FIXED-b ASYMPTOTICS FOR BLOCKWISE EMPIRICAL LIKELIHOOD

Xianyang Zhang and Xiaofeng Shao

University of Missouri-Columbia and
University of Illinois at Urbana-Champaign

Abstract: We describe an extension of the fixed-b approach introduced by Kiefer and Vogelsang (2005) to the empirical likelihood estimation framework. Under fixed-b asymptotics, the empirical likelihood ratio statistic evaluated at the true parameter converges to a nonstandard yet pivotal limiting distribution that can be approximated numerically. The impact of the bandwidth parameter and kernel choice is reflected in the fixed-b limiting distribution. Compared to the χ^2-based inference procedure used by Kitamura (1997) and Smith (2011), the fixed-b approach provides a better approximation to the finite sample distribution of the empirical likelihood ratio statistic. Correspondingly, as shown in our simulation studies, the confidence region based on the fixed-b approach has more accurate coverage than its traditional counterpart.

Key words and phrases: Blocking, empirical likelihood, fixed-b asymptotics, time series.

1. Introduction

Empirical likelihood (EL) (Owen (1988, 1990)) is a nonparametric technique for conducting inference for parameters in nonparametric settings. EL has been studied extensively in the statistics and econometrics literature (see Owen (2001), Kitamura (2006), and Chen and Van Keilegom (2009) for comprehensive reviews). One striking property of EL is the nonparametric version of Wilks' theorem that states that the EL ratio statistic evaluated at the true parameter converges to a χ^2 liming distribution. This property was first demonstrated for the mean parameter by Owen (1990) and was further extended to the estimating equation framework by Qin and Lawless (1994). However, Wilks' phenomenon fails to hold for stationary time series because the dependence within the observations is not taken into account in EL. Kitamura (1997) proposed blockwise empirical likelihood (BEL), which is able to accommodate the dependence of the data, and Wilks' theorem continues to hold for the BEL ratio statistic under suitable weak dependence assumptions. The BEL can be viewed as a special case of the generalized empirical likelihood (GEL) with smoothed moment conditions (Smith (2011)).

The performance of BEL and its variations (Nordman (2009) and Smith (2011)) can depend crucially on the choice of the bandwidth parameter for which no sound guidance is available. Kiefer and Vogelsang (2005) proposed the so-called fixed-b asymptotic theory in the heteroscedasticity-autocorrelation robust (HAR) testing context. It was found that the asymptotic distribution obtained by treating the bandwidth as a fixed proportion (say b) of the sample size provides a better approximation to the sampling distribution of the studentized test statistic than the traditional χ^2-based approximation. See Jansson (2004), Sun, Phillips, and Jin (2008), and Zhang and Shao (2013) for rigorous theoretical justifications. The fixed-b approach has the advantage of accounting for the effect of the bandwidth and the kernel, as different bandwidth parameters and kernels correspond to different limiting (null) distributions (also see Shao and Politis (2013) for a recent extension to the subsampling and block bootstrap context). The main thrust of the present paper is the development of a new asymptotic theory in the BEL estimation framework made possible by the fixed-b approach. We consider the problem in the moment condition model (Qin and Lawless (1994) and Smith (2011)) that is a fairly general framework used by both statisticians and econometricians. Under the fixed-b asymptotic framework, we show that the asymptotic null distribution of the EL ratio statistic evaluated at the true parameter is nonstandard yet pivotal, and that it can be approximated numerically. It is interesting to note that the fixed-b limiting distribution coincides with the χ^2 distribution as b gets close to zero. We also illustrate the idea in the GEL estimation framework and demonstrate the usefulness of the fixed-b approach through simulation studies.

For notation, let $D[0,1]$ be the space of functions on $[0,1]$ which are right-continuous and have left limits, endowed with the Skorokhod topology (see Billingsley (1999)). Weak convergence in $D[0,1]$, or more generally in the $\mathbb{R}^m$-valued function space $D^m[0,1]$, is denoted by " $\Rightarrow$ ", where $m \in \mathbb{N}$. Convergence in probability and convergence in distribution are denoted by " $\rightarrow^p$ " and " $\overset{d}{\rightarrow}$ " respectively. Let $C[0,1]$ be the space of continuous functions on $[0,1]$. Denote by $\lfloor a \rfloor$ the integer part of $a \in \mathbb{R}$.

2. Methodology

2.1. Empirical likelihood

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

$$\mathbb{E}[f(y_t, \theta_0)] = 0, \quad t = 1, 2, \ldots, n, \qquad (2.1)$$

where $f(y, \theta) : \mathbb{R}^{l+p} \to \mathbb{R}^k$ is a map that is differentiable with respect to θ and rank$(E[\partial f(y_t, \theta_0)/\partial \theta']) = p$ with $k \geq p$. To deal with time series data, we consider the smoothed moment conditions introduced by Smith (2011),

$$f_{tn}(\theta) = \frac{1}{S_n} \sum_{s=t-n}^{t-1} \mathcal{K}\left(\frac{s}{S_n}\right) f(y_{t-s}, \theta), \tag{2.2}$$

where $\mathcal{K}(\cdot)$ is a kernel function and $S_n = bn$ with $b \in (0, 1)$ is the bandwidth parameter. Smoothing of the moment conditions induces a heteroskedasticity and autocorrelation consistent (HAC) covariance estimator of the long run variance matrix of $\{f(y_t, \theta)\}_{t=1}^n$. Let $f_t(\theta) = f(y_t, \theta)$ and $\tilde{f}_n(\theta) = \sum_{t=1}^n f_{tn}(\theta)/n$, where $f_{tn}(\theta)$ is defined in (2.2).

Consider the profile empirical log-likelihood function based on the smoothed moment restrictions,

$$\mathcal{L}_n(\theta) = \sup\left\{\sum_{t=1}^n \log(p_t) : p_t \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t f_{tn}(\theta) = 0\right\}. \tag{2.3}$$

Standard Lagrange multiplier arguments imply that the maximum is attained when

$$p_t = \frac{1}{n\{1 + \lambda' f_{tn}(\theta)\}}, \quad \text{with} \quad \sum_{t=1}^n \frac{f_{tn}(\theta)}{1 + \lambda' f_{tn}(\theta)} = 0.$$

The maximum empirical likelihood estimate (MELE) is then given by $\hat{\theta}_{el} = \text{argmax}_{\theta \in \Theta} \mathcal{L}_n(\theta)$. Following Kitamura (2006), the empirical log-likelihood function can also be derived by considering the dual problem (see e.g., Borwein and Lewis (1991)),

$$\mathcal{L}_n(\theta) = \min_{\lambda \in \mathbb{R}^k} -\sum_{t=1}^n \log(1 + \lambda' f_{tn}(\theta)) - n \log n, \tag{2.4}$$

where $\log(x) = -\infty$ for $x < 0$. Here (2.4) has a natural connection with the generalized empirical likelihood (GEL), and it facilitates our theoretical derivation under the fixed-b asymptotics. To introduce the fixed-b approach, we define the empirical log-likelihood ratio function

$$elr(\theta) = 2 \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^n \frac{\log(1 + \lambda' f_{tn}(\theta))}{S_n}, \tag{2.5}$$

for $\theta \in \Theta$ and $S_n = bn$. Under the traditional small-b asymptotics, $nb^2 + 1/(nb) \to 0$ as $n \to \infty$, and suitable weak dependence assumptions (see e.g., Smith (2011)), it can be shown that

$$elr(\theta_0) = n\tilde{f}_n(\theta_0)'\left(b\sum_{t=1}^n f_{tn}(\theta_0)f_{tn}(\theta_0)'\right)^{-1}\tilde{f}_n(\theta_0) + o_p(1) \xrightarrow{d} \frac{\kappa_1^2}{\kappa_2}\chi_k^2,$$

where $\kappa_1 = \int_{-\infty}^{+\infty} \mathcal{K}(x)dx$ and $\kappa_2 = \int_{-\infty}^{+\infty} \mathcal{K}^2(x)dx$ (assuming that $\kappa_1, \kappa_2 < \infty$). However, the χ^2-based approximation can be poor, especially when the dependence is strong and the bandwidth parameter is large (see Section 3). To derive the fixed-b limiting distribution, we make an assumption that is standard in moment condition models.

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

Assumption 1 can be verified under suitable moment and weak dependence assumptions on $f_t(\theta_0)$ (see e.g., Phillips (1987)). For the kernel function, we assume the following.

Assumption 2. *The kernel* $\mathcal{K} : \mathbb{R} \to [-c_0, c_0]$ *for some* $0 < c_0 < \infty$*, is piecewise continuously differentiable.*

Fix $b \in (0, 1)$, where $b = S_n/n$. Using summation by parts, the Continuous Mapping Theorem and Itô's formula, it is not hard to show that, for $t = \lfloor nr \rfloor$ with $r \in [0, 1]$,

$$\sqrt{n}f_{tn}(\theta_0) = \frac{\sqrt{n}}{S_n} \sum_{s=t-n}^{t-1} \mathcal{K}\left(\frac{s}{S_n}\right)f_{t-s}(\theta_0) \Rightarrow \frac{\Lambda D_k(r; b)}{b}, \qquad (2.6)$$

where $D_k(r; b) = \int_0^1 \mathcal{K}((r-s)/b)dW_k(s)$. Let $C^{\otimes k}[0, 1] = \{(f_1, f_2, \ldots, f_k) : f_i \in C[0, 1]\}$. For any $g \in C^{\otimes k}[0, 1]$, take $G_{el}(g) = \max_{\lambda \in \mathbb{R}^k} \int_0^1 \log(1 + \lambda'g(t))dt$. We show in the Appendix that the functional $G_{el}(\cdot)$ is continuous under the sup norm. Therefore, by the Continuous Mapping Theorem, we can characterize the asymptotic behavior of $elr(\theta_0)$.

Theorem 1. *Suppose Assumptions 1−2 hold. For $n \to +\infty$ and b fixed,*

$$elr(\theta_0) \xrightarrow{d} U_{el,k}(b; \mathcal{K}) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^1 \log\left(1 + \lambda' \int_0^1 \mathcal{K}\left(\frac{r-s}{b}\right)dW_k(s)\right) dr. \quad (2.7)$$

The proof of Theorem 1 is given in the supplementary material. Theorem 1 shows that the fixed-b limiting distribution of $elr(\theta_0)$ is nonstandard yet pivotal for a given bandwidth and kernel, and its critical values can be obtained via simulation or iid bootstrap (because the bootstrapped sample satisfies the Functional Central Limit Theorem). Let $u_{el,k}(b; \mathcal{K}; 1 - \alpha)$ be the $100(1 - \alpha)\%$ quantile of $U_{el,k}(b; \mathcal{K})/(1 - b)$. Given $b \in (0, 1)$, a $100(1 - \alpha)\%$ confidence region for the parameter θ_0 is then given by

$$CI(1 - \alpha; b) = \left\{\theta \in \mathbb{R}^p : \frac{elr(\theta)}{1 - b} \le u_{el,k}(b; \mathcal{K}; 1 - \alpha)\right\}. \qquad (2.8)$$

When $\mathcal{K}(x) = \mathbf{I}(x \geq 0)$, we have $D_k(r; b) = W_k(r)$ and $\mathcal{A} := \{\lambda \in \mathbb{R}^k : \min_{r \in [0,1]}(1 + \lambda' D_k(r; b)) \geq 0\} = \{\lambda \in \mathbb{R}^k : \min_{r \in [0,1]}(1 + \lambda' W_k(r)) \geq 0\}$. By Lemma 1 of Nordman, Bunzel, and Lahiri (2013), we know that $\mathcal{A}$ is bounded with probability one, which implies that $P(U_{el,k}(b; \mathcal{K}) = \infty) = 0$. We conjecture that $P(U_{el,k}(b; \mathcal{K}) = \infty)$ can be positive for particular $\mathcal{K}(\cdot)$ and $b \in (0,1)$. In our simulations, critical values are calculated based on the cases where $U_{el,k}(b; \mathcal{K}) < \infty$ (when b is close to zero, $P(U_{el,k}(b; \mathcal{K}) = \infty)$ is rather small, as seen from our unreported simulation results). The nonstandard limiting distribution also provides some insights on how likely the origin is not contained in the convex hull of $\{f_{tn}(\theta_0)\}_{t=1}^n$ when the sample size n is large.

Remark 1. To capture the dependence within the observations, one can employ the commonly used blocking technique first applied to the EL by Kitamura (1997). To illustrate, we consider the fully overlapping smoothed moment condition given by $f_{tn}(\theta) = (1/m) \sum_{j=t}^{t+m-1} f(y_j, \theta)$ with $t = 1, 2, \ldots, n - m + 1$ and $m = \lfloor nb \rfloor$ for $b \in (0,1)$. Under suitable weak dependence assumptions, we have $\sqrt{n} f_{tn}(\theta_0) \Rightarrow \Lambda\{W_k(r + b) - W_k(r)\}/b$ for $t = \lfloor nr \rfloor$. Using similar arguments to those in Theorem 1, we can show that

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

We generate the critical values of $U_{el,k}(b)/(1 - b)$ (conditioning on $U_{el,k}(b) < \infty$) for b from 0.01 to 0.3 with spacing 0.01, and further approximate the critical values by a cubic function of b following the practice of Kiefer and Vogelsang (2005). The estimates of the coefficients of the corresponding cubic functions are given in Table 1. Similarly, we summarize the critical values of $U_{el,k}(b; \mathcal{K})/(1 - b)$ (conditioning on $U_{el,k}(b; \mathcal{K}) < \infty$) with $\mathcal{K}(x) = (5\pi/8)^{1/2}(1/x)J_1(6\pi x/5)$ for b from 0.01 to 0.2 in Table 2, where $J_1(\cdot)$ denotes the Bessel function of the first kind.

Remark 2. A natural question to ask is whether the fixed-b asymptotics is consistent with the traditional small-b asymptotics when b is close to zero. We provide an affirmative answer by showing that $U_{el,k}(b, \mathcal{K})$ converges to a scaled χ_k^2 distribution as $b \to 0$. We assume $\mathcal{K}$ satisfies certain regularity conditions (see Assumption 2.2 in Smith (2011)). Using the Taylor expansion and some standard arguments for EL, it is not hard to show that

$$U_{el,k}(b; \mathcal{K}) = \frac{1}{b} \int_0^1 D_k(r, b)' dr \left(\int_0^1 D_k(r; b) D_k(r; b)' dr \right)^{-1} \int_0^1 D_k(r, b) dr + o_p(1).$$

Table 1. Critical value function coefficients.

	a_0	a_1	a_2	a_3	R^2
$u_{el,1}(b; 0.90)$	2.661	6.547	12.819	-8.329	0.9984
$u_{el,1}(b; 0.95)$	3.917	5.819	34.483	-14.192	0.9976
$u_{el,1}(b; 0.99)$	6.593	4.631	231.740	-484.131	0.9855
$u_{el,2}(b; 0.90)$	4.827	-1.521	212.177	-477.642	0.9983
$u_{el,2}(b; 0.95)$	6.586	-18.806	469.034	-1076.779	0.9945
$u_{el,2}(b; 0.99)$	9.928	-36.918	1028.429	-2530.245	0.9906
$u_{el,3}(b; 0.90)$	6.424	-1.099	405.193	-1072.778	0.9962
$u_{el,3}(b; 0.95)$	7.783	3.125	560.737	-1552.979	0.9909
$u_{el,3}(b; 0.99)$	10.138	22.209	1080.359	-3330.819	0.9565

THE CRITICAL VALUE $u_{el,k}(b; 1-\alpha)$ IS APPROXIMATED BY A CUBIC FUNCTION $a_0 + a_1 b + a_2 b^2 + a_3 b^3$ OF b. THE ESTIMATED COEFFICIENTS AND MULTIPLE R^2 ARE REPORTED. THE BROWNIAN MOTION IS APPROXIMATED BY A NORMALIZED PARTIAL SUM OF $1,000$ I.I.D. STANDARD NORMAL RANDOM VARIABLES AND THE NUMBER OF MONTE CARLO REPLICATION IS $5,000$.

Table 2. Critical value function coefficients.

	a_0	a_1	a_2	a_3	R^2
$u_{el,1}(b; \mathcal{K}; 0.90)$	3.324	8.243	112.149	-155.159	0.9992
$u_{el,1}(b; \mathcal{K}; 0.95)$	5.116	-9.585	533.935	-1459.216	0.9989
$u_{el,1}(b; \mathcal{K}; 0.99)$	9.612	-85.001	2237.407	-6595.415	0.9955
$u_{el,2}(b; \mathcal{K}; 0.90)$	5.650	8.652	799.757	-2928.350	0.9930
$u_{el,2}(b; \mathcal{K}; 0.95)$	7.709	-23.367	1833.534	-6752.999	0.9868
$u_{el,2}(b; \mathcal{K}; 0.99)$	11.708	-54.068	4501.733	-17748.759	0.9802
$u_{el,3}(b; \mathcal{K}; 0.90)$	6.860	48.990	1373.570	-6288.066	0.9804
$u_{el,3}(b; \mathcal{K}; 0.95)$	7.656	105.088	1714.933	-8718.140	0.9731
$u_{el,3}(b; \mathcal{K}; 0.99)$	4.963	505.832	346.210	-9711.193	0.9502

THE CRITICAL VALUE $u_{el,k}(b; \mathcal{K}; 1-\alpha)$ IS APPROXIMATED BY A CUBIC FUNCTION $a_0 + a_1 b + a_2 b^2 + a_3 b^3$ OF b. THE ESTIMATED COEFFICIENTS AND MULTIPLE R^2 ARE REPORTED. THE BROWNIAN MOTION IS APPROXIMATED BY A NORMALIZED PARTIAL SUM OF $1,000$ IID STANDARD NORMAL RANDOM VARIABLES AND THE NUMBER OF MONTE CARLO REPLICATION IS $5,000$.

Under Assumption 2, we derive that

$$\frac{1}{b} \int_0^1 D_k(r, b) dr = \int_0^1 \int_0^1 \mathcal{K}\left(\frac{r-s}{b}\right) dr d\frac{W_k(s)}{b}$$
$$= \int_0^1 \int_{-s/b}^{(1-s)/b} \mathcal{K}(t) dt dW_k(s) \xrightarrow{d} \kappa_1 W_k(1).$$

Define the semi-positive definite kernel $\mathcal{K}_b^*(r, s) = \int_0^1 \mathcal{K}((t-r)/b)\mathcal{K}((t-s)/b) dt/(b\kappa_2)$.

Then $V_k(b) = (1/b)\int_0^1 D_k(r;b)D_k(r;b)'dr = \kappa_2\int_0^1\int_0^1 \mathcal{K}_b^*(r,s)dW_k(r)dW_k(s) \overset{d}{=} \kappa_2\sum_{j=1}^{+\infty}\lambda_{j,b}\eta_j\eta_j'$, where $\{\eta_j\}_{j=1}^{+\infty}$ is an independent sequence of $N_k(0,I_k)$ random vectors and the $\lambda_{j,b}$ are the eigenvalues associated with $\mathcal{K}_b^*(r,s)$. Note that

$$\frac{\mathbb{E}\{V_k(b)\}}{\kappa_2} = I_k\int_0^1 \mathcal{K}_b^*(r,r)dr = \frac{I_k}{\kappa_2 b}\int_0^1\int_0^1 \mathcal{K}^2\Big(\frac{t-r}{b}\Big)dtdr \to I_k.$$

Let $\eta_j = (\eta_{j1},\ldots,\eta_{jk})$ and denote by $V_k^{(l,m)}(b)$ the (l,m)th element of $V_k(b)$ with $1 \le l,m \le k$. Since $\sum_{j=1}^{+\infty}\lambda_{j,b}^2 = \int_0^1\int_0^1\{\mathcal{K}_b^*(r,s)\}^2 drds \to 0$ as $b \to 0$ (see e.g., Sun (2010)), we get

$$\mathbb{E}\Big\{\frac{V_k^{(l,m)}(b)}{\kappa_2}\Big\}^2 = \sum_{j=1}^{+\infty}\sum_{j'=1}^{+\infty}\lambda_{j,b}\lambda_{j',b}\mathbb{E}\eta_{jl}\eta_{jm}\eta_{j'l}\eta_{j'm}$$

$$= \begin{cases} \sum_{j=1}^{+\infty}\lambda_{j,b}^2 \to 0, & l \neq m; \\ \big(\sum_{j=1}^{+\infty}\lambda_{j,b}\big)^2 + 2\sum_{j=1}^{+\infty}\lambda_{j,b}^2 \to 1, & l = m, \end{cases}$$

which implies that $V_k(b) \to^p \kappa_2 I_k$. Therefore ,we have $U_{el,k}(b,\mathcal{K}) \overset{d}{\to} (\kappa_1^2/\kappa_2)\chi_k^2$ as $b \to 0$. Compared to the χ^2-approximation, the fixed-*b* limiting distribution that captures the choice of the kernel and the bandwidth is expected to provide better approximation to the finite sample distribution of the BEL ratio statistic at the true parameter when b is relatively large.

2.2. Generalized empirical likelihood

We extend the fixed-*b* approach to the Generalized empirical likelihood (GEL) estimation framework (Newey and Smith (2004)). To describe GEL, we let ρ be a concave function defined on an open set $\mathcal{I}$ that contains the origin. Set $\rho(x) = -\infty$ for $x \notin \mathcal{I}$, and let $\rho_j(x) = \partial^j\rho(x)/\partial x^j$ and $\rho_j = \rho_j(0)$ for $j = 0,1,2$. We normalize ρ so that $\rho_1 = \rho_2 = -1$. Consider the set $\Pi_n(\theta) = \{\lambda : \lambda'f_{tn}(\theta) \in \mathcal{I}, t = 1,2,\ldots,n\}$. The GEL estimator is the solution to a saddle point problem,

$$\hat{\theta}_{gel} = \underset{\theta\in\Theta}{\operatorname{argmin}} \sup_{\lambda\in\mathbb{R}^k} \hat{P}(\theta,\lambda) = \underset{\theta\in\Theta}{\operatorname{argmin}} \sup_{\lambda\in\Pi_n(\theta)} \hat{P}(\theta,\lambda),$$

where $\hat{P}(\theta,\lambda) = \frac{1}{S_n}\sum_{t=1}^n\{\rho(\lambda'f_{tn}(\theta)) - \rho_0\}$. The GEL ratio function is given by

$$gelr(\theta) = 2\sup_{\lambda\in\mathbb{R}^k}\hat{P}(\theta,\lambda). \tag{2.9}$$

The GEL estimator includes a number of special cases that have been well studied in the statistics and econometrics literature. The EL, exponential tilting (ET), and continuous updating (CUE) are special cases of the GEL. Thus

$\rho(x) = \log(1 - x)$ and $\mathcal{I} = (-\infty, 1)$ for EL, $\rho(x) = -e^x$ and $\mathcal{I} = \mathbb{R}$ for ET, and $\rho(x) = -(1+x)^2/2$ and $\mathcal{I} = \mathbb{R}$ for CUE. More generally, members of the Cressie-Read power divergence family of discrepancies discussed by Imbens, Spady, and Johnson (1998) are included in the GEL class with $\rho(x) = -(1+\gamma x)^{(\gamma+1)\gamma}/(\gamma+1)$ (see Newey and Smith (2004)).

Let $G_{gel}(f) = \max_{\lambda \in \mathbb{R}^k} \int_0^1 \{\rho(\lambda' g(t)) - \rho_0\} dt$ for $g \in C^{\otimes k}[0,1]$. If $\rho(\cdot)$ is strictly concave and twice continuously differentiable, under suitable assumptions it can be shown that $G_{gel}(\cdot)$ is a continuous functional under the sup norm. Since the argument follows from that presented in the appendix with a minor modification, we skip the details (see Remark 1.1 in the supplementary material). Therefore, we have

$$gelr(\theta_0) \xrightarrow{d} U_{\rho,k}(b; \mathcal{K}) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^1 \left\{ \rho\left(\lambda' \int_0^1 \mathcal{K}((r-s)/b)dW_k(s)\right) - \rho_0 \right\} dr.$$

The GEL-based confidence region for the parameter θ_0 is

$$\widetilde{CI}(1-\alpha; b) = \left\{ \theta \in \mathbb{R}^p : \frac{gelr(\theta)}{1-b} \leq u_{\rho,k}(b; \mathcal{K}; 1-\alpha) \right\}, \qquad (2.10)$$

where $u_{\rho,k}(b; \mathcal{K}; 1-\alpha)$ is the $100(1-\alpha)\%$ quantile of $U_{\rho,k}(b; \mathcal{K})/(1-b)$, which can again be obtained via simulation or iid bootstrap.

3. Numerical Studies

We conducted two sets of simulation studies to compare and contrast the finite sample performance of the inference procedure based on the fixed-b approximation and the BEL of Kitamura (1997) and Smith (2011). The simulation results presented below are based on the simulation runs where the origin is contained in the convex hull of $\{f_{tn}(\theta_0)\}$.

3.1. Mean and quantiles

Consider the time series models AR(1), $Y_t = \rho Y_{t-1} + \epsilon_t$ with $\rho = -0.5$, 0.2, 0.5, 0.8, and AR(2), $Y_t = (5/6)Y_{t-1} - (1/6)Y_{t-2} + \epsilon_t$. The latter was used in Chen and Wong (2009) where the focus was to compare the finite sample coverages of the quantile delivered by BEL. In both models, $\{\epsilon_t\}$ is a sequence of iid standard normal random variables. We focus on the inference for the mean, the median and the 5% quantile. For the mean, $f(y_t, \theta) = y_t - \theta$. For the q-th quantile, we consider the moment condition $f_q(y_t, \theta) = \int_{-\infty}^{(\theta-y_t)/h} K(x)dx - q$, where $K(\cdot)$ is an r-th order window that satisfies

$$\int u^j K(u)du = \begin{cases} 1, & j = 0; \\ 0, & 1 \leq j \leq r-1; \\ \kappa_0, & j = r, \end{cases}$$

for some integer $r \geq 2$, and h is a bandwidth such that $h \to 0$ as $n \to +\infty$. When $h = 0$, we have $f_q(y_t, \theta) = \mathbf{I}(y_t \leq \theta) - q$. To accommodate dependence, we consider the BEL with fully overlapping moment conditions $f_{tn}(\theta) = (1/m)\sum_{j=t}^{t+m-1} f(y_t, \theta)$ for $t = 1, 2, \ldots, n - m + 1$ and $m = \lfloor nb \rfloor$ with $b \in (0, 1)$. For comparisons, we consider smoothed EL with the kernel $\mathcal{K}(x) = (5\pi/8)^{1/2} (1/x)J_1(6\pi x/5)$, where $J_1(\cdot)$ is the Bessel function of the first kind. The HAC covariance estimator induced by using $\mathcal{K}(\cdot)$ is essentially the same as the nonparametric long run variance estimator with the Quadratic spectral kernel (see Example 2.3 of Smith (2011)). The sample sizes considered were $n = 100$ and 400, and b was chosen from 0.02 to 0.2. To draw inference for the quantiles, we employed the second order Epanechnikov window with bandwidth $h = cn^{-1/4}$ for $c = 0, 1$, following Chen and Wong (2009). The coverage probabilities and corresponding interval widths for the mean and quantiles delivered by the fixed-b approximation and the χ^2-based approximation are depicted in Figures 1−4.

For the mean, undercoverage occurs for both the fixed-b calibration and the χ^2-based approximation when the dependence is positive, and becomes more severe as the dependence strengthens. Inference based on the fixed-b calibration provided uniformly better coverage probabilities in all cases, and was quite robust to the choice of b. The improvement was significant, especially for large bandwidth. On the other hand, the fixed-b based interval was slightly wider than the χ^2-based interval. For negative dependence with $\rho = -0.5$, the fixed-b calibration tended to provide overcoverage, but the improvement over the χ^2-based approximation could be seen for relatively large b. These findings are consistent with the intuition that the larger b is, the more accurate the fixed-b based approximation is relative to the χ^2-based approximation used by Kitamura (1997) and Smith (2011). The results for the median and 5% quantile were qualitatively similar to those in the mean case. The choice of $h = 1$ tended to provide slightly shorter interval widths as compared to the unsmoothed counterpart, $h = 0$ in some cases (see Chen and Wong (2009)). A comparison of Figure 1 with Figure 3 (Figure 2 with Figure 4) has the coverage probabilities for the EL based on the kernel $\mathcal{K}(x)$ generally closer to the nominal level than the BEL counterpart with the corresponding interval widths wider. This phenomenon is consistent with the finding that QS kernel provides better coverage, but wider interval widths, compared to the Bartlett kernel in Kiefer and Vogelsang (2005) under the GMM framework. Our unreported simulation results also demonstrate the usefulness of fixed-b calibration under the GEL estimation framework. The results for ET are available upon request.
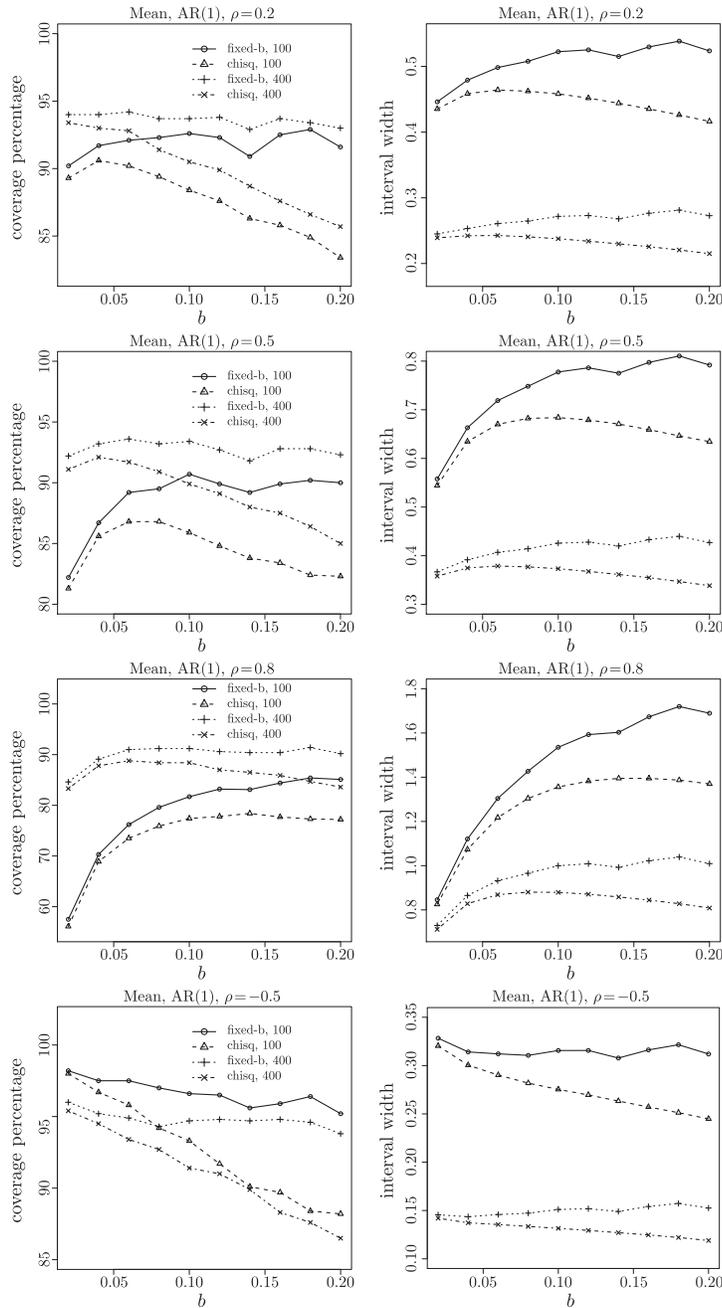
Figure 1. Coverage probabilities for the mean delivered by the BEL based on the fixed-b approximation and the χ^2-based approximation. The nominal level is 95% and the number of Monte Carlo replications is 1,000.
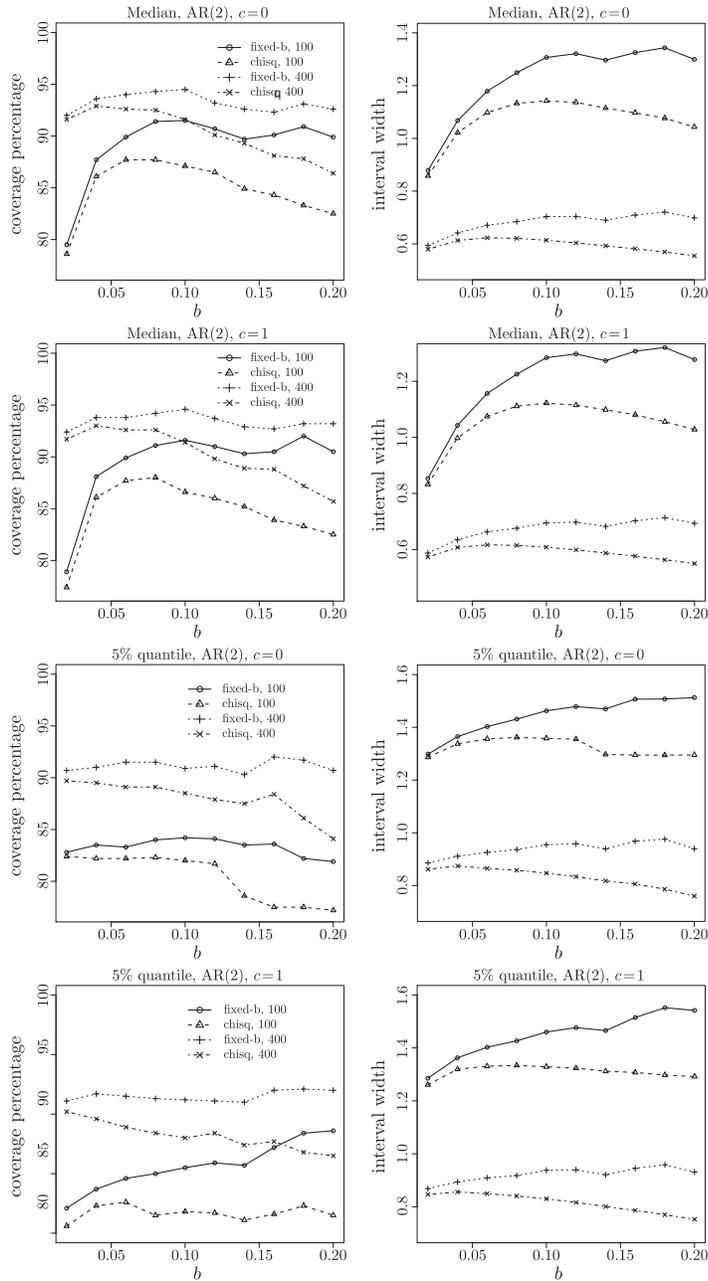
Figure 2. Coverage probabilities for the median and 5% quantile delivered by the BEL based on the fixed-*b* approximation and the χ^2-based approximation. The nominal level is 95% and the number of Monte Carlo replications is 1,000.
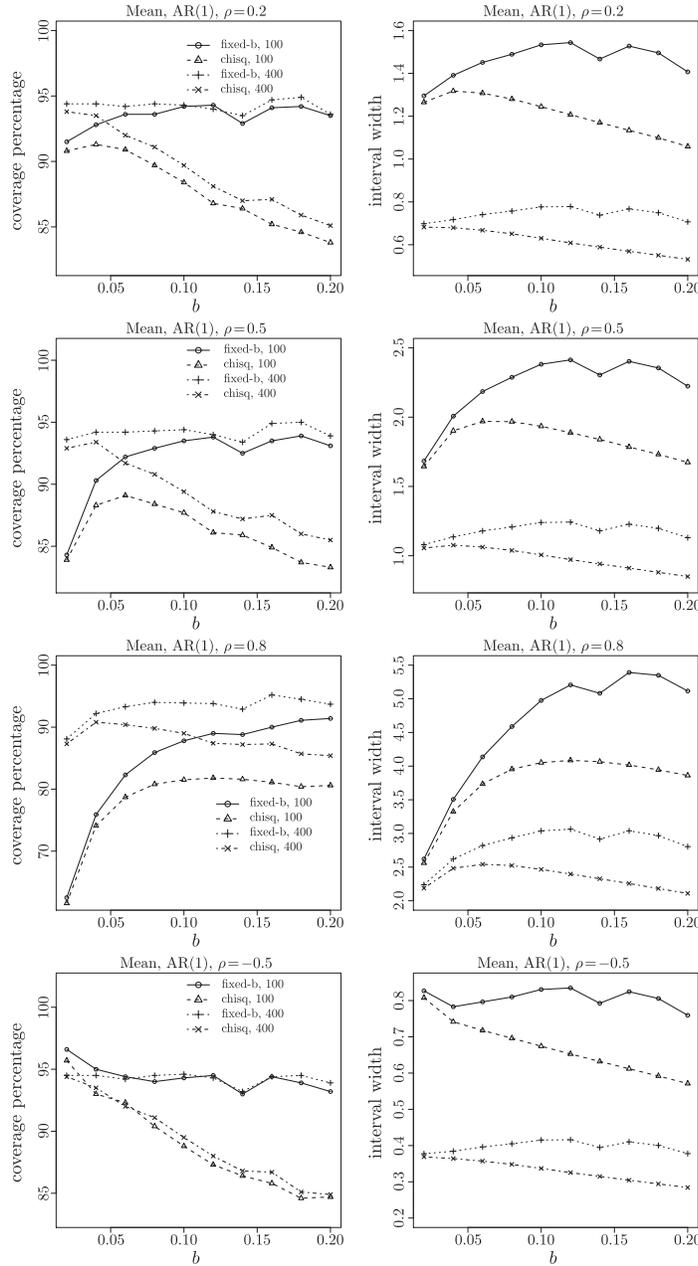
Figure 3. Coverage probabilities for the mean delivered by the smoothed EL based on the fixed-b approximation and the χ^2-based approximation. The corresponding kernel is $\mathcal{K}(x) = (5\pi/8)^{1/2} (1/x) J_1 (6\pi x/5)$. The nominal level is 95% and the number of Monte Carlo replications is 1,000.
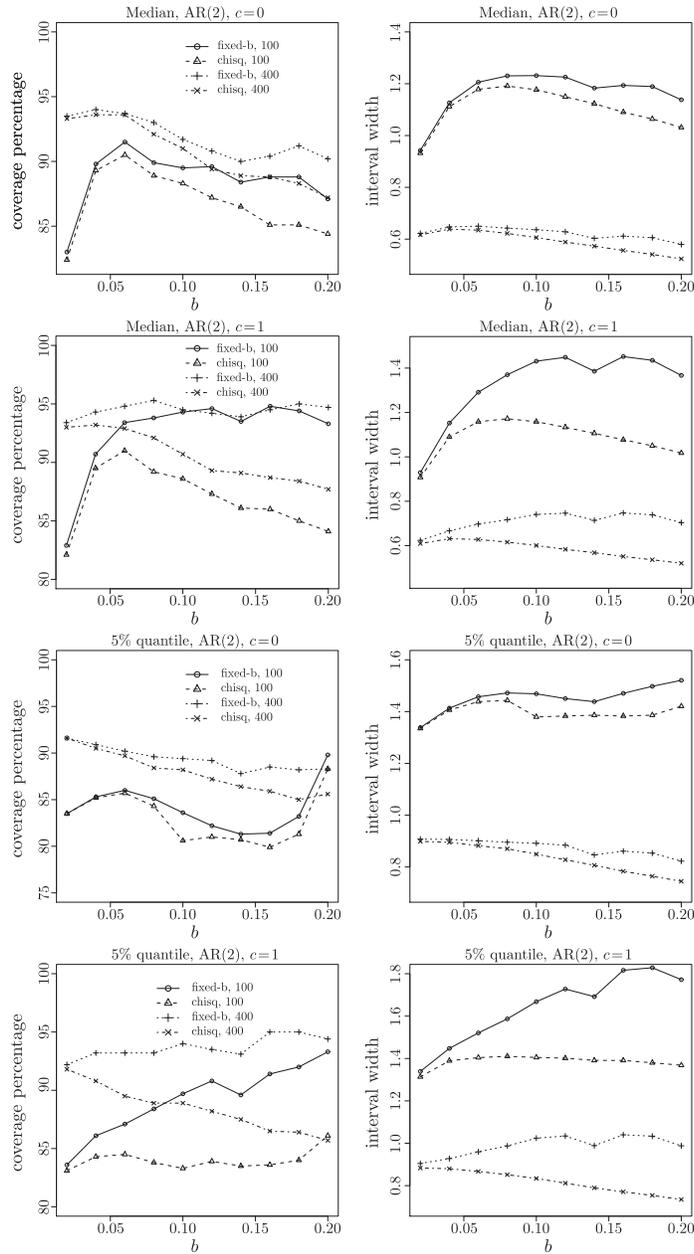
Figure 4. Coverage probabilities for the median and 5% quantile delivered by the smoothed EL based on the fixed-b approximation and the χ^2-based approximation. The corresponding kernel is $\mathcal{K}(x) = (5\pi/8)^{1/2}(1/x)J_1(6\pi x/5)$. The nominal level is 95% and the number of Monte Carlo replications is 1,000.
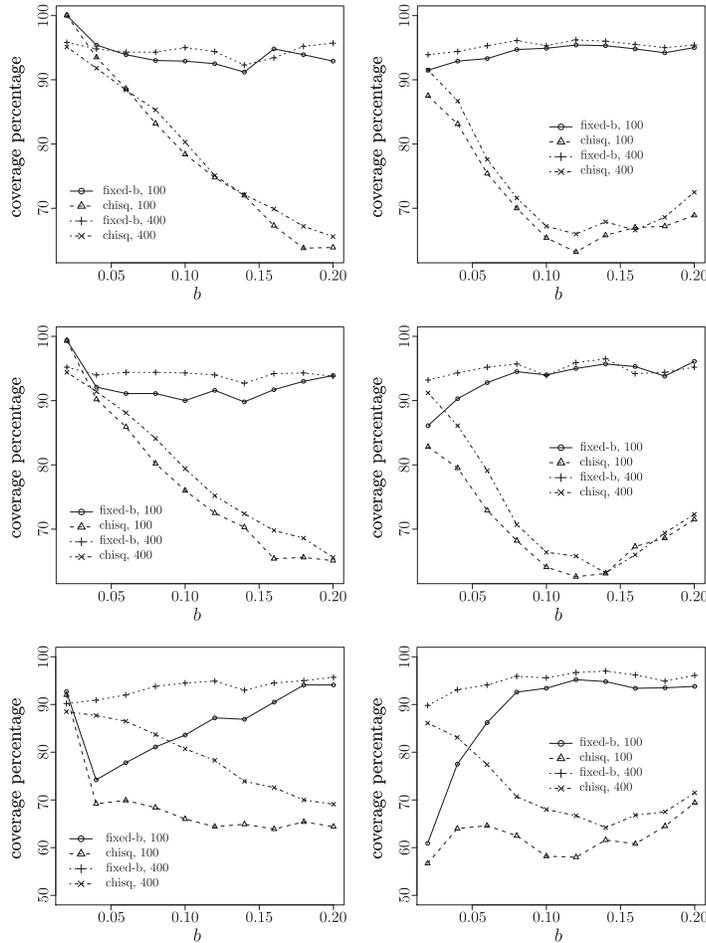
Figure 5. Coverage probabilities delivered by the BEL (left panels) and the smoothed EL (right panels) based on the fixed-b approximation and the χ^2-based approximation. The corresponding kernel for the smoothed EL is $\mathcal{K}(x) = (5\pi/8)^{1/2}(1/x)J_1(6\pi x/5)$. The nominal level is 95% and the number of Monte Carlo replications is 1,000.

3.2. Time series regression

We consider the stylized linear regression model with an intercept and a regressor x_t: $y_t = \beta_1 + \beta_2 x_t + u_t$ for $1 \leq t \leq n$, where $\{x_t\}$ and $\{u_t\}$ are generated independently from an AR(1) model with common coefficient $\tilde{\rho}$. We set the true parameter $\beta_0 = (\beta_{10}, \beta_{20}) = (0, 0)$ and chose $\tilde{\rho} \in \{0.2, 0.5, 0.8\}$. We are interested in constructing confidence contour for β_0. Consider the moment conditions $f_t(\beta) = (u_t(\beta), x_t u_t(\beta), x_{t-1} u_t(\beta), x_{t-2} u_t(\beta))$ with $u_t(\beta) = y_t - \beta_1 - \beta_2 x_t$ and $3 \leq t \leq n$. We report the coverage probabilities for the BEL and the

smoothed EL with kernel $\mathcal{K}(x)$ based on the fixed-b approximation and the χ^2-based approximation in Figure 5. As the dependence strengthens, the fixed-b and χ^2-based approximations deteriorate. The coverage probabilities obtained from the fixed-b calibration are consistently closer to the nominal level, and the improvement is significant for large bandwidths. In contrast, the coverage probabilities based on the χ^2 approximation are severely downward biased for relatively large b.

To sum up, the fixed-b approximation provides a uniformly better approximation to the sampling distribution of the EL ratio statistic for a wide range of b, and it tends to deliver more accurate coverage probability in confidence interval construction and size in testing. From a practical viewpoint, the choice of the bandwidth parameter has a great impact on the finite sample performance of the EL ratio statistic; it is of interest to consider the optimal bandwidth under the fixed-b paradigm.

Acknowledgements

References

Billingsley, P. (1999). *Convergence of Probability Measures*. Second edition. Wiley, New York.

Borwein, J. M. and Lewis, A. S. (1991). Duality relationships for entropy-type minimization problems. *SIAM J. Control Optim.* **29**, 325-338.

Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18**, 415-447.

Chen, S. X. and Wong, C. (2009). Smoothed block empirical likelihood for quantiles of weakly dependent processes. *Statist. Sinica* **19**, 71-82.

Cressie, N. and Read, T. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46**, 440-464.

Imbens, G. W., Spady, R. H. and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica* **66**, 333-357.

Jansson, M. (2004). On the error of rejection probability in simple autocorrelation robust tests. *Econometrica* **72**, 937-946.

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

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

Kitamura, Y. (2006). Empirical likelihood methods in econometrics: theory and practice. Cowles Foundation Discussion Paper 1569.

Newey, W. K. and Smith, R. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219-255.

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

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

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

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

Owen, A. (2001). *Empirical Likelihood.* Chapman and Hall, New York.

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

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

Shao, X. and Politis, D. N. (2013). Fixed-b subsampling and block bootstrap: improved confidence sets based on p-value calibration. *J. Roy. Statist. Soc. Ser. B* **75**, 161-184.

Smith, R. (2011). GEL criteria for moment condition models. *Econometric Theory* **27**, 1192-1235.

Sun, Y. (2010). Let's fix it: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust Inference. working paper.

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

Zhang, X. and Shao, X. (2013). Fixed-smoothing asymptotics for time series. *Ann. Statist.* **41**, 1329-1349.

Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, USA.

E-mail: zhangxiany@missouri.edu

Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright St., Champaign, IL 61820, USA.

E-mail: xshao@illinois.edu