AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

A Non-standard Empirical Likelihood for Time Series

Daniel J. Nordman, Helle Bunzel and Soumendra N. Lahiri

CREATES Research Paper 2012-55

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

A Non-standard Empirical Likelihood for Time Series

Daniel J. Nordman, Helle Bunzel and Soumendra N. Lahiri

Dept. of Statistics, Iowa State University;

Dept. of Economics, Iowa State University

& CREATES, Aarhus University;

Dept. of Statistics, Texas A&M University

Abstract

Standard blockwise empirical likelihood (BEL) for stationary, weakly dependent time series requires specifying a fixed block length as a tuning parameter for setting confidence regions. This aspect can be difficult and impacts coverage accuracy. As an alternative, this paper proposes a new version of BEL based on a simple, though non-standard, data-blocking rule which uses a data block of *every* possible length. Consequently, the method involves no block selection and is also anticipated to exhibit better coverage performance. Its non-standard blocking scheme, however, induces non-standard asymptotics and requires a significantly different development compared to standard BEL. We establish the large-sample distribution of log-ratio statistics from the new BEL method for calibrating confidence regions for mean or smooth function parameters of time series. This limit law is not the usual chi-square one, but is distribution-free and can be reproduced through straightforward simulations. Numerical studies indicate that the proposed method generally exhibits better coverage accuracy than standard BEL.

Keywords: Brownian motion; Confidence Regions; Stationarity; Weak Dependence

MSC subject classiciation: Primary 62G09; Secondary 62G20, 62M10

Acknowledgments: This research is partially supported by NSF grant DMS-0906588, and NSF grant DMS-1007703 and NSA grant H98230-11-1-0130 (Lahiri).

1 Introduction

For independent, identically distributed data (iid), Owen (1988, 1990) introduced empirical likelihood (EL) as a general methodology for re-creating likelihood-type inference without a joint distribution for the data, as typically specified in parametric likelihood. However, the iid formulation of EL fails for dependent data by ignoring the underlying dependence structure. As a remedy, Kitamura (1997) proposed so-called blockwise empirical likelihood (BEL) methodology for stationary, weakly dependent processes, which has been shown to provide valid inference in various scenarios with time series (cf. Lin and Zhang, 2001; Bravo, 2005, 2009; Zhang, 2006; Nordman, Sibbertsen and Lahiri, 2007; Chen and Wong, 2009; Nordman, 2009; Wu and Cao, 2011; Lei and Qin, 2011). Similarly to the iid EL version, BEL creates an EL log-ratio statistic having a chi-square limit for inference, but the BEL construction crucially involves blocks of consecutive observations in time, rather than individual observations. This data-blocking serves to capture the underlying time dependence and related concepts have also proven important in defining resampling methodologies for dependent data, such as the moving block bootstrap of Hall (1985), Künsch (1989) and Liu and Singh (1992), and time subsampling methods of Carlstein (1986), Politis and Romano (1993), and Politis, Romano and Wolf (1999). However, the coverage accuracy of BEL can depend crucially on the block length selection, which is a fixed value $1 \leq b \leq n$ for a given sample size n, and appropriate choices can vary with the underlying process (a point briefly illustrated at the end of this section).

To advance the BEL methodology in a direction away from block selection with a goal of improved coverage accuracy, we propose an alternative version of BEL for stationary, weakly dependent time series, called an expansive block empirical likelihood (EBEL). The EBEL method involves a non-standard, but simple, data-blocking rule where a data block of every possible length is used. Consequently, the method does not require a block length choice. We investigate EBEL in the prototypical problem of inference about the process mean or a smooth function of means. For setting confidence regions for such parameters, we establish the limiting distribution of log-likelihood ratio statistics from the EBEL method. Because of the non-standard blocking scheme, the justification of this limit distribution requires a new and substantially different treatment compared to that of standard BEL (which closely resembles that of EL for iid data in its large-sample development, cf. Owen, 1990; Qin and Lawless, 1994). In fact, unlike with standard BEL or EL for iid data, the limiting distribution involved is non-standard and *not* chi-square. However, the EBEL limit law is distribution-free, corresponding to a special integral of standard Brownian motion on $[0, 1]$, and so can be easily approximated through simulation to obtain appropriate quantiles for calibrating confidence regions. In addition to avoiding block selection, we anticipate that the EBEL method may have generally better coverage accuracy than standard BEL methods, though formally establishing and comparing convergence rates is beyond the scope of this manuscript (and, in fact, optimal rates and block sizes for even standard BEL remain to be determined). Simulation studies, though, suggest that interval estimates from the EBEL method can perform much better than the standard BEL approach, especially when the later employs a poor block choice, and be less sensitive to the dependence strength of the underlying process.

The rest of manuscript is organized as follows. We end this section by briefly recalling the standard BEL construction with overlapping blocks and its distributional features. In Section 2, we separately describe the EBEL method for inference on process means and smooth function model parameters, and establish the main distributional results in both cases. These results require introducing a new type of limit law based on Brownian motion, which is also given in Section 2. As an additional feature with the approach, we also consider the possibility of "weighting" the data blocks in the EBEL construction, which influences the distributional limit. The idea of using weights with data blocks has parallels with other resampling methods for time series, such as the tapered block bootstrap (Paparoditis and Politis, 2001) and tapered BEL (Nordman, 2009). Section 3 provides a numerical study of the coverage accuracy of the EBEL method and comparisons to several existing versions of BEL. Section 4 offers some concluding remarks and heuristic arguments on the expected performance of EBEL. Proofs of the main results are deferred to Section 5.

To motivate what follows, we briefly recall the BEL construction, considering, for concreteness, inference about the mean $\mathrm{E}X_t = \mu \in \mathbb{R}^d$ of a vector-valued stationary stretch $X_1, \ldots, X_n$. Upon choosing an integer block length $1 \leq b \leq n$, a collection of maximally overlapping (OL) blocks of length b is given by $\{(X_i, \ldots, X_{i+b-1}) : i = 1, \ldots, N_b \equiv n - b + 1\}$. For a given $\mu \in \mathbb{R}^d$ value, each block in the collection provides a centered block sum $B_{i,\mu} \equiv \sum_{j=i}^{i+b-1}(X_j - \mu)$ for defining a BEL function

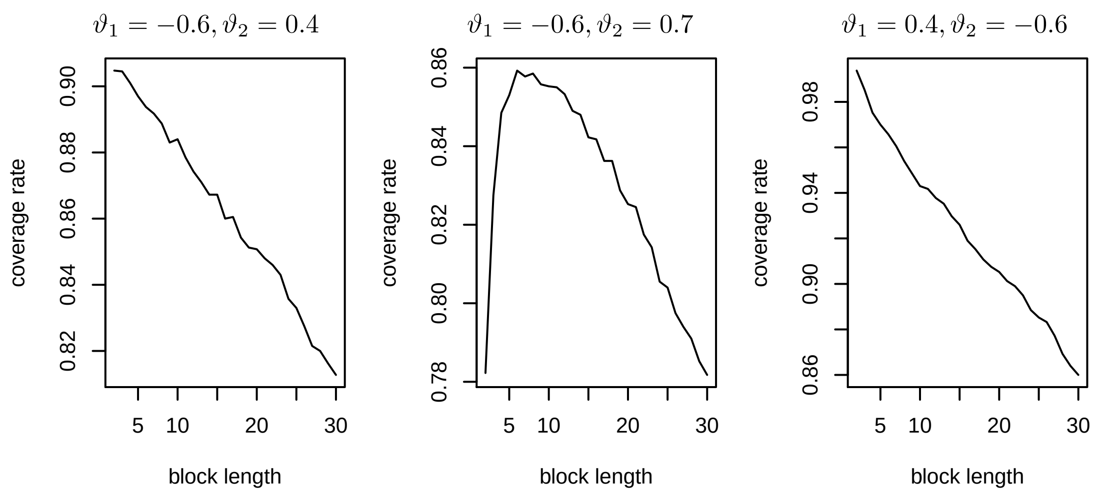
$$L_{\mathrm{BEL},n}(\mu) = \sup\left\{\prod_{i=1}^{N_b} p_i : p_i \geq 0, \sum_{i=1}^{N_b} p_i = 1, \sum_{i=1}^{N_b} p_i B_{i,\mu} = 0_d\right\} \tag{1}$$

and corresponding BEL ratio $R_{\mathrm{BEL},n}(\mu) = L_n(\mu)/N_b^{-N_b}$, where $0_d = (0, \ldots, 0)' \in \mathbb{R}^d$. The function $L_{\mathrm{BEL},n}(\mu)$ assesses the plausibility of a value μ by maximizing a multinomial likelihood from probabilities $\{p_i\}_{i=1}^{N_b}$ assigned to the centered block sums $B_{i,\mu}$ under a zero-expectation constraint. Without the linear mean constraint in (1), the multinomial product is maximized when each $p_i = 1/N_b$ (i.e., the empirical distribution on blocks), defining the ratio $R_{\mathrm{BEL},n}(\mu)$. Under certain mixing and moment conditions entailing weak dependence, and if the block b grows with the sample size n but at a smaller rate (i.e., $b^{-1} + b^2/n \to 0$ as $n \to \infty$), the log-EL ratio of the standard BEL has chi-square limit

$$-\frac{2}{b} \log R_{\mathrm{BEL},n}(\mu_0) \xrightarrow{d} \chi^2_d, \tag{2}$$

at the true mean parameter $\mu_0 \in \mathbb{R}^d$ (cf. Kitamura, 1997). Here b^{-1} represents an adjustment in (2) to account for OL blocks and, for iid data, a block length $b = 1$ above produces the EL distributional result of Owen (1988, 1990). To illustrate the connection between block selection and performance, Figure 1 shows the coverage rate of nominal 90% BEL confidence intervals $\{\mu \in \mathbb{R} : -2/b \log R_{\text{BEL},n}(\mu_0) \le \chi^2_{1,0.9}\}$, as a function of the block size b, for estimating the mean of three different MA(2) processes based on samples of size $n = 100$. One observes that the coverage accuracy of BEL varies with the block length and and that the best block size can depend on the underlying process. The EBEL method described next is a type of generalization of the OL BEL version, without a particular block selection.

Figure 1: Plot of coverage rates for 90% BEL intervals for the process mean $\mathrm{E}X_t = \mu$ over various blocks $b = 2, \ldots, 30$, based on samples of size $n = 100$ from three MA(2) processes $X_t = Z_t + \vartheta_1 Z_{t-2} + \vartheta_2$ with iid standard normal innovations $\{Z_t\}$ (from 4,000 simulations).



2 Expansive block empirical likelihood

2.1 Mean inference

Suppose $X_1, \ldots, X_n$ represents a sample from a strictly stationary process $\{X_t : t \in \mathbb{Z}\}$ taking values in $\mathbb{R}^d$ and consider problem about inference on the process mean $\mathrm{E}X_t = \mu \in \mathbb{R}^d$. While the BEL uses data blocks of a fixed length b for a given sample size n, the EBEL uses overlapping data blocks $\{(X_1), (X_1, X_2), \ldots, (X_1, \ldots, X_n)\}$ that vary in length up to the longest block consisting of the entire time series. Hence, this block collection, which constitutes a type of forward "scan" in the block subsampling language of McElroy and Politis (2007), contains a data block of every possible length b for a given sample size n.

Let $w : [0, 1] \to [0, \infty)$ denote a nonnegative weighting function. To assess the likelihood of a given value of μ, we create centered block sums $T_{i,\mu} = w(i/n) \sum_{j=1}^{i} (X_j - \mu)$, $i = 1, \ldots, n$, and define a EBEL function $L_n(\mu)$ and ratio $R_n(\mu)$ as

$$L_n(\mu) = \sup\left\{\prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i T_{i,\mu} = 0_d\right\}, \qquad R_n(\mu) = \frac{L_n(\mu)}{n^{-n}}. \tag{3}$$

After defining the block sums, the computation of $L_n(\mu)$ is analogous to the BEL version and essentially the same as that described by Owen (1988, 1990) for iid data. Namely, when the zero 0_d vector lies in the interior convex hull of $\{T_{i,\mu} : i = 1, \ldots, n\}$, then $L_n(\mu)$ is the uniquely achieved maximum at probabilities $p_i = 1/[n(1 + \lambda'_{n,\mu} T_{i,\mu})] > 0$, $i = 1, \ldots, n$, with $\lambda_{n,\mu} \in \mathbb{R}^d$ (a Lagrange multiplier) satisfying

$$\sum_{i=1}^n \frac{T_{i,\mu}}{n(1 + \lambda'_{n,\mu} T_{i,\mu})} = 0_d; \tag{4}$$

see Owen (1990) for these and other computational details. Regarding the weight function above in the EBEL formulation, more details are provided below and in Sections 2.2 with several examples studied numerically in Section 3.

The next section establishes the limiting distribution of the log-EL ratio from the EBEL method for setting confidence regions for the process mean μ parameter. However, it is helpful to initially describe how the subsequent developments of EL differ from previous ones with iid or weakly dependent data (cf. Kitamura, 1997). The standard arguments for developing EL results, due to Owen (1990, p. 101), typically begin from algebraically re-writing (4) to express the Lagrange multiplier. If we consider the real-valued case $d = 1$ for simplicity, this becomes

$$\lambda_{n,\mu} = \frac{\sum_{i=1}^n T_{i,\mu}}{\sum_{i=1}^n T_{i,\mu}^2} + \frac{\lambda_{n,\mu}^2}{\sum_{i=1}^n T_{i,\mu}^2} \sum_{i=1}^n \frac{T_{i,\mu}^3}{1 + \lambda'_{n,\mu} T_{i,\mu}}.$$

In the usual independence or weak dependence cases of EL (e.g., where $B_{i,\mu}$ from (1) replaces $T_{i,\mu}$ in the Lagrange multiplier above), the first right-side term dominates the second, which gives a substantive form for $\lambda_{n,\mu}$ as a ratio of sample means and consequently drives the large sample results (i.e., producing chi-square limits). However, in the EBEL case here, both terms on the right side above have the *same* order, implying that the standard approach to developing EL results breaks down under the EBEL blocking scheme.

The large sample results for the EBEL method require two mild assumptions stated below. Let $\mathcal{C}_d[0,1]$ denote the metric space of all $\mathbb{R}^d$-valued continuous functions on $[0,1]$ with the supremum metric $\rho(g_1, g_2) \equiv \sup_{0 \leq t \leq 1} \|g_1(t) - g_2(t)\|$, and let $B(t) = (B_1(t), \ldots, B_d(t))'$, $0 \leq t \leq 1$, denote a $\mathcal{C}_d[0,1]$-valued random variable where $B_1(t), \ldots, B_d(t)$ are iid copies of standard Brownian motion on $[0,1]$.

Assumptions

(A.1) The weight function $w : [0, 1] \to [0, \infty)$ is continuous on $[0, 1]$ and is strictly positive on an interval $(0, c)$ for some $c \in (0, 1]$.

(A.2) Let $\mathrm{E}X_t = \mu_0 \in \mathbb{R}^d$ denote the true mean of the stationary process $\{X_t\}$ and suppose $d \times d$ matrix $\Sigma = \sum_{j=-\infty}^{\infty} \mathrm{Cov}(X_0, X_j)$ is positive definite. For the empirical process $S_n(t)$ on $t \in [0, 1]$ defined by linear interpolation of $\{S_n(i/n) = \sum_{j=1}^{i}(X_j - \mu_0) : i = 0, \ldots, n\}$ with $S_n(0) = 0$, it holds that $S_n(\cdot)/n^{1/2} \xrightarrow{d} \Sigma^{1/2} B(\cdot)$ in $\mathcal{C}_d[0, 1]$.

Assumption (A.1) is used to guarantee that, in probability, the EBEL ratio $R_n(\mu_0)$ positively exists at the true mean, which holds for uniformly weighted blocks $w(t) = 1$, $t \in [0, 1]$, for example. Assumption (A.2) is a functional central limit theorem for weakly dependent data, which holds, for example, under appropriate mixing and moment conditions on the process $\{X_t\}$ (cf. Herrndorf, 1984).

2.2 Main distributional results

To state the limit law for the log-EBEL ratio (3), we first require a result regarding a vector $B(t) = (B_1(t), \ldots, B_d(t))'$, $0 \le t \le 1$, of iid copies $B_1(t), \ldots, B_d(t)$ of standard Brownian motion on $[0, 1]$. Indeed, the limit distribution of $-2 \log R_n(\mu_0)$ is a non-standard functional of the vector of Brownian motion $B(\cdot)$. Theorem 1 below identifies the key elements of the limit law and describes some of its basic structural properties.

Theorem 1 *Suppose that $B(t) = (B_1(t), \ldots, B_d(t))'$, $0 \le t \le 1$, is defined on a probability space and let $f(t) = w(t)B(t)$, $0 \le t \le 1$, where $w(\cdot)$ satisfies Assumption (A.1). Then, with probability 1 (w.p.1), there exists an $\mathbb{R}^d$-valued random vector Y_d satisfying the following:*

(i) Y_d is the unique minimizer of

$$g_d(a) \equiv -\int_0^1 \log(1 + a' f(t))dt \ \text{ for } a \in \overline{K}_d \equiv \{y \in \mathbb{R}^d : \min_{0 \le t \le 1}(1 + y' f(t)) \ge 0\};$$

the latter set is the closure of $K_d \equiv \{y \in \mathbb{R}^d : \min_{0 \le t \le 1}(1 + a' f(t)) > 0\}$, which is open, bounded and convex in $\mathbb{R}^d$ (w.p.1). On K_d, g_d is also real-valued, strictly convex, and infinitely differentiable (w.p.1).

(ii) $$-\infty < g_d(Y_d) < 0, \qquad Y_d' \int_0^1 f(t)dt > 0, \qquad 0 \le \int_0^1 \frac{Y_d' f(t)}{1 + Y_d' f(t)} dt < \infty$$

(iii) If $Y_d \in K_d$, then Y_d is the unique solution to $\displaystyle\int_0^1 \frac{f(t)}{1 + a' f(t)} dt = 0_d$ for $a \in K_d$; and if $\displaystyle\int_0^1 \frac{f(t)}{1 + a' f(t)} dt = 0_d$ has a solution $a \in K_d$, then this solution is uniquely Y_d.

To comment on Theorem 1, we use the subscript d in Theorem 1 to denote the dimension of either the random vector Y_d, the space K_d or the arguments of g_d. The function g_d is well-defined and convex on $\overline{K}_d$, though possibly $g_d(a) = +\infty$ for $a \in \partial K_d = \{y \in \mathbb{R}^d : \min_{0\leq t\leq 1}(1+y'f(t)) = 0\}$ on the boundary of K_d. Importantly, the probability law of $g_d(Y_1)$ is distribution-free and, because standard Brownian motion is fast and straightforward to simulate, the distribution of $g_d(Y_d)$ can be approximately numerically. Parts (ii) and (iii) provide properties for characterizing and identifying the minimizer Y_d. For example, considering the real-valued case $d = 1$, it holds that $K_1 = (m, M)$ where $m = -[\max_{0\leq t\leq 1} f(t)]^{-1} < 0 < M = -[\min_{0\leq t\leq 1} f(t)]^{-1}$ and the derivative $dg_1(a)/da$ is strictly increasing on K_1 by convexity. Because the derivative of g_1 at 0 is $-\int_0^1 f(x)dx$, parts (ii)-(iii) imply that if $-\int_0^1 f(x)dx < 0$ then either $Y_1 = m$ or Y_1 solves $dg_1(a)/da = 0$ on $m < a \leq 0$; alternatively, if $-\int_0^1 f(x)dx > 0$, then $Y_1 = M$ or Y_1 solves $dg_1(a)/da = 0$ on $0 \leq a < M$. Additionally, the scale of the weight function $w(\cdot)$ does not influence the distribution of $g_d(Y_d)$; that is, defining f with w or cw, for a non-zero $c \in \mathbb{R}$, produces the same minimized value $g_d(Y_d)$.

We may now state the main result on the large-sample behavior of the EBEL log-ratio evaluated at the true process mean $\mathrm{E}X_t = \mu_0 \in \mathbb{R}^d$. Recall that, when $L_n(\mu_0) > 0$ in (3), the EBEL log-ratio admits an expansion (4) at μ_0 in terms of the Lagrange multiplier $\lambda_{n,\mu_0} \in \mathbb{R}^d$.

Theorem 2 *Under Assumptions A.1-A.2, as $n \to \infty$,*

(i) $n^{1/2}\Sigma^{1/2}\lambda_{n,\mu_0} \xrightarrow{d} Y_d$,

(ii) $-\dfrac{1}{n}\log R_n(\mu_0) \xrightarrow{d} -g_d(Y_d)$,

recalling $\Sigma = \sum_{j=-\infty}^{\infty} \mathrm{Cov}(X_0, X_j)$, *and that* Y_d *and* $g_d(Y_d)$ *are defined as in Theorem 1.*

From Theorem 2(i), it is interesting to note that the Lagrange multiplier in the EBEL method exhibits a faster order convergence $O_p(n^{-1/2})$ compared to that $O_p(bn^{-1/2})$ in the standard BEL case, where $b \to \infty$ as $n \to \infty$, and its limiting distribution is also not the typical normal one. This has a direct impact on the limit law of the EBEL ratio statistic. As Theorem 2(ii) shows, the negative logarithm of the EBEL ratio statistic, scaled by the inverse of the sample size, has a non-standard limit, given by the functional $-g_d(Y_d)$ of the vector of Brownian motion $B(\cdot)$ (cf. Theorem 1), that critically depends on the limit Y_d of the scaled Lagrange multiplier. Although non-standard, the distribution of $-g_d(Y_d)$ is free of any population parameters. Hence, quantiles of $-g_d(Y_d)$, which are easy to compute numerically (cf. Section 3), can be used to calibrate the EBEL confidence regions. In contrast to the standard BEL (2), EBEL confidence regions do not require a choice of block size. As $-g_d(Y_d)$ is a strictly positive random variable, an approximate $100(1-\alpha)\%$ confidence region for μ_0 can be computed as

$$\{\mu \in \mathbb{R}^d : -n^{-1}\log R_n(\mu_0) \leq a_{d,1-\alpha}\},$$

where $a_{d,1-\alpha}$ is the lower $(1-\alpha)$ percentile of $-g_d(Y_d)$. When $d = 1$, the confidence region is an interval; for $d > 2$, the region is guaranteed to be connected without voids in $\mathbb{R}^d$.

We next provide an additional result, which shows the size of a EBEL confidence region will be no larger than $O_p(n^{-1/2})$ in diameter around the true mean $\mathrm{E}X_t = \mu_0$. Let

$$G_n \equiv \{\mu \in \mathbb{R}^d : R_n(\mu) \geq R_n(\mu_0) > 0\} \tag{5}$$

be the collection of mean parameter values which are at least as likely as μ_0, and therefore elements of a EBEL confidence region whenever the true mean is.

Corollary 1 *Under the assumptions of Theorem 2, $Z_n = O_p(n^{-1/2})$ holds, where $Z_n \equiv \sup\{\|\mu - \mu_0\| : \mu \in G_n\}$ for G_n in (5).*

We note that Theorem 2 remains valid for potentially negative-valued weight functions $w(\cdot)$ as well (i.e., assuming w is continuous, real-valued, and that either w or $-w$ is strictly positive on some $(0, c] \subset [0, 1]$ in place of A.1). Simulations have shown that, with weight functions oscillating between positive and negative values on $[0, 1]$ (e.g., $w(t) = \sin(2\pi t)$), EBEL intervals for the process mean perform consistently well in terms of coverage accuracy. However, with weight functions $w(\cdot)$ that vary in sign, a result as in Corollary 1 fails to hold. For this reason, the weight functions $w(\cdot)$ considered are non-negative as stated in Assumption A.1.

2.3 Smooth function model parameters

We next consider extending the EBEL method for inference on a broad class of parameters under the so-called "smooth function model" of Bhattacharya and Ghosh (1978) and Hall (1992). For independent and time series data, respectively, Hall and La Scala (1990) and Kitamura (1997) have considered EL inference for similar parameters; see also Owen (1990, sec. 4).

If $\mathrm{E}X_t = \mu_0 \in \mathbb{R}^d$ again denotes the true mean of the process, the target parameter of interest is given by

$$\theta_0 = H(\mu_0) \in \mathbb{R}^p, \tag{6}$$

based on a smooth function $H(\mu) = (H_1(\mu), \ldots, H_p(\mu))'$ of the mean parameter μ, where $H_i : \mathbb{R}^d \to \mathbb{R}$ for $i = 1, \ldots, p$ and $p \leq d$. This framework allows a large variety of parameters to be considered such as sums, differences, products and ratios of means, which can be used, for example, to formulate parameters such as covariances and autocorrelations as functions of the m-dimensional moment structure (for a fixed m) of a time series. For a univariate stationary series $U_1, \ldots, U_n$, for instance, one can define a multivariate series X_t based on transformations of $(U_t, \ldots, U_{t+m-1})$ and estimate parameters for the process $\{U_t\}$ based on appropriate functions H of the mean of X_t. The correlations $\theta_0 = H(\mu_0)$ of $\{U_t\}$ at lags m and $m_1 < m$, for example, can be formulated in (6) by $H(x_1, x_2, x_3, x_4) = (x_3 - x_1^2, x_4 - x_1^2)'/[x_2 - x_1^2]$ and $\mathrm{E}X_t = \mu_0$ for $X_t = (U_t, U_t^2, U_t U_{t+m_1}, U_t U_{t+m})' \in \mathbb{R}^4$. Künsch (1989) and Lahiri (2003, Ch. 4) provide further examples of smooth function parameters.

For inference on the parameter $\theta = H(\mu)$, the EBEL ratio is defined as

$$R_n(\theta) = \sup\left\{\prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i T_{i,\mu} = 0_d, \mu \in \mathbb{R}^d, H(\mu) = \theta\right\},$$

and its limit distribution is provided next.

Theorem 3 *In addition to the assumptions of Theorem 2, suppose H from (6) is continuously differentiable in a neighborhood of μ_0 and that ∇_{μ_0} has rank $p \leq d$, where $\nabla_\mu \equiv [\partial H_i(\mu)/\partial \mu_j]_{i=1,\ldots,p;j=1,\ldots,d}$ denotes the $p \times d$ matrix of first-order partial derivatives of H. Then, at the true parameter $\theta_0 = H(\mu_0)$, as $n \to \infty$*

$$-\frac{1}{n} \log R_n(\theta_0) \xrightarrow{d} -g_p(Y_p)$$

with Y_p and $g_p(Y_p)$ as defined in Theorem 1.

Theorem 3 shows that the log-EBEL ratio statistic for the parameter $\theta_0 = H(\mu_0) \in \mathbb{R}^p$ under the smooth function model continues to have a limit of the same form as that in the case of the EBEL for the mean parameter $\mu_0 \in \mathbb{R}^d$ itself. The main difference is that the functional $g_p(Y_p)$ is now defined in terms of a p-dimensional Brownian motion as in Theorem 1, but with $p \leq d$, where p denotes the dimension of the parameter θ_0. It is interesting to note that, similarly to the traditional profile likelihood theory in a parametric set-up with iid observations, the limit law here does not depend on the function H as long as the matrix ∇_{μ_0} of the first order partial derivatives of H at $\mu = \mu_0$ has full rank p. Due to the non-standard blocking, the proof of this EBEL result again requires a different development compared to standard BEL (cf. Kitamura, 1997), which follows similarly to the iid EL case (cf. Owen, 1990, sec. 4; Hall & La Scala, 1990).

3 Numerical studies

Here we summarize the results of a simulation study to investigate the performance of the EBEL method, considering the coverage accuracy of confidence intervals (CIs) for the process mean. We considered several real-valued ARMA(1,2) processes $X_t = \phi X_t + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$ defined with respect to an underlying iid innovation series $\{\varepsilon_t\}$; Table 1 lists the parameter combinations considered, denoted as Models $1, \ldots, 9, 0$, which allow a variety of dependence structures with ranges of weak and strong dependence. We also considered several AR(1) processes $X_t = \phi X_t + Y_t$ defined by a (stationary) first-order threshold moving average innovations $Y_t = \varepsilon_t +$

Table 1: Parameters for ARMA(1,2) processes are denoted as Models $1, \ldots, 9, 0$. Parameters in AR(1) processes ,defined by threshold moving average innovations, are denoted as Models $a, \ldots, h$.

Parameters	Process Models									
	1 & a	2 & b	3 & c	4 & d	5 & e	6 & f	7 & g	8 & h	9	0
ϕ	-0.3	-0.3	0.0	0.0	0.0	0.6	0.9	0.6	0.9	0.6
θ_1	0.1	0.7	-0.6	0.4	0.7	0.7	-0.6	-0.3	0.0	0.0
θ_2	0.7	1.0	0.4	-0.6	1.0	0.4	-0.3	0.7	0.0	0.0

$[\theta_1 + \theta_2 \mathbb{I}(Y_{t-1} \leq 4)]\varepsilon_{t-1}$ with respect to an iid innovation series $\{\varepsilon_t\}$ (cf. Tong, 1978); Table 1 again lists parameters for defining these processes, denoted as Models $a, \ldots, h$.

For each process (and depending on an innovation $\{\varepsilon_t\}$ type), we generated 2000 samples of size $n = 250, 500, 1000$ for comparing the coverage accuracy of 90% CIs from various EL procedures. For the EBEL method, we present results for 11 different weighting functions $w(\cdot)$ listed in Table 2, some of which are based on the following forms, for $c \in (0, 1]$,

$$\begin{aligned} w_{linear.to.flat,c} &= (t/c)\mathbb{I}(t \leq c) + \mathbb{I}(t > c) \\ w_{trap,c}(t) &= t/c\mathbb{I}(t \leq c) + \mathbb{I}(c \leq t \leq 1-c) + (1-t)/c\mathbb{I}(1-c \leq t \leq 1) \\ w_{quad.to.flat,c} &= (t/c)^2\mathbb{I}(t \leq c) + \mathbb{I}(t > c) \\ w_{step,c} &= 0.01t/c\mathbb{I}(t \leq c) + [99(t-c) + 0.01]\mathbb{I}(c < t \leq c + 0.01) + \mathbb{I}(t > c + 0.01). \end{aligned}$$

The weighting functions in Table 2 differ in their initial curvature on $[0, 1]$ and at the point which these potentially "top off" at value of 1. The "step function" $w_{step,c}$ is continuous, but essentially equals zero initially with a large jump to 1 near c. Functions 5 and 6 in Table 2 down-weight at both ends of the interval $[0, 1]$ in a symmetrical fashion. We had also tried $w_{linear.to.flat,c}$ for $c = 0.5, 0.75, 1$, $w_{quad.to.flat,c}$ for $c = 0.25, 0.5$, $w_{step,c}(t)$ for $c = 0.3, 0.5, 0.7$, and $w_{trap,.4}(t)$, but, within each group, the results were quantitatively similar to weight functions in Table 2. Additionally, for each weight function w, the limiting distribution of the EBEL ratio $-g_1(Y_1)$ under Theorem 2 was approximated by 50000 simulations in order to determine the 90th (or 95th) percentile for calibrating intervals. Table 2 lists these approximated percentiles along with Monte Carlo error bounds.

For comparison, we also include coverage results for the standard BEL method, with OL blocks (BEL) or non-overlapping blocks (NBEL), as well as a tapered BEL (TBEL) method. The NBEL uses a subset of the data blocks from the BEL method (cf. Kitamura, 1997), and the TBEL resembles BEL but uses a trapezoidal taper $w_{trap,0.43}$ to down-weight observations at the

Table 2: Weight functions used the simulation study for the EBEL method along with approximated 90th and 95th lower percentiles of the limit law $-g_1(Y_1)$ of the log-EBEL ratio (Theorem 2), based on 50000 simulations. To indicate Monte Carlo error, the approximated percentile $\pm$ the parenthetical quantity gives a 95% CI for the true percentile of $-g_1(Y_1)$.

	Weight function $w(t)$, $t \in [0,1]$	90th percentile		95th percentile	
1	$w(t) = 1$	2.51	(0.03)	3.28	(0.04)
2	$w_{linear.to.flat,0.9}(t)$	5.64	(0.09)	7.77	(0.14)
3	$w_{linear.to.flat,0.25}(t)$	5.04	(0.08)	7.17	(0.13)
4	$w(t) = (1-\cos(2\pi t))/2\mathbb{I}(t \leq 0.5) + \mathbb{I}(t > 0.5)$	7.59	(0.13)	11.06	(0.19)
5	$w(t) = (1-\cos(2\pi t))/2$	7.00	(0.15)	10.47	(0.19)
6	$w_{trap,0.5}(t)$	5.06	(0.09)	7.19	(0.11)
7	$w(t) = t^2$	8.56	(0.15)	12.20	(0.20)
8	$w_{quad.to.flat,0.75}(t)$	8.47	(0.15)	12.05	(0.22)
9	$w_{quad.to.flat,0.9}(t)$	8.75	(0.15)	12.38	(0.19)
10	$w_{step,0.1}(t)$	8.32	(0.09)	10.51	(0.14)
11	$w_{step,0.9}$	6.24	(0.09)	8.44	(0.14)

ends of each block $(X_i \cdot w_{trap,0.43}[(1-0.5)/b], \ldots, X_{i+b-1} \cdot w_{trap,0.43}[(b-0.5)/b])$, $i = 1, \ldots, n-b+1$, of length b; see Nordman (2009) for details. (Note that EBEL weights data blocks of varying length, not observations within each data block as in the TBEL method.) For each of these BEL methods, we considered block choices $b = Cn^{1/3}$, $C = 0.5, 1, 2$, with a block order $n^{1/3}$ based on its consideration by Kitamura (1997, p. 2093) and scaling C around 1 borrowed from implementations in the block bootstrap literature (cf. Lahiri, 2003, ch. 5).

Figure 2 shows the coverage accuracy of 90% EL CIs for the mean of the threshold-based Models *a-h*, with either standard normal $N(0,1)$ or centered χ^2_1 innovations $\{\varepsilon_t\}$. Figure 3 shows the same for Models 0-9 with centered Bernoulli ($Ber(0.5)$) or Pareto innovations $\{\varepsilon_t\}$, the latter having probability density function $2.1x^{-3.1}$, $x > 1$. These figures suggest that the EBEL method generally performs as well as a BEL procedure using a good block choice, and much better when the latter employ a bad block choice. Additionally, the EBEL method tends to be less sensitive to the underlying dependence in its performance (e.g., the strong positive dependence Model 9 in Figure 3 where standard BEL methods can exhibit extreme under-coverage, or the negative dependence Model 4 where over-coverage occurs). Also, the coverage results in the EBEL method are fairly similar across quite different weighting functions $w(\cdot)$; that is, the method is largely insensitive to the weight function, with the one exception being that constant weighting $w(t) = 1$ tended to perform generally worse. To give an idea of the coverage rates of 95% intervals for the process mean, Figure 4 contrasts coverages of 90% and 95% CIs for Models $1, \ldots, 9, 0$ with χ_1^2 innovations. The pattern of coverages is qualitatively similar for both nominal coverage levels.

4 Conclusions

The proposed expansive block empirical likelihood (EBEL) is a type of variation on standard blockwise empirical likelihood (BEL) for time series which, instead of using a fixed block length b for a given sample size n, involves a non-standard blocking scheme to capture the dependence structure. As the coverage accuracy of standard BEL methods depends intricately on the block choice b (where the best b can vary with the underlying process), the EBEL has an advantage in eliminating block selection. As mentioned in the Introduction, we also anticipate that the EBEL method will generally have better rates of coverage accuracy compared to other existing versions of BEL, such as the overlapping tapered and non-tapered versions of BEL. The simulations of Section 3 lend support to this notion, along with suggesting that the EBEL can be less sensitive to the strength of the underlying time dependence. While asymptotic coverage rates for BEL methods remain to be determined, we may offer the following heuristic based on analogs drawn to so-called "fixed-b asymptotic" (cf. Keifer, Vogelsang and Bunzel, 2000; Bunzel, Kiefer and Vogelsang, 2001; Kiefer and Vogelsang, 2002), or related "self-normalization" (cf. Lobato, 2001; Shao, 2011) schemes.

In asymptotic expansions of log-likelihood statistics from standard BEL formulations, the data blocks serve to provide a type of subsampling variance estimator (cf. Carlstein, 1986; Politis and Romano, 1993) for purposes of normalizing scale and obtaining chi-square limits for log-BEL ratio statistics. Such variance estimators are consistent, requiring block sizes b which grow at a smaller rate than the sample size n (i.e., $b^{-1} + b/n \to \infty$ as $n \to \infty$), and are known to have equivalences to variance estimators formulated as lag window estimates involving kernel functions and bandwidths b with similar behavior to block lengths $b^{-1} + b/n \to \infty$ (cf. Künsch, 1989; Politis, 2003). That is, standard BEL intervals have parallels with normal theory intervals based on normalization with consistent lag window estimates. However, considering interval inference with sample means for example, there is some numerical and theoretical evidence (cf. Bunzel et al 2001; Sun, Philips and Jin, 2008) that normalizing scale with inconsistent lag window estimates having fixed bandwidth ratios (e.g., $b/n = C$ for some $C \in (0, 1]$) results in better coverage accuracy compared to normalization with consistent ones, though the former case requires calibrating intervals with non-normal limit laws. Shao (2011, sec 2.1) provides a nice summary of these points as well as the form of some of these distribution-free limit laws, which typically involve ratios of random variables defined by Brownian motion (cf. Kiefer and Vogelsang, 2002). While the EBEL method is not immediately analogous to normalizing with inconsistent variance estimators (as mentioned in Section 2.1, the usual EL expansions do not hold for EBEL), there are parallels in that the EBEL method does not use block lengths satisfying standard bandwidth conditions and confidence region calibration involves non-normal limits based on Brownian motion. This heuristic in the mean case suggests that better coverage rates associated with fixed-b asymptotics over standard normal theory asymptotics may be anticipated to carry over to comparisons of EBEL to standard BEL formulations.

5 Proofs of main results

To establish Theorem 1, we first require a lemma regarding a standard Brownian motion. For concreteness, suppose $B(t) \equiv B(\omega, t) = (B_1(\omega, t), \ldots, B_d(\omega, t))'$, $\omega \in \Omega$, $t \in [0,1]$ is a random $\mathcal{C}_d[0,1]$-valued element defined on some probability space $(\Omega, \mathcal{F}, P)$, where $B_1, \ldots, B_d$ are again distributed as iid copies of standard Brownian motion on $[0,1]$. In the following, we use the basic fact that each $B_i(\cdot)$ is continuous on $[0,1]$ with probability 1 (w.p.1) along with the fact that increments of standard Brownian motion are independent (cf. Freedman, 1983).

Lemma 1 *With probability 1, it holds that*

(i) $\min_{0\le t<\epsilon} a'B(t) < 0 < \max_{0\le t<\epsilon} a'B(t)$ *for all* $\epsilon > 0$ *and* $a \in \mathbb{R}^d$, $\|a\| = 1$.

(ii) 0_d *is in the interior of the convex hull of* $B(t)$, $0 \le t \le 1$.

(iii) There exists a positive random variable M *such that, for all* $a \in \mathbb{R}^d$, *it holds that* $\min_{0\le t\le 1} a'B(t) \le -M\|a\|$ *and* $M\|a\| \le \max_{0\le t\le 1} a'B(t)$.

(iv) If Assumption A.1 holds in addition, (i), (ii), (iii) above hold upon replacing $B(t)$ *with* $f(t) = w(t)B(t)$, $t \in [0,1]$.

Proof of Lemma 1. For real-valued Brownian motion, it is known that $\min_{0\le t<\epsilon} B_i(t) < 0 < \max_{0\le t<\epsilon} B_i(t)$ holds for all $\epsilon > 0$ w.p.1. (cf. Freedman, 1983, Lemma 55); we modify the proof of this. Let $\{t_n\} \subset (0,1)$ be a decreasing sequence where $t_n \downarrow 0$ as $n \to \infty$. Pick and fix $c_1, \ldots, c_d \in \{-1, 1\}$ and define the event $A_n \equiv A_{n,c_1,\ldots,c_d} = \{\omega \in \Omega : c_i B_i(\omega, t_n) > 0, i = 1, \ldots, d\}$. Then, $P(A_n) = 2^{-d}$ for all $n \ge 1$ by normality and independence. As the events $B_n = \bigcup_{k=n}^{\infty} A_k$, $n \ge 1$, are decreasing, it holds that

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n) \ge \lim_{n\to\infty} P(A_n) = 2^{-d}.$$

Since $\bigcap_{n=1}^{\infty} B_n$ is a tail event generated by the independent random variables $B_i(t_1)-B_i(t_2), B_i(t_2)-B_i(t_3), \ldots$ for $i = 1, \ldots, d$, (i.e., increments of Brownian motion are independent and $B_i(0) = 0$), it follows from Kolmogorov's 0-1 law that $1 = P\left(\bigcap_{n=1}^{\infty} B_n\right) = P\left(A_n \text{ infinitely often (i.o.)}\right)$. Hence, $P\left(A_{n,c_1,\ldots,c_d} \text{ i.o. for any } c_i \in \{1,-1\},\ i = 1,\ldots,d\right) = 1$ must hold, which implies part (i).

For part (ii), if 0_d is not in the interior convex hull of $B(t)$, $t \in [0,1]$, then the supporting/separating hyperplane theorem would imply that, for some $a \in \mathbb{R}^d$, $\|a\| = 1$, it holds that $a'B(t) \geq 0$ for all $t \in [0,1]$, which contradicts part (i).

To show part (iii), we use the events developed in part (i) and define $n_{c_1,\ldots,c_d} = \min\{n : A_{n,c_1,\ldots,c_d} \text{ holds}\}$. Define $M = \min\{|B_i(t_{n_{c_1,\ldots,c_d}})| : c_1,\ldots,c_d \in \{-1,1\}, i = 1,\ldots,d\} > 0$. For $a = (a_1,\ldots,a_d)' \in \mathbb{R}^d$, let $c_i^a = \max\{-\text{sign}(a_i), 1\}$, $i = 1,\ldots,d$. Then, $a'B(t_{n_{c_1^a,\ldots,c_d^a}}) = -\sum_{i=1}^d |a_i B_i(t_{n_{c_1^a,\ldots,c_d^a}})| \leq -M\|a\|$, and likewise $a'B(t_{n_{-c_1^a,\ldots,-c_d^a}}) = \sum_{i=1}^d |a_i B_i(t_{n_{-c_1^a,\ldots,-c_d^a}})| \geq M\|a\|$. This establishes (iii).

Part (iv) follows from the fact that $w(t) > 0$ for $t \in (0,c)$ and we may make take the positive sequence $\{t_n\} \subset (0,c)$ in the proof of part (i). Then, the results for $B(t)$ imply the same hold upon substituting $f(t) = w(t)B(t)$, $t \in [0,1]$. $\square$

Proof of Theorem 1. The set $K_d = \{a \in \mathbb{R}^d : \min_{0\leq t\leq 1}(1 + a'f(t)) > 0\}$ is open, bounded and convex (w.p.1), where boundedness follows from Lemma 1(iii,iv). Likewise, the closure $\overline{K}_d = \{a \in \mathbb{R}^d : \min_{0\leq t\leq 1}(1+a'f(t)) \geq 0\}$ is convex and bounded. Since $\min_{0\leq t\leq 1}(1+a'f(t))$ is a continuous function in $a \in \mathbb{R}^d$, one may apply the Dominated Convergence Theorem (DCT) (with the fact that $\min_{0\leq t\leq 1}(1 + a'f(t))$ is bounded away from 0 on closed balls inside K_d around a) to show that partial derivatives of $g_d(\cdot)$ at $a \in K_d$ (of all orders) exist, with first and second partial derivatives given by

$$\frac{\partial g_d(a)}{\partial a} = -\int_0^1 \frac{f(t)}{1 + a'f(t)}dt, \qquad \frac{\partial^2 g_d(a)}{\partial a \partial a'} = \int_0^1 \frac{f(t)f(t)'}{[1 + a'f(t)]^2}dt.$$

Because $\int_0^1 f(t)f(t)'dt$ is positive definite by Lemma 1(i,iv) and the continuity of f, the matrix $\partial^2 g_d(a)/\partial a\partial a'$ is also positive definitive for all $a \in K_d$, implying g_d is strictly convex on K_d. By Jensen's inequality, it also holds that g_d is convex on $\overline{K}_d$.

Note for $a \in \overline{K}_d$, $g_d(a) \geq -\int_0^1 \log(1 + \sup_{a\in\overline{K}_d} \|a\| \cdot \sup_{0\leq t\leq 1} \|f(t)\|) > -\infty$ holds, so that $I \equiv \inf_{a\in\overline{K}_d} g_d(a)$ exists. Additionally, $0_d \in K_d$ with $g_d(0_d) = 0$ and $\partial g_d(0_d)/\partial a = -\int_0^1 f(t)dt$, where the components of $\int_0^1 f(t)dt$ are all non-zero (w.p.1) by normality and independence; by the continuity of partial derivatives on the open set K_d, there then exists $\bar{a} \in K_d$ such that $\bar{a}'\int_0^t f(t)dt > 0$ holds with the components of $-\int_0^1 f(t)$ and $\partial g_d(\bar{a})/\partial a$ having the same sign. By strict convexity, $g_d(0_d) - g_d(\bar{a}) > [\partial g_d(\bar{a})/\partial a]'(0_d - \bar{a}) > 0$ follows, implying $I < 0$ and $I = \inf_{a\in\tilde{K}_d} g_d(a)$ for the level set $\tilde{K}_d \equiv \{a \in \overline{K}_d : g_d(a) \leq 0\}$.

Then, there exists a sequence $a_n \in \tilde{K}_d$ such that $g_d(a_n) < I + n^{-1}$ for $n \geq 1$. Since $\{a_n\}$ is bounded, we may extract a subsequence such that $a_{n_k} \to Y_d \in \tilde{K}_d$, for some $Y_d \in \tilde{K}_d$. Pick $\delta \in (0,1)$. Then, by the DCT,

$$\begin{aligned}
\underline{\lim} g_d(a_{n_k}) &\geq \underline{\lim} \int_{\{t: a_{n_k}' f(t) > -1+\delta\}} -\log(1 + a_{n_k}' f(t)) dt \\
&= \int_{\{t: Y_d' f(t) > -1+\delta\}} -\log(1 + Y_d' f(t)) dt \\
&= g_d(Y_d) + \int_{\{t: Y_d' f(t) \leq -1+\delta\}} \log(1 + Y_d' f(t)) dt.
\end{aligned}$$

Note that because $g_d(Y_d) \in (-\infty, 0]$, it follows that $-\int_{\{t: Y_d' f(t) < 0\}} \log(1 + Y_d' f(t)) dt < \infty$ and $\{t \in [0,1] : Y_d' f(t) = -1\}$ has Lebesgue measure zero. Hence, the DCT yields

$$\lim_{\delta \to 0} -\int_{\{t: Y_d' f(t) \leq -1+\delta\}} \log(1 + Y_d' f(t)) dt = 0.$$

Consequently,

$$I \geq \overline{\lim}\, g_d(a_{n_k}) \geq \underline{\lim} g_d(a_{n_k}) \geq g_d(Y_d) \geq I,$$

establishing the existence of a minimizer Y_d of g_d on $\overline{K}_d$ such that $-\infty < I = g_d(Y_d) < 0$.

For part (ii) of Theorem 1, note $y_n = (1 - n^{-1}) Y_d + n^{-1} 0_d \in K_d$, $n \geq 1$, by convex geometry, as K_d is the convex interior of $\overline{K}_d$. Then, $g_d(y_n) \leq (1 - n^{-1}) g_d(Y_d)$ holds by convexity of g_d and $g_d(0_d) = 0$, implying $0 \leq n[g_d(y_n) - g_d(Y_d)] \leq -g_d(Y_d) < \infty$, from which it follows that $g_d(y_n) \to g_d(Y_d)$ and, by the mean value theorem,

$$0 \leq n[g_d(y_n) - g_d(Y_d)] = \int_0^1 \frac{Y_d' f(t)}{1 + c_n Y_d' f(t)} dt \leq -g_d(Y_d)$$

holds for some $(1 - n^{-1}) < c_n < 1$ (note $c_n Y_d \in K_d$ so $\min_{0 \leq t \leq 1}(1 + c_n Y_d' f(t)) > 0$ for all n); the latter implies $0 \leq \int_{\{t: Y_d' f(t) < 0\}} -Y_d' f(t)/[1 + c_n Y_d' f(t)] dt \leq \int_{\{t: Y_d' f(t) > 0\}} Y_d' f(t) < \infty$ so that Fatou's lemma yields

$$0 \leq \int_{\{t: Y_d' f(t) < 0\}} -\frac{Y_d' f(t)}{1 + Y_d' f(t)} dt < \infty$$

as $n \to \infty$, and consequently $\int_0^1 1/[1 + Y_d' f(t)] dt < \infty$. We may then apply the DCT to find

$$\lim_{n \to \infty} \int_0^1 \frac{Y_d' f(t)}{1 + c_n Y_d' f(t)} dt = \int_0^1 \frac{Y_d' f(t)}{1 + Y_d' f(t)} dt \in [0, \infty).$$

Also by convexity and $0_d \in K_d$, $0 > g_d(Y_d) - g_d(0_d) > [\partial g_d(0_d)/\partial a]'(Y_d - 0_d)$ holds (w.p.1), implying $Y_d' \int_0^1 f(t) dt > 0$ from $\partial g_d(0_d)/\partial a = -\int_0^1 f(t) dt$. This establishes part (ii) of Theorem 1.

To show uniqueness of the minimizer, we shall construct sequences with the same properties in the proof of part (ii) above. Suppose $x \in \tilde{K}_d$ such that $g_d(x) = I = g_d(Y_d)$. Defining

$x_n = (1-n^{-1})x + n^{-1}0_d \in K_d$ and $y_n = (1-n^{-1})Y_d + n^{-1}0_d \in K_d$ for $n \geq 1$, by convexity we have $0 \geq g_d(x) - g_d(y_n) > [\partial g_d(y_n)/\partial a]'(x - y_n)$, so that taking limits yields $0 \geq -\int_0^1 (x - Y_d)' f(t)/[1 + Y_d' f(t)]dt$, and, by symmetry, $0 \geq -\int_0^1 (Y_d - x)' f(t)/[1 + x' f(t)]dt$ as well. Adding these terms gives

$$0 \geq \int_0^1 \frac{[(x - Y_d)' f(t)]^2}{(1 + x' f(t))(1 + Y_d' f(t))} dt,$$

implying that $x = Y_d$ by Lemma 1(iv) and the continuity of f.

Finally, to establish part (iii), if $Y_d \in K_d$, then $0_d = \partial g_d(Y_d)/\partial a = -\int_0^1 f(t)/[1 + Y_d' f(t)]dt$ must hold. If there exists another $b \in \overline{K}_d$ satisfying $\int_0^1 f(t)/[1 + b' f(t)]dt = 0_d$, then adding $\partial g_d(Y_d)/\partial a$ to this integral and multiplying by $(Y_d - b)'$ yields $0 = \int_0^1 [(b - Y_d)' f(t)]^2/[(1 + b' f(t))(1 + Y_d' f(t))]dt$, implying that $b = Y_d$. Also, if $0_d = \int_0^1 f(t)/[1 + b' f(t)]dt = -\partial g_d(b)/\partial a$ holds for some $b \in K_d$, then strict convexity implies $g_d(a) - g_d(b) > [\partial g_d(b)/\partial a]'(a - b) = 0$ for all $a \in \overline{K}_d$, implying $b = Y_d$ is the unique minimizer of g_d. □

Proof of Theorem 2. Under Assumption A.2, we use Skorohod's embedding theorem (cf. Theorem 1.1.04, van der Vaart and Wellner, 1996) to embed $\{S_n(\cdot)\}$ and $\{B(\cdot)\}$ in a larger probability space $(\Omega, \mathcal{F}, P)$ such that $\sup_{0 \leq t \leq 1} \|\Sigma^{-1/2} S_n(t)/n^{1/2} - B(t)\| \to 0$ w.p.1(P). Defining $T_n(t) = w(t) S_n(t)$ and $f(t) = w(t) B(t)$, $t \in [0,1]$, the continuity of w under Assumption A.1 then implies

$$\sup_{0 \leq t \leq 1} \left\| \frac{\Sigma^{-1/2} T_n(t)}{n^{1/2}} - f(t) \right\| \to 0 \quad \text{w.p.1.} \tag{7}$$

Note that $T_{i,\mu_0} = w(i/n) \sum_{j=1}^{i} (X_j - \mu_0) = T_n(i/n)$, $i = 1, \ldots, n$. By (7) and Lemma 1, 0_d is in the interior convex hull of $\{T_{i,\mu_0} : i = 1, \ldots, n\}$ eventually (w.p.1) so that $L_n(\mu_0) > 0$ eventually (w.p.1.). (That is, by Lemma 1(iv), there exists $A \in \mathcal{F}$ with $P(A) = 1$ and, for $\omega \in A$, $\min_{0 \leq t \leq} a' f(\omega, t) \leq -M(\omega)$ and $\max_{0 \leq t \leq} a' f(\omega, t) \geq M(\omega)$ hold for some $M(\omega) > 0$ and all $a \in \mathbb{R}^d$, $\|a\| = 1$. Then, (7) implies $\min_{1 \leq i \leq n} a' \Sigma^{-1/2} T_n(\omega, i/n) < 0 < a' \max_{1 \leq i \leq n} \Sigma^{-1/2} T_n(\omega, i/n)$ holds for all $a \in \mathbb{R}^d$, $\|a\| = 1$ eventually, implying 0_d is in the interior convex hull of $\{\Sigma^{-1/2} T_n(i/n) : i = 1, \ldots, n\}$.) Hence, eventually (w.p.1), as in (4), we can write

$$\frac{1}{n} R_n(\mu_0) = -\frac{1}{n} \sum_{i=1}^{n} \log(1 + \lambda_{n,\mu_0}' T_{i,\mu_0}) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + \ell_n' T_{i,n})$$

where $T_{i,n} \equiv \Sigma^{-1/2} T_n(i/n)/n^{1/2}$, $i = 1, \ldots, n$ and $\ell_n = n^{1/2} \Sigma^{1/2} \lambda_{n,\mu_0}$ and

$$\min_{i=1,\ldots,n} (1 + \ell_n' T_{i,n}) > 0, \qquad \sum_{i=1}^{n} \frac{1}{n(1 + \ell_n' T_{i,n})} = 1, \qquad \sum_{i=1}^{n} \frac{T_{i,n}}{n(1 + \ell_n' T_{i,n})} = 0_d. \tag{8}$$

From here, all considered convergence will be pointwise along some fixed $\omega \in A$ where $P(A) = 1$, and we suppress the dependence of terms f, T_n, etc. on ω. Then, (8) [i.e.,

$\min_{i=0,ldots,n} \ell_n'(\Sigma^{-1/2}T_n(i/n)/n^{1/2}) > -1]$ with (7) and Lemma 1(iv) implies that $\|\ell_n\|$ is bounded eventually. For any subsequence $\{n_j\}$ of $\{n\}$, we may extract a further subsequence $\{n_k\} \subset \{n_j\}$ such that $\ell_{n_k} \to b$ for some $b \in \overline{K}_d$. For simplicity, write $n_k \equiv k$ in the following. We will show below that $k^{-1} \log R_k(\mu_0) \to g_d(Y_d)$ and that $\ell_k \to Y_d$, where $Y_d \in \overline{K}_d$ denotes the minimizer of $g_d(a) = -\int_0^1 \log(1 + a' f(t))dt$, $a \in \overline{K}_d$ under Theorem 1. Since the subsequence $\{n_j\}$ is arbitrary, we then have $n^{-1} \log R_n(\mu_0) \to g_d(Y_d)$ and $\ell_n \to Y_d$ w.p.1, implying the distributional convergence in Theorem 2.

Define $Y_\epsilon = (1-\epsilon)Y_d + \epsilon 0_d \in K_d$ (since $0_d \in K_d$, the interior of $\overline{K}_d$) for $\epsilon \in (0,1)$. From $Y_\epsilon \in K_d$, $\min_{0 \le t \le 1}(1 + Y_\epsilon' f(t)) > \delta$ holds for some $\delta > 0$ (dependent on ϵ) so that $\min_{1 \le i \le k}(1 + Y_\epsilon' T_{i,k}) > \delta$ holds eventually by (7). Then, because

$$g_{d,k}(a) \equiv -\frac{1}{k} \sum_{i=1}^{k} \log(1 + a' T_{i,k})$$

is strictly convex on $a \in \{y \in \mathbb{R}^d : \min_{1 \le i \le k}(1 + y' T_{i,n}) > 0\}$ with a unique minimizer at ℓ_k by (8) (i.e., $\partial g_{d,k}(\ell_k)/\partial a = 0_d$ holds and strict convexity follows when $k^{-1} \sum_{i=1}^k T_{i,k} T_{i,k}'$ is positive definite, which holds eventually from $k^{-1} \sum_{i=1}^k T_{i,k} T_{i,k}' \to \int_0^1 f(t) f(t)' dt$ by (7) and the DCT, with the latter matrix being positive definite w.p.1 by Lemma 1(iv) and continuity of f), we have that

$$g_{d,k}(Y_\epsilon) \ge g_{d,k}(\ell_k) = \frac{1}{k} \log R_k(\mu_0).$$

Define $\bar{g}_{d,k}(a) \equiv -k^{-1} \sum_{i=1}^k \log(1 + a' f(i/k))$, $a \in K_d$. Then, by Taylor expansion (recalling $\min_{0 \le t \le 1}(1 + Y_\epsilon' f(t)) > \delta$, $\min_{1 \le i \le k}(1 + Y_\epsilon' T_{i,k}) > \delta$),

$$\begin{aligned} |g_{d,k}(Y_\epsilon) - \bar{g}_{d,k}(Y_\epsilon)| &\le \frac{1}{k} \sum_{i=1}^{k} |Y_\epsilon'(T_{i,k} - f(i/k))| \left(\frac{1}{1 + Y_\epsilon' T_{i,k}} + \frac{1}{1 + Y_\epsilon' f(i/k)} \right) \\ &\le \|Y_d\| 2\delta^{-1} \max_{1 \le i \le k} \|T_{i,k} - f(i/k)\| \to 0 \end{aligned}$$

from (7) and Theorem 1. Also, by the DCT, $\bar{g}_{d,k}(Y_\epsilon) \to g_d(Y_\epsilon)$ as $k \to \infty$. Hence, $g_d(Y_\epsilon) \ge \overline{\lim}\, g_{d,k}(\ell_k)$ holds and, since $g_d(Y_\epsilon) \le (1-\epsilon) g_d(Y_d)$ by convexity and $g_d(0_d) = 0$, we have, letting $\epsilon \to 0$, that

$$g_d(Y_d) \ge \overline{\lim}\, g_{d,k}(\ell_k). \tag{9}$$

Recalling $\ell_k \to b \in \overline{K}_d$, define $b_\epsilon = (1-\epsilon)b + \epsilon 0_d \in K_d$, so that $\min_{0 \le t \le 1}(1 + b_\epsilon' f(t)) > 0$. Then, $\bar{g}_{d,k}(b_\epsilon) \to g_d(b_\epsilon)$ by (7) and the DCT. And, by Taylor expansion and using (8),

$$\begin{aligned} \overline{\lim}\, |g_{d,k}(\ell_k) - \bar{g}_{d,k}(b_\epsilon)| &\le \overline{\lim} \max_{1 \le i \le k} |\ell_k' T_{i,k} - b_\epsilon' f(i/k)| \left(1 + \frac{1}{k} \sum_{i=1}^{k} \frac{1}{1 + b_\epsilon' f(i/k)} \right) \\ &\le \epsilon \sup_{0 \le t \le 1} |b' f(t)| \left(1 + \int_0^1 \frac{1}{1 + b_\epsilon' f(t)} dt \right) \equiv C(\epsilon), \end{aligned}$$

following from (7) and the DCT. Hence, we have

$$\underline{\lim} g_{d,k}(\ell_k) \geq g_d(b_\epsilon) - C(\epsilon). \tag{10}$$

We will show below that

$$\int_0^1 \frac{1}{1 + b' f(t)} dt < \infty \tag{11}$$

holds. In which case, $\lim_{\epsilon \to 0} \int_0^1 [1 + b'_\epsilon f(t)]^{-1} dt = \int_0^1 [1 + b' f(t)]^{-1} dt < \infty$ by the DCT and so that $C(\epsilon) \to 0$ as $\epsilon \to 0$ (noting $\sup_{0 \leq t \leq 1} |b' f(t)| < \infty$ since f is continuous and $\overline{K}_d$ is bounded by Theorem 1). By Fatou's lemma and the DCT, $\underline{\lim}_{\epsilon \to 0} g_d(b_\epsilon) \geq g_d(b)$ holds also. Hence, by (9)-(10), we then have

$$g_d(Y_d) \geq \overline{\lim}\, g_{d,k}(\ell_k) \geq \underline{\lim} g_{d,k}(\ell_k) \geq g_d(b) \geq g_d(Y_d),$$

implying $b = Y_d$ by the uniqueness of the minimizer and $\lim_{k \to \infty} k^{-1} \log R_k(\mu_0) = g_d(Y_d)$.

To finally show (11), let $A = \{t \in [0,1] : 1 + b' f(t) \leq d\}$ for some $0 < d \leq 1/2$ chosen so that $\{t \in [0,1] : 1 + b' f(t) = d\}$ has Lebesgue measure zero (since f is continuous). Let $A^c = [0,1] \setminus A$. Using the indicator function $\mathbb{I}(\cdot)$, define a simple function

$$h_k(t) \equiv \sum_{i=1}^{k} \frac{\ell'_k T_{i,k}}{1 + \ell'_k T_{i,k}} \mathbb{I}\left(t \in \left(\frac{i-1}{k}, \frac{i}{k}\right]\right), \quad t \in [0,1].$$

From (8), note that

$$\int_A h_k(t) dt + \int_{A^c} h_k(t) dt = \frac{1}{k} \sum_{i=1}^{k} \frac{\ell'_k T_{i,k}}{1 + \ell'_k T_{i,k}} = 0_d.$$

From (7), $\mathbb{I}(t \in A^c) h_k(t) \to \mathbb{I}(t \in A^c) b' f(t)/(1 + b' f(t))$ (almost everywhere (a.e.) Lebesgue measure) and for large k, $\mathbb{I}(t \in A^c)|h_k(t)| \leq 2C/d$ holds for $t \in [0,1]$, since eventually $\max_{1 \leq i \leq k} |\ell'_k T_{i,k}|$ is bounded by a constant $C > 0$ and also $1 + b' f(t) + (\ell'_k T_{i,k} - b' f(t)) > d/2$ for $t \in A^c$, $(i-1)/k < t \leq i/k$. Then, by the DCT, $\int_{A^c} h_k(t) dt \to \int_{A^c} b' f(t)/(1 + b' f(t)) dt$. And for $\delta \in (0,1)$, note

$$-\mathbb{I}(t \in A) h_k(t) \geq h_{1,k}(t) \equiv \sum_{i=1}^{k} \frac{-\ell'_k T_{i,k}}{1 + \ell'_k T_{i,k} + \delta \mathbb{I}(\text{sign}(\ell'_k T_{i,k}) < 0)} \mathbb{I}\left(t \in \left(\frac{i-1}{k}, \frac{i}{k}\right] \cap A\right)$$

Since $|h_{1,k}(t)| \leq C/\delta$ and $h_{1,k}(t) \to -\mathbb{I}(t \in A) b' f(t)/(1 + b' f(t) + \delta)$ (a.e. Lebesgue measure), by the DCT

$$0 \leq \int_A \frac{-b' f(t)}{1 + b' f(t) + \delta} dt = \lim_{k \to \infty} \int_A h_{1,k}(t) dt \leq \lim_{k \to \infty} \int_A -h_k(t) dt = \int_{A^c} \frac{b' f(t)}{1 + b' f(t)} dt$$

using $\int_A -h_k(t) dt = \int_{A^c} h_k(t) dt$. Letting $\delta \to 0$, Fatou's lemma gives

$$0 \leq \int_A \frac{-b' f(t)}{1 + b' f(t)} dt \leq \int_{A^c} \frac{b' f(t)}{1 + b' f(t)} dt < \infty.$$

Because $-b'f(t) \geq 1/2$ on A, $\int_A [1 + b'f(t)]^{-1} dt < \infty$ holds, implying (11). $\square$

Proof of Corollary 1. As in the proof of Theorem 2, we embed $\{S_n(\cdot)\}$ and $\{B(\cdot)\}$ in a larger probability space $(\Omega, \mathcal{F}, P)$ such that (7) holds, recalling $f(t) = w(t)B(t)$, $t \in [0,1]$. For simplicity and without loss of generality, we suppose that $\Sigma = I_{d\times d}$ (the identity matrix) and that $w(t) \in [0,1]$ in the following. For $\delta > 0$, there exists $L \equiv L(\delta) > 1$ and $\epsilon \equiv \epsilon(\delta) \in (0,1)$ such that

$$P\left(\max_{0\leq t\leq 1} \|B(t)\| > L\right) < \delta, \quad P(\|Y_d\| > L) < \delta, \quad P\left(\max_{0\leq t\leq \epsilon} \|B(t)\| > \frac{1}{2\max_{0\leq t\leq 1} w(t)}\right) < \delta,$$

with tightness following from Theorem 1 and the probability properties of the maximum of standard Brownian motion (cf. Athreya and Lahiri, 2006, ch. 15.2). Let $A \in \mathcal{F}$ be the event that (7) holds along with $\max_{0\leq t\leq 1} \|B(t)\| \leq L$, $\|Y_d\| \leq L$, and $\max_{0\leq t\leq \epsilon} \|B(t)\| \leq 1/[2\max_{0\leq t\leq 1} w(t)]$. We will show that there exists a fixed $M > 0$ (to be specified later but depending on δ) such that $\overline{\lim}\, Z_n(\omega)n^{1/2} \leq M$ holds for any $\omega \in A$. In which case,

$$\begin{aligned}
\lim_{n\to\infty} \sup_{m\geq n} P(Z_m m^{1/2} > M) &\leq P(A^c) + \lim_{n\to\infty} P\left(A \cap \bigcup_{m=n}^{\infty} \{\omega : Z_m(\omega)m^{1/2} > M\}\right) \\
&= P(A^c) + P\left(A \cap \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\omega : Z_m(\omega)m^{1/2} > M\}\right) \\
&\leq 3\delta + P(A, \overline{\lim}\, Z_n n^{1/2} > M) = 3\delta.
\end{aligned}$$

Since δ can be made arbitrarily small and M can be chosen depending on δ, $Z_n n^{1/2}$ is tight, establishing Corollary 1.

Fix $\omega \in A$ and we will suppress the dependence of random variables on ω. Suppose $\overline{\lim}\, Z_n n^{1/2} > M$. Then, there exists a subsequence $\{n_j\}$ and $\mu_{n_j} \in G_{n_j}$ such that $\|\mu_{n_j} - \mu_0\| > n_j^{-1/2}M$. Since G_{n_j} is connected without voids and $\mu_0 \in G_{n_j}$ (note $L_n(\mu_0) > 0$ eventually as in the proof of Theorem 2 by (7)), there exists $b_{n_j} \in \mathbb{R}^d$, $\|b_{n_j}\| = 1$ such that $v_{n_j} = n_j^{-1/2} M b_{n_j} + \mu_0 \in G_{n_j}$. We may exact a further subsequence $\{n_k\}$ such that $b_{n_k} \to b \in \mathbb{R}^d$, $\|b\| = 1$. In the following, denote $n_k \equiv k$ for simplicity. For $\mu \in \mathbb{R}^d$, let $T_{k,\mu}(0) = 0$ and $T_{k,\mu}(i/k) \equiv T_{i,\mu} = w(i/k)\sum_{j=1}^{i}(X_j - \mu)$, $i = 1, \ldots, k$, and define the process $T_{k,\mu}(t)$, $t \in [0,1]$ by linear interpolation of $\{T_{k,\mu}(i/k) : i = 0, \ldots, k\}$. Note, in (7), that $T_k(i/k) = T_{i,\mu_0} = T_{i,\mu} + iw(i/k)(\mu - \mu_0)$. Hence, by (7), $\sup_{0\leq t\leq 1} \|T_{k,v_k}/k^{1/2} - f_b(t)\| \to 0$ w.p.1(P), where $f_b(t) = w(t)[B(t) + Mtb]$. One can then show, as in the proof of Theorem 2, that $k^{-1}R_k(\mu_0) \to g_d(Y_d)$ and $k^{-1}R_k(v_k) \to C(M)$ for

$$C(M) \equiv \min\left\{-\int_0^1 \log(1 + a'f_b(t))dt : a \in \mathbb{R}^d, \min_{0\leq t\leq 1}(1 + a'f_b(t)) \geq 0\right\}.$$

On A, $g_d(Y_d) \geq -\log(1+L^2)$ and $C(M) \leq -\int_0^1 \log(1+b'f_b(t))dt$, where $\min_{0\leq t\leq \epsilon}(1+b'f_b(t)) \geq 1/2$ and, if $M = \epsilon^{-1}(L+[L^q/\min_{\epsilon\leq t\leq c} w(t)])$, $\min_{\epsilon\leq t\leq 1}(1+b'f_b(t)) \geq 1$ and $\min_{\epsilon\leq t\leq c}(1+b'f_b(t)) \geq 1+L^q$ hold for $q\geq 1$, where $w(t)>0$ for $t\in(0,c)$ by Assumption A.1. By choosing q large in this M, $C(M) \leq \log 2 - (c-\epsilon)\log(1+L^q) < -2\log(1+L^2)$ follows. Then, for large k, it holds that $k^{-1}\log R_k(v_k) < -2\log(1+L^2) < k^{-1}\log R_k(\mu_0)$, implying that $v_k \notin G_k$, which is a contradiction. Hence, on A, $\overline{\lim}\, Z_n n^{1/2} \leq M$ holds. $\square$

Proof of Theorem 3. As in the proof of Theorem 2, we again embed $\{S_n(\cdot)\}$ and $\{B(\cdot)\}$ in a larger probability space $(\Omega, \mathcal{F}, P)$ such that (7) holds with $f(t) = w(t)B(t)$, $t\in[0,1]$, implying 0_d is in the interior convex hull of $\{T_{i,\mu_0} : i = 1,\ldots,n\}$ and $R_n(\mu_0) > 0$ eventually w.p.1(P) (as in the proof of Theorem 2).

Using the proof of Theorem 1, one can show that there exists $\tilde{Y}_p \in \mathbb{R}^p$ which is the unique minimizer of $\tilde{g}_p(a) = -\int_0^1 \log(1+a'\tilde{f}(t))dt$ for $a \in \{y\in\mathbb{R}^p : \min_{0\leq t\leq 1}(1+y'\tilde{f}(t)) \geq 0\}$, where $\tilde{f}(t) = w(t)\tilde{B}(t)$ for $\tilde{B}(t) = [\nabla_{\mu_0}\Sigma\nabla'_{\mu_0}]^{-1/2}\nabla_{\mu_0}\Sigma^{1/2}B(t)$. Here, $\tilde{B}(\cdot)$ is distributed as vector of p iid components of standard Brownian motion, so that $\tilde{Y}_p$ and $\tilde{g}_p(\tilde{Y}_p)$ have the same distribution as Y_p and $g_p(Y_p)$ in Theorem 1, and all properties stated for Y_p, g_p, f in the Theorem 1 apply to $\tilde{Y}_p, \tilde{g}_p, \tilde{f}$. In particular, Theorem 1(iv) and Theorem 2(ii) imply $0 < \int_0^1 [1+\tilde{Y}_p'\tilde{f}(t)]^{-1}dt < \infty$ and $0 < \tilde{F}_1 \equiv \int_0^1 tw(t)[1+\tilde{Y}_p'\tilde{f}(t)]^{-1}dt < \infty$ w.p.1 (while f is bounded with w.p.1) so that $\tilde{F} \equiv \int_0^1 \Sigma^{1/2}f(t)[1+\tilde{Y}_p'\tilde{f}(t)]^{-1}dt/\tilde{F}_1$ is a $\mathbb{R}^p$-valued random variable. Hence, as in the proof of Corollary 1, we may construct an event $A\in\mathcal{F}$, with arbitrarily large probability, such that (7), $\|\tilde{F}\| \leq M/2$ and $\overline{\lim}\, Z_n n^{1/2} < M$ hold for $\omega \in A$ for some $M>0$ (suppressing the dependence of the variables on ω). On A, we shall show $n^{-1}\log R_n(\theta_0) \to \tilde{g}_p(\tilde{Y}_p)$, where again $\tilde{g}_p(\tilde{Y}_p)$ has the same distribution as $g_p(Y_p)$ in Theorem 1. Since $P(A)$ can be made arbitrarily large, the distributional convergence in Theorem 3 will then follow.

Fix $\omega\in A$. From the proof of Theorem 2, recall 0_d is in the interior convex hull of $\{T_{i,\mu_0} : i = 1,\ldots,n\}$ and $R_n(\mu_0) > 0$ eventually. Also, $Z_n n^{1/2} < M$ holds eventually, and we may write $R_n(\theta_0) = \sup D_n$ for

$$D_n = \left\{\prod_{i=1}^n np_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i T_{i,\mu} = 0_d, H(\mu) = \theta_0, \|\mu-\mu_0\| \leq Mn^{-1/2}\right\}$$

Note that, if $\sum_{i=1}^n p_i T_{i,\mu} = 0_d$ holds with $\min_{1\leq i\leq n} p_i > 0$ for some $\mu\in\mathbb{R}^q$ where $H(\mu) = \theta_0$ and $\|\mu-\mu_0\| \leq Mn^{-1/2}$, then $\mu = \sum_{i=1}^n p_i w(i/n) S_n(i/n)/\sum_{i=1}^n p_i i w(i/n)$ where $\sum_{i=1}^n p_i i w(i/n) > 0$ by Assumption A.1; by Taylor expansion around μ_0, $\nabla_{\mu^*}(\mu-\mu_0) = 0_p$ holds for some $\mu^*\in\mathbb{R}^d$ with $\|\mu^*-\mu_0\| \leq Mn^{-1/2}$, which implies $0_p = [\sum_{i=1}^n p_i i w(i/n)]\nabla_{\mu^*}(\mu-\mu_0) = \nabla_{\mu^*}\sum_{i=1}^n p_i T_{i,\mu_0}$.

Hence, $D_n \subset E_n$, where

$$E_n = \left\{ \prod_{i=1}^n np_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \nabla_\mu \sum_{i=1}^n p_i T_{i,\mu_0} = 0_p, \|\mu - \mu_0\| \leq Mn^{-1/2} \right\}$$

so that $\log R_n(\theta_0) = \log \sup D_n \leq \log \sup E_n$

Writing $\tilde{R}_n(\mu) = \sup\{\prod_{i=1}^n np_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \nabla_\mu \sum_{i=1}^n p_i T_{i,\mu_0} = 0_p\}$ for $\|\mu - \mu_0\| \leq Mn^{-1/2}$, it holds that $\log \tilde{R}_n(\mu) = -\sum_{i=1}^n (1 + \lambda_{n,\mu}' \nabla_\mu T_{i,\mu_0})$ where $\lambda_{n,\mu}$ is the minimizer of the strictly convex function $-\sum_{i=1}^n (1 + a' \nabla_\mu T_{i,\mu_0})$, $a \in \{y \in \mathbb{R}^d : \min_{1\leq i \leq n}(1 + y'\nabla_\mu T_{i,\mu_0}) > 0\}$ (following from $\tilde{R}_n(\mu) > 0$ because 0_d is in the interior convex hull of $\{T_{i,\mu_0} : i = 1, \ldots, n\}$ and that $n^{-2}\sum_{i=1}^n \nabla_{\mu_0} T_{i,\mu_0} T_{i,\mu_0}' \nabla_{\mu_0}'$ is eventually positive definite by $n^{-2}\sum_{i=1}^n \nabla_{\mu_0} T_{i,\mu_0} T_{i,\mu_0}' \nabla_{\mu_0}' \to \int_0^1 [\nabla_{\mu_0} \Sigma^{1/2} f(t)][\nabla_{\mu_0} \Sigma^{1/2} f(t)]' dt$ from (7) and the DCT). One can then show, as in the proof of Theorem 2, that

$$\frac{1}{n} \log \tilde{R}_n(\mu_0) = \sum_{i=1}^n \log(1 + \lambda_{n,\mu_0}' \nabla_{\mu_0} T_{i,\mu_0}) \to \tilde{g}_p(\tilde{Y}_p),$$

and $n^{1/2}[\nabla_{\mu_0} \Sigma \nabla_{\mu_0}']^{1/2} \lambda_{n,\mu_0} \to \tilde{Y}_p$. Also, for $\tilde{Y}_\epsilon = (1-\epsilon)\tilde{Y}_p + \epsilon 0_p$, it holds that $\min_{0 \leq t \leq 1}(1 + \tilde{Y}_\epsilon' \tilde{f}(t)) > 0$, which implies $\inf_{\|\mu - \mu_0\| \leq Mn^{-1/2}} \min_{1 \leq i \leq n}(1 + \tilde{Y}_\epsilon' \nabla_\mu T_{i,\mu_0}) > 0$ eventually. Since $n^{-1} \log \sup E_n = n^{-1} \sup_{\|\mu - \mu_0\| \leq Mn^{-1/2}} \log \tilde{R}_n(\mu)$ and, for each n, there exists $\|\mu_n - \mu_0\| \leq Mn^{-1/2}$ such that

$$-\frac{1}{n} + \frac{1}{n} \log \sup E_n \leq \frac{1}{n} \log \tilde{R}_n(\mu_n) = -\frac{1}{n} \sum_{i=1}^n (1 + \lambda_{n,\mu_n}' \nabla_{\mu_n} T_{i,\mu_0}) \leq -\frac{1}{n} \sum_{i=1}^n (1 + \tilde{Y}_\epsilon' \nabla_{\mu_n} T_{i,\mu_0}),$$

we have

$$\overline{\lim} \frac{1}{n} \log \sup D_n \leq \overline{\lim} \frac{1}{n} \log \sup E_n \leq \overline{\lim} \; -\frac{1}{n} \sum_{i=1}^n (1 + \tilde{Y}_\epsilon' \nabla_{\mu_n} T_{i,\mu_0}) = \tilde{g}_p(\tilde{Y}_\epsilon)$$

by (7) and the DCT. Since $\lim_{\epsilon \to 0} \tilde{g}_p(\tilde{Y}_\epsilon) = \tilde{g}_p(\tilde{Y}_p)$ by convexity of $\tilde{g}_p$ (i.e., $0 \leq \tilde{g}_p(\tilde{Y}_\epsilon) - \tilde{g}_p(\tilde{Y}_p) \leq -\epsilon \tilde{g}_p(\tilde{Y}_p)$), it holds that $\overline{\lim}\, n^{-1} \log \sup D_n \leq \tilde{g}_p(\tilde{Y}_p)$.

By Taylor expansion for $\|\mu - \mu_0\| \leq Mn^{-1/2}$, we can write $H(\mu) - \theta_0 = J_\mu(\mu - \mu_0)$, for the $p \times d$ matrix $J_\mu = \int_0^1 \nabla_{\mu_0 + t(\mu - \mu_0)} dt$. Now similarly to $\tilde{R}_n(\mu)$ above, for $\bar{R}_n(\mu) = \sup\{\prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, J_\mu \sum_{i=1}^n p_i T_{i,\mu_0} = 0_p\}$ with $\|\mu - \mu_0\| \leq Mn^{-1/2}$, we can write $\log \bar{R}_n(\mu) = -\sum_{i=1}^n (1 + \bar{\lambda}_{n,\mu}' J_\mu T_{i,\mu_0})$ where $\bar{\lambda}_{n,\mu}$ satisfies

$$\min_{1 \leq i \leq n} (1 + \bar{\lambda}_{n,\mu}' J_\mu T_{i,\mu_0}) > 0, \quad \sum_{i=1}^n \frac{1}{n(1 + \bar{\lambda}_{n,\mu}' J_\mu T_{i,\mu_0})} = 1, \quad \sum_{i=1}^n \frac{J_\mu T_{i,\mu_0}}{1 + \bar{\lambda}_{n,\mu}' J_\mu T_{i,\mu_0}} = 0_p. \quad (12)$$

Write

$$h_n(\mu) = \sum_{i=1}^n \frac{S_n(i/n) w(i/n)}{n^{3/2}(1 + \bar{\lambda}_{n,\mu}' J_\mu T_{i,\mu_0})} \Bigg/ \sum_{i=1}^n \frac{i w(i/n)}{n^{3/2}(1 + \bar{\lambda}_{n,\mu}' J_\mu T_{i,\mu_0})},$$

which is a continuous function of μ, $\|\mu - \mu_0\| \leq Mn^{-1/2}$, because $\bar{\lambda}_{n,\mu}$ is as well by the implicit function theorem (following from the fact that, as mentioned above, that $n^{-2}\sum_{i=1}^{n} \nabla_{\mu_0} T_{i,\mu_0} T'_{i,\mu_0} \nabla'_{\mu_0}$ is eventually positive definite); see Qin and Lawless (1994, p. 305). Note also that

$$n^{1/2}(h_n(\mu) - \mu_0) = \sum_{i=1}^{n} \frac{T_{i,\mu_0}}{n^{3/2}(1 + \bar{\lambda}'_{n,\mu} J_\mu T_{i,\mu_0})} \Bigg/ \sum_{i=1}^{n} \frac{iw(i/n)}{n^2(1 + \bar{\lambda}'_{n,\mu} J_\mu T_{i,\mu_0})}.$$

Using (7) and DCT, one can show $\sup_{\|\mu-\mu_0\| \leq Mn^{-1/2}} \|n^{1/2}(h_n(\mu) - \mu_0) - \tilde{F}\| \to 0$ so that, on A (for which $\|\tilde{F}\| \leq M/2$), $\sup_{\|\mu-\mu_0\| \leq Mn^{-1/2}} \|h_n(\mu) - \mu_0\| \leq Mn^{-1/2}$ holds eventually. That is, for all large n, $h_n(\mu) - \mu_0$ is a continuous mapping from the closed ball $\|\mu - \mu_0\| \leq Mn^{-1/2}$ to itself. By Brouwer's fixed point theorem, there then exists μ_n^* such that $h_n(\mu_n^*) = \mu_n^*$ for $\|\mu_n^* - \mu_0\| \leq Mn^{-1/2}$. This implies $0_p = h_n(\mu_n^*) - \mu_n^*$, which when multiplied by $\sum_{i=1}^{n} iw(i/n)/[1 + \bar{\lambda}'_{n,\mu_n^*} J_{\mu_n^*} T_{i,\mu_0}] > 0$ (positivity by (12) and Assumption A.1), further implies that

$$0_p = \sum_{i=1}^{n} \frac{T_{i,\mu_n^*}}{n(1 + \bar{\lambda}'_{n,\mu_n^*} J_{\mu_n^*} T_{i,\mu_0})} \tag{13}$$

and also

$$\sum_{i=1}^{n} \frac{iw(i/n)}{1 + \bar{\lambda}'_{n,\mu_n^*} J_{\mu_n^*} T_{i,\mu_0}} \cdot J_{\mu_n^*}(\mu_n^* - \mu_0) = \sum_{i=1}^{n} \frac{J_{\mu_n^*} T_{i,\mu_0}}{1 + \bar{\lambda}'_{n,\mu_n^*} J_{\mu_n^*} T_{i,\mu_0}} = 0_p$$

by $\mu_n^* - \mu_0 = h_n(\mu_n^*) - \mu_0$ and (12); from $\sum_{i=1}^{n} iw(i/n)/[1 + \bar{\lambda}'_{n,\mu_n^*} J_{\mu_n^*} T_{i,\mu_0}] > 0$, it holds that $J_{\mu_n^*}(\mu_n^* - \mu_0) = 0_p$ or equivalently $H(\mu_n^*) = H(\mu_0) = \theta_0$. In other words, this last fact combined with (12)-(13) entail that all conditions are satisfied for $\prod_{i=1}^{n}[1 + \bar{\lambda}'_{n,\mu_n^*} J_{\mu_n^*} T_{i,\mu_0}]^{-1} \in D_n$. Hence,

$$\underline{\lim} \frac{1}{n} \log \bar{R}_n(\mu_n^*) \leq \underline{\lim} \frac{1}{n} \log \sup D_n \leq \overline{\lim} \frac{1}{n} \log \sup D_n \leq \tilde{g}_p(\tilde{Y}_p),$$

the last inequality being established earlier. One can then show $\frac{1}{n} \log \bar{R}_n(\mu_n^*) \to \tilde{g}_p(\tilde{Y}_p)$ using (7) so that, on A, $n^{-1} \log R_n(\theta_0) = n^{-1} \log \sup D_n \to \tilde{g}_p(\tilde{Y}_p)$, completing the proof of Theorem 3. □

References

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory.* Springer, New York.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.*, **6**, 434-451.

Bravo, F. (2005). Blockwise empirical entropy tests for time series regressions. *J. Time Series Anal.*, **26**, 185-210.

Bravo, F. (2009). Blockwise generalized empirical likelihood inference for non-linear dynamic moment conditions models. *Econom. J.*, **12**, 208-231.

Bunzel, H., Kiefer, N. M. and Vogelsang, T. J. (2001). Simple robust testing of hypotheses in nonlinear models. *J. Am. Statist. Assoc.*, **96**, 1088-1096.

Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.*, **14**, 1171-1179.

Chen, S. X. and Wong, C. M. (2009). Smoothed block empirical likelihood for quantiles of weakly dependent processes. *Statist. Sinica*, **19**, 71-81.

Freedman, D. (1983). *Brownian motion and diffusion.* Springer, New York.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.*, **20**, 231-246.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York: Springer- Verlag.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.*, **58**, 109-127.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, **12**, 141-153.

Kiefer, N. M. and Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, **70**, 2093-2095.

Kiefer, N. M., Vogelsang, T. J. and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, **68**, 695-714.

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

Künsch, H. R. (1989). The jackknife and bootstrap for general stationary observations.*Ann. Statist.*, **17**, 1217-1261.

Lahiri, S. N. (2003). *Resampling methods for dependent data.* Springer, New York.

Lei, Q. and Qin, Y. (2011). Empirical likelihood for quantiles under negatively associated samples. *J. Statist. Plann. Inference*, **141**, 1325-1332.

Lin, L. and Zhang, R. (2001). Blockwise empirical Euclidean likelihood for weakly dependent processes. *Statist. Prob. Lett*, **53**, 143-152.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap,* 225-248, R. LePage and L. Billard (editors). John Wiley & Sons, New York.

Lobato, I. N. (2001). Testing that a dependent process is correlated. *J. Am. Statist. Assoc.*, **96**, 1066-1076.

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

Nordman, D. J., Sibbertsen, P. and Lahiri, S. N. (2007). Empirical likelihood for the mean under long-range dependence. *J. Time Series Anal.*, **28**, 576-599.

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

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.

Paparoditis, E. and Politis, D. N. (2001). Tapered block bootstrap. *Biometrika*, **88**, 1105-1119.

Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statist. Sci.*, **18**, 219-230.

Politis, D. N. and Romano, J. P. (1993). On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes Appl.*, **45**, 155-167.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300-325.

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

Sun, Y., Phillips, P. C. B. and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, **76**, 175-194.

Tong, H. (1978). On a threshold model. In C.H. Chen (ed.), *Pattern Recognition and Signal Processing*. Amsterdam: Sijthoff and Noordhoff.

van der Vaart, A. W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

Wu, R. and Cao, J. (2011). Blockwise empirical likelihood for time series of counts. *J. Multivariate Anal.* **102**, 661-673.

Zhang, J. (2006). Empirical likelihood for NA series. *Statist. Prob. Lett.*, **76**, 153-60.

Figure 2: Coverage rates of nominal 90% CIs for the process mean with Models *a*-*h* using $N(0,1)$ (left) or centered χ_1^2 (right) innovations for sample sizes $n = 250$, 500 or 1000. Axis labels for EBEL denote the weight functions in Table 2; labels 1-3 for NBEL/BEL/TBEL denote increasing block sizes.

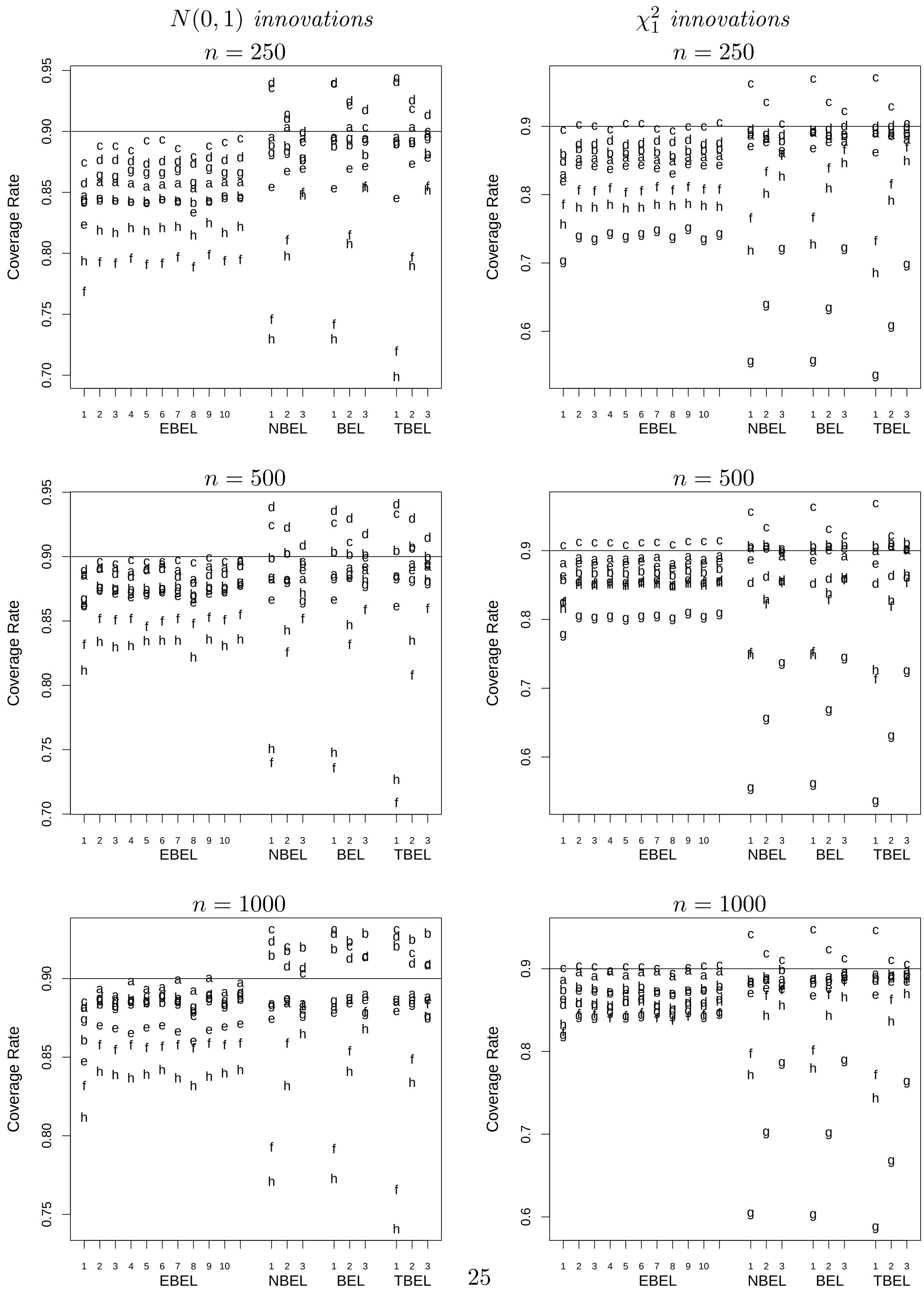


Figure 3: Coverage rates of nominal 90% CIs for the process mean with Models $1, \ldots, 9, 0$ using centered Pareto (left) or Ber(0.5) (right) innovations, for sample sizes $n = 250$, 500 or 1000.

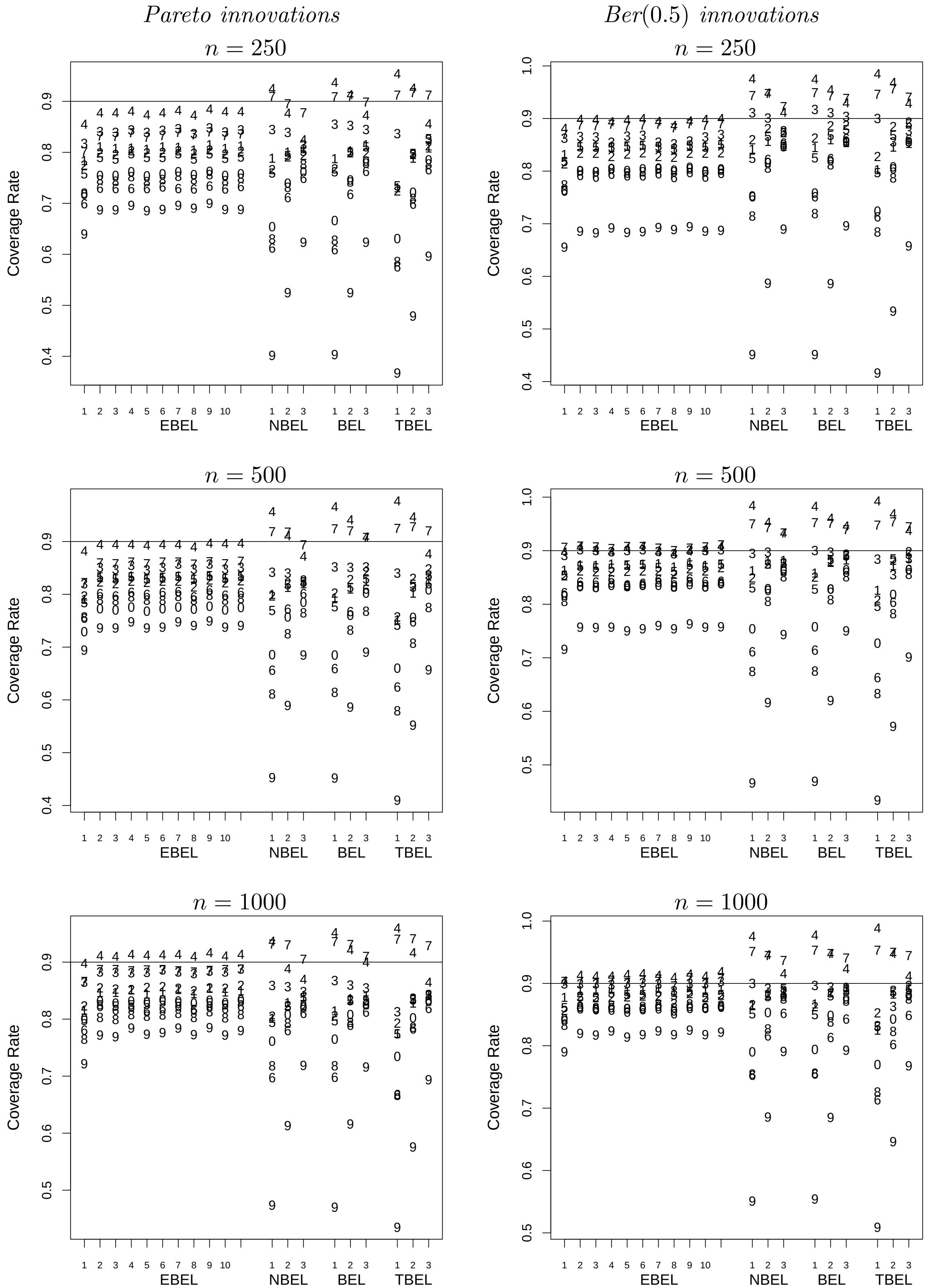
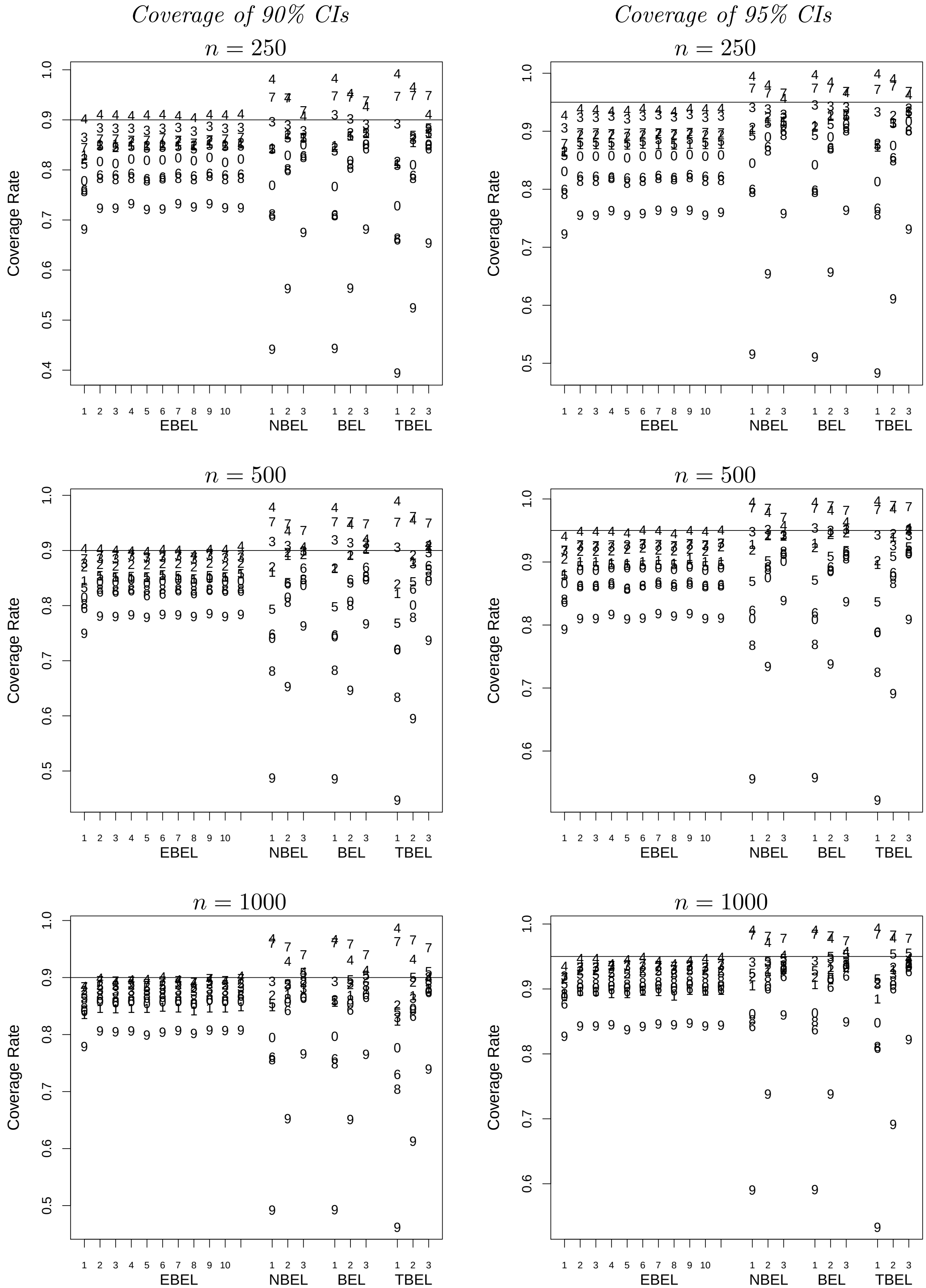


Figure 4: Coverage rates of nominal 90% (left) and 95% (right) CIs for the process mean with Models $1, \ldots, 9, 0$ using centered χ_1^2 innovations, for sample sizes $n = 250$, 500 or 1000.



Research Papers 2012

Center for Research in Econometric Analysis of Time Series

2012-39: H. Peter Boswijk, Michael Jansson and Morten Ørregaard Nielsen: Improved Likelihood Ratio Tests for Cointegration Rank in the VAR Model

2012-40: Mark Podolskij, Christian Schmidt and Johanna Fasciati Ziegel: Limit theorems for non-degenerate U-statistics of continuous semimartingales

2012-41: Eric Hillebrand, Tae-Hwy Lee and Marcelo C. Medeiros: Let's Do It Again: Bagging Equity Premium Predictors

2012-42: Stig V. Møller and Jesper Rangvid: End-of-the-year economic growth and time-varying expected returns

2012-43: Peter Reinhard Hansen and Allan Timmermann: Choice of Sample Split in Out-of-Sample Forecast Evaluation

2012-44: Peter Reinhard Hansen and Zhuo Huang: Exponential GARCH Modeling with Realized Measures of Volatility Statistics

2012-45: Peter Reinhard Hansen and Allan Timmermann: Equivalence Between Out-of-Sample Forecast Comparisons and Wald

2012-46: Søren Johansen, Marco Riani and Anthony C. Atkinson: The Selection of ARIMA Models with or without Regressors

2012-47: Søren Johansen and Morten Ørregaard Nielsen: The role of initial values in nonstationary fractional time series models

2012-48: Peter Christoffersen, Vihang Errunza, Kris Jacobs and Hugues Langlois: Is the Potential for International Diversi…cation Disappearing? A Dynamic Copula Approach

2012-49: Peter Christoffersen, Christian Dorion , Kris Jacobs and Lotfi Karoui: Nonlinear Kalman Filtering in Affine Term Structure Models

2012-50: Peter Christoffersen, Kris Jacobs and Chayawat Ornthanalai: GARCH Option Valuation: Theory and Evidence

2012-51: Tim Bollerslev, Lai Xu and Hao Zhou: Stock Return and Cash Flow Predictability: The Role of Volatility Risk

2012-52: José Manuel Corcuera, Emil Hedevang, Mikko S. Pakkanen and Mark Podolskij: Asymptotic theory for Brownian semi-stationary processes with application to turbulence

2012-53: Rasmus Søndergaard Pedersen and Anders Rahbek: Multivariate Variance Targeting in the BEKK-GARCH Model

2012-54: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperature Trends and Co-Shifting: A Shifting Mean Vector Autoregressive Analysis

2012-55: Daniel J. Nordman, Helle Bunzel and Soumendra N. Lahiri: A Non-standard Empirical Likelihood for Time Series