts (e.g., chi-square limits), but often without explicit assumptions about the data-generating mechanism.

Because EL is often intended to be used nonparametrically, extending the methodology to dependent data can be challenging. At issue, an EL formulation often needs to accommodate the unknown (and potentially complex) dependence structure in such data. There have various attempts to develop EL for dependent data, mostly in time series applications. This manuscript attempts to summarize differing approaches for EL inference with time series. We briefly review EL for iid data in Section 2 along with extensions to high-dimensional data. Section 3 presents EL versions for time series which essentially follow the iid data EL formulation with similar distributional properties, despite data dependence. These are typically tailored to a particular model structure or special inference problem, in a way that serial correlation in a time series is not an issue. However, across many general inference problems with time series, the iid data version of EL will generally *fail* and a valid EL formulation needs to nonparametrically accommodate the underlying dependence structure. Data-blocking is a broad technique for this, and Sections 4–6 summarize and compare different block-based EL approaches (e.g., block EL, tapered block EL, regenerative block EL). Further generalizations of block-based EL are described in Section 7. As an alternative to data-blocking, Section 8 describes a frequency domain EL for time series based on a data-transformation. Section 9 then illustrates different EL for time series with a numerical example. Section 10 outlines EL for long-range dependent processes, and Section 11 provides some concluding remarks and open research problems with EL for dependent data.

This EL review is meant to complement some existing ones. Owen's (2001) book provides an accessible account of many developments of EL, including some for dependent data. Chen and Van Keilegom (2009) review EL methods for regression problems, often for independent data but with some connections to dependent data (Section 3.2 here). Both Kitamura (2006) and Bravo (2007) provide summaries of important features of EL for time series inference, with connections to econometrics. Kitamura's (2006) review outlines generalized versions of EL (based on different discrepancy statistics) and their properties, along with associations between EL and general methods of moments estimation (Hansen, 1982). Much of this discussion is related to iid data, but includes a detailed description of distributional properties of a block EL method (Section 4 here). Bravo's (2007) review also describes important uses of EL with general estimating functions in econometric applications (e.g., parameter and moment condition testing) and extends generalized-discrepancy versions of EL mentioned by Kitamura (2006) to time series (cf. Section 7 here). Because Kitamura (2006) and Bravo (2007) detail EL point estimation and moment testing with general possibly over-identifying estimating functions, we do not consider these EL aspects here; these are time series extensions of EL features available for iid data (Qin and Lawless, 1994) as briefly mentioned in Section 2. Rather, we attempt to consolidate general formulations of EL for time series, expanding upon of Velasco's (2009) nice sketch of EL for dependent data.

2. Empirical likelihood under independence

As mentioned in the Introduction, EL inference about a parameter is based on a non-parametric likelihood function built by probability profiling data. For a prototypical example, suppose $X_1, \ldots, X_n$ are iid $\mathbb{R}^d$-valued random vectors and consider inference about their unknown mean $EX_1 = \mu_0$. Consider a distribution $F(x) = \sum_{i=1}^n p_i \mathbb{I}(X_i \le x)$, $x \in \mathbb{R}^d$, supported on the data and created by assigning a probability p_i to data value X_i, $i = 1, \ldots, n$, such that $\sum_{i=1}^n p_i = 1$, where $\mathbb{I}(\cdot)$ denotes the indicator function. Given the data, a likelihood function for this distribution would be $L(F) \equiv \prod_{i=1}^n [F(X_i) - F(X_i-)] = \prod_{i=1}^n p_i$, which is maximized when F is the empirical distribution $F_n(x) = \sum_{i=1}^n n^{-1}\mathbb{I}(X_i \le x)$ (i.e., $p_i = n^{-1}$). To judge the plausibility of a hypothesized mean value $\mu \in \mathbb{R}^d$, the EL method uses an EL function defined as

$$\begin{aligned} L_n(\mu) &\equiv \sup\{L(F) : F \text{ has mean } \mu \text{ \& is supported on } X_1, \ldots, X_n\} \\ &= \sup\left\{\prod_{i=1}^n p_i : 0 \le p_1, \ldots, p_n \le 1, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i G_i(\mu) = 0_d\right\}, \end{aligned} \tag{1}$$

where $G_i(\mu) \equiv (X_i - \mu)$, $i = 1, \ldots, n$ and 0_k is a vector of k zeros, $k \geq 1$. Above $\prod_{i=1}^n p_i$ represents a likelihood formed by assigning probabilities $p_1, \ldots, p_n$ to $X_1, \ldots, X_n$ under a "mean μ" constraint $\sum_{i=1}^n p_i X_i = \mu$. As $L(F_n) = n^{-n}$ maximizes $L(F)$ (without a mean constraint), an EL ratio for the population mean at $\mu \in \mathbb{R}^d$ is defined as

$$R_n(\mu) \equiv L_n(\mu)/n^{-n} \in [0, 1]. \tag{2}$$

Owen (1988, 1990) established a remarkable result that, at the true mean parameter $\mu_0 \in \mathbb{R}^d$, the EL log-ratio statistic satisfies a nonparametric version of Wilks (1938) theorem by having a chi-square limit distribution:

$$-2\log R_n(\mu_0) \xrightarrow{d} \chi_d^2 \quad \text{as } n \to \infty, \tag{3}$$

assuming $\mathrm{Var}(X_1)$ is positive definite. Consequently, (3) provides a basis for conducting large sample tests and calibrating approximate $100(1-\alpha)\%$ confidence regions as $C_{n,1-\alpha} \equiv \{\mu \in \mathbb{R}^d : -2\log R_n(\mu) \leq \chi^2_{d;1-\alpha}\}$, similarly to the parametric case.

For what follows, we also recall an extension of the iid EL method to "just-identified" estimating equations (i.e., same number of estimating functions as parameters); see Qin and Lawless (1994). In this case, a target parameter $\theta \in \Theta \subset \mathbb{R}^p$ is linked to data values $X_t \in \mathbb{R}^d$ by p estimating functions $g = [g_1, \ldots, g_p]' : \mathbb{R}^d \times \Theta \to \mathbb{R}^p$, which satisfy a moment condition $\mathrm{E}g(X_t; \theta_0) = 0_p$ at the true parameter value θ_0. The mean inference case reduces to function $g(x; \theta) = x - \theta$ ($\theta = \mu$) and, in general, an EL function $L_n(\theta)$ and ratio statistic $R_n(\theta)$ for a given value of θ are defined by substituting $G_i(\theta) \equiv g(X_i; \theta)$ for $G_i(\mu)$ in (1)–(2); see Owen (1990, 2001), Qin and Lawless (1994) and Kitamura (2006) for computational details with EL. For this generalization of EL, a Wilks theorem (3) continues to hold: $-2\log R_n(\theta) \xrightarrow{d} \chi_p^2$ as $n \to \infty$ at the true $\theta_0 \in \mathbb{R}^p$ (for non-singular $\mathrm{Var}[g(X_1; \theta_0)]$).

The basic EL method has been extended to more complex inference problems involving independent data and we mention only a few works here. Hall and La Scala (1990) summarize geometric and accuracy properties of EL confidence regions for parameters as smooth functions of means. Qin and Lawless (1994, 1995) expand the properties of EL based on estimating functions $g : \mathbb{R}^d \times \Theta \to \mathbb{R}^r$ (where $r \geq p$ is permitted with $\theta \in \mathbb{R}^p$), showing that the maximizer θ_n of $L_n(\theta)$ is a consistent and asymptotically normal estimator for θ_0; that nuisance parameters in EL functions can be profiled out; and that log-ratio statistics $-2\log[R_n(\theta_0)/R_n(\hat{\theta}_n)]$ and $-2\log R_n(\hat{\theta}_n)$ can be used to, respectively, test $H_0 : \theta = \theta_0$ and "$H_0 : \mathrm{E}g(X_t; \theta_0) = 0_r$ holds for some true θ_0" (i.e., data-generation satisfies the moment conditions). Such EL properties are shared by generalized method of moments estimators (cf. Newey and Smith, 2004) and largely carry over to EL formulations with time series (cf. Kitamura, 2006; Bravo, 2007). For EL confidence regions for mean parameters with iid data, DiCiccio et al. (1991) show an approximation rate $P(\mu_0 \in C_{n,1-\alpha}) = 1 - \alpha + O(n^{-1})$ and that the scaled log-EL ratio admits a Bartlett correction that improves the coverage rate to $O(n^{-2})$. Similar corrections have been established for general estimating functions by Chen and Cui (2006, 2007).

Recently, the EL methodology has been studied in high dimensional problems by several authors. Hjort et al. (2009) consider the performance of the standard EL method based on p-dimensional estimating equations with a sample of size n, when $p \to \infty$ with n. They allow the rate $p = o(n^{1/3})$ to establish a non-degenerate limit distribution of the EL ratio statistic in high dimensions. Chen et al. (2009) improved upon the rate restriction in Hjort et al. (2009), allowing $p = o(n^{1/2})$ under some regularity conditions. An important result of Tsao (2004) shows that the definition of EL for a p-dimensional population mean based on a sample size n breaks down on a set of positive probability whenever $p \geq n/2$. As a result, extension of EL to high dimensional problems is itself a challenging task. A variant of the EL in such situations, called the *Adjusted Empirical Likelihood* (AEL), is given by Chen et al. (2008), which has been further modified by Emerson and Owen (2009). The AEL method adds additional pseudo-observations – one in the case of Chen et al. (2008)'s formulation and two in the case of Emerson and Owen (2009)'s – so that a hypothesized value of the mean parameter is contained in the convex hull of the augmented data set. Bartolucci (2007) advocated a version of the *Penalized Empirical Likelihood* (PEL) approach that is based on the Mahalanobis distance between the convex hull of the n data values and the hypothesized value of the p-dimensional mean vector. The PEL of Bartolucci (2007) is well defined for all values of $p \leq n$, as long as the sample covariance matrix is nonsingular. Properties of Bartolucci (2007)'s method have been studied by Lahiri and Mukhopadhyay (2012a). Other important papers on the PEL in the high dimensional set up are Otsu (2007) and Tang and Leng (2010), who add a penalty function to the standard EL, and thus, are subject to Tsao (2004)'s bound, $p \leq n/2$, for their validity. A further extension of the PEL to $p > n$ problems has been recently studied by Lahiri and Mukhopadhyay (2012b).

3. Non-adjusted empirical likelihood for time series

A general feature of EL for iid data is the property of so-called "self-studentization" (cf. Hall and La Scala, 1990), meaning that the EL log-ratio does not require any direct steps of variance estimation to obtain chi-square limits (3) and correct studentization occurs automatically within the EL method for iid data. Recall that the iid version of EL is characterized by probability profiling individual observations. However, in a time series context, the iid EL formulation may generally fail without suitable modifications to adjust the EL method to accommodate process serial correlation (cf. Kitamura, 1997). That is, iid EL version uses a type of "self-studentization" which is generally invalid for dependent data, particularly when the asymptotic properties of EL involve the infinite dimensional distribution of a time process (e.g., non-zero autocovariances at all lags).

However, the iid formulation of EL and its associated Wilks theorem as in (3) *are* valid for certain inference problems with time series and appropriate estimating functions where large-sample estimation depends only on some finite dimensional aspect of the process distribution. For simplicity, we refer to such EL methods for time series as "non-adjusted" for serial correlation and describe two general applications in Sections 3.1–3.2.

3.1. Model-based EL

A model-based version of EL assumes a structural model for the time series $\{X_t\} \in \mathbb{R}^d$ (e.g., autoregressive (AR)), which may involve unknown parameters $\theta \in \Theta \subset \mathbb{R}^p$ of interest as well as time innovations with unknown distributional form. Commonly, by assuming a process model, an estimating function $g : \mathbb{R}^{d(q+1)} \times \Theta \to \mathbb{R}^p$ can be specified based on an observation $X_t \in \mathbb{R}^d$ and its last q neighbors $X_{t-1}, \ldots, X_{t-q}$ (for some $q \geq 1$) so that

$$g(X_t, X_{t-1}, \ldots, X_{t-q}; \theta_0) = \varepsilon_t \in \mathbb{R}^p \tag{4}$$

holds at the true parameter $\theta_0 \in \mathbb{R}^p$ for a mean-zero process $\{\varepsilon_t\}$ which is iid (or, more generally, constitutes a martingale difference array (MDA), cf. Athreya and Lahiri (2006, ch. 16.1). Note that the moment condition $Eg(X_t, X_{t-1}, \ldots, X_{t-q}; \theta_0) = 0_p$ is implied by (4) and $E\varepsilon_t = 0_p$. Because the large-sample properties of EL here often depend only on the marginal distribution of ε_t, the iid EL formulation (i.e., individual probability profiling) is typically valid for estimating the model parameters θ, where EL function $L_n(\theta)$ is built by profiling the $(n-q-1)$ quantities $g(X_t, X_{t-1}, \ldots, X_{t-q}; \theta)$, $t = q+1, \ldots, n$.

For example, consider a causal, real-valued AR(p) process $X_t = \theta_1 X_{t-1} + \cdots + \theta_p X_{t-p} + e_t$, where the errors $\{e_t\}$ are iid mean-zero with $Ee_1^2 < \infty$. Then, for inference on the AR parameters $\theta = (\theta_1, \ldots, \theta_p)' \in \mathbb{R}^p$, the estimating functions

$$g(X_t, X_{t-1}, \ldots, X_{t-p}; \theta) = \left\{ X_t - \sum_{i=1}^{p} \theta_i X_{t-i} \right\} \cdot (X_{t-1}, \ldots, X_{t-p})' \in \mathbb{R}^p, \quad t = p+1, \ldots, n,$$

satisfy (4) with the martingale difference $\varepsilon_t = e_t(X_{t-1}, \ldots, X_{t-p})'$ at the true parameter θ_0 (e_t being independent of $\{X_j : j < t\}$). In this problem, Chuang and Chan (2002) proved the validity of iid version of EL with a chi-square limit (allowing AR innovations $\{e_t\}$ to be a martingale difference). For the same inference problem, Bravo (2010) established a similar "non-adjusted" result with a generalized version of EL (Section 7). Examples of other model-based EL works falling into the "non-adjusted" category include EL for

1. *Infinite variance AR models* (Li et al., 2010), resembling the above AR(p) model but with $Ee_t^2 = +\infty$ so that estimating functions change.
2. *AR models with explanatory variables* (Zhao and Wang, 2011), where $X_t = \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{j=1}^{q} \alpha_i Z_{j,t} + e_t$ such that $\{e_t\}$ is a MDA, independently $(Z_{1,t}, \ldots, Z_{q,t})$ are random explanatory variables, and the parameters of interest are the β_i, α_j's.
3. *Self-citing threshold AR models* (Chen et al., 2012a) with iid innovations, where particular estimating functions lead to a MDA as in (4).
4. *Partially time varying coefficient models* (Fan et al., 2012) given as $X_t = \sum_{i=1}^{p} \beta_i W_{i,t} + \sum_{j=1}^{q} \alpha_i(t) Z_{j,t} + e_t$, $t = 1/n, \ldots, n/n$, where $W_{i,t}$'s are fixed (or random) design points, $Z_{i,t}$'s are random regressors, each $\alpha_i(\cdot)$ is a smooth function, and the β_j's are the parameters of interest.
5. *Generalized random coefficient models* (Zhao and Wang, 2012) given by $X_t = \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{j=1}^{p} Z_{j,t} X_{t-i} + e_t$, where $W_t = (e_t, Z_{1,t}, \ldots, Z_{p,t})'$ are iid (mean-zero and independent of the past X_t's), and the goal is to estimate the fixed β_i's and the covariance of W_t.
6. *Partially linear regression models* $X_i = \sum_{j=1}^{p} \beta_j z_{j,i} + g(t_i) + e_i$, $1 \leq i \leq n$, where the t_i's are fixed design points, the $z_{j,i}$'s are fixed regressors, $g(\cdot)$ is an unknown smooth function, and target parameters as β_i's. Chen and Cui (2008) treated the case where errors $\{e_t\}$ are a MDA and the design points t_i are general, while Fan and Liang (2010) considered $\{e_t\}$ to have a moving average representation and equally spaced points t_i.
7. *Semiparametric varying coefficient linear models* with error-prone linear covariates (Huang et al., 2010).
8. *Regular generalized autoregressive conditional heteroskedasticity (GARCH) models* (Chan and Ling, 2006) stated as $X_t = e_t\sqrt{Z_t}$, $Z_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j Z_{t-j}$ for iid e_t's with $Ee_t = 0$, $Ee_t^2 = 1$ and target parameters ω, α_i's and β_j's.
9. *AR-ARCH models* (Li et al., 2012a) given by $X_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + e_t\sqrt{Z_t}$, $Z_t = \omega + \sum_{j=1}^{q} \beta_j Z_{t-j}$ for iid e_t's with $Ee_t = 0$, $Ee_t^2 = 1$; estimation again concerns the parameters ω, α_i's and β_j's.

As several references above involve martingale assumptions, see also Mykland (1995) dual likelihood with martingales and Bravo (2007, ch. 6.1). For completeness, there are two exceptions in the works above where $-2\log R_n(\theta_0)$ does not have a chi-square limit: unstable (unit root) AR(p) models (Chuang and Chan, 2002; Chan and Ling, 2006; Bravo, 2010) and partially time varying coefficient models with random $W_{i,t}$'s (Fan et al., 2012).

3.2. EL formulations with kernel-smoothers

For inference on probability density functions or certain regression functions from time series, nonparametric kernel-based estimators are often natural and have large-sample properties determined solely by a marginal distribution of the process (cf. Velasco, 2009; Chen and Wong, 2009, p. 73). Using kernel smoothers to define estimating functions, the implication is that iid version of EL can be applied to these inference problems even in the presence of data dependence, without adjustments for serial correlation.

As a concrete example, suppose a real-valued strictly stationary process $\{X_t\}$ has a probability density function $f(\cdot)$ and consider EL inference of the parameter $\theta = f(x)$ (for some given $x \in \mathbb{R}$) based on a sample $X_1, \ldots, X_n$. Using a kernel function $K(\cdot)$ and a sequence of bandwidths $h = h_n \to 0$ as $n \to \infty$, define an estimating function $g(X_i; \theta) = h^{-1}K([x-X_i]/h) - \theta$ for each $i = 1, \ldots, n$. These estimating functions satisfy $\mathrm{E}\sum_{i=1}^n g(X_i; \theta)/n = \mathrm{E}f_n(x) - f(x) \approx 0$ for a kernel density estimator $f_n(x) = n^{-1}h^{-1}\sum_{i=1}^n K([x - X_i]/h)$. Using these kernel-based estimating functions in the iid EL construction, Xiong and Lin (2012) have shown that usual a Wilks result holds, $-2\log R_n(\theta_0) \xrightarrow{d} \chi_1^2$, for weakly dependent time processes exhibiting either positive association (PA) or negative association (NA). (For the definitions of PA or NA, which cover processes that may not be mixing, see Joag-Dev and Proschan (1983).)

Similar time series applications of the iid EL framework with kernel-smoothing include the nonparametric regression-type tests such as from Chen et al. (2003) where, from stationary pairs $(Y_t, X_t) \in \mathbb{R} \times \mathbb{R}^d$, the goal is to assess whether the conditional mean $E(Y_t|X_t = x) = m(x)$, $x \in [0, 1]^d$, belongs to a parametric family $m_\theta(x)$, $\theta \in \Theta$. Using a kernel-smoothed parametric estimate $\tilde{m}_\theta(x)$, estimating functions $g(Y_t, X_t; x) = K([x - X_t]/h)[Y_t - \tilde{m}_\theta(x)]$ are used in the iid EL formulation to provide a statistic $-2\log R_n(x)$ as in (2) for a given $x \in [0, 1]^d$. This construction is valid because the asymptotic properties of EL depend on the marginal distribution of (Y_t, X_t), not the infinite dimensional distribution of the process (cf. Velasco, 2009). Chen and Gao (2007) developed a bandwidth adapted version of this EL test, which has also been extended for testing a conditional variance $\mathrm{Var}(Y_t|X_t = x) = \sigma^2(x)$ (Chen et al., 2012b). Lian (2009) considered inference about $m(x)$ using estimating functions with kernel smoothers, producing chi-square limits as in (3); Bravo (2007, ch. 5.2) established a similar EL result. Other instances of kernel smoothing within an iid EL construction involve tests about conditional probability densities (cf. Tripathi and Kitamura, 2003; Su and White, 2012).

4. Block empirical likelihood

As mentioned in Section 3, for many general inference problems with time series, the iid formulation of EL (i.e., individual observation-based) typically fails because the automatic EL "self-studentization" occurring under independence is incorrect for capturing the correlation structure in dependent data. As a remedy, Kitatura (1997) introduced a blockwise version of EL for weakly dependent time processes, in which individual observations are replaced by blocks of consecutive data points in time. Such data blocking is a general strategy for preserving the underlying dependence among neighboring time observations, and similar blocking approaches apply for extending resampling methods to time series (e.g., block bootstraps; Lahiri, 2003). We describe a common version of the block empirical likelihood (BEL) method of Kitamura (1997) for inference with estimating equations.

Let $X_1, \ldots, X_n$ represent a realization from a stationary process $\{X_t\}$ of $\mathbb{R}^d$-valued random vectors and $g : \mathbb{R}^d \times \Theta \to \mathbb{R}^p$ be a function satisfying the moment condition

$$\mathrm{E}g(X_t; \theta_0) = 0_p, \tag{5}$$

where θ_0 is the true value of a parameter of interest $\theta \in \Theta \subset \mathbb{R}^p$; here p denotes the dimension of θ and the number of estimation functions g. For example, $g(x; \theta) = x - \theta$ gives an estimating function for the process mean $\theta = \mathrm{E}X_t$. To define the BEL method, we require a collection of data blocks. Let $\ell_n = \ell$ be an integer sequence of block lengths satisfying $\ell^{-1} + \ell/n \to 0$ as $n \to \infty$, ensuring blocks are small relative to n but increase in size for larger samples. Let $(X_i, \ldots, X_{i+\ell-1})$, for $i = 1, \ldots, N \equiv n - \ell + 1$, denote a collection of length ℓ blocks which are maximally overlapping; BEL can be alternatively defined with other block collections (e.g., non-overlapping blocks). To create a BEL function to assess a given parameter value θ, each block contributes an average $G_i(\theta) = \ell^{-1}\sum_{j=i}^{i+\ell-1} g(X_j; \theta)$ of variables computed from the ith data block, $i = 1, \ldots, N$. By probability profiling these block averages, we obtain the BEL function $L_n(\theta) = N^{-N}R_n(\theta)$ and ratio

$$R_n(\theta) = \sup\left\{\prod_{i=1}^N Np_i : 0 \le p_1, \ldots, p_N \le 1, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i G_i(\theta) = 0_p\right\}, \tag{6}$$

analogously to the iid case (1)–(2) but using the N blocks here. The following result of Kitamura (1997) establishes a Wilks theorem for the BEL log-ratio statistic at the true parameter θ_0, under general conditions entailing weak time dependence. Define the strong mixing coefficient of stationary $\{X_t\}$ as $\alpha(k) = \sup\{P(A \cap B) - P(A)P(B)| A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty\}$, where $\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty$ respectively denote σ-algebras generated by $\{X_j : j \le 0\}$ and $\{X_j : j \ge k\}$ (cf. Athreya and Lahiri, 2006, ch. 16.2).

Theorem 1. *Suppose θ_0 satisfies (5); that $\mathrm{E}\|g(X_t; \theta_0)\|^{2+\delta} < \infty$ and $\sum_{k=1}^\infty \alpha(k)^{\delta/(2+\delta)} < \infty$ hold for some $\delta > 0$; and that the $p \times p$ matrix $\sum_{k=-\infty}^\infty \mathrm{Cov}\big[g(X_0; \theta_0), g(X_k; \theta_0)\big]$ is positive definite. If, in addition, $\ell^{-1} + \ell^2/n \to 0$ as $n \to \infty$, then*

$$-2\ell^{-1}\log R_n(\theta_0) \xrightarrow{d} \chi_p^2. \tag{7}$$

Unlike EL with iid data, the block averages involved in BEL are usually locally correlated under time dependence, which necessitates a block adjustment ℓ^{-1} for the BEL log-ratio to have a proper limit distribution. A block length $\ell = 1$ in Theorem 1 reproduces the EL distributional result (3) for iid data. Similarly to the iid case, Theorem 1 allows tests and confidence regions for time series parameters θ from the BEL ratio statistic and a chi-square calibration. However, the coverage performance of BEL depends on the block length choice ℓ for a given sample size n and the underlying process. Some strategies for block selection are illustrated with a numerical example in Section 9.

Remark 1. Kitamura (1997) showed that BEL method with general estimating functions exhibits same important properties described in Section 2 for the iid EL case, e.g., asymptotically normal point estimators $\hat{\theta}_n$ as the maximizer of $R_n(\theta)$ along with parameter and moment tests based on $-2\ell^{-1}\log[R_n(\theta_0)/R_n(\hat{\theta}_n)]$ and $-2\ell^{-1}\log R_n(\hat{\theta}_n)$, respectively; see Kitamura (1997, 2006) or Bravo (2007) for more details.

Remark 2. For inference about a stationary mean $\mu = EX_t$, Zhang (2006, 2007) proved that Theorem 1 holds for time series under NA or PA, respectively (see Section 3.2). For quantile estimation $\theta_q \equiv \{x : F(x) \ge q\} \in \mathbb{R}$ (given $q \in (0, 1)$) with a stationary, mixing process $\{X_t\}$, Chen and Wong (2009) showed that the estimating function $g(X_t; \theta_q) = \mathbb{I}(X_t \le \theta_q) - q$ fulfilling (5) can be replaced by a kernel smoothed version in BEL for improved effects; Lei and Qin (2011) considered the same approach for NA series. Qin et al. (2011) and Li et al. (2012a) established a BEL method for estimating probability density functions under NA and PA, respectively; as pointed out by Xiong and Lin (2012), however, data-blocking is unnecessary in this particular problem and the unadjusted EL approach is valid (Section 3.2). The data-blocking in Lei and Qin (2011), Qin and Li (2011), Qin et al. (2011)and Li et al. (2012b) is a bit complex (i.e., Bernstein's big-blocks–little-blocks), and Kitamura's blocking schemes should also be valid.

Remark 3. While the BEL method here is intended for weakly dependent processes, assumptions of stationarity can be relaxed for BEL inference within an estimating function framework. Bravo (2009) established the validity of BEL and other block nonparametric likelihoods for a class of non-stationary series $\{X_t\}$. BEL applies as well to linear regressions $X_t = z_t'\theta + \varepsilon_t$ involving a stationary error process $\{\varepsilon_t\}$ and fixed regressors z_t under very mild conditions (Nordman, 2008a,b). In a similar regression problem with NA errors $\{\varepsilon_t\}$, Qin and Li (2011) also established a BEL method and Bravo (2005) also considered BEL-related tests for time series regressions. Wu and Cao (2011) extended BEL to count data with generalized linear estimating functions in a non-stationary setting (their BEL statistics, though, may be missing block adjustments). Zhang et al. (2012) proposed a jackknife-blockwise EL approach to reduce the computational burden in profiling out nuisance parameters in the BEL method, which applies to mixing, non-stationary processes.

5. Tapered block empirical likelihood

We introduce data tapers here to define a modified version of BEL for inference about "smooth function model" parameters (Hall, 1992, Section 2.4) of an $\mathbb{R}^d$-valued stationary process $\{X_t\}$. Suppose that the target parameter is given by

$$\theta_0 = H(\mu_0) \in \mathbb{R}^p \tag{8}$$

where $H : \mathbb{R}^d \to \mathbb{R}^p$, $p \le d$, is a function of the true process mean $EX_t = \mu_0 \in \mathbb{R}^d$. This formulation allows a wide range of parameters given that X_t and its mean are $\mathbb{R}^d$-valued; see Künsch (1989) or Lahiri (2003, Ch. 4) for examples as well as Section 9. Hall and La Scala (1990) and Kitamura (1997) considered EL inference for smooth function parameters with independent and time series data, respectively. Our goal here is to describe a Tapered Block Empirical Likelihood (TBEL) for θ.

We first define TBEL for the mean parameter $\mu \in \mathbb{R}^d$. For mean inference, recall that the BEL approach from Section 4 involves block averages $G_i(\mu) = \sum_{j=i}^{i+\ell-1}(X_j - \mu)/\ell$ from the collection $(X_i, \ldots, X_{i+\ell-1})$, $i = 1, \ldots, N = n - \ell + 1$ of overlapping length ℓ data blocks. For a sequence $w_\ell(1), \ldots, w_\ell(\ell) \in [0, 1]$ of weights with $\|w_\ell\|_1 \equiv \sum_{j=1}^{\ell} w_\ell(j) > 0$, the TBEL method substitutes the tapered average $T_i(\mu) = \sum_{j=i}^{i+\ell-1} w_\ell(j)(X_j - \mu)/\|w_\ell\|_1$ for $G_i(\mu)$ with each block. With this change, the TBEL ratio statistic $R_n(\mu)$ for μ is defined like the BEL version, replacing $G_i(\mu)$ with $T_i(\mu)$ in (6). Using a set-up familiar in spectral estimation with time series (Brillinger, 1981, Ch. 3), define the weights as

$$w_\ell(i) \equiv w([i - 0.5]/\ell), \quad i = 1, \ldots, \ell, \tag{9}$$

using a tapering window $w : \mathbb{R} \to [0, 1]$ that is zero for $t \notin [0, 1]$, symmetric about $t = 1/2$, non-decreasing for $t \in (0, 1/2]$, and positive in a neighborhood of $t = 1/2$. Note that for $w(t) = \mathbb{I}(t \in [0, 1])$, where $\mathbb{I}(\cdot)$ denotes the indicator function, all block observations receive the same weight, and the TBEL reduces to BEL. However, for tapers $w(t)$ decreasing to zero at $t = 0$ or 1, such as a trapezoidal taper $w_{trap}(t) = 2t\mathbb{I}(t \in [0, 1/2]) + 2(1 - t)\mathbb{I}(t \in [1/2, 1])$ or cosine-bell taper $w_{\cos}(t) = \mathbb{I}(t \in [0, 1])[1 - \cos(2\pi t)]/2$, both edges of a data block are down-weighted in the tapered average $T_i(\mu)$, which generally leads to better "self-studentization" steps in TBEL under dependence (cf. Paparoditis and Politis, 2001). From the TBEL ratio for $\mu \in \mathbb{R}^d$, the TBEL ratio statistic for the parameter θ under (8) is given by profiling

$$R_n(\theta) \equiv \sup\{R_n(\mu) : \mu \in \mathbb{R}^d, H(\mu) = \theta\}.$$

The following TBEL result for smooth function parameters is an extension of Nordman (2009).

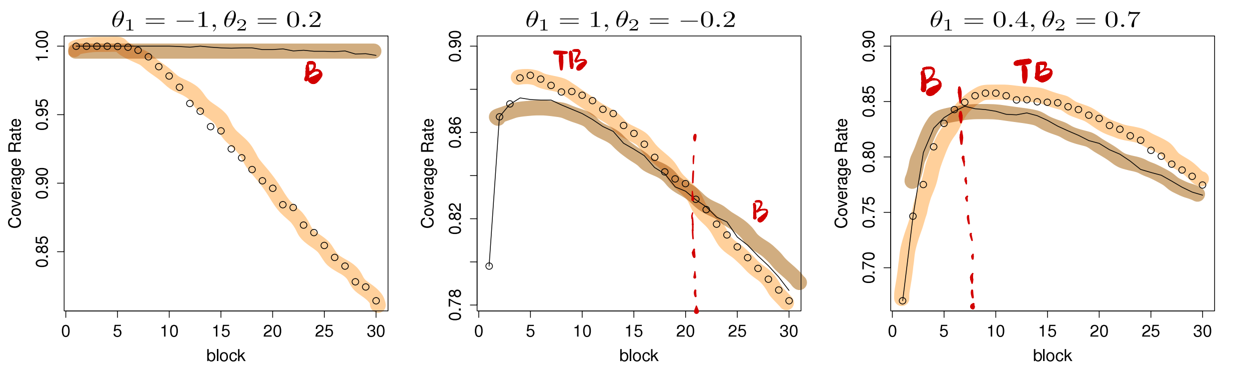

Fig. 1. As a function of block length ℓ, observed coverages of 90% CIs for the process mean of $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$ (iid standard normal $\{Z_t\}$) using TBEL (∘) and BEL (—) methods with a sample size $n = 100$ (based on 4000 simulations).

Theorem 2. *Suppose* $\mathrm{E}\|X_t\|^{6+\delta} < \infty$ *and* $\sum_{k=1}^{\infty} k^2 \alpha(k)^{\delta/(6+\delta)} < \infty$ *hold for some* $\delta > 0$, *and that* $H(\cdot)$ *from* (8) *is continuously differentiable in a neighborhood of* $\mu_0 \in \mathbb{R}^d$ *and that* $\nabla[\sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_0, X_k)]\nabla'$ *has rank* $p \le d$, *where* $\nabla \equiv [\partial H_i(\mu_0)/\partial \mu_j]_{i=1,\ldots,p;j=1,\ldots,d}$ *denotes the* $p \times d$ *matrix of first-order partial derivatives of* H *at* μ_0. *If the taper weights satisfy* (9) *and* $\ell^{-1} + \ell^2/n \to 0$ *as* $n \to \infty$, *then*

$$-2a_n \log R_n(\mu_0) \xrightarrow{d} \chi_p^2, \tag{10}$$

where $a_n = \sum_{i=1}^{\ell} w_\ell^2(j)/(\sum_{j=1}^{\ell} w_\ell(j))^2$.

The TBEL log-ratio statistic requires an adjustment a_ℓ to account for both overlapping blocks and the taper. For a constant taper $w(t) = \mathbb{I}(t \in [0, 1])$, the TBEL and BEL methods match and the adjustment becomes $a_n = \ell^{-1}$ as in Theorem 1. However, for non-constant tapers, the TBEL method with an appropriate block length can lead to better coverage accuracy than BEL; a small numerical example of this is presented in Fig. 1.

Remark 4. We have presented the TBEL method for smooth function parameters (8), which provide an alternative parameter formulation to estimation functions. While not formally established in the literature, a TBEL result with general estimating functions should hold analogously to the BEL result in Theorem 1 (e.g., replacing ℓ^{-1} with a_n there).

6. Regenerative block empirical likelihood

Harari-Kermadec (2011) introduced a novel version of block-based EL, called the Regenerative Block Empirical Likelihood (RBEL), intended for time series which are *Markov chains* (e.g., ARMA, ARCH, GARCH, bilinear processes). The main idea of RBEL is that observations $X_1, \ldots, X_n$ from certain Markov chains can be partitioned into data blocks which are *iid*, though random in length. RBEL then leads to a block-based EL formulation which resembles the EL version for iid data.

To provide a basic framework for Markov chains on a general state space (cf. Athreya and Lahiri, 2006, Ch. 14), let $\{X_t\}_{t\ge0}$ be a sequence of random variables assuming values in some space $\mathbb{S}$ (e.g., a subset of $\mathbb{R}^d$) and let $\mathcal{S}$ denote a corresponding σ-algebra for $\mathbb{S}$. Then $\{X_t\}_{t\ge0}$ is a Markov chain (MC) if

$$P(X_{t+1} \in A|X_t, \ldots, X_0) = P(X_{t+1} \in A|X_t) \equiv p(X_t; A) \quad \text{w.p.1}$$

holds for any $A \in \mathcal{S}$, any $n \ge 0$, and any initial distribution ν of X_0. The function $p(x; A) = P(X_1 \in A|X_0 = x) \in [0, 1]$ of $x \in \mathbb{S}, A \in \mathcal{S}$ is call the transition probability function (TPF). Hence, conditioned on $X_0, \ldots, X_t$, the distribution of X_{t+1} depends only on X_t through $P(\cdot; \cdot)$. As an example, let $\{\varepsilon_t\}_{t\ge1}$ be iid with distribution μ on $\mathbb{R}$ and, independently, let X_0 be a random variable with distribution ν; then, a "waiting time chain" $\{X_t\}_{t\ge0}$, defined by $X_t = \max\{X_{t-1} + \varepsilon_t, 0\}$ for $t \ge 1$, is a $\mathbb{S} = [0, \infty)$-valued MC with TPF $p(x; A) = P(\max\{x + \varepsilon_1, 0\} \in A)$ and initial distribution ν. In the following, let $P_x \equiv p(x; \cdot)$ denote the TPF given $X_0 = x \in \mathbb{S}$.

The main MC features underlying the RBEL method are chains which are *regenerative sequences* and have a *stationary distribution*. In general, a sequence of random variables $\{X_t\}_{t\ge0}$ is called "regenerative" if there exist a sequence of random times $0 < T_1 < T_2 < T_3 < \cdots$ such that the "excursion blocks" $\{\eta_j\}_{j\ge1}$ are iid, where $\eta_j \equiv (X_{T_j+1}, \ldots, X_{T_{j+1}}, T_{j+1} - T_j)$ represents both the number $T_{j+1} - T_j$ of variables and their values $X_t, T_j < t \le T_{j+1}$, between two times T_j and T_{j+1}. Hence, the regenerative times T_j cut the series $\{X_t\}_{t\ge T_1}$ into iid pieces (excluding the initial variables $X_0, \ldots, X_{T_1}$). For certain MCs $\{X_t\}_{t\ge0}$, regenerative times $\{T_j\}_{j\ge1}$ are definable as

$$T_1 = \min\{t > 0 : X_t \in \Delta\}, \qquad T_{i+1} = \min\{t > T_i : X_t \in \Delta\}, \quad i \ge 1; \tag{11}$$

representing the consecutive "hitting times" of a special set $\Delta \in \mathcal{S}$. For now, we assume that a singleton set Δ can be chosen to which the MC is recurrent in the sense that $P_x(T_1 < \infty) = 1$ for any $x \in \mathbb{S}$, where T_1 denotes the first entrance time to

Δ (with $T_1 = +\infty$ if $X_t \notin \Delta$ for all $t \geq 1$). That is, if the MC is guaranteed to return to Δ in finite time from any starting point $x \in \mathbb{S}$, then the successive times at which the chain returns to Δ define a set of iid block excursions. As an example, the waiting time chain with $\mathrm{E}\varepsilon_1 < 0$ is recurrent to 0 and is regenerative with T_j's defined by successive returns of $\{X_t\}_{t\geq 0}$ to $\Delta = \{0\}$ (Athreya and Lahiri, 2006, ch. 14). Supposing that the chain possess a recurrent singleton set Δ, we also assume that the chain is aperiodic and $\mathrm{E}_\Delta T_1 = \mathrm{E}[T_1|X_0 \in \Delta] < \infty$, implying a unique stationary distribution π on $\mathbb{S}$ exists (by definition, satisfying $\pi(A) = \int_{\mathbb{S}} p(x; A)\pi(dx)$ for any $A \in \mathcal{S}$) prescribed by the occupation measure $\pi(A) = \mathrm{E}_\Delta[\sum_{t=1}^{T_1} \mathbb{I}(X_t \in A)]/\mathrm{E}_\Delta T_1$, $A \in \mathcal{S}$ (cf. Meyn and Tweedie, 2009, Theorem 10.2.2). For example, the waiting time chain with $\mathrm{E}\varepsilon_1 < 0$ has a such recurrent hitting set $\Delta = \{0\}$ and a stationary distribution π matching that of $\sup_{j\geq 1} \sum_{t=1}^{j} \varepsilon_t$.

We may now state the RBEL construction in terms of general estimating functions, supposing the MC possesses a recurrent singleton set Δ with properties as above and a stationary distribution π. Let $G : \mathbb{S} \times \mathbb{R}^p \to \mathbb{R}^p$ be estimating functions where a target parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$ satisfies

$$\mathrm{E}G(X; \theta_0) = 0_p \quad \text{for } X \sim \pi; \tag{12}$$

that is, under the MC *stationary distribution*, the estimating function fulfills a moment condition at the true parameter θ_0. If the MC $\{X_t\}_{t\geq 0}$ is itself stationary, then the moment condition based on π in (12) would essentially match that for the BEL version (5) so that BEL and RBEL could apply to the *same* parameters in this case. In general though, the parameters treatable in the BEL and RBEL frameworks may not have a direct correspondence, as the MC $\{X_t\}_{t\geq 0}$ could generally be non-stationary (if $\nu \neq \pi$.) In the RBEL implementation based on observations $X_1, \ldots, X_n$ from the chain, suppose there are $l + 1 = \sum_{t=1}^{n} \mathbb{I}(X_t \in \Delta)$ hitting times of Δ among $\{1, \ldots, n\}$, denoted as $1 \leq T_1 < T_2 < \cdots T_{l+1} \leq n$ $(l \equiv l_n)$. Make l data blocks from the excursions $(X_{T_i+1}, \ldots, X_{T_{i+1}})$ for $i = 1, \ldots, l$, throwing out a first block $X_1, \ldots, X_{T_1}$ and last $X_{T_{l+1}+1}, \ldots, X_n$ if necessary. To assess a potential parameter value θ, each block contributes a block sum $G_i(\theta) = \sum_{t=T_i+1}^{T_{i+1}} G(X_t; \theta)$, $i = 1, \ldots, l$, which are then probability profiled to give a RBEL ratio statistic

$$R_n(\theta) = \sup\left\{ \prod_{i=1}^{l} lp_i : 0 \leq p_1, \ldots, p_l \leq 1, \sum_{i=1}^{l} p_i = 1, \sum_{i=1}^{l} p_i G_i(\theta) = 0_p \right\}. \tag{13}$$

Note that the linear expectation constraint in (13) on the *iid* quantities $G_i(\theta)$ mimics the moment condition (12), which is equivalent to $\mathrm{E}_\Delta \sum_{t=1}^{T_1} G(X_t; \theta_0) = 0_p$ using the occupation measure representation of the stationary distribution π.

The next result, due to Harari-Kermadec (2011), gives the distribution of the RBEL log-ratio statistic; assumptions about the first entrance time T_1 to Δ in (11) involve E_Δ and E_ν denoting a conditional expectation given $X_0 \in \Delta$ or an expectation given $X_0 \sim \nu$.

Theorem 3. *Suppose that the Markov chain $\{X_t\}_{t\geq 0}$ on $(\mathbb{S}, \mathcal{S})$, with initial distribution ν for X_0 and probability transition function $p(\cdot; \cdot)$, has a stationary distribution satisfying (12) with $\mathrm{E}\|G(X; \theta_0)\|^2 < \infty$ and non-singular $\mathrm{E}G(X; \theta_0)G(X; \theta_0)'$ for $X \sim \pi$. Suppose also that a singleton recurrent set $\Delta \in \mathcal{S}$ exists with $p(x; \Delta) = 1$ for all $x \in \mathbb{S}$ with $\mathrm{E}_\Delta[T_1^2] < \infty$ and $\mathrm{E}_\nu[T_1] < \infty$. Then, at the true parameter θ_0,*

$$-2 \log R_n(\theta_0) \xrightarrow{d} \chi_p^2 \quad \text{as } n \to \infty. \tag{14}$$

Unlike the BEL version (7), the distributional result (14) for the RBEL log-ratio statistic involves no block correction and resembles the iid EL case (3) due to the independence of the regenerative block excursions.

The above RBEL result assumes that a special (i.e., singleton, recurrent) hitting set Δ is known for the MC, which may be unfeasible in many cases. Harari-Kermadec (2011) also extended the RBEL method to Harris recurrent MCs with stationary distributions. Harris recurrent MCs are characterized by a so-called "small" set $\Delta \in \mathcal{S}$ (i.e., a hitting set) which can be used to construct regeneration times by a split chain technique (cf. Athreya and Lahiri, 2006, ch. 14). Using this technique, it is possible to *estimate* regeneration times for a sample $X_1, \ldots, X_n$ from a Harris recurrent MC, using a small set Δ and a kernel estimate of a density for the TPF $p(\cdot; \cdot)$; the RBEL construction (13) then proceeds as before, dividing the sample into (estimated) generative blocks. Under some additional assumptions, the REBEL result (14) continues to hold along with other inference through estimating functions (Harari-Kermadec, 2011).

Remark 5. While not involving block length choices as in BEL approaches, the general RBEL method is not tuning parameter free in that small set Δ selection and kernel density estimation are required; the order of the MC may require estimation as well.

7. Further generalizations of block empirical likelihood

Here we mention two generalizations of the BEL method with general estimating functions from Section 4. The first is that the choice of likelihood distance, or divergence measure, may be varied in defining the nonparametric likelihood ratio $R_n(\cdot)$. The second is observational data blocks may be replaced by kernel-smoothed windows of observations.

An EL log-ratio statistic is one particular criterion, among a variety, for assessing the distances between empirical distributions (Section 2). For example, if $X_1, \ldots, X_n \in \mathbb{R}^d$ are iid, instead of using EL log-ratio statistic

$$-2\log R_n(\mu) = \inf\left\{-2\sum_{i=1}^{n}\log(p_i n) : 0 \le p_1, \ldots, p_n \le 1, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i(X_i - \mu) = 0_d\right\}$$

from (2) for mean parameter $\mu \in \mathbb{R}^d$ inference, one could use Kullback–Leibler (Efron, 1981) or Euclidean (Owen, 1991) criteria by replacing $-2\sum_{i=1}^{n}\log(p_i n)$ above with discrepancies:

$$2n^{-1}\sum_{i=1}^{n} p_i \log(p_i n) \quad \text{or} \quad \sum_{i=1}^{n}(p_i n - 1)^2,$$

respectively. More generally, substituting the Cressie–Read divergence $2\sum_{i=1}^{n}[(p_i n)^{-\lambda} - 1]/[\lambda(1+\lambda)]$ for $-2\sum_{i=1}^{n}\log(p_i n)$ produces a discrepancy statistic $D_{n,\lambda}(\mu)$ for a given $\lambda \in \mathbb{R}$ that includes log-EL, log-Kullback–Leibler and log-Euclidean likelihood statistics as special cases ($\lambda = 0, -1$ or -2, defined by taking limits for $\lambda = 0, -1$) and for which $D_{n,\lambda}(\mu_0) \xrightarrow{d} \chi_d^2$ as $n \to \infty$ at the true mean μ_0 under mild conditions in the iid data case (Baggerly, 1998). As in Section 2, one may also replace the mean estimating function $X_t - \mu$ with a more general one $g(x; \theta) : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^p$ satisfying $\mathrm{E}g(X_t; \theta_0) = 0_p$ at a true $\theta_0 \in \mathbb{R}^p$ and form analogous test statistics $D_{n,\lambda}(\theta)$ for hypothesized parameter values θ. Additionally, such Cressie–Read statistics with estimating functions can then be embedded into a larger class of nonparametric likelihood-based discrepancies called generalized EL statistics (Smith, 1997; Newey and Smith, 2004). See Kitamura (2006) and Bravo (2007) for reviews of these alternative nonparametric likelihoods (e.g., generalized EL statistics with possibly overidentifying estimating functions) for independent data, where Kitamura (2006) provides several justifications that favor EL among this list based on accuracy considerations in estimation and testing. Our goal here is to briefly mention some important references for extending these alternative nonparametric likelihood statistics to time series where, as with EL, data blocking has often played an important role. For inference on the mean of a mixing, stationary series $\{X_t\} \in \mathbb{R}^d$, Bravo (2002) established a chi-square limit for the above Cressie–Read statistics using Kitamura's (1997) blocking method; that is, if the data block averages $G_i(\mu) = \sum_{j=i}^{i+\ell-1}(X_j - \mu)/\ell$ from Section 4 are substituted and probability profiled to define the discrepancy statistic $D_{n,\lambda}(\mu_0)$, then a Wilks theorem $\ell^{-1}W_n(\mu_0) \xrightarrow{d} \chi_d^2$ as $n \to \infty$ holds using the same block adjustment as BEL; Kitamura and Stutzer (1997) earlier considered the case $\lambda = -1$. For stationary, mixing processes, Lin and Zhang (2001) proved the BEL results with general estimating functions also hold for block-based Euclidean likelihood (e.g., chi-square limits for $\ell^{-1}W_{n,-2}(\theta_0)$); Chen and Zhang (2010) extended these results to include stationary $\{X_t\}$ exhibiting NA. Bravo (2009) developed a block-based generalized EL for potentially non-stationary and non-linear time series, extending Kitamura (1997)'s BEL to other block-based nonparametric likelihoods.

Data-blocks may also be replaced with kernel-smoothed data windows within an estimating function-based framework of generalized EL for time series (cf. Smith, 1997). Instead of using block averages of estimating functions (cf. Section 4) from a time series $X_1, \ldots, X_n$, one uses a window smoothed average of estimating functions $G_i(\theta) = h^{-1}\sum_{j=i-n}^{i-1} K(j/h) g(X_{i-j}; \theta)$ for each observation $i = 1, \ldots, n$, where $K(\cdot)$ is a real-valued kernel and $h = h_n$ is a bandwidth parameter. For the commonly used truncation-kernel $K(x) = \mathbb{I}(|x| \le 1)$, $x \in \mathbb{R}$, and a bandwidth $h = (2\ell + 1)/2 \approx \ell \to \infty$ as $n \to \infty$, the window smoothed averages $G_i(\theta) = h^{-1}\sum_{j=\max\{i-n,-\ell\}}^{\min\{i-1,\ell\}} g(X_{i-j}; \theta)$ resemble previous BEL block averages except at the edges of the observed time stretch $i = 1, \ldots, n$. With such window smoothing, Guggenberger and Smith (2008) developed generalized EL inference for possibly non-stationary time series with tests statistics having chi-square limits, extending results of Kitamura and Stutzer (1997); see also Otsu (2006) for generalized EL developments for time series based on window smoothing in place of data blocking.

8. Empirical likelihood in the frequency domain

To handle serial correlation in an alternative manner, another version of EL for time series can be formulated in the frequency domain. Instead of attempting to capture the time dependence through data blocking, a frequency domain version of EL applies a *data transformation*, via the discrete Fourier transform (DFT), intended to weaken the dependence. Because DFTs at distinct Fourier frequencies are known to be asymptotically independent (cf. Brillinger, 1981; Yajima, 1989), transforming the data allows one to exploit the approximate independence structure of the DFTs and formulate a Frequency Domain Empirical Likelihood (FDEL) method by mimicking the EL version for independent data. Unlike BEL with block lengths, FDEL has the advantage of no tuning parameter selection, but at a price. Due to its frequency domain formulation, FDEL is effective for a *restricted* class of process parameters that can be expressed as functionals of the process spectral density (e.g., autocorrelations). That is, BEL can apply for inference about the general mean structure of a process, while FDEL does not. Additionally, the FDEL method assumes that the time series $\{X_t\}$ has a linear representation

$$X_t = \mu + \sum_{j=\infty}^{\infty} b_j \varepsilon_{t-j} \in \mathbb{R} \tag{15}$$

in terms of an iid sequence of mean-zero innovations $\{\varepsilon_t\}$ and a linear filter sequence $\{b_j\}$, $\sum_{j=-\infty}^{\infty} b_j^2 < \infty$ with $b_0 = 1$. Monti (1997) first introduced a FDEL version for Whittle inference. In the following, we present a FDEL formulation based on a general framework from Nordman and Lahiri (2006), which includes results of Monti (1997).

Let $\{X_t\}$ be a stationary process as in (15) having mean $\mu \in \mathbb{R}$, autocovariance function $r(k) = \text{Cov}(X_0, X_k)$, $k \geq 0$, and spectral density $f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} r(k)e^{-\imath\lambda k}$, $\lambda \in \Pi \equiv [-\pi, \pi]$ where $\imath = \sqrt{-1}$. Let $d_n(\lambda) = \sum_{t=1}^{n} X_t e^{-\imath\lambda t}$, $\lambda \in \Pi$, denote the DFT of a data sample $X_1, \ldots, X_n$ and let $I_n(\lambda) = (2\pi n)^{-1} d_n(\lambda) d_n(-\lambda)$, $\lambda \in \Pi$, denote the periodogram. Suppose that we are interested in inference about a parameter $\theta \in \Theta \subset \mathbb{R}^p$ defined by a spectral estimating equation

$$\int_0^{\pi} G(\lambda; \theta_0) f(\lambda) = a_0 \in \mathbb{R}^p \tag{16}$$

for some known $a_0 \in \mathbb{R}^p$, where $\theta_0 \in \mathbb{R}^p$ denotes the true value of the parameter and $G : \Pi \times \Theta \to \mathbb{R}^p$ denotes a vector of even estimating functions. Typically, the vector of constants a_0 must be the zero vector 0_p, but other choices of a_0 may be allowable in special cases. Common parameters that fit into the framework (16) with $a_0 = 0_p$ include ratios of spectral means (cf. Lahiri, 2003, ch. 8); for example, if $\theta_0 = (r(k_1), \ldots, r(k_p))'/r(0)$ is a vector of autocorrelations at given lags $k_1, \ldots, k_p \in \mathbb{Z}$, then (16) holds with $a_0 = 0_p$ and $G(\lambda; \theta) = \big(\cos(k_1), \ldots, \cos(k_p)\big)' - \theta \in \mathbb{R}^p$. Another example is given next.

Example 8.1 (*Whittle Estimation*). Consider a parametric family of spectral densities $\mathcal{F} \equiv \{f_\theta : \theta \in \Theta\}$ such that f_θ is positive on Π and, for any $\theta_1 \neq \theta_2$, the set $\{\lambda \in \Pi : f_{\theta_1}(\lambda) \neq f_{\theta_2}(\lambda)\}$ has positive Lebesgue measure. Whittle estimation aims to find a $\theta_0 \in \Theta$ which minimizes a theoretical distance $W(\theta) = \int_0^{\pi} \log f_\theta(\lambda) + [f(\lambda)/f_\theta(\lambda)] d\lambda$, where f is the true spectral density (which may not belong to $\mathcal{F}$). Considering a common formulation (cf. Hannan, 1973; Fox and Taqqu, 1986), suppose spectral densities in $\mathcal{F}$ can be written as

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} k_\beta(\lambda), \quad \lambda \in \Pi, \tag{17}$$

where $\theta = (\sigma^2, \beta')' \in (0, \infty) \times \mathbb{R}^{p-1}$ and k_β is a density kernel on Π involving parameters $\beta = (\beta_1, \ldots, \beta_{p-1})' \in \mathbb{R}^{p-1}$ such that Kolmogorov's formula holds: $\int_0^{\pi} \log k_\beta(\lambda) d\lambda = 0$. Then, under mild conditions, the minimizer θ_0 of $W(\theta)$ solves the equations

$$\int_0^{\pi} k(\lambda; \beta) f(\lambda) = 0_{p-1}, \qquad \int_0^{\pi} \frac{f(\lambda)}{f_\theta(\lambda)} = \pi, \tag{18}$$

where $f_\theta^{-1} = 1/f_\theta$ and $k(\lambda; \beta) \equiv \partial k_\beta^{-1}(\lambda)/\partial\beta$ is an $\mathbb{R}^{p-1}$-vector of first-order partial derivatives of $k_\beta^{-1} = 1/k_\beta$. Here (16) holds with non-zero $a_0 = (\pi, 0, \ldots, 0)' \in \mathbb{R}^p$ and $G^{whit}(\lambda; \theta) = (f_\theta^{-1}(\lambda), k(\lambda; \beta)')' \in \mathbb{R}^p$. Note that, to target inference on parameters β in the kernel k_β, $\tilde{G}^{whit}(\lambda; \beta) = k(\lambda; \beta) \in \mathbb{R}^{p-1}$ may be used satisfying (16) with $a_0 = 0_{p-1}$.

Next we define the FBEL ratio statistic for a parameter θ defined by the estimating equation (16). Let $\lambda_{i,n} = 2\pi i/n$, $i = 1, \ldots, n_0$ denote the discrete Fourier frequencies with $n_0 = \lfloor (n-1)/2 \rfloor$. The FDEL function for θ is $L_n(\theta) = n_0^{-n_0} R_n(\theta)$ and the FDEL ratio is

$$R_n(\theta) = \left\{ \prod_{i=1}^{n_0} n_0 p_i : 0 \leq p_1, \ldots, p_{n_0} \leq 1, \sum_{i=1}^{n_0} p_i = 1, \pi \sum_{i=1}^{n_0} p_i G(\lambda_{i,n}; \theta) I_n(\lambda_{i,n}) = a_0 \right\}; \tag{19}$$

here values p_i are assigned to each periodogram value $G(\lambda_{i,n}; \theta) I_n(\lambda_{i,n})$ under a constraint that is a discretized version of the moment condition (16). The following result gives the limit distribution of the FDEL log-ratio statistic (19); see Nordman and Lahiri (2006).

Theorem 4. *Suppose $\{X_t\}$ is a linear process as in (15) with $E(\varepsilon_1^2) > 0$, $E(\varepsilon_1^8) < \infty$ and $\sum_{j=-\infty}^{\infty} |b_j| < \infty$. Assume also that (16) holds; that each component of $G(\cdot; \theta_0)$ is Lipschitz of order greater than $1/2$ on $[0, \pi]$; and that the $p \times p$ matrix $W_{\theta_0} \equiv \int_0^{\pi} f^2(\lambda) G(\lambda; \theta_0) G(\lambda; \theta_0)'$ is non-singular. If $a_0 = 0_p$, then*

$$-4 \log R_n(\theta_0) \xrightarrow{d} \chi_p^2 \quad \text{as } n \to \infty. \tag{20}$$

The FDEL result is similar to the EL case with iid data, except that a scalar -4 appears rather than -2; this is because $n_0 \approx n/2$ values are probability profiled in (19) rather than n. Several important spectral estimating functions, including the Whittle-type function $\tilde{G}^{whit}$ in Example 8.1, satisfy (16) with a zero-valued a_0. Assumptions on the innovations in (15) and the estimating functions can be weakened in Theorem 4, and it is possible to use a tapered periodogram $I_n^{tap}(\lambda) = |\sum_{t=1}^{n} w_n(t) X_t e^{-\imath t\lambda}|^2 / [2\pi \sum_{t=1}^{n} w_n^2(t)]$, $\lambda \in \Pi$, in place of I_n for defining (19), where the taper weights are of the form (9). Then, $-4t_n \log R_n(\theta_0)$ has a χ_p^2-limit where $t_n = n^{-1}[\sum_{t=1}^{n} w_n^2(t)]^2 / \sum_{t=1}^{n} w_n^4(t)$ is a scaling correction for the taper, where $t_n = 1$ holds as in (20) in the untapered case $w_n(\cdot) = 1$; see Nordman (2009).

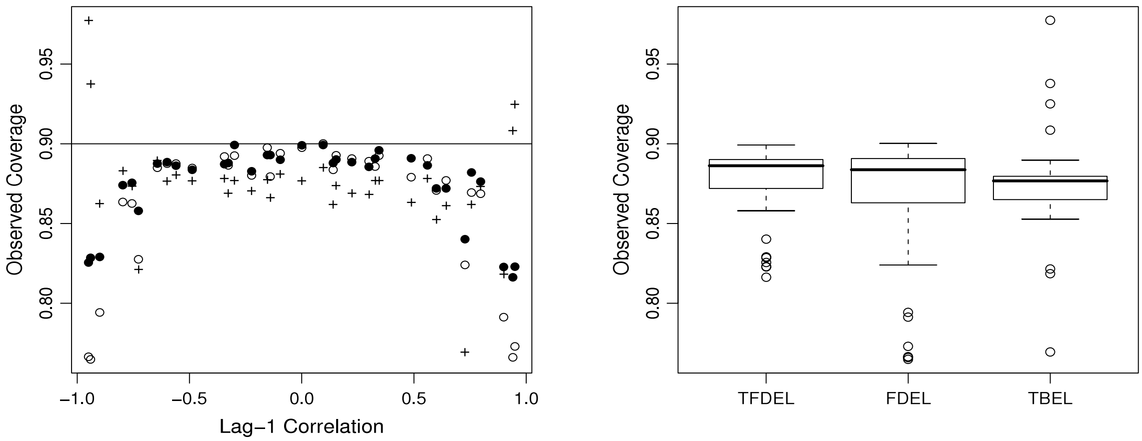

Fig. 2. Boxplots of observed coverages of 90% CIs for the lag-1 autocorrelation θ over 35 ARMA(1, 1) processes with sample size $n = 250$ as well as a plot of coverages for tapered-FDEL/TFDEL (•), FDEL (○) and TBEL (+) methods for each process/θ value.

As a small illustration, we considered 90% CIs for the lag-1 autocorrelation $\theta = r(1)/r(0)$ from 35 different, real-valued ARMA(1, 1) processes $X_t = \phi X_{t-1} + Z_t + \varphi Z_{t-1}$ with iid $N(0, 1)\{Z_t\}$ and parameter combinations $|\phi| \in \{0, 0.3, 0.6, 0.9\}$, $|\varphi| \in \{0, 0.4, 0.8\}$. For size $n = 300$ samples, Fig. 2 shows coverage results using tapered-FDEL, FDEL and TBEL (with $\ell = 2n^{1/3}$) with a taper window w_{trap} (Section 5) to define tapered-FDEL/TBEL. While both FDEL and TBEL methods apply to the autocorrelation parameter, the coverage results for FDEL are generally better due to the method's data transformation.

An alternative model-based version of FDEL exists for inference on the parameter θ when the true spectral density f is assumed to lie in a parametric family $\mathcal{F} \equiv \{f_\theta : \theta \in \Theta\}$, but the joint probability distribution of $X_1, \ldots, X_n$ is otherwise unspecified. Define the model-based version of the FDEL ratio statistic as

$$R_{n,\mathcal{F}}(\theta) = \left\{ \prod_{i=1}^{n_0} n_0 p_i : 0 \le p_i, \sum_{i=1}^{n_0} p_i = 1, \pi \sum_{i=1}^{n_0} p_i G(\lambda_{i,n}; \theta)[I_n(\lambda_{i,n}) - f_\theta(\lambda_{i,n})] = 0_p \right\}.$$

The key difference between $R_n(\theta)$ and $R_{n,\mathcal{F}}(\theta)$ is that the model-based version makes use of the fact that $\mathrm{E}I_n(\lambda)$ is approximately equal to $f_\theta(\lambda)$ under θ, while $R_n(\theta)$ in (19) characterizes the true value θ_0 purely by the moment condition (16). The following gives the limit distribution of the model-based FDEL statistic $R_{n,\mathcal{F}}(\theta)$ (cf. Nordman and Lahiri, 2006).

Theorem 5. *In addition to the assumptions of Theorem 4, suppose that $f = f_{\theta_0} \in \mathcal{F}$; and that each component of $f_{\theta_0} G(\cdot; \theta_0)$ is Lipschitz of order greater than $1/2$ on $[0, \pi]$. Then, if $a_0 = 0_p$ in (16),*

$$-2 \log R_{n,\mathcal{F}}(\theta_0) \xrightarrow{d} \chi_p^2 \quad \text{as } n \to \infty. \tag{21}$$

Further, (21) holds even for $a_0 \neq 0_p$ if $\kappa_{4,\varepsilon} \equiv \mathrm{E}(\varepsilon_1^4) - 3[\mathrm{E}(\varepsilon_1^2)]^2 = 0$.

There are important differences in the distributional results for $-4 \log R_n(\theta_0)$ and $-2 \log R_{n,\mathcal{F}}(\theta_0)$ from Theorems 4–5. One is that $-2 \log R_{n,\mathcal{F}}(\theta_0)$ *requires* the true spectral density f to belong to a model class $\mathcal{F}$ (i.e, $f = f_{\theta_0}$) in addition to the spectral moment condition (16), involving $a_0 \in \mathbb{R}^p$. Additionally, the chi-square limit for the model-based $-2 \log R_{n,\mathcal{F}}(\theta_0)$ holds for non-zero a_0 in (16) *only if* the iid innovations $\{\varepsilon_t\}$ in (15) have a fourth order cumulant $\kappa_{4,\varepsilon}$ which equals zero (i.e., Gaussian processes $\{X_t\}$). The condition $a_0 = 0_p$ is generally required in FDEL to avoid invalid self-studentization steps, which arise by treating the complete collection $\{I_n(\lambda_{j,n})\}_{j=1}^{n_0}$ of periodogram ordinates as independent when they are generally not (cf. Dahlhaus and Janas, 1996, p. 1939). Differences between the two FDEL formulations are particularly important in Whittle estimation.

Remark 6 (*Whittle Estimation, continued*)**.** Monti (1997) proposed a FDEL version for Whittle estimation with linear processes as in (15), using the previously mentioned parametric model class $\mathcal{F} \equiv \{f_\theta = (2\pi)^{-1}\sigma^2 k_\beta : \theta = (\sigma^2, \beta')' \in (0, \infty) \times \mathbb{R}^{p-1}\}$ from (17). In our presentation, Monti (1997)'s FDEL corresponds to the model-based $-2 \log R_{n,\mathcal{F}}(\theta)$ using the estimating functions G^{whit} from Example 8.1. These estimating functions involve a non-zero spectral moment a_0 in (16) due to the innovation σ^2 parameter in (18). Hence, Monti's (1997) chi-square distributional result generally requires $\kappa_{4,\varepsilon} = 0$ (i.e., Gaussian processes) in addition to the model class $\mathcal{F}$ being correctly specified. In contrast, inference about the model parameters $\beta \in \mathbb{R}^{p-1}$ is more broadly valid with the FDEL version $-4 \log R_n(\beta)$ using the estimating functions $\tilde{G}^{whit}(\cdot; \beta)$ for which (16) holds with $a_0 = 0_{p-1}$ and $-4 \log R_n(\beta_0)$ has a χ_{p-1}^2 limit. This approach does not directly consider σ^2, but an estimate $\hat{\sigma}^2 = 2 \int_0^\pi \sigma^2 I_n(\lambda) k_{\hat{\beta}}(\lambda) d\lambda$ follows from estimated values of β. Applying $-4 \log R_n(\beta)$ also does not require the model class to be correctly specified, only that the spectral moment condition (16) hold for $\tilde{G}^{whit}(\cdot; \beta)$ with $a_0 = 0_{p-1}$.

Independently, Ogata (2005) also suggested the FDEL version based estimating functions $\tilde{G}^{whit}(\cdot;\beta)$ for Whittle estimation and established a distributional result as in Theorem 4. Chan and Liu (2012) proved that Monti's version of FDEL for Whittle estimation is Bartlett correctable (i.e., improving the chi-square calibration with a mean adjustment, cf. Hall and La Scala, 1990), assuming that the process is Gaussian. Ogata and Taniguchi (2010) have extended FDEL-based Whittle estimation to multivariate linear processes.

Remark 7. The FDEL framework also allows point estimation as well as tests of parameter values and moment conditions with possibly over-identifying estimating functions; see Nordman and Lahiri (2006). Ogata (2012) further extended a FDEL framework with general estimating functions to locally stationary, multivariate linear processes.

9. Data example

Our goal here is to illustrate and compare some of the EL methods described in previous sections. Fig. 3 displays a time series $X_1,\ldots,X_{67}$ of annual U.S. unemployment rates over the years 1947–2013 (given as a percentage of the civilian work force of age 16 years or older), which we consider to be a realization of a stationary process; data are available from the U.S. Bureau of Labor Statistics.

We first consider estimating the mean $\mu = \mathrm{E}X_t$ unemployment rate with a 90% confidence interval (CI) based on BEL or TBEL methods from Sections 4–5, respectively. This requires selecting an appropriate block length ℓ for each method. As a compounding issue, little is currently known about theoretically optimal block lengths for coverage accuracy. However, we describe some general approaches for block selection in the following.

One strategy for choosing a block length for BEL borrows from spectral density estimation. In its asymptotic mechanics, BEL uses data blocks to form a block-based variance estimator for purposes of studentization which is asymptotically equivalent to a spectral density estimator (at the origin) based on Bartlett's lag-window kernel (Kitamura, 1997, p. 2093); the TBEL analogously involves a block-based variance estimator related to a tapered block bootstrap estimator of Paparoditis and Politis (2001). In either case, data-driven rules exists for selecting block sizes (e.g., bandwidths) which are MSE-optimal for spectral density/variance estimation, and these block selections can then be applied to BEL/TBEL. For example, applying a procedure of Andrews (1991, pp. 834–835), we fit an approximating AR(1) model to the unemployment data and use this to estimate an optimal Bartlett kernel bandwidth as $\ell = 11$, which then produces a BEL 90% CI for the mean unemployment rate as (5.16, 6.49); estimating the same Bartlett kernel bandwidth with a different method by Politis and White (2004) gives a block length $b = 3$ and then a 90% BEL interval as (5.31, 6.35). The block selection rule of Paparoditis and Politis (2001) (with a pilot bandwidth $n^{1/5}$, $n = 67$) yields a block $b = 5$ for TBEL and 90% CI as (5.25, 6.39). While the motivation by spectral estimation is reasonable, there are no theoretical guarantees that such block choices are indeed "optimal" for block-based EL methods, and different rules for block selection may lead to different CIs. Another approach for choosing EL block lengths is the "minimum volatility" method of Politis et al. (1999, Section 9.3.2). While purely heuristic, its basic principle is that approximately correct block lengths for inference might be characterized by confidence regions with stable behavior as a function of ℓ. Hence, in creating EL intervals/regions over a range of ℓ, an adequate block length might then be chosen by visual inspection. To illustrate, Fig. 3 shows the widths of 90% BEL CIs as a function of ℓ which become stable around $\ell = 11$, suggesting this as an appropriate block choice and CI for μ as (5.16, 6.49) by minimum volatility; for TBEL, the minimum volatility suggests a block $\ell = 14$ and 90% CI (5.16, 6.52) in agreement with the BEL method. While also having no theoretical guarantees, an advantage of minimum volatility is that the block selection approach applies generally, whereas spectral bandwidth estimators often require tailoring to each inference problem.

We next demonstrate EL inference with smooth function model parameters and general estimating functions. Fig. 3 shows the sample partial autocovariance function (PACF) for the unemployment data, which appears large at the first lag and, to a lesser extent, at the second lag. An AR(2) model approximation may be appropriate if a test of the process PACF at lag 2, given by $\theta = [\gamma(2)\gamma(0) - \gamma(1)^2]/[\gamma(0)^2 - \gamma(1)^2]$ for $\gamma(k) = \mathrm{Cov}(X_1, X_{1+|k|})$, $k \in \mathbb{Z}$, indicates a difference from zero. Defining observations

$$Y_t = \left(\sum_{i=0}^{2} X_{t-i}/3, \sum_{i=0}^{2} X_{t-i}^2/3, \sum_{i=0}^{1} X_{t-i}X_{t-i+1}/2, X_tX_{t+2}\right)', \quad t = 3,\ldots,67,$$

the parameter θ fits into the smooth function framework (8) as $\theta = H(\mathrm{E}Y_t)$ for $H(x_1,x_2,x_3,x_4) = [(x_4 - x_1^2)(x_2 - x_1^2) - (x_3 - x_1^2)^2]/[(x_2 - x_1^2)^2 - (x_3 - x_1^2)^2]$. To test $H_0 : \theta = 0$, we use a chi-square (df $= 1$) distribution from Theorem 2 to compute p-values from log-BEL ratio statistics $-2\ell^{-1}\log R_n(\theta)$ at $\theta = 0$ over a series of ℓ, as shown in Fig. 3. By minimum volatility, we select a block $\ell = 4$ with a resulting BEL p-value of 5e − 06, indicating a significant PACF at lag 2.

We next consider fitting an AR(2) model $X_t = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \epsilon_t$, with white noise errors ε_t. To estimate the AR coefficients (ϕ_1, ϕ_2), we apply BEL to the sample-mean centered data $Z_t = X_t - \bar{X}_{67}$, $t = 1,\ldots,67$ with estimating functions

$$g(Z_t, Z_{t-1}, Z_{t-2}; \phi_1, \phi_2) = [Z_{t-1},\ Z_{t-2}]'\,(Z_t - \phi_1 Z_{t-1} + \phi_2 Z_{t-2}), \quad t = 3,\ldots,67,$$

which have mean (approximately) zero under an AR(2) assumption. (Technically, an additional estimating function can be included to account for the unknown mean μ but we ignore this for simplicity; additionally, one could apply the

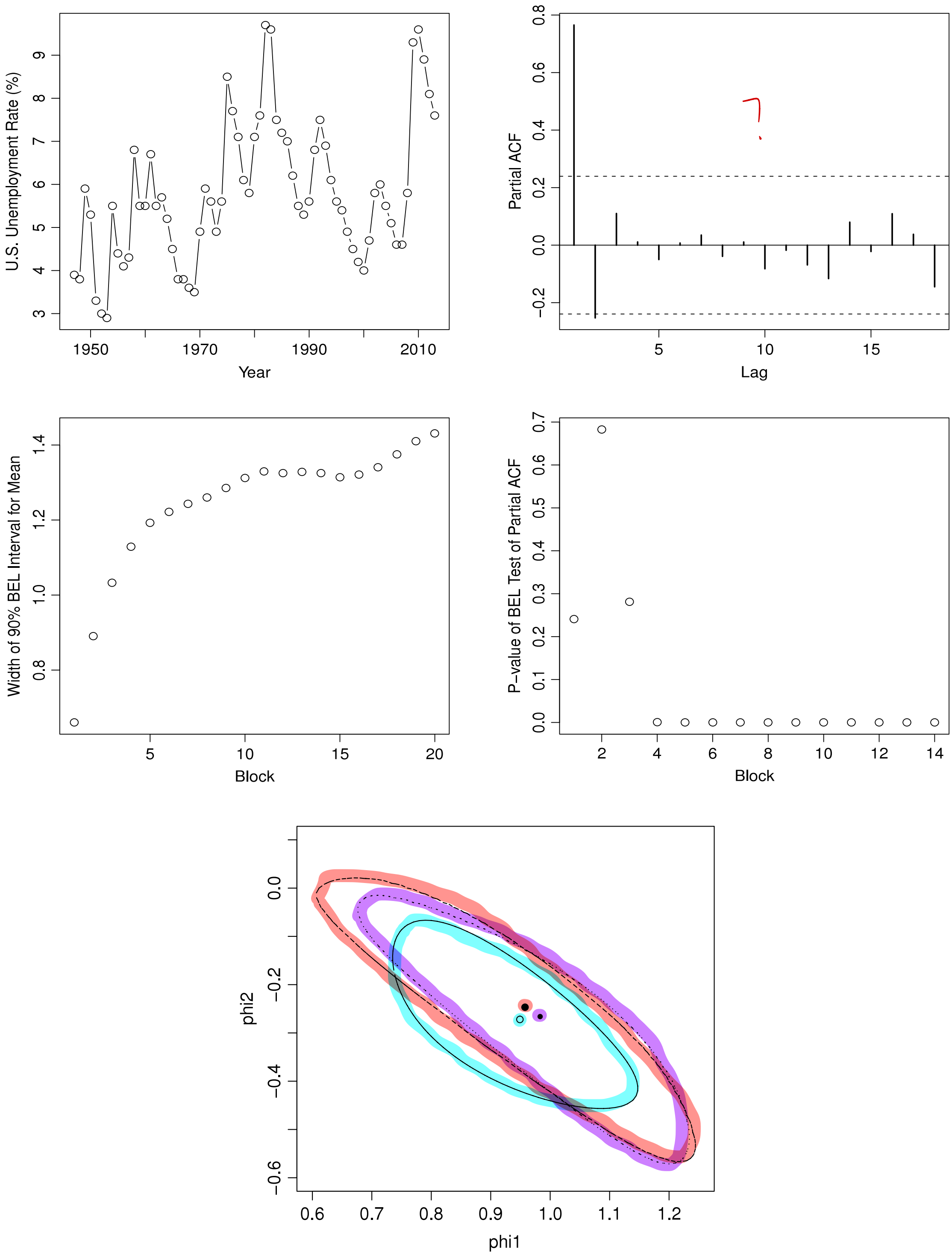

Fig. 3. (1st row, left) Plot of U.S. Annual Unemployment Rates from 1947–2013; (1st row, right) Sample partial ACF; (2nd row, left) Plot of widths of 90% BEL CIs for μ as a function of block length ℓ; (2nd row, right) Plot of p-values of BEL tests for a zero (lag 2) partial ACF as a function of ℓ; (Bottom) 90% confidence regions for AR(2) parameters (ϕ_1, ϕ_2) based on BEL (−−), TBEL ($\cdots$) and FDEL (−) methods; corresponding maximum EL point estimates denoted as •, · and ◦, respectively.

"unadjusted" EL version of Section 3.1 for AR models, but BEL applies under more general innovation assumptions.) By minimal volatility on the volumes of the confidence regions, we select a BEL block size of $\ell = 6$ and set a 90% confidence region for (ϕ_1, ϕ_2) as shown in Fig. 3; analogously, using TBEL, we select a block $\ell = 5$ producing a smaller confidence region in Fig. 3. To estimate the AR coefficients (ϕ_1, ϕ_2), we can also apply the FDEL method described in Example 8.1 of Section 8 along with two Whittle-based spectral estimating functions

$$\tilde{G}^{whit}(\lambda; \phi_1, \phi_2) = [\phi_1 + (\phi_2 - 1)\cos x,\ \ \phi_2 - \cos(2x) + \phi_1 \cos x]', \quad \lambda \in [-\pi, \pi];$$

these are the partial derivatives $\partial[1/k_{\phi_1,\phi_2}(\lambda)]/\partial\phi_i$ where $k_{\phi_1,\phi_2}(\lambda) = |1 - \phi_1 e^{\imath\lambda} - \phi_2 e^{\imath 2\lambda}|^2$ is proportional to the spectral density of an AR(2) model. Using the FDEL log-ratio $-4\log R_n(\phi_1, \phi_2)$ and Theorem 4, we calibrate a 90% confidence region shown in Fig. 3. Due to the data transformation, we might expect the FDEL region to be the most precise here, as evidenced in the figure.

We conclude with EL tests to assess the quality of an AR(2) model fit. Using FDEL with estimated residuals $\hat{\varepsilon}_t = Z_t - \hat{\phi}_1 Z_{t-1} - \hat{\phi}_2 Z_{t-2}$, $t = 3, \ldots, 67$ and estimating functions $g(\lambda) = [\cos(\lambda), \cos(2\lambda)]'$, we test whether the innovation autocorrelations appear to be zero at lags 1 and 2 (i.e., assessing if the moment condition (16) holds with $a_0 = (0, 0)'$); the resulting FDEL p-value is 0.83 using a χ^2_2 calibration, supporting a white noise assumption of the residuals. Similarly, using BEL to test whether the residual covariances appear to be zero at lags 1 and 2 with two estimating functions $g(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}) = [\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} + \hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-2},\ \ \hat{\varepsilon}_t\hat{\varepsilon}_{t-2}]$, $t = 3, \ldots, 67$, we select a block $\ell = 7$ by minimal volatility and a similar p-value of 0.87. Changing the estimating functions in the BEL method to $g(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}) = [\hat{\varepsilon}_t\hat{\varepsilon}^2_{t-1} + \hat{\varepsilon}_{t-1}\hat{\varepsilon}^2_{t-2},\ \ \hat{\varepsilon}_t\hat{\varepsilon}^2_{t-2}]$, $t = 3, \ldots, 67$, and testing if these have approximately mean zero gives an assessment of independence in the innovations; a BEL p-value of 0.0013 ($\ell = 7$) suggests the innovations are not independent.

10. Empirical likelihood under long-range dependence

Suppose $\{X_t\}$ is a real-valued, stationary process having an integrable spectral density function $f(\lambda)$, $\lambda \in \Pi = [-\pi, \pi]$ that satisfies

$$f(\lambda) \sim \mathcal{D}_\alpha |\lambda|^{-\alpha} \quad \text{as } \lambda \to 0 \tag{22}$$

for some $\alpha \in [0, 1)$ and positive constant $\mathcal{D}_\alpha > 0$ (where $\sim$ denotes that ratio of terms equals one in the limit). When $\alpha = 0$, we refer to the process $\{X_t\}$ as weakly or short-range dependent (SRD) and mixing assumptions typically imply $\alpha = 0$. When $\alpha > 0$, we say $\{X_t\}$ is strongly or long-range dependent (LRD). This classification of $\{X_t\}$ as SRD or LRD is common, in which long-range dependence (LRD) entails a pole of f at the origin (Hosking, 1981; Beran, 1994). Time series exhibiting LRD have a characteristically slow decay of process autocovariances $r(k) = \mathrm{Cov}(X_0, X_k)$ given by

$$r(k) \sim \mathcal{C}_\alpha |k|^{-(1-\alpha)} \quad \text{as } |k| \to \infty, \tag{23}$$

for some non-zero real constant $\mathcal{C}_\alpha$. (The behavior (22) of the spectral density f and the autocovariance decay (23) are closely related under LRD; see Robinson (1995a, p. 1634).) Hence, while the sum of autocovariances $\sum_{k=0}^{\infty} r(k)$ diverges under LRD (23), short-range dependence (SRD) is typically associated with rapidly decaying covariances such that $\sum_{k=0}^{\infty} |r(k)| < \infty$. Until now, we have focused on EL methods for weakly dependent or SRD time series, and Sections 10.1–10.2 summarize some EL extensions to LRD.

10.1. EL in the time domain

Suppose $X_1, \ldots, X_n \in \mathbb{R}$ arise from a stationary process $\{X_t\}$ having true mean $\mathrm{E}X_t = \mu_0 \in \mathbb{R}$, which could be either SRD or LRD under (22). Consider inference on the process mean parameter using the BEL ratio $R_n(\mu)$, as given in (6), based on length ℓ blocks. The main distributional result $-2\ell^{-1}R_n(\mu_0) \xrightarrow{d} \chi^2_1$ for BEL under SRD (cf. Section 4) *fails* to hold under LRD. This is because BEL again involves a type of self-studentization intended for SRD, but variances of samples averages $\mathrm{Var}(n^{-1}\sum_{t=1}^{n} X_t)$ exhibit a much slower decay $O(n^{-\alpha})$, $\alpha \in (0, 1)$, under LRD compared to $O(n^{-1})$ under SRD. However, BEL may be modified for LRD as described next; see Nordman et al. (2007) for generalizations.

Define a scaling factor $c_n = \ell^{-1}(\ell/n)^\alpha$ based on the memory exponent $\alpha \in [0, 1)$ in (22)–(23) and a version $\hat{c}_n = \ell^{-1}(\ell/n)^{\hat{\alpha}_n}$ based on an estimator $\hat{\alpha}_n$ of α from $X_1, \ldots, X_n$.

Theorem 6. *Suppose the process $\{X_t\}$ has a spectral density f satisfying (22) with $\alpha \in [0, 1)$ and has a linear form as in (15) with iid mean-zero innovations $\{\varepsilon_t\}$ such that $\mathrm{E}(\varepsilon^q) < \infty$ holds for an even integer $q \geq 4/(1-\alpha)$. When $\alpha > 0$, assume (23) holds; when $\alpha = 0$, suppose f is bounded on any compact subinterval of $(0, \pi]$. Additionally, if $\ell^{-1} + \ell^2/n \to 0$ as $n \to \infty$, then at the true process mean $\mathrm{E}X_t = \mu_0$,*

$$-2c_n \log R_n(\mu_0) \xrightarrow{d} \chi^2_1 \quad \text{as } n \to \infty. \tag{24}$$

Additionally, if $|\hat{\alpha}_n - \alpha| \log n \xrightarrow{p} 0$, then (24) holds upon replacing c_n with $\hat{c}_n$.

The scaling c_n corrects for the effect of the strong dependence among the data blocks. Under SRD ($\alpha = 0$), this becomes $c_n = \ell^{-1}$ and the distributional result (24) reduces to the previous Wilks theorem (7) for BEL. The several estimators $\hat{\alpha}$ of the memory exponent α exist (cf. Robinson, 1995a,b) with convergence rates exceeding the condition in Theorem 6.

10.2. EL in the frequency domain

For a real-valued process $\{X_t\}$ as in (15), a FDEL method under LRD can be formulated in the exact same manner as in the SRD case (Section 8) based on a vector of spectral estimating functions $G(\lambda; \theta) : \Pi \times \Theta \to \mathbb{R}^p$ which target a true parameter value $\theta_0 \in \Theta \subset \mathbb{R}^p$ satisfying a spectral moment condition (16). The main difference is that, under LRD, the process spectral density f is unbounded at the origin (22) and some care is needed in selecting estimating functions for valid FDEL inference. In particular, if process spectral density f is continuous on $(0, \pi]$ with $f(\lambda) \le C_1|\lambda|^{-\alpha}$, $\lambda \in \Pi$, for some $\alpha \in [0, 1)$ and $C_1 > 0$, we require $\|G(\lambda; \theta_0)\| \le C_2|\lambda|^{\gamma}$, $\lambda \in \Pi$, to hold at the true θ_0 for some $\gamma \in [0, 1)$, $C_2 > 0$ where $\alpha - \gamma < 1/2$ to ensure $\int_0^{\pi} f^2(\lambda)\|G(\lambda; \theta_0)\|^2 d\lambda < \infty$. Whittle estimation (Example 8.1, Section 8) is an important application of FDEL, for which these conditions are often satisfied.

Recall that two FDEL versions are possible. One involves the FDEL ratio statistic $R_n(\theta)$ in (19) based on the spectral estimating functions and the moment condition (16); the other FDEL version is based on a parametric spectral density class $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ (assumed to contain the true spectral density as $f = f_{\theta_0}$) and uses a model-based ratio statistic $R_{n,\mathcal{F}}(\theta)$. The following provides distributional results for FDEL log-ratio statistics under LRD (and SRD for $\alpha = 0$), which are extensions of Section 8 results; see Nordman and Lahiri (2006) for precise regularity conditions.

Theorem 7. *Suppose that $\{X_t\}$ is a linear process as in (15) with $E(\varepsilon_1^2) > 0$, $E(\varepsilon_1^8) < \infty$; that $f(\lambda) \le C_1|\lambda|^{-\alpha}$ and $\|G(\lambda; \theta_0)\| \le C_2|\lambda|^{\gamma}$, $\lambda \in \Pi$ for some $\alpha, \gamma \in [0, 1)$ with $\alpha - \gamma < 1/2$; that $\int_0^{\pi} f^2(\lambda)G(\lambda; \theta_0)G(\lambda; \theta_0)'$ is non-singular; and that $G(\lambda; \theta_0)$ satisfies some regularity conditions. Then, if (16) holds $a_0 = 0_p$,*

$$-4 \log R_n(\theta_0) \xrightarrow{d} \chi_p^2 \quad \text{as } n \to \infty.$$

Assuming additionally for the model-based FDEL that $f = f_{\theta_0} \in \mathcal{F}$, then

$$-2 \log R_{n,\mathcal{F}}(\theta_0) \xrightarrow{d} \chi_p^2 \quad \text{as } n \to \infty. \tag{25}$$

Further, (25) holds even for $a_0 \neq 0_p$ if $\kappa_{4,\varepsilon} \equiv \mathrm{E}(\varepsilon_1^4) - 3[\mathrm{E}(\varepsilon_1^2)]^2 = 0$.

The same FDEL framework for Whittle estimation in Example 8.1 [involving spectral densities $f_\theta = (2\pi)^{-1}\sigma^2 k_\beta$ of the form (17) (i.e., $\theta = (\sigma^2, \beta')'$, $\sigma^2 \in (0, \infty)$, $\beta \in \mathbb{R}^{p-1}$) and estimating functions $G^{whit}(\cdot; \theta)$ or $G^{whit}(\cdot; \beta)$ satisfying (18)] applies to important classes of LRD time processes, such as fractional Gaussian processes (Mandelbrot and Van Ness, 1968) and the fractional autoregressive integrated moving average (FARIMA) processes (Granger and Joyeux, 1980; Hosking, 1981). Recently, Yau (2012) extended Monti (1997)'s model-based FDEL result for Whittle estimation in FARIMA densities, showing $-2 \log R_{n,\mathcal{F}}(\theta)$ based on the estimating functions $G^{whit}(\cdot; \theta)$ to have a chi-square limit. As in Remark 6 (Section 8), these estimating functions have a non-zero spectral moment a_0 in (16) due to the innovation σ^2 parameter, so that a chi-square limit for $-2 \log R_{n,\mathcal{F}}(\theta_0)$ is generally valid only when $\kappa_{4,\varepsilon} = 0$ (i.e., Gaussian FARIMA processes). However, as in Remark 6, the FDEL version $-4 \log R_n(\beta_0)$ can be generally applied using the estimating functions $\tilde{G}^{whit}(\cdot; \beta)$ for the (non-σ^2) FARIMA parameters, which satisfy (16) with zero-valued a_0.

11. Concluding remarks and open research questions

A main challenge in extending EL to dependent data is creating a likelihood function by probability profiling suitable functions of data in a way accommodates the dependence structure for a given inference situation. At issue, an EL method needs to be correctly self-studentize, or intrinsically estimate variances in its mechanics, to produce non-degenerate distributional limits for log-ratio statistics (e.g., chi-square limits). The iid version of EL (Section 2) probability profiles individual observations and a similar approach can apply to specialized inference problems with time series (cf. Section 3). More broadly, we have reviewed two general strategies for extending EL to a variety of time series inference in the presence of autocorrelation: *data blocking* to locally capture time dependence (Sections 4–7) and *data transformations* (Section 8) to modify the dependence. We conclude by mentioning several open research questions regarding EL for dependent data.

1. *Block length selection.* A significant open problem is determining a large-sample form, or asymptotic expansion, of the theoretically optimal block length ℓ for the coverage accuracy of BEL regions. At an optimal block size, it is of interest to understand and compare the coverage accuracy rate for BEL against other interval methods. A larger issue is developing well-founded data-driven rules for selecting block lengths, and theoretical guidance may help this effort. Compared to EL, the literature on optimal block size and block selection is far more developed for block bootstraps (cf. Lahiri, 2003, ch. 5; Politis and White, 2004).
2. *Effects of data smoothing windows.* Data blocks in the BEL method can be replaced with smoothed windows of data (cf. Smith, 1997; Kitamura and Stutzer, 1997) or tapered block versions (Sections 5, 7). One important topic, which has received little attention, is investigating and characterizing properties of the smoothing windows/tapers to improve the performance of BEL. This is akin to studies of the effect of different kernels in spectral estimation and bootstrap methodology (cf. Künsch, 1989; Andrews, 1991; Paparoditis and Politis, 2001).

3. *Higher order accuracy*. For iid data, the coverage accuracy of EL regions can be improved with Bartlett corrections (cf. Section 2) and the extent of the improvement has been well-studied. Much less is known about similar adjustments for time series. Can the higher order accuracy of BEL/TBEL regions be improved by Bartlett corrections or other resampling-based calibrations? If so, can the extent of this be quantified? Bartlett corrections should be possible, quite generally, for frequency domain versions of EL (cf. Section 8) due to approximate independent of periodogram ordinates; a general and formal treatment is potentially useful but currently lacking.
4. *Fixed-bandwidth versions of EL*. In asymptotic expansions of BEL log-ratio statistics, data blocks provide a type of block-based variance estimator for purposes of normalizing scale and obtaining chi-square limits. Such variance estimators have equivalences to lag window estimators involving sample covariances, kernel functions, and bandwidths ℓ with similar behavior to block lengths $\ell^{-1} + n/\ell \to 0$ as $n \to \infty$. That is, BEL intervals share connections to normal theory intervals based on studentization with consistent lag window estimates. However, some numerical and theoretical evidence exists that normalizing scale with inconsistent lag window estimates (i.e., based on a fixed bandwidth ratio $\ell/n = C \in (0, 1]$) may provide more accurate interval estimators (cf. Kiefer and Vogelsang, 2005; Shao, 2010). This suggests that development of EL methods with non-standard block formulations, based fixed bandwidth ratios, may lead to improved coverage properties.
5. *Data transformations in EL*. The frequency domain EL (FDEL) of Section 8 is based one type of data transformation involving the discrete Fourier transform. However, this approach is generally limited to inference about spectral parameters in ratio form (e.g., auto-correlations not auto-covariances) based on spectral estimating functions with mean zero (16); see Lahiri (2003, ch. 8) for more details. It may be possible to extend FDEL to non-zero-mean estimating functions and include a larger variety of spectral parameters; Kreiss and Paparoditis (2003) have developed analogous generalizations for the frequency domain bootstrap. FDEL for multivariate data has not been generally considered. Finally, other data transformations may be possible for EL (cf. Kitamura, 2006). One approach is to fit an appropriate time series model, such as an AR(p) model with AR order $p \to \infty$ as $n \to \infty$, to approximate the original series and apply EL methods to the approximation; a similar principle underlies the sieve bootstrap (cf. Kreiss et al., 2011).
6. *EL generalizations to other dependent data*. Many versions of EL have been developed with a focus on stationary time series, and less is known about EL for non-stationary data (though some treatments exist, cf. Remark 3, Section 4). Non-stationarity may be difficult to characterize generally, but EL methods developed to handle local (rather than global) stationarity may be potentially very useful, applying to time series with a slowly evolving stochastic structure. A framework for such processes is given by Dahlhaus (1996, 1997).

   Some EL extensions have ventured into spatial data. Nordman (2008a) developed a spatial BEL version for stationary, mixing spatial processes observed on a partial grid in $\mathbb{R}^d$, extending Kitamura's (1997) time series results ($d = 1$). Nordman (2008b) considered BEL spatial regression problems with general non-stochastic regressors, and Nordman and Caragea (2008)introduced a spatial BEL for estimating variogram model parameters. Recently, Kostov (2013) proposed a smoothed EL method for inference in spatial quantile regressions, and Kaiser and Nordman (2012) developed BEL goodness-of-fit tests for spatial Markov models. All of these works are limited to spatial data collected on a grid or lattice. However, far more diverse sampling structures exist for spatial data (e.g., stochastic spatial locations, infill sampling), for which extensions of the EL methodology could be quite useful; see Bandyopadhyay et al. (2012) for one EL example with irregularly located spatial data.

Acknowledgments

The authors are grateful to Editor and Professor Anirban DasGupta for his encouragement, patience, and hard work which made this manuscript possible.

References

Andrews, D.K.B., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Athreya, K.B., Lahiri, S.N., 2006. Measure Theory and Probability Theory. Springer, New York.
Baggerly, K., 1998. Empirical likelihood as a goodness of fit measure. Biometrika 85, 535–547.
Bandyopadhyay, S., Lahiri, S.N., Nordman, D., 2012. Frequency domain empirical likelihood method for irregularly spaced spatial data. Unpublished manuscript, Lehigh University, PA, USA.
Bartolucci, F., 2007. A penalized version of the empirical likelihood ratio for the population mean. Statist. Probab. Lett. 77, 104–110.
Beran, J., 1994. Statistical Methods for Long Memory Processes. Chapman and Hall, London.
Bravo, F., 2007. Empirical Likelihood Methods With Applications to Econometrics. In: Lecture Notes, Center for Applied Economics and Policy Research, Department of Economics, Indiana University.
Bravo, F., 2010. Nonparametric likelihood inference for general autoregressive models. Stat. Methods Appl. 19, 79–106.
Bravo, F., 2009. Blockwise generalized empirical likelihood inference for non-linear dynamic moment conditions models. Econom. J. 12, 208–231.
Bravo, F., 2005. Blockwise empirical entropy tests for time series regressions. J. Time Series Anal. 26, 185–210.
Bravo, F., 2002. Blockwise empirical Cressie–Read test statistics for α-mixing processes. Statist. Probab. Lett. 58, 319–325.
Brillinger, D.R., 1981. Time Series: Data Analysis and Theory. Holden-Day, San Francisco.
Chan, N.H., Ling, S., 2006. Empirical likelihood for GARCH models. Econometric Theory 22, 403–428.
Chan, N.H., Liu, L., 2012. Bartlett correctability of empirical likelihood for time series. J. Japan Statist. Soc. 40, 221–238.
Chen, S.X., Cui, H.J., 2006. On Bartlett correction of empirical likelihood in the presence of nuisance parameters. Biometrika 93, 215–220.
Chen, S.X., Cui, H.J., 2007. On the second order properties of empirical likelihood for generalized estimation equations. J. Econometrics 141, 492–516.
Chen, X., Cui, W., 2008. Empirical likelihood inference for partial linear models under martingale difference sequence. Statist. Probab. Lett. 78, 2895–2901.

Chen, S.X., Gao, J., 2007. An adaptive empirical likelihood test for parametric time series regression models. J. Econometrics 141, 950–972.
Chen, S.X., Gao, J., Li, M., 2012a. Simultaneous specification tests for the mean and variance structures of regression with applications to testing of diffusion models. Preprint Department of statistics, Iowa State University, Ames, IA.
Chen, S.X., Härdle, W., Li, M., 2003. An empirical likelihood goodness-of-fit test for time series. J. R. Stat. Soc. Ser. B 65, 663–678.
Chen, S.X., Peng, L., Qin, Y., 2009. Empirical likelihood methods for high dimension. Biometrika 96, 711–722.
Chen, Y., Song, L., Feng, J., 2012b. Empirical likelihood for threshold autoregressive models. Comm. Statist. Theory Methods 41, 970–982.
Chen, S.X., Van Keilegom, I., 2009. A review on empirical likelihood methods for regression. Test 18, 415–447.
Chen, J., Variyath, A.M., Abraham, B., 2008. Adjusted empirical likelihood and its properties. J. Comput. Graph. Statist. 17, 426–443.
Chen, S.X., Wong, C.M., 2009. Smoothed block empirical likelihood for quantiles of weakly dependent processes. Statist. Sinica 19, 71–81.
Chen, Y.-Y., Zhang, L.-X., 2010. Empirical Euclidean likelihood for general estimating functions under association dependence. Appl. Math. J. Chinese Univ. 25, 437–446.
Chuang, C.-S., Chan, N.H., 2002. Empirical likelihood for autoregressive models with applications to unstable time series. Statist. Sinica 12, 387–407.
Dahlhaus, R., 1996. On the Kulback-Leibler information divergence of locally stationary processes. Stochastic Process. Appl. 62, 139–168.
Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Ann. Statist. 25, 1–37.
Dahlhaus, R., Janas, D., 1996. A frequency-domain bootstrap for ratio statistics in time series analysis. Ann. Statist. 24, 1934–1963.
DiCiccio, T.J., Hall, P., Romano, J.P., 1991. Empirical likelihood is Bartlett-correctable. Ann. Statist. 19, 1053–1061.
Efron, B., 1981. Nonparametric standard errors and confidence intervals (with discussion). Canad. J. Statist. 9, 139–172.
Emerson, S.C., Owen, A., 2009. Calibration of the empirical likelihood method for a vector mean. Electron. J. Stat. 3, 1161–1192.
Fan, G.-L., Liang, H.-Y., 2010. Empirical likelihood inference for semiparametric model with linear process errors. J. Korean Statist. Soc. 39, 55–65.
Fan, G.-L., Liang, H.-Y., Huang, Z.-S., 2012. Empirical likelihood inference for partially time-varying coefficient models with dependent observations. J. Nonparametr. Stat. 24, 71–84.
Fox, R., Taqqu, M.S., 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Statist. 14, 517–532.
Granger, C.W.J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. J. Time Series Anal. 1, 15–29.
Guggenberger, P., Smith, R.J., 2008. Generalized empirical likelihood tests in time series models with potential identification failure. J Econometrics 142, 134–161.
Hall, P., 1992. The Bootstrap and Edgeworth expansion. Springer-Verlag, New York.
Hall, P., La Scala, B., 1990. Methodology and algorithms of empirical likelihood. Internat. Statist. Rev. 58, 109–127.
Hannan, E.J., 1973. The asymptotic theory of linear time-series models. J. Appl. Probab. 10, 130–145.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Harari-Kermadec, H., 2011. Regenerative block empirical likelihood for Markov chains. J. Nonparametr. Stat. 23, 781–802.
Hjort, N.L., McKeague, I.W., Van Keilegom, I., 2009. Extending the scope of empirical likelihood. Ann. Statist. 37, 1079–1111.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165–176.
Huang, Z., Zhou, Z., Jiang, R., Qian, W., Zhang, R., 2010. Empirical likelihood based inference for semiparametric varying coefficient partially linear models with error-prone linear covariates. Statist. Probab. Lett. 80, 497–504.
Joag-Dev, K., Proschan, F., 1983. Negative association of random variables with applications. Ann. Statist. 25, 2084–2102.
Kaiser, M.S., Nordman, D.J., 2012. Empirical likelihood for spatial model assessment. Stat. Interface 5, 303–318.
Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21, 1130–1164.
Kitamura, Y., 2006. Empirical likelihood methods in econometrics: Theory and practice. Cowles foundation discussion paper 1569.
Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Ann. Statist. 25, 2084–2102.
Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.
Künsch, H.R., 1989. The jackknife and bootstrap for general stationary observations. Ann. Statist. 17, 1217–1261.
Kostov, P., 2013. Empirical likelihood estimation of the spatial quantile regression. J. Geogr. Syst. 15, 51–69.
Kreiss, J.-P., Paparoditis, E., 2003. Autoregressive-aided periodogram bootstrap for time series. Ann. Statist. 31, 1923–1955.
Kreiss, J.-P., Paparoditis, E., Politis, D.N., 2011. On the range of validity of the autoregressive sieve bootstrap. Ann. Statist. 39, 2103–2130.
Lahiri, S.N., 2003. Resampling Methods for Dependent Data. Springer, New York.
Lahiri, S.N., Mukhopadhyay, S., 2012a. On the Mahalanobis-distance based penalized empirical likelihood method in high dimensions. Stat. Interface 5, 331–338.
Lahiri, S.N., Mukhopadhyay, S., 2012b. A penalized empirical likelihood method in high dimensions. Ann. Statist. 40, 2511–2540.
Lei, Q., Qin, Y., 2011. Empirical likelihood for quantiles under negatively associated samples. J. Statist. Plann. Inference 141, 1325–1332.
Lian, H., 2009. Empirical likelihood confidence intervals for nonparametric functional data analysis. Preprint, Division of Mathematical Sciences, Nanyang Technological University, Singapore.
Li, J., Liang, W., He, S., 2012a. Empirical likelihood for AR-ARCH models based on LAD estimation. Acta Math. Appl. Sin. 28, 371–382.
Li, J., Liang, W., He, S., Wu, X., 2010. Empirical likelihood for the smooth LAD estimator in infinite variance autoregressive models. Statist. Probab. Lett. 80, 1420–1430.
Lin, L., Zhang, R., 2001. Blockwise empirical Euclidean likelihood for weakly dependent processes. Statist. Probab. Lett. 53, 143–152.
Li, Y., Qin, Y., Lei, Q., 2012b. Confidence intervals for probability density functions under associated samples. J. Statist. Plann. Inference 142, 1516–1524.
Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications. SIAM Rev. 10, 422–437.
Meyn, S.P., Tweedie, R.L., 2009. Markov Chains and Stochastic Stability. Cambridge University Press, Cambridge, UK.
Monti, A.C., 1997. Empirical likelihood confidence regions in time series models. Biometrika 84, 395–405.
Mykland, P.A., 1995. Dual likelihood. Ann. Statist. 23, 395–405.
Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219–255.
Nordman, D.J., 2008a. Empirical likelihood for spatial regression. Metrika 68, 351–363.
Nordman, D.J., 2008b. A blockwise empirical likelihood for spatial data. Statist. Sinica 18, 1111–1129.
Nordman, D.J., 2009. Tapered empirical likelihood for time series data in time- and frequency-domains. Biometrika 96, 119–132.
Nordman, D.J., Caragea, P.C., 2008. Point and interval estimation of variogram models using spatial empirical likelihood. J. Amer. Statist. Assoc. 103, 350–361.
Nordman, D.J., Lahiri, S.N., 2006. A frequency domain empirical likelihood for short- and long-range dependence. Ann. Statist. 34, 3019–3050.
Nordman, D.J., Sibbertsen, P., Lahiri, S.N., 2007. Empirical likelihood for the mean under long-range dependence. J. Time Series Anal. 28, 576–599.
Ogata, H., 2005. Empirical likelihood approach for non-Gaussian stationary processes. Sci. Math. Jpn. 62, 429–438.
Ogata, H., 2012. Estimation for non-Gaussian locally stationary processes with empirical likelihood method. Adv. Decis. Sci. 2012/704693.
Ogata, H., Taniguchi, M., 2010. An empirical likelihood approach for non-Gaussian vector stationary processes and its application to minimum contrast estimation. Aust. N. Z. J. Stat. 52, 451–468.
Otsu, T., 2007. Penalized empirical likelihood estimation of semiparametric models. J. Multivariate. Anal. 98, 1923–1954.
Otsu, T., 2006. Generalized empirical likelihood inference for nonlinear and time series models under weak identification. Econometric Theory 22, 513–527.
Owen, A.B., 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. Chapman & Hall, London.
Owen, A.B., 1991. Empirical likelihood for linear models. Ann. Statist. 19, 1725–1747.
Paparoditis, E., Politis, D.N., 2001. Tapered block bootstrap. Biometrika 88, 1105–1119.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer-Verlag, New York.
Politis, D.N., White, H., 2004. Automatic block-length selection for the dependent bootstrap. Econometric Rev. 23, 53–70.
Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Ann. Statist. 22, 300–325.

Qin, J., Lawless, J., 1995. Estimating equations, empirical likelihood and constraints on parameters. Canad. J. Statist. 23, 145–159.
Qin, Y., Li, Y., 2011. Empirical likelihood for linear models under negatively associated errors. J. Multivariate Anal. 102, 153–163.
Qin, Y., Li, Y., Lei, Q., 2011. Empirical likelihood for probability density functions under negatively associated samples. J. Statist. Plann. Inference 141, 373–381.
Robinson, P.M., 1995a. Gaussian semiparametric estimation of long range dependence. Ann. Statist. 23, 1630–1661.
Robinson, P.M., 1995b. Log-periodogram regression of time series with long range dependence. Ann. Statist. 23, 1048–1072.
Shao, X., 2010. A self-normalized approach to confidence interval construction in time series. J. R. Stat. Soc. Ser. B Stat. Methodol. 72, 343–366.
Smith, R.J., 1997. Alternative semi-parametric likelihood approaches to generalized method of moments. Econ. J. 107, 503–519.
Su, L., White, H., 2012. Testing conditional independence via empirical likelihood. (Preprint) School of Economics, Singapore Management University.
Tang, C.Y., Leng, C., 2010. Penalized high-dimensional empirical likelihood. Biometrika 97, 905–920.
Tripathi, G., Kitamura, Y., 2003. Testing conditional moment restrictions. Ann. Statist. 31, 2059–2095.
Tsao, M., 2004. Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. Ann. Statist. 32, 1215–1221.
Velasco, C., 2009. Comments on: A review on empirical likelihood methods for regression. Test 18, 455–457.
Wilks, S.S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Statist. 9, 60–62.
Wu, R., Cao, J., 2011. Blockwise empirical likelihood for time series of counts. J. Multivariate Anal. 102, 661–673.
Xiong, X., Lin, Z., 2012. Empirical likelihood inference for probability density functions under association. J. Statist. Plann. Inference 142, 986–992.
Yajima, Y., 1989. A central limit theorem of Fourier transforms of strongly dependent stationary processes. J. Time Series Anal. 10, 375–383.
Yau, C.Y., 2012. Empirical likelihood in long-memory time series models. J. Time Series Anal. 33, 269–275.
Zhang, J., 2006. Empirical likelihood for NA series. Statist. Probab. Lett. 76, 153–160.
Zhang, J., 2007. Empirical likelihood ratio confidence interval for positively associated series. Acta Math. Appl. Sin. 23, 245–254.
Zhang, R., Peng, L., Qi, Y., 2012. Jackknife-blockwise empirical likelihood methods under dependence. J. Multivariate Anal. 104, 56–72.
Zhao, Z.-W., Wang, D.-H., 2011. Empirical likelihood for an autoregressive model with explanatory variables. Comm. Statist. Theory Methods 40, 559–570.
Zhao, Z.-W., Wang, D.-H., 2012. Statistical inference for generalized random coefficient autoregressive model. Math. Comput. Modelling 56, 152–166.