cemmap

centre for microdata methods and practice

Quasi-maximum likelihood and the kernel block bootstrap for nonlinear dynamic models

Paulo M.D.C. Parente
Richard J. Smith

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP60/19

An ESRC Research Centre

Quasi-Maximum Likelihood and The Kernel Block Bootstrap for Nonlinear Dynamic Models*

Paulo M.D.C. Parente
REM and CEMAPRE
ISEG, School of Economics and Management, Universidade de Lisboa
pparente@iseg.ulisboa.pt

Richard J. Smith
cemmap, U.C.L and I.F.S.
Faculty of Economics, University of Cambridge
Department of Economics, University of Melbourne
ONS Economic Statistics Centre of Excellence
rjs27@econ.cam.ac.uk

This Draft: October 2019

Abstract

This paper applies a novel bootstrap method, the kernel block bootstrap, to quasi-maximum likelihood estimation of dynamic models with stationary strong mixing data. The method first kernel weights the components comprising the quasi-log likelihood function in an appropriate way and then samples the resultant transformed components using the standard "m out of n" bootstrap. We investigate the first order asymptotic properties of the kernel block bootstrap method for quasi-maximum likelihood demonstrating, in particular, its consistency and the first-order asymptotic validity of the bootstrap approximation to the distribution of the quasi-maximum likelihood estimator. A set of simulation experiments for the mean regression model illustrates the efficacy of the kernel block bootstrap for quasi-maximum likelihood estimation.

*We are grateful to seminar participants for their comments on earlier versions of this paper presented at the London School of Economics, Monash University, Universidad de Alicante and the Universities of Melbourne, Queensland, Sydney and York. Paulo Parente is grateful for the financial support of Fundação para a Ciência e a Tecnologia (Project PTDC/IIM-ECO/3097/2014).

Address for correspondence: Paulo M.D.C. Parente, ISEG, School of Economics and Management, Universidade de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal.

JEL Classification: C14, C15, C22
Keywords: Bootstrap; heteroskedastic and autocorrelation consistent inference; quasi-maximum likelihood estimation.

1 Introduction

This paper applies the kernel block bootstrap (KBB), proposed in Parente and Smith (2019), PS henceforth, to quasi-maximum likelihood estimation with stationary and weakly dependent data. The basic idea underpinning KBB arises from earlier papers, see, e.g., Kitamura and Stutzer (1997) and Smith (1997, 2011), which recognise that a suitable kernel function-based weighted transformation of the observational sample with weakly dependent data preserves the large sample efficiency for randomly sampled data of (generalised) empirical likelihood, (G)EL, methods. In particular, the mean of and, moreover, the standard random sample variance formula applied to the transformed sample are respectively consistent for the population mean [Smith (2011, Lemma A.1, p.1217)] and a heteroskedastic and autocorrelation (HAC) consistent and automatically positive semidefinite estimator for the variance of the standardized mean of the original sample [Smith (2005, Section 2, pp.161-165, and 2011, Lemma A.3, p.1219)].

In a similar spirit, KBB applies the standard "m out of n" nonparametric bootstrap, originally proposed in Bickel and Freedman (1981), to the transformed kernel-weighted data. PS demonstrate, under appropriate conditions, the large sample validity of the KBB estimator of the distribution of the sample mean [PS Theorem 3.1] and the higher order asymptotic bias and variance of the KBB variance estimator [PS Theorem 3.2]. Moreover, [PS Corollaries 3.1 and 3.2], the KBB variance estimator possesses a favourable higher order bias property, a property noted elsewhere for consistent variance estimators using tapered data [Brillinger (1981, p.151)], and, for a particular choice of kernel function weighting and choice of bandwidth, is optimal being asymptotically close to one based on the optimal quadratic spectral kernel [Andrews (1991, p.821)] or Bartlett-Priestley-Epanechnikov kernel [Priestley (1962, 1981, pp. 567-571), Epanechnikov (1969) and Sacks and Ylvisacker (1981)]. Here, though, rather than being applied to the original data as in PS, the KBB kernel function weighting is applied to the individual observational components of the quasi-log likelihood criterion function itself. The asymptotic validity of the KBB bootstrap follows from an adaptation of the general results on resampling methods for extremum estimators given in Gonçalves and White (2004).

Myriad variants for dependent data of the bootstrap method proposed in the landmark article Efron (1979) also make use of the standard "m out of n" nonparametric bootstrap, but, in contrast to KBB, applied to "blocks" of the original data. See, *inter alia*, the moving blocks bootstrap (MBB) [Künsch (1989), Liu and Singh (1992)], the circular block bootstrap [Politis and Romano (1992a)], the stationary bootstrap [Politis and Romano (1994)], the external bootstrap for m-dependent data [Shi and Shao (1988)], the frequency domain bootstrap [Hurvich and Zeger (1987), see also Hidalgo (2003)], and its generalization the transformation-based bootstrap [Lahiri (2003)], and the autoregressive sieve bootstrap [Bühlmann (1997)]; for further details on these methods, see, e.g., the monographs Shao and Tu (1995) and Lahiri (2003). Whereas the block length of these other methods is typically a declining fraction of sample size, the implicit KBB block length is dictated by the support of the kernel function and, thus, with unbounded support as in the optimal case, would be the sample size itself.

When the object of inference is the stochastic process mean, the KBB method bears comparison with the tapered block bootstrap (TBB) of Paparoditis and Politis (2001). Indeed, in this case, KBB may be regarded as a generalisation and extension of TBB. TBB is also based on a reweighted sample of the observations but with weight function with bounded support and, so, whereas each KBB data point is in general a transformation of all original sample data, those of TBB use a fixed block size and, implicitly thereby, a fixed number of data points. More generally then, the TBB weight function class is a special case of that of KBB but is more restrictive; a detailed comparison of KBB and TBB is provided in PS Section 4.1. TBB is extended in Paparoditis and Politis (2002) to approximately linear statistics but differs from the KBB method introduced here for quasi-maximum likelihood estimation.

The paper is organized as follows. After outlining some preliminaries Section 2 introduces KBB and reviews the results in PS. Section 3 demonstrates how KBB can be applied in the quasi-maximum likelihood framework and, in particular, details the consistency of the KBB estimator and its asymptotic validity for quasi-maximum likelihood. Section 4 reports a Monte Carlo study on the performance of KBB for the mean regression model. Finally section 5 concludes. Proofs of the results in the main text are provided in Appendix B with intermediate results required for their proofs given in Appendix A.

2 Kernel Block Bootstrap

To introduce the kernel block bootstrap (KBB) method, consider a sample of T observations, $z_1, ..., z_T$, on the strictly stationary real valued d_z-dimensional vector sequence $\{z_t, t \in \mathbb{Z}\}$ with unknown mean $\mu = \mathrm{E}[z_t]$ and autocovariance sequence $R(s) =$

$\mathrm{E}[(z_t-\mu)(z_{t+s}-\mu)']$, $(s=0,\pm1,...)$. Under suitable conditions, see Ibragimov and Linnik (1971, Theorem 18.5.3, pp. 346, 347), the limiting distribution of the sample mean $\bar{z}=\sum_{t=1}^{T}z_t/T$ is described by $T^{1/2}(\bar{z}-\mu)\xrightarrow{d}N(0,\Sigma_\infty)$, where $\Sigma_\infty=\lim_{T\to\infty}\mathrm{var}[T^{1/2}\bar{z}]=\sum_{s=-\infty}^{\infty}R(s)$.

Let $k_j=\int_{-\infty}^{\infty}k(x)^j dx$ with sample counterpart $\hat{k}_j=\sum_{s=1-T}^{T-1}k(s/S_T)^j/S_T$, $(j=1,2)$, where $k(\cdot)$ denotes a suitable kernel function. The KBB approximation to the distribution of the sample mean $\bar{z}$ randomly samples the kernel-weighted centred observations

$$z_{tT}=\frac{1}{(\hat{k}_2S_T)^{1/2}}\sum_{r=t-T}^{t-1}k(\frac{r}{S_T})(z_{t-r}-\bar{z}),(t=1,...,T), \tag{2.1}$$

where S_T is a bandwidth parameter.

Remark 2.1. The definition of z_{tT} (2.1) rescales that in Kitamura and Stutzer (1997) and Smith (1997, 2011) by $(S_T/\hat{k}_2)^{1/2}$ with k_2 replaced without loss by $\hat{k}_2$, see PS Corollary K.2, p.31.

Let $\bar{z}_T=T^{-1}\sum_{t=1}^{T}z_{tT}$ denote the sample mean of z_{tT}, $(t=1,...,T)$. Under appropriate conditions, $\bar{z}_T\xrightarrow{p}0$ and $\Sigma_\infty^{-1/2}(T/S_T)^{1/2}\bar{z}_T\xrightarrow{d}N(0,I_{d_z})$; see, e.g., Smith (2011, Lemmas A.1 and A.2, pp.1217-19). Moreover, the KBB variance estimator, defined in standard random sampling outer product form,

$$\hat{\Sigma}_{\text{KBB}}=T^{-1}\sum_{t=1}^{T}(z_{tT}-\bar{z}_T)(z_{tT}-\bar{z}_T)'\xrightarrow{p}\Sigma_\infty; \tag{2.2}$$

and is thus an automatically positive semidefinite heteroskedastic and autocorrelation consistent (HAC) variance estimator; see Smith (2011, Lemma A.3, p.1219).

KBB applies the standard "m out of n" non-parametric bootstrap method to the index set $\mathcal{T}_T=\{1,...,T\}$; see Bickel and Freedman (1981). That is, the indices t_s^* and, thereby, $z_{t_s^*}$, $(s=1,...,m_T)$, are a random sample of size m_T drawn from, respectively, $\mathcal{T}_T$ and $\{z_{tT}\}_{t=1}^{T}$, where $m_T=[T/S_T]$, the integer part of T/S_T.

Remark 2.2. The KBB sample mean $\bar{z}_{m_T}^*=\sum_{s=1}^{m_T}z_{t_s^*T}/m_T$ may be regarded as that from a random sample of size m_T taken from the blocks $\mathcal{B}_t=\{k\{(t-r)/S_T\}(z_r-\bar{z})/(\hat{k}_2S_T)^{1/2}\}_{r=1}^{T}$, $(t=1,...,T)$. See PS Remark 2.2, p.3. Note that the blocks $\{\mathcal{B}_t\}_{t=1}^{T}$ are overlapping and, if the kernel function $k(\cdot)$ has unbounded support, the block length is T.

Let $\mathcal{P}^*_\omega$ denote the bootstrap probability measure conditional on $\{z_{tT}\}_{t=1}^T$ (or, equivalently, the observational data $\{z_t\}_{t=1}^T$) with E^* and var^* the corresponding conditional expectation and variance respectively. Under suitable regularity conditions, see PS Assumptions 3.1-3.3, pp.3-4, the bootstrap distribution of the scaled and centred KBB sample mean $m_T^{1/2}(\bar{z}^*_{m_T} - \bar{z}_T)$ converges uniformly to that of $T^{1/2}(\bar{z} - \mu)$, i.e.,

$$\sup_{x \in \mathcal{R}} \left| \mathcal{P}^*_\omega\{m_T^{1/2}(\bar{z}^*_{m_T} - \bar{z}_T)/k_1 \leq x\} - \mathcal{P}\{T^{1/2}(\bar{z} - \mu) \leq x\} \right| \to 0, \text{ prob-}\mathcal{P}; \tag{2.3}$$

see PS Theorem 3.1, p.5.

Given the stricter additional requirement PS Assumption 3.4, p.5, PS Theorem 3.2, p.6, provides higher order results on moments of the KBB variance estimator $\hat{\Sigma}_{\mathrm{KBB}}$ (2.2). Let $k^*_{(q)} = \lim_{y \to 0} \{1 - k^*(y)\} / |y|^q$, where the induced self-convolution kernel $k^*(y) = \int_{-\infty}^{\infty} k(x-y)k(x)dx/k_2$. Define

$$\mathrm{MSE}(T/S_T, \hat{\Sigma}_{\mathrm{KBB}}, W_T) = (T/S_T)\mathrm{E}[vec(\hat{\Sigma}_{\mathrm{KBB}} - J_T)' W_T vec(\hat{\Sigma}_{\mathrm{KBB}} - J_T)]$$

where W_T is a positive semi-definite weight matrix and $J_T = \sum_{s=1-T}^{T-1} (1 - |s|/T)R(s)$. Let K_{pp} denote the $p^2 \times p^2$ commutation matrix $\sum_{i=1}^p \sum_{j=1}^p e_i e_j' \otimes e_j e_i'$, where e_i is the ith elementary vector, $(i = 1, ..., p)$, (Magnus and Neudecker, 1979, Definition 3.1, p.383).

Bias. $\mathrm{E}[\hat{\Sigma}_{\mathrm{KBB}}] = J_T + S_T^{-2}(\Gamma_{k^*} + o(1)) + U_T$, where $\Gamma_{k^*} = -k^*_{(2)} \sum_{s=-\infty}^{\infty} |s|^2 R(s)$ and $U_T = O((S_T/T)^{b-1/2}) + o(1/S_T^2) + O(S_T^{b-2}/T^b) + O(S_T/T) + O((S_T/T)^{3/2-\varepsilon})$ with $b > 1$ and $\varepsilon \in (0, 1/2)$;

Variance. if $S_T^5/T \to \gamma \in (0, \infty)$, then $(T/S_T)\mathrm{var}[\hat{\Sigma}_{\mathrm{KBB}}] = \Delta_{k^*} + o(1)$, where $\Delta_{k^*} = (I_{pp} + K_{pp})(\Sigma_\infty \otimes \Sigma_\infty) \int_{-\infty}^{\infty} k^*(y)^2 dy$;

Mean squared error. if $S_T^5/T \to \gamma \in (0, \infty)$, then $\mathrm{MSE}(T/S_T, \hat{\Sigma}_{\mathrm{KBB}}, W_T) = tr(W(I_{pp} + K_{pp})(\Sigma_\infty \otimes \Sigma_\infty)) \int_{-\infty}^{\infty} k^*(y)^2 dy + vec(\Gamma_{k^*})' W vec(\Gamma_{k^*})/\gamma + o(1)$.

Remark 2.3. The bias and variance results are similar to Parzen (1957, Theorems 5A and 5B, pp.339-340) and Andrews (1991, Proposition 1, p.825), when the Parzen exponent q equals 2. The KBB bias, cf. the tapered block bootstrap (TBB), is $O(1/S_T^2)$, an improvement on $O(1/S_T)$ for the moving block bootstrap (MBB). The expression $\mathrm{MSE}(T/S_T, \hat{\Sigma}_{\mathrm{KBB}}, W_T)$ is identical to that for the mean squared error of the Parzen (1957) estimator based on the induced self-convolution kernel $k^*(y)$.

Optimality results for the estimation of Σ_∞ are an immediate consequence of PS Theorem 3.2, p.6, and the theoretical results of Andrews (1991) for the Parzen (1957) estimator. Consider the kernel function

$$k(x) = (\frac{5\pi}{8})^{1/2}\frac{1}{x}J_1(\frac{6\pi x}{5}) \text{ if } x \neq 0 \text{ and } (\frac{5\pi}{8})^{1/2}\frac{3\pi}{5} \text{ if } x = 0; \tag{2.4}$$

here $J_v(z) = \sum_{k=0}^{\infty}(-1)^k (z/2)^{2k+v} / \{\Gamma(k+1)\Gamma(k+2)\}$, a Bessel function of the first kind (Gradshteyn and Ryzhik, 1980, 8.402, p.951) with $\Gamma(\cdot)$ the gamma function. Smith (2011, Example 2.3, p.1204) shows that the quadratic spectral (QS) kernel $k^*_{\text{QS}}(y)$ is the induced self-convolution kernel $k^*(y)$ associated with the kernel $k(x)$ (2.4), where the QS kernel

$$k^*_{\text{QS}}(y) = \frac{3}{(ay)^2}\left(\frac{\sin ay}{ay} - \cos ay\right), a = 6\pi/5, \tag{2.5}$$

Remark 2.4. The QS kernel $k^*_{\text{QS}}(y)$ (2.5) is well-known to possess optimality properties, e.g., for the estimation of spectral densities (Priestley, 1962; 1981, pp. 567-571) and probability densities (Epanechnikov, 1969, Sacks and Ylvisacker, 1981).

PS Corollary 3.1, p.7, establishes an optimality result for the KBB variance estimator $\hat{\Sigma}_{\text{KBB}}(S_T)$ (2.2) computed with the kernel function (2.4) which is denoted as $\tilde{\Sigma}_{\text{KBB}}(S_T)$. For sensible comparisons, the requisite bandwidth parameter is $S_{Tk^*} = S_T / \int_{-\infty}^{\infty} k^*(y)^2 dy$, see Andrews (1991, (4.1), p.829), if the respective asymptotic variances scaled by T/S_T are to coincide; see Andrews (1991, p.829). Then, for any bandwidth sequence S_T such that $S_T \to \infty$ and $S_T^5/T \to \gamma \in (0, \infty)$, $\lim_{T\to\infty} \text{MSE}(T/S_T, \hat{\Sigma}_{\text{KBB}}, W_T) - \text{MSE}(T/S_T, \tilde{\Sigma}_{\text{KBB}}, W_T) \geq 0$ with strict inequality if $k^*(y) \neq k^*_{\text{QS}}(y)$ with positive Lebesgue measure; see PS Corollary 3.1, p.7.

The bandwidth $S_T^* = [4vec(\Gamma_{k^*})' W vec(\Gamma_{k^*}) / \operatorname{tr}(W(I_{pp}+K_{pp})(\Sigma_\infty \otimes \Sigma_\infty)) \int_{-\infty}^{\infty} k^*(y)^2 dy]^{1/5} T^{1/5}$ is also optimal in the following sense. For any bandwidth sequence S_T such that $S_T \to \infty$ and $S_T^5/T \to \gamma \in (0, \infty)$, $\lim_{T\to\infty} \text{MSE}(T^{4/5}, \hat{\Sigma}_{\text{KBB}}(S_T), W_T) - \text{MSE}(T^{4/5}, \hat{\Sigma}_{\text{KBB}}(S_T^*), W_T) \geq 0$ with strict inequality unless $S_T = S_T^* + o(1/T^{1/5})$; see PS Corollary 3.2, p.7.

3 Quasi-Maximum Likelihood

This section applies the KBB method briefly outlined above to parameter estimation in the quasi-maximum likelihood (QML) setting. In particular, under the regularity conditions detailed below, KBB may be used to construct hypothesis tests and confidence intervals. The proofs of the results basically rely on verifying a number of the conditions required for several general lemmas established in Gonçalves and White (2004) on resampling methods for extremum estimators. Indeed, although the focus of Gonçalves and White (2004) is MBB, the results therein also apply to other block bootstrap schemes such as KBB.

To describe the set-up, let the d_z-vectors z_t, $(t = 1, ..., T)$, denote a realisation from the stationary and strong mixing stochastic process $\{z_t\}_{t=1}^{\infty}$. The d_θ-vector θ of parameters is of interest where $\theta \in \Theta$ with the compact parameter space $\Theta \subseteq \mathcal{R}^{d_\theta}$. Consider the log-density $\mathcal{L}_t(\theta) = \log f(z_t; \theta)$ and its expectation $\mathcal{L}(\theta) = \mathrm{E}[\mathcal{L}_t(\theta)]$. The true value θ_0 of θ is defined by

$$\theta_0 = \arg\max_{\theta\in\Theta} \mathcal{L}(\theta)$$

with, analogously, the QML estimator $\hat{\theta}$ of θ_0

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \bar{\mathcal{L}}(\theta),$$

where the sample mean $\bar{\mathcal{L}}(\theta) = \sum_{t=1}^{T} \mathcal{L}_t(\theta)/T$.

The KBB method for QML makes use of the kernel smoothed log density function

$$\mathcal{L}_{tT}(\theta) = \frac{1}{(\hat{k}_2 S_T)^{1/2}} \sum\nolimits_{r=t-T}^{t-1} k(\frac{r}{S_T}) \mathcal{L}_{t-r}(\theta), (t = 1, ..., T),$$

cf. (2.1). As in Section 2, the indices t_s^* and the consequent bootstrap sample $\mathcal{L}_{t_s^* T}(\theta)$, $(s = 1, ..., m_T)$, denote random samples of size m_T drawn with replacement from the index set $\mathcal{T}_T = \{1, ..., T\}$ and the bootstrap sample space $\{\mathcal{L}_{tT}(\theta)\}_{t=1}^{T}$, where $m_T = [T/S_T]$ is the integer part of T/S_T. The bootstrap QML estimator $\hat{\theta}^*$ is then defined by

$$\hat{\theta}^* = \arg\max_{\theta\in\Theta} \bar{\mathcal{L}}_{m_T}^*(\theta)$$

where the bootstrap sample mean $\bar{\mathcal{L}}_{m_T}^*(\theta) = \sum_{s=1}^{m_T} \mathcal{L}_{t_s^* T}(\theta)/m_T$.

Remark 3.1. Note that, because $\mathrm{E}[\partial\mathcal{L}_t(\theta_0)/\partial\theta] = 0$, it is unnecessary to centre $\mathcal{L}_t(\theta)$, $(t = 1, ..., T)$, at $\bar{\mathcal{L}}(\theta)$; cf. (2.1).

The following conditions are imposed to establish the consistency of the bootstrap estimator $\hat{\theta}^*$ for θ_0. Let $f_t(\theta) = f(z_t; \theta)$, $(t = 1, 2, ...)$.

Assumption 3.1 **(a)** $(\Omega, \mathcal{F}, \mathcal{P})$ *is a complete probability space;* **(b)** *the finite d_z-dimensional stochastic process z_t: $\Omega \longmapsto \mathcal{R}^{d_z}$, $(t = 1, 2, ...)$, is stationary and strong mixing with mixing numbers of size $-v/(v-1)$ for some $v > 1$ and is measurable for all t, $(t = 1, 2, ...)$.*

Assumption 3.2 **(a)** *f: $\mathcal{R}^{d_z} \times \Theta \longmapsto \mathcal{R}^+$ is $\mathcal{F}$-measurable for each $\theta \in \Theta$, Θ a compact subset of $\mathcal{R}^{d_\theta}$;* **(b)** *$f_t(\cdot)$: $\Theta \longmapsto \mathcal{R}^+$ is continuous on Θ a.s.-$\mathcal{P}$;* **(c)** *$\theta_0 \in \Theta$ is the unique maximizer of* $\mathrm{E}[\log f_t(\theta)]$, $\mathrm{E}[\sup_{\theta\in\Theta} |\log f_t(\theta)|^\alpha] < \infty$ *for some $\alpha > v$;* **(d)** $\log f_t(\theta)$ *is global Lipschitz continuous on Θ, i.e., for all $\theta, \theta^0 \in \Theta$, $|\log f_t(\theta) - \log f_t(\theta^0)| \leq L_t \|\theta - \theta^0\|$ a.s.-$\mathcal{P}$ and $\sup_T \mathrm{E}[\sum_{t=1}^T L_t/T] < \infty$;*

Let $\mathbb{I}(x \geq 0)$ denote the indicator function, i.e., $\mathbb{I}(A) = 1$ if A true and 0 otherwise.

Assumption 3.3 **(a)** *$S_T \to \infty$ and $S_T = o(T^{\frac{1}{2}})$;* **(b)** *$k(\cdot)$: $\mathcal{R} \longmapsto [-k_{\max}, k_{\max}]$, $k_{\max} < \infty$, $k(0) \neq 0$, $k_1 \neq 0$, and is continuous at 0 and almost everywhere;* **(c)** *$\int_{-\infty}^{\infty} \bar{k}(x)dx < \infty$ where $\bar{k}(x) = \mathbb{I}(x \geq 0)\sup_{y\geq x} |k(y)| + \mathbb{I}(x < 0)\sup_{y\leq x} |k(y)|$;* **(d)** *$|K(\lambda)| \geq 0$ for all $\lambda \in \mathcal{R}$, where $K(\lambda) = (2\pi)^{-1}\int k(x)\exp(-ix\lambda)dx$.*

Theorem 3.1. Let Assumptions 3.1-3.3 hold. Then, if $T^{1/\alpha}/m_T \to 0$, **(a)** $\hat{\theta} - \theta_0 \to 0$, prob-$\mathcal{P}$; **(b)** $\hat{\theta}^* - \hat{\theta} \to 0$, prob-$\mathcal{P}^*$, prob-$\mathcal{P}$.

To prove consistency of the KBB distribution requires a strengthening of the above assumptions.

Assumption 3.4 **(a)** *$(\Omega, \mathcal{F}, \mathcal{P})$ is a complete probability space;* **(b)** *the finite d_z-dimensional stochastic process z_t: $\Omega \longmapsto \mathcal{R}^{d_z}$, $(t = 1, 2, ...)$, is stationary and strong mixing with mixing numbers of size $-3v/(v-1)$ for some $v > 1$ and is measurable for all t, $(t = 1, 2, ...)$.*

Assumption 3.5 **(a)** *f: $\mathcal{R}^{d_z} \times \Theta \longmapsto \mathcal{R}^+$ is $\mathcal{F}$-measurable for each $\theta \in \Theta$, Θ a compact subset of $\mathcal{R}^{d_\theta}$;* **(b)** *$f_t(\cdot)$: $\Theta \longmapsto \mathcal{R}^+$ is continuously differentiable of order 2 on Θ a.s.-P, $(t = 1, 2, ...)$;* **(c)** *$\theta_0 \in \mathrm{int}(\Theta)$ is the unique maximizer of* $\mathrm{E}[\log f_t(\theta)]$.

Define $A(\theta) = \mathrm{E}[\partial^2 \mathcal{L}_t(\theta)/\partial\theta\partial\theta']$ and $B(\theta) = \lim_{T\to\infty} \mathrm{var}[T^{1/2}\partial\bar{\mathcal{L}}(\theta)/\partial\theta]$.

Assumption 3.6 **(a)** *$\partial^2 \mathcal{L}_t(\theta)/\partial\theta\partial\theta'$ is global Lipschitz continuous on Θ;* **(b)** $\mathrm{E}[\sup_{\theta\in\Theta} \|\partial\mathcal{L}_t(\theta)/\partial\theta\|^\alpha] < \infty$ *and* $\mathrm{E}[\sup_{\theta\in\Theta} \|\partial^2\mathcal{L}_t(\theta)/\partial\theta\partial\theta'\|^\alpha] < \infty$ *for some $\alpha > \max[4v, 1/\eta]$;* **(c)** *$A_0 = A(\theta_0)$ is non-singular and $B_0 = \lim_{T\to\infty} \mathrm{var}[T^{1/2}\partial\bar{\mathcal{L}}(\theta_0)/\partial\theta]$ is positive definite.*

Remark 3.2. Assumption 3.6(b) obviates the condition $T^\alpha/m_T \to 0$ of Theorem 3.1 required by the bootstrap pointwise WLLN Lemma A.2 in Appendix A.

Under these regularity conditions,

$$B_0^{-1/2} A_0 T^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_{d_\theta});$$

see the Proof of Theorem 3.2. Moreover,

THEOREM 3.2. Suppose Assumptions 3.2-3.6 are satisfied. Then, if $S_T \to \infty$ and $S_T = O(T^{\frac{1}{2}-\eta})$ with $0 < \eta < \frac{1}{2}$,

$$\sup_{x \in \mathcal{R}^{d_\theta}} \left| \mathcal{P}_\omega^* \{ T^{1/2}(\hat{\theta}^* - \hat{\theta})/k^{1/2} \leq x \} - \mathcal{P}\{ T^{1/2}(\hat{\theta} - \theta_0) \leq x \} \right| \to 0, \text{ prob-}\mathcal{P},$$

where $k = k_2/k_1^2$.

REMARK 3.3. The factor k may be replaced without loss by $\hat{k} = \hat{k}_2/\hat{k}_1^2$, see PS Corollary K.2, p.31. Cf. Remark 2.1.

Let $\bar{\mathcal{L}}_T(\theta) = \sum_{t=1}^T \mathcal{L}_{tT}(\theta)/T$. An alternative less computationally intensive centred bootstrap may be based on the next result.

COROLLARY 3.1. Under the conditions of Theorem 3.2, if $S_T \to \infty$ and $S_T = O(T^{\frac{1}{2}-\eta})$ with $0 < \eta < \frac{1}{2}$,

$$\sup_{x \in \mathcal{R}^{d_\theta}} \left| \mathcal{P}_\omega^* \{ -[\frac{\partial^2 \bar{\mathcal{L}}_{m_T}^*(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1} T^{1/2} (\frac{\partial \bar{\mathcal{L}}_{m_T}^*(\hat{\theta})}{\partial\theta} - \frac{\partial \bar{\mathcal{L}}_T(\hat{\theta})}{\partial\theta})/k^{1/2} \leq x \} - \mathcal{P}\{ T^{1/2}(\hat{\theta} - \theta_0) \leq x \} \right| \to 0, \text{ prob-}\mathcal{P},$$

where $k = k_2/k_1^2$.

REMARK 3.4. From the bootstrap UWL Lemma A.1 in Appendix A, suitably adapted, the matrix $\partial^2 \bar{\mathcal{L}}_T(\hat{\theta})/\partial\theta\partial\theta'$ may substitute for $\partial^2 \bar{\mathcal{L}}_{m_T}^*(\hat{\theta})/\partial\theta\partial\theta'$ in Corollary 3.1, *viz.* $\sup_{\theta\in\Theta} (k_2/S_T)^{1/2} \left\| \partial^2 \bar{\mathcal{L}}_{m_T}^*(\theta)/\partial\theta\partial\theta' - \partial^2 \bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta' \right\| \to 0$, prob-$\mathcal{P}_\omega^*$, prob-$\mathcal{P}$. A similar argument together with the UWLs $\sup_{\theta\in\Theta} \left\| (k_2/S_T)^{1/2} \partial^2 \bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta' - k_1 A(\theta) \right\| \to 0$, prob-$\mathcal{P}$, and $\sup_{\theta\in\Theta} \left\| \partial^2 \bar{\mathcal{L}}(\theta)/\partial\theta\partial\theta' - A(\theta) \right\| \to 0$, prob-$\mathcal{P}$, yields

$$\sup_{x \in \mathcal{R}^{d_\theta}} \left| \mathcal{P}_\omega^* \{ -[\frac{\partial^2 \bar{\mathcal{L}}(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1} m_T^{1/2} (\frac{\partial \bar{\mathcal{L}}_{m_T}^*(\hat{\theta})}{\partial\theta} - \frac{\partial \bar{\mathcal{L}}_T(\hat{\theta})}{\partial\theta}) \leq x \} - \mathcal{P}\{ T^{1/2}(\hat{\theta} - \theta_0) \leq x \} \right| \to 0, \text{ prob-}\mathcal{P},$$

similarly to Paparoditis and Politis (2002, Theorem 2.2, p.135) for the QML implementation of TBB expressed in terms of the influence function corresponding to the QML criterion $\bar{\mathcal{L}}(\theta)$, *viz.*,

$$IF(z, \hat{F}_T) = -[\frac{\partial^2 \bar{\mathcal{L}}(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1}(\frac{\partial \mathcal{L}_t(\hat{\theta})}{\partial\theta} - \frac{\partial \bar{\mathcal{L}}(\hat{\theta})}{\partial\theta}), (t = 1, ..., T),$$

noting that $\partial\bar{\mathcal{L}}(\hat{\theta})/\partial\theta = 0$.

Remark 3.5. As noted in Remark 3.4, $\partial^2\bar{\mathcal{L}}(\hat{\theta})/\partial\theta\partial\theta'$ may be replaced by $(k/S_T)^{1/2}\partial^2\bar{\mathcal{L}}_T(\hat{\theta})/\partial\theta\partial\theta'$. Hence,

$$\sup_{x\in\mathcal{R}^{d_\theta}} \left| \mathcal{P}^*_\omega\{-[\frac{\partial^2 \bar{\mathcal{L}}_T(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1}T^{1/2}(\frac{\partial \bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta} - \frac{\partial \bar{\mathcal{L}}_T(\hat{\theta})}{\partial\theta})/k^{1/2} \leq x\} - \mathcal{P}\{T^{1/2}(\hat{\theta} - \theta_0) \leq x\} \right| \to 0, \text{ prob-}\mathcal{P}.$$

Remark 3.6. It follows from the first order condition $\partial\bar{\mathcal{L}}(\hat{\theta})/\partial\theta = 0$ for $\hat{\theta}$ that the term $\partial\bar{\mathcal{L}}_T(\hat{\theta})/\partial\theta$ in Corollary 3.1 and Remarks 3.4 and 3.5 may be omitted. The corresponding uncentred bootstrap from Remark 3.4 is the KBB version of Gonçalves and White (2004, Corollary 2.1, p.203) for MBB applied to the one-step QML estimator; also see Davidson and MacKinnon (1999).

Remark 3.7. The KBB variance estimator for the large sample variance matrix $A_0^{-1}B_0A_0^{-1}$ of the QML estimator $\hat{\theta}$ (or $\hat{\theta}^*$) is given by the outer product form

$$\frac{1}{T}[\frac{\partial^2 \bar{\mathcal{L}}_T(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1}\sum_{t=1}^{T}\frac{\partial \mathcal{L}_{tT}(\hat{\theta})}{\partial\theta}\frac{\partial \mathcal{L}_{tT}(\hat{\theta})}{\partial\theta'}[\frac{\partial^2 \bar{\mathcal{L}}_T(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1}.$$

Cf. (2.2); see PS (2.2), p.2, and Smith (2005, Theorem 2.1, p.165, and 2011, Lemma A.3, p.1219).

4 Simulation Results

In this section we report the results of a set of Monte Carlo experiments comparing the finite sample performance of different methods for the construction of confidence intervals for the parameters of the mean regression model when there is autocorrelation in the data. We investigate KBB, MBB and confidence intervals based on HAC covariance matrix estimators.

4.1 Design

We consider the same simulation design as that of Andrews (1991, Section 9, pp.840-849) and Andrews and Monahan (1992, Section 3, pp.956-964), i.e., linear regression with an intercept and four regressor variables. The model studied is

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \sigma_t u_t, \tag{4.1}$$

where σ_t is a function of the regressors $x_{i,t}$, $(i = 1, ..., 4)$, to be specified below. The interest concerns 95% confidence interval estimators for the coefficient β_1 of the first non-constant regressor.

The regressors and error term u_t are generated as follows. First,

$$u_t = \rho u_{t-1} + \varepsilon_{0,t},$$

with initial condition $u_{-49} = \varepsilon_{0,-49}$. Let

$$\tilde{x}_{i,t} = \rho \tilde{x}_{i,t-1} + \varepsilon_{i,t}, (i = 1, ..., 4),$$

with initial conditions $\tilde{x}_{i,-49} = \varepsilon_{i,-49}$, $(i = 1, ..., 4)$. As in Andrews (1991), the innovations ε_{it}, $(i = 0, ..., 4)$, $(t = -49, ..., T)$, are independent standard normal random variates. Define $\tilde{x}_t = (\tilde{x}_{1,t}, ..., \tilde{x}_{4,t})'$ and $\bar{x}_t = \tilde{x}_t - \sum_{s=1}^{T} \tilde{x}_s / T$. The regressors $x_{i,t}$, $(i = 1, ..., 4)$, are then constructed as in

$$\begin{aligned} x_t &= (x_{1,t}, ..., x_{4,t})' \\ &= [\sum_{s=1}^{T} \bar{x}_s \bar{x}_s' / T]^{-1/2} \bar{x}_t, (t = 1, ..., T). \end{aligned}$$

The observations on the dependent variable y_t are obtained from the linear regression model (4.1) with the true parameter values by invariance set as $\beta_i = 0$, $(i = 0, ..., 4)$, without loss of generality.

The values of ρ are 0.0, 0.2, 0.5, 0.7 and 0.9. Homoskedastic, $\sigma_t = 1$, and heteroskedastic, $\sigma_t = |x_{1t}|$, regression errors are examined. Sample sizes $T = 64$, 128 and 256 are considered.

The number of bootstrap replications for each experiment was 1000 with 5000 random samples generated.

4.2 Bootstrap Methods

Confidence intervals based on KBB are compared with those obtained for MBB [Fitzenberger (1997), Gonçalves and White (2004)] and TBB [Paparoditis and Politis (2002)] for least squares (LS) estimation of (4.1). For succinctness, only the results on the standard percentile bootstrap confidence intervals, Efron (1979), are presented.[1]

To describe the standard percentile KBB method, let $\hat{\beta}_1$ denote the LS estimator of β_1 and $\hat{\beta}_1^*$ its bootstrap counterpart. Because the asymptotic distribution of the LS estimator $\hat{\beta}_1$ is normal and hence symmetric about β_1, in large samples the distributions of $\hat{\beta}_1 - \beta_1$ and $\beta_1 - \hat{\beta}_1$ are the same. From the uniform consistency of the bootstrap, Theorem 3.2, the distribution of $\beta_1 - \hat{\beta}_1$ is well approximated by the distribution of $\hat{\beta}_1^* - \hat{\beta}_1$. Therefore, the bootstrap percentile confidence interval for β_1 is given by

$$\left([1-\frac{1}{k^{1/2}}]\hat{\beta}_1 + \frac{\hat{\beta}^*_{1,0.025}}{k^{1/2}}, [1-\frac{1}{k^{1/2}}]\hat{\beta}_1 + \frac{\hat{\beta}^*_{1,0.975}}{k^{1/2}}\right),$$

where $\hat{\beta}^*_{1,\alpha}$ is the 100α percentile of the distribution of $\hat{\beta}_1^*$ and, recall, $k = k_2/k_1^2$.[2] For MBB, $k = 1$.

KBB confidence intervals are constructed with the following choices of kernel function $k(\cdot)$: truncated, Bartlett and (2.4) kernel functions, which respectively induce the Bartlett [BT], Smith (2011, Example 2.1, p.1203), Parzen [PZ], Smith (2011, Example 2.2, pp.1203-1204), and the optimal quadratic spectral [QS] (2.5), Smith (2011, Example 2.3, p.1204), kernel functions $k^*(\cdot)$ as the associated convolutions, and the kernel function [PP] based on the optimal trapezoidal taper of Paparoditis and Politis (2001), see Paparoditis and Politis (2001, p.1111). The respective confidence interval estimators are denoted by KBB_{J}, where J $=$ BT, PZ, QS and PP. Percentile bootstrap confidence intervals based on Corollary 3.1 are denoted KBB^a_{J}, while those based on Remarks 3.4 and 3.5 are denoted by KBB^b_{J} and KBB^c_{J} respectively.[3,4] A similar notation is adopted for bootstrap confidence intervals based on MBB and TBB where the latter is computed using the optimal Paparoditis and Politis (2001) trapezoidal taper. The validity of the MBB confidence intervals follows from results to be found in Fitzenberger (1997) and Gonçalves and White (2004). Although Paparoditis and Politis (2002) only provides a theoretical justification for TBB^b, the validity of the other TBB confidence intervals follows using

[1]The standard percentile method is valid here because the asymptotic distribution of the LS estimator is symmetric; see Politis (1998, p.45). Empirical rejection rates for bootstrap confidence intervals based on the symmetric percentile and the equal-tailed methods, Hall (1992, p.12), were also computed and are available upon request.

[2]Bootstrap intervals based on $\hat{k}$ were also computed with results similar to those obtained with k; see Remark 3.3.

[3]Uncentred bootstrap confidence intervals, cf. Remark 3.6, were also computed with results similar to the respective centred versions from Corollary 3.1 and Remarks 3.4 and 3.5.

[4]Since $\hat{\theta}^* - \hat{\theta} = -[\partial^2 \bar{\mathcal{L}}^*_{m_T}(\hat{\theta})/\partial\theta\partial\theta']^{-1}\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})/\partial\theta$ for the LS estimator, bootstrap confidence intervals based on Theorem 3.2 are numerically identical to those based on the uncentred Corollary 3.1 bootstrap.

versions of results in this paper adapted for TBB.

Standard t-statistic confidence intervals using heteroskedastic autocorrelation consistent (HAC) estimators for the asymptotic variance matrix B_0 are considered based on truncated [TR], Bartlett [BT], Parzen [PZ], Tukey-Hanning [TH] and quadratic spectral [QS] kernel functions $k^*(\cdot)$; see Andrews (1991). Alternative t-statistic confidence intervals based on the Smith (2005) HAC estimator of B_0, cf. Remark 3.7, are also examined which use kernel functions $k(\cdot)$ that induce Bartlett [$\mathrm{S_{BT}}$], Parzen [$\mathrm{S_{PZ}}$] and quadratic spectral [$\mathrm{S_{QS}}$] kernels $k^*(\cdot)$ respectively and the optimal Paparoditis and Politis (2001) trapezoidal taper [$\mathrm{S_{PP}}$].[5]

4.3 Bandwidth Choice

The accuracy of the bootstrap approximation in practice is particularly sensitive to the choice of the bandwidth or block size S_T. Gonçalves and White (2004) suggests basing the choice of MBB block size on the optimal automatic bandwidth, see Andrews (1991, Section 5, pp.830-832), appropriate for HAC variance matrix estimation using the Bartlett kernel, noting that the MBB bootstrap variance estimator is asymptotically equivalent to the Bartlett kernel variance estimator. Smith (2011, Lemma A.3, p.1219) obtained a similar equivalence between the KBB variance estimator and the corresponding HAC estimator based on the induced kernel function $k^*(\cdot)$; see also Smith (2005, Lemma 2.1, p.164). We adopt a similar approach to that of Gonçalves and White (2004) to the choice of the bandwidth for KBB confidence interval estimators. However, rather than using the method suggested in Andrews (1991) for estimation of the optimal automatic bandwidth for the induced kernel function $k^*(\cdot)$, a non-parametric estimator of this bandwidth is adopted; see Politis and Romano (1995). Despite lacking a theoretical justification, the results discussed below indicate that this procedure fares well for the simulation designs studied here.

The infeasible optimal bandwidth for HAC variance matrix estimation based on the kernel $k^*(\cdot)$ is given by

$$S_T^* = \left(q(k^*_{(q)})^2 \alpha(q) T \Big/ \int_{-\infty}^{\infty} k^*(x)^2 dx \right)^{1/(2q+1)},$$

where $k^*_{(q)} = \lim_{x\to 0}[1-k^*(x)]/|x|^q$ and $\alpha(q) = 2vec(\sum_{s=-\infty}^{\infty} |s|^2 R(s))' W vec(\sum_{s=-\infty}^{\infty} |s|^2 R(s))$

[5]The HAC estimator of B_0 of Andrews (1991) is given by $\sum_{s=1-T}^{T-1} k^*(s/S_T)\hat{R}_T(s)$ where the sample autocovariance $\hat{R}_T(s) = T^{-1}\sum_{t=\max[1,1-s]}^{\min[T,T-s]}(\partial\mathcal{L}_t(\hat{\theta})/\partial\theta)(\partial\mathcal{L}_{t-s}(\hat{\theta})/\partial\theta)'$, $(s = 1-T, ..., T-1)$, while the Smith (2005) HAC estimator of B_0 is $T^{-1}\sum_{t=1}^{T}(\partial\mathcal{L}_{tT}(\hat{\theta})/\partial\theta)(\partial\mathcal{L}_{tT}(\hat{\theta})/\partial\theta)'$, cf. (2.2).

$/\,\mathrm{tr}(W(I_{pp}+K_{pp})(\Sigma_\infty\otimes\Sigma_\infty))$, $q\in[0,\infty)$, cf. the optimal KBB bandwidth S_T^* of section 2 when $q=2$; see Andrews (1991, (5.2), p.830). Note that $q=1$ for the Bartlett [BT] kernel and $q=2$ for the Parzen [PZ], quadratic spectral [QS] kernels and the optimal Paparoditis and Politis (2001) taper [PP]. In the linear regression model (4.1) context, with diagonal weight matrix $W=\mathrm{diag}(w_1,...,w_4)$,

$$\alpha(q)=\frac{\sum_{i=1}^4 w_i[\sum_{s=-\infty}^{\infty}|s|^q R_i(s)]^2}{\sum_{i=1}^4 w_i[\sum_{s=-\infty}^{\infty} R_i(s)]^2},$$

where $R_i(s)$ is the sth autocovariance of $x_{it}(y_t-\sum_{k=1}^4 x_{kt}\beta_k)$, $(i=1,...,4)$. As in the Monte Carlo study Andrews (1991, section 9, pp.840-849), unit weights $w_i=1$, $(i=1,...,4)$, are chosen.

The optimal bandwidth S_T^* requires the estimation of the parameters $\alpha(1)$ and $\alpha(2)$. Rather then base estimation of $\alpha(q)$ on a particular ARMA model as suggested in Andrews (1991, Section 6, pp.832-837), a feasible non-parametric estimator of the Andrews (1991) optimal bandwidth replaces $\alpha(q)$ by a consistent estimator based on the flat-top lag-window of Politis and Romano (1995), *viz.*

$$\hat\alpha(q)=\frac{\sum_{i=1}^4[\sum_{j=-M_i}^{M_i}|j|^q\lambda(\frac{j}{M_i})\hat R_i(j)]^2}{\sum_{i=1}^4[\sum_{j=-M_i}^{M_i}\lambda(\frac{j}{M_i})\hat R_i(j)]^2},(q=1,2),$$

where $\lambda(t)=\mathbb{I}(|t|\in[0,1/2])+2(1-|t|)\mathbb{I}(|t|\in(1/2,1])$, $\hat R_i(j)$ is the sample jth autocovariance of $\{x_{it}(y_t-\sum_{k=1}^4 x_{kt}\beta_k)\}$, $(i=1,...,4)$, using LS estimation of β_k, $(k=1,...,4)$, and M_i, $(i=1,...,4)$, are computed using the method described in Politis and White (2004, ftn. c, p.59). The feasible optimal bandwidth estimator is then $\hat S_T^*=\left(q(k_{(q)}^*)^2\hat\alpha(q)T\Big/\int_{-\infty}^{\infty}k^*(x)^2dx\right)^{1/(2q+1)}$ whereas the bandwidth formula $S_T^*=0.6611\,(\hat\alpha(2)T)^{1/5}$ is used for the truncated kernel [TR] HAC estimator, see Andrews (1991, ftn.5, p. 834).

Bootstrap sample sizes are defined as $m_T=\max[\lfloor T/\hat S_T^*\rfloor,1]$, where $\lfloor\cdot\rfloor$ is the floor function. MBB and TBB block sizes are given by $\min[\lceil\hat S_T^*\rceil,T]$, where $\lceil\cdot\rceil$ is the ceiling function and S_T^* the optimal bandwidth estimator for the Bartlett kernel $k^*(\cdot)$ for MBB and for the kernel $k^*(\cdot)$ induced by the optimal Paparoditis and Politis (2001) trapezoidal taper for TBB.

4.4 Results

Tables 1 and 2 provide the empirical coverage rates for 95% confidence interval estimates obtained using the methods described above for the homoskedastic and heteroskedastic cases respectively.

Tables 1 and 2 around here

Overall, to a lesser or greater degree, all confidence interval estimates display under-coverage for the true value $\beta_1 = 0$ but especially for high values of ρ, a feature found in previous studies of MBB, see, e.g., Gonçalves and White (2004), and confidence intervals based on t-statistics with HAC variance matrix estimators, see Andrews (1991). As should be expected from the theoretical results of Section 3, as T increases, empirical coverage rates approach the nominal rate of 95%.

Additionally, Tables 1 and 2 reveal that the empirical coverage rates of the bootstrap confidence intervals based on Corollary 3.1 are very similar to those based on Theorem 3.2, although the former corresponds to a centred version of the latter, see ftn. 4, and is intuitively expected to yield improvements, cf. Paparoditis and Politis (2001, p.1108). Furthermore, the empirical coverage rates of the bootstrap confidence intervals constructed using the results in Remarks 3.4 and 3.5 are systematically lower across KBB, MBB and TBB than those based on Theorem 3.2 and Corollary 3.1. With a few exceptions, all bootstrap confidence interval estimates outperform those based on HAC t-statistics for all values of ρ and for all sample sizes except for $T = 256$ when, for lower and moderate values, both bootstrap and HAC t-statistic methods produce similarly satisfactory results. The following discussion is therefore conducted based solely on KBB, MBB and TBB.

A comparison of the various KBB confidence interval estimates for the homoskedastic design in Table 1 for $T = 64$ with those using MBB reveals that generally, for low values of ρ, the coverage rates for KBB_{BT} are closer to the nominal 95% than those of MBB, although both are based on the truncated kernel, and other KBB methods. Furthermore, KBB_{PP} is superior to KBB_{BT}, KBB_{PZ} and KBB_{QS} for high values of ρ, although not dramatically so for moderate ρ. While both bootstraps use the same kernel function, MBB has similar coverage rates to KBB_{BT} for low to moderate ρ but higher coverage rates for the higher values of ρ. TBB coverage is poorer than MBB at low values of ρ and is dominated by KBB_{PP} at all values of ρ even though both methods use the same taper/kernel. A similar pattern is repeated for the larger sample size $T = 128$ although the differences across bootstrap methods narrow. For sample size $T = 256$, all bootstrap and HAC t-statistic confidence intervals display similar coverage rates except for $\rho = 0.9$ when KBB_{PP} is superior. Overall, the results with homoskedastic innovations in Table 1 indicate that KBB_{PP} is the superior bootstrap method at moderate to high values of ρ at all sample sizes with KBB_{BT}, KBB_{QS}, KBB_{PP} and TBB reasonably competitive for the lower ρ at the larger sample sizes.

In Table 2, for heteroskedastic innovations, the differences in coverage rates between the various methods narrow and are more varied. For sample size $T = 64$, all KBB bootstrap confidence intervals display similar coverage for low ρ but KBB_{QS} and KBB_{PP} are superior and perform similarly for moderate to high values of ρ and for all sample sizes. MBB is again dominated by KBB_{BT} and, likewise, KBB_{PP} is superior to TBB at all sample sizes.

4.5 Summary

In general, for homoskedastic innovations, confidence interval estimates based on KBB_{PP} provide the best coverage rates for all values of ρ and sample sizes whereas, under heteroskedasticity, the performance of KBB_{BT}, KBB_{QS} and KBB_{PP} confidence intervals are similar and dominate for low and moderate values of ρ and the larger sample sizes. KBB_{QS} is broadly competitive at all values of ρ except at $\rho = 0.9$ for homoskedastic innovations.

5 Conclusion

This paper applies the kernel block bootstrap method to quasi-maximum likelihood estimation of dynamic models under stationarity and weak dependence. The proposed bootstrap method is simple to implement by first kernel-weighting the components comprising the quasi-log likelihood function appropriately and then sampling the resultant transformed components using the standard "m out of n" bootstrap.

We investigate the first order asymptotic properties of the kernel block bootstrap for quasi-maximum likelihood demonstrating, in particular, its consistency and the first-order asymptotic validity of the bootstrap approximation to the distribution of the quasi-maximum likelihood estimator. A number of first order equivalent kernel block bootstrap schemes are suggested of differing computational complexities. A set of simulation experiments for the mean linear regression model illustrates the efficacy of the kernel block bootstrap for quasi-maximum likelihood estimation. Indeed, in these experiments, the kernel block bootstrap outperforms other bootstrap methods for the sample sizes considered, especially if the induced KBB kernel function is chosen appropriately as either the Bartlett kernel or the quadratic spectral kernel or the optimal taper of Paparoditis and Politis (2001) is used to kernel-weight the quasi-log likelihood function.

Appendix

Throughout the Appendices, C and Δ denote generic positive constants that may be different in different uses with C, M, and T the Chebyshev, Markov, and triangle inequalities respectively. UWL is a uniform weak law of large numbers such as Newey and McFadden (1994, Lemma 2.4, p.2129) for stationary and mixing (and, thus, ergodic) processes.

A similar notation is adopted to that in Gonçalves and White (2004). For any bootstrap statistic $T^*(\cdot,\omega)$, $T^*(\cdot,\omega) \to 0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$ if, for any $\delta > 0$ and any $\xi > 0$, $\lim_{T\to\infty} \mathcal{P}\{\omega : \mathcal{P}^*_\omega\{\lambda : |T^*(\lambda,\omega)| > \delta\} > \xi\} = 0$.

To simplify the analysis, the appendices consider the transformed uncentred observations

$$\mathcal{L}_{tT}(\theta) = \frac{1}{(k_2 S_T)^{1/2}} \sum_{s=t-1}^{t-T} k(\frac{s}{S_T})\mathcal{L}_{t-s}(\theta)$$

with k_2 substituting for $\hat{k}_2 = S_T^{-1} \sum_{t=1-T}^{T-1} k(t/S_T)^2$ in the main text since $\hat{k}_2 - k_2 = o(1)$, cf. PS Supplement Corollary K.2, p.S.21.

For simplicity, where required, it is assumed T/S_T is integer.

Appendix A: Preliminary Lemmas

Assumption A.1. (Bootstrap Pointwise WLLN.) For each $\theta \in \Theta \subset \mathcal{R}^{d_\theta}$, Θ a compact set, $S_T \to \infty$ and $S_T = o(T^{-1/2})$

$$(k_2/S_T)^{1/2}[\bar{\mathcal{L}}^*_{m_T}(\theta) - \bar{\mathcal{L}}_T(\theta)] \to 0, \text{ prob-}\mathcal{P}^*_\omega, \text{ prob-}\mathcal{P}.$$

Remark A.1. See Lemma A.2 below.

Assumption A.2. (Uniform Convergence.)

$$\sup_{\theta\in\Theta} \left|(k_2/S_T)^{1/2}\bar{\mathcal{L}}_T(\theta) - k_1\bar{\mathcal{L}}(\theta)\right| \to 0 \text{ prob-}\mathcal{P}.$$

Remark A.2. The hypotheses of the UWLs Smith (2011, Lemma A.1, p.1217) and Newey and McFadden (1994, Lemma 2.4, p.2129) for stationary and mixing (and, thus, ergodic) processes are satisfied under Assumptions 3.1-3.3. Hence, noting $\sup_{\theta\in\Theta}\left\|\bar{\mathcal{L}}(\theta)-\mathcal{L}(\theta)]\right\| \to 0$, prob-$\mathcal{P}$, where $\mathcal{L}(\theta)=\mathrm{E}[\mathcal{L}_t(\theta)]$, $\sup_{\theta\in\Theta}\left\|(k_2/S_T)^{1/2}\bar{\mathcal{L}}_T(\theta)-k_1\mathcal{L}(\theta)]\right\|\to 0$, prob-$\mathcal{P}$. Thus, Assumption A.2 follows by T and k_1, $k_2=O(1)$.

Assumption A.3. (Global Lipschitz Continuity.) For all $\theta,\theta^0\in\Theta$, $|\mathcal{L}_t(\theta)-\mathcal{L}_t(\theta^0)|\leq L_t\left\|\theta-\theta^0\right\|$ a.s.$\mathcal{P}$ where $\sup_T E[\sum_{t=1}^T L_t/T]<\infty$.

Remark A.3. Assumption A.3 is Assumption 3.2(c).

Lemma A.1. (Bootstrap UWL.) Suppose Assumptions A.1-A.3 hold. Then, for $S_T\to\infty$ and $S_T=o(T^{1/2})$,

$$\sup_{\theta\in\Theta}\left|(k_2/S_T)^{1/2}\bar{\mathcal{L}}^*_{m_T}(\theta)-k_1\bar{\mathcal{L}}(\theta)\right|\to 0,\text{ prob-}\mathcal{P}^*_\omega,\text{ prob-}\mathcal{P}.$$

Proof. From Assumption A.2 the result is proven if

$$\sup_{\theta\in\Theta}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right|\to 0,\text{ prob-}\mathcal{P}^*_\omega,\text{ prob-}\mathcal{P}.$$

The following preliminary results are useful in the later analysis. By global Lipschitz continuity of $\mathcal{L}_t(\cdot)$ and by T, for T large enough,

$$\begin{aligned}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}_T(\theta))-\bar{\mathcal{L}}_T(\theta^0)\right| &\leq \frac{1}{T}\sum_{t=1}^{T}\frac{1}{S_T}\sum_{s=t-T}^{t-1}\left|k\left(\frac{s}{S_T}\right)\right|\left|\mathcal{L}_{t-s}(\theta)-\mathcal{L}_{t-s}(\theta^0)\right|\\ &= \frac{1}{T}\sum_{t=1}^{T}\left|\mathcal{L}_t(\theta)-\mathcal{L}_t(\theta^0)\right|\left|\frac{1}{S_T}\sum_{s=1-t}^{T-t}k\left(\frac{s}{S_T}\right)\right|\\ &\leq C\left\|\theta-\theta^0\right\|\frac{1}{T}\sum_{t=1}^{T}L_t \qquad\text{(A.1)}\end{aligned}$$

since for some $0<C<\infty$

$$\left|\frac{1}{S_T}\sum_{s=1-t}^{T-t}k\left(\frac{s}{S_T}\right)\right|\leq O(1)<C$$

uniformly t for T large enough, see Smith (2011, eq. (A.5), p.1218). Next, for some

$0 < C^* < \infty$,

$$
\begin{aligned}
(k_2/S_T)^{1/2}E^*[\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}^*_{m_T}(\theta^0)\right|] &= \frac{1}{m_T}\sum_{s=1}^{m_T}\frac{1}{S_T}E^*[\sum_{r=t_s^*-T}^{t_s^*-1}\left|k\left(\frac{r}{S_T}\right)\right|\left|\mathcal{L}_{t_s^*-r}(\theta)-\mathcal{L}_{t_s^*-r}(\theta^0)\right|] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left|\mathcal{L}_t(\theta)-\mathcal{L}_t(\theta^0)\right|\frac{1}{S_T}\sum_{r=t-T}^{t-1}\left|k\left(\frac{r}{S_T}\right)\right| \\
&\leq C^*\left\|\theta-\theta^0\right\|\frac{1}{T}\sum_{t=1}^{T}L_t.
\end{aligned}
$$

Hence, by M, for some $0 < C^* < \infty$ uniformly t for large enough T,

$$
\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}^*_{m_T}(\theta^0)\right|>\epsilon\}\leq\frac{C^*}{\epsilon}\left\|\theta-\theta^0\right\|\frac{1}{T}\sum_{t=1}^{T}L_t. \tag{A.2}
$$

The remaining part of the proof is identical to Gonçalves and White (2000, Proof of Lemma A.2, pp.30-31) and is given here for completeness; cf. Hall and Horowitz (1996, Proof of Lemma 8, p.913). Given $\varepsilon > 0$, let $\{\eta(\theta_i,\varepsilon),(i=1,...,I)\}$ denote a finite subcover of Θ where $\eta(\theta_i,\varepsilon)=\{\theta\in\Theta:\|\theta-\theta_i\|<\varepsilon\}$, $(i=1,...,I)$. Now

$$
\sup_{\theta\in\Theta}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right|=\max_{i=1,\ldots,I}\sup_{\theta\in\eta(\theta_i,\varepsilon)}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right|.
$$

The argument $\omega\in\Omega$ is omitted for brevity as in Gonçalves and White (2000). It then follows that, for any $\delta > 0$ (and any fixed ω),

$$
\mathcal{P}^*_\omega\{\sup_{\theta\in\Theta}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right|>\delta\}\leq\sum\nolimits_{i=1}^{I}\mathcal{P}^*_\omega\{\sup_{\theta\in\eta(\theta_i,\varepsilon)}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right|>\delta\}.
$$

For any $\theta\in\eta(\theta_i,\varepsilon)$, by T,

$$
\begin{aligned}
(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right| &\leq (k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta_i)-\bar{\mathcal{L}}_T(\theta_i)\right|+(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}^*_{m_T}(\theta_i)\right| \\
&\quad +(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}_T(\theta)-\bar{\mathcal{L}}_T(\theta_i)\right|.
\end{aligned}
$$

Hence, for any $\delta > 0$ and $\xi > 0$,

$$
\begin{aligned}
\mathcal{P}\{\mathcal{P}^*_\omega\{\sup_{\theta\in\eta(\theta_i,\varepsilon)}(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)\right|>\delta\}>\xi\} &\leq \mathcal{P}\{\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta_i)-\bar{\mathcal{L}}_T(\theta_i)\right|>\frac{\delta}{3}\}>\frac{\xi}{3}\} \\
&\quad +\mathcal{P}\{\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}^*_{m_T}(\theta_i)\right|>\frac{\delta}{3}\}>\frac{\xi}{3}\} \\
&\quad +\mathcal{P}\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}_T(\theta)-\bar{\mathcal{L}}_T(\theta_i)\right|>\frac{\delta}{3}\}.
\end{aligned} \tag{A.3}
$$

By Assumption A.1

$$\mathcal{P}\{\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta_i)-\bar{\mathcal{L}}_T(\theta_i)\right|>\frac{\delta}{3}\}>\frac{\xi}{3}\}<\frac{\xi}{3}$$

for large enough T. Also, by M (for fixed ω) and Assumption A.3, noting $L_t \geq 0$, $(t=1,...,T)$, from eq. (A.2),

$$\begin{aligned}\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}^*_{m_T}(\theta_i)\right| &> \frac{\delta}{3}\}\leq\frac{3C^*}{\delta}\left\|\theta-\theta_i\right\|\frac{1}{T}\sum\nolimits_{t=1}^T L_t\\ &\leq \frac{3C^*\varepsilon}{\delta}\frac{1}{T}\sum\nolimits_{t=1}^T L_t.\end{aligned}$$

As a consequence, for any $\delta>0$ and $\xi>0$, for T sufficiently large,

$$\begin{aligned}\mathcal{P}\{\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}^*_{m_T}(\theta_i)\right| &> \frac{\delta}{3}\}>\frac{\xi}{3}\}\leq\mathcal{P}\{\frac{3C^*\varepsilon}{\delta}\frac{1}{T}\sum\nolimits_{t=1}^T L_t>\frac{\xi}{3}\}\\ &= \mathcal{P}\{\frac{1}{T}\sum\nolimits_{t=1}^T L_t>\frac{\xi\delta}{9C^*\varepsilon}\}\\ &\leq \frac{9C^*\varepsilon}{\xi\delta}\mathrm{E}[\frac{1}{T}\sum\nolimits_{t=1}^T L_t]\\ &\leq \frac{9C^*\varepsilon\Delta}{\xi\delta}<\frac{\xi}{3}\end{aligned}$$

for the choice $\varepsilon<\xi^2\delta/27C^*\Delta$, where, since, by hypothesis $\mathrm{E}[\sum_{t=1}^T L_t/T]=O(1)$, the second and third inequalities follow respectively from M and Δ a sufficiently large but finite constant such that $\sup_T \mathrm{E}[\sum_{t=1}^T L_t/T]<\Delta$. Similarly, from eq. (A.1), for any $\delta>0$ and $\xi>0$, by Assumption A.3, $\mathcal{P}\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}_T(\theta)-\bar{\mathcal{L}}_T(\theta_i)\right|>\delta/3\}\leq\mathcal{P}\{C\varepsilon\sum_{t=1}^T L_t/T>\delta/3\}\leq 3C\varepsilon\Delta/\delta<\xi/3$ for T sufficiently large for the choice $\varepsilon<\xi\delta/9C\Delta$.

Therefore, from eq. (A.3), the conclusion of the Lemma follows if

$$\varepsilon=\frac{\xi\delta}{9\Delta}\max\left(\frac{1}{C},\frac{\xi}{3C^*}\right).\blacksquare$$

Lemma A.2. (Bootstrap Pointwise WLLN.) Suppose Assumptions 3.1, 3.2(a) and 3.3 are satisfied. Then, if $T^{1/\alpha}/m_T\to 0$ and $\mathrm{E}[\sup_{\theta\in\Theta}\left|\log f_t(\theta)\right|^\alpha]<\infty$ for some $\alpha>v$, for each $\theta\in\Theta\subset\mathcal{R}^{d_\theta}$, Θ a compact set,

$$(k_2/S_T)^{1/2}[\bar{\mathcal{L}}^*_{m_T}(\theta)-\bar{\mathcal{L}}_T(\theta)]\to 0,\text{ prob-}\mathcal{P}^*_\omega,\text{ prob-}\mathcal{P}.$$

Proof: The argument θ is suppressed throughout for brevity. First, cf. Gonçalves and White (2004, Proof of Lemma A.5, p.215),

$$(k_2/S_T)^{1/2}(\bar{\mathcal{L}}^*_{m_T}-\bar{\mathcal{L}}_T)=(k_2/S_T)^{1/2}(\bar{\mathcal{L}}^*_{m_T}-\mathrm{E}^*[\bar{\mathcal{L}}^*_{m_T}])-(k_2/S_T)^{1/2}(\bar{\mathcal{L}}_T-\mathrm{E}^*[\bar{\mathcal{L}}^*_{m_T}]).$$

Since $\mathrm{E}^*[\bar{\mathcal{L}}^*_{m_T}] = \bar{\mathcal{L}}_T$, cf. PS (Section 2.2, pp.2-3), the second term $\bar{\mathcal{L}}_T - \mathrm{E}^*[\bar{\mathcal{L}}^*_{m_T}]$ is zero. Hence, the result follows if, for any $\delta > 0$ and $\xi > 0$ and large enough T, $\mathcal{P}\{\mathcal{P}^*_\omega\{(k_2/S_T)^{1/2}\left|\bar{\mathcal{L}}^*_{m_T} - \mathrm{E}^*[\bar{\mathcal{L}}^*_{m_T}]\right| > \delta\} > \xi\} < \xi$.

Without loss of generality, set $\mathrm{E}^*[\bar{\mathcal{L}}^*_{m_T}] = 0$. Write $\mathcal{K}_{tT} = (k_2/S_T)^{1/2}\mathcal{L}_{tT}$, $(t = 1, ..., T)$. First, note that

$$\begin{aligned}\mathrm{E}^*[\left|\mathcal{K}_{t^*_s T}\right|] &= \frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{K}_{tT}| = \frac{1}{T}\sum\nolimits_{t=1}^{T}\left|\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})\mathcal{L}_{t-s}\right| \\ &\leq O(1)\frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{L}_t| = O_p(1),\end{aligned}$$

uniformly, $(s = 1, ..., m_T)$, by WLLN and $\mathrm{E}[\sup_{\theta\in\Theta}|\log f_t(\theta)|^\alpha] < \infty$, $\alpha > 1$. Also, for any $\delta > 0$,

$$\begin{aligned}\frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{K}_{tT}| - \frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{K}_{tT}|\,\mathbb{I}(|\mathcal{K}_{tT}| &< m_T\delta) = \frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{K}_{tT}|\,\mathbb{I}(|\mathcal{K}_{tT}| \geq m_T\delta) \\ &\leq \frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{K}_{tT}|\max_t \mathbb{I}(|\mathcal{K}_{tT}| \geq m_T\delta).\end{aligned}$$

Now, by M,

$$\max_t|\mathcal{K}_{tT}| = O(1)\max_t|\mathcal{L}_t| = O_p(T^{1/\alpha});$$

cf. Newey and Smith (2004, Proof of Lemma A1, p.239). Hence, since, by hypothesis, $T^{1/\alpha}/m_T = o(1)$, $\max_t \mathbb{I}(|\mathcal{K}_{tT}| \geq m_T\delta) = o_p(1)$ and $\sum_{t=1}^{T}|\mathcal{K}_{tT}|/T = O_p(1)$,

$$\mathrm{E}^*[\left|\mathcal{K}_{t^*_s T}\right|\mathbb{I}(\left|\mathcal{K}_{t^*_s T}\right| \geq m_T\delta)] = \frac{1}{T}\sum\nolimits_{t=1}^{T}|\mathcal{K}_{tT}|\,\mathbb{I}(|\mathcal{K}_{tT}| \geq m_T\delta) = o_p(1). \tag{A.4}$$

The remaining part of the proof is similar to that for Khinchine's WLLN given in Rao (1973, pp.112-114). For each s define the pair of random variables

$$V_{t^*_s T} = \mathcal{K}_{t^*_s T}\mathbb{I}(\left|\mathcal{K}_{t^*_s T}\right| < m_T\delta), W_{t^*_s T} = \mathcal{K}_{t^*_s T}\mathbb{I}(\left|\mathcal{K}_{t^*_s T}\right| \geq m_T\delta),$$

yielding $\mathcal{K}_{t^*_s T} = V_{t^*_s T} + W_{t^*_s T}$, $(s = 1, ..., m_T)$. Now

$$\mathrm{var}^*[V_{t^*_s T}] \leq \mathrm{E}^*[V^2_{t^*_s T}] \leq m_T\delta\mathrm{E}^*[\left|V_{t^*_s T}\right|]. \tag{A.5}$$

Write $\bar{V}^*_{m_T} = \sum\nolimits_{s=1}^{m_T} V_{t^*_s T}/m_T$. Thus, from eq. (A.5), using C,

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{V}^*_{m_T} - \mathrm{E}^*[V_{t^*_s T}]\right| &\geq \varepsilon\} \leq \frac{\mathrm{var}^*[V_{t^*_s T}]}{m_T\varepsilon^2} \\ &\leq \frac{\delta\mathrm{E}^*[\left|V_{t^*_s T}\right|]}{\varepsilon^2}.\end{aligned}$$

Also $\left|\bar{\mathcal{K}}_T - \mathrm{E}^*[V_{t_s^*T}]\right| = o_p(1)$, i.e., for any $\varepsilon > 0$, T large enough, $\left|\bar{\mathcal{K}}_T - \mathrm{E}^*[V_{t_s^*T}]\right| \leq \varepsilon$, since by T, noting $\mathrm{E}^*[V_{t_s^*T}] = \sum_{t=1}^{T} \mathcal{K}_{tT}\mathbb{I}(\,|\mathcal{K}_{tT}| < m_T\delta)/T$,

$$\begin{aligned}\left|\bar{\mathcal{K}}_T - \mathrm{E}^*[V_{t_s^*T}]\right| &= \left|\frac{1}{T}\sum_{t=1}^{T} \mathcal{K}_{tT} - \frac{1}{T}\sum_{t=1}^{T} \mathcal{K}_{tT}\mathbb{I}(|\mathcal{K}_{tT}| < m_T\delta)\right| \\ &\leq \frac{1}{T}\sum_{t=1}^{T} |\mathcal{K}_{tT}|\,\mathbb{I}(|\mathcal{K}_{tT}| \geq m_T\delta) = o_p(1)\end{aligned}$$

from eq. (A.4). Hence, for T large enough,

$$\mathcal{P}^*\{\left|\bar{V}^*_{m_T} - \bar{\mathcal{K}}_T\right| \geq 2\varepsilon\} \leq \frac{\delta\mathrm{E}^*[\left|V_{t_s^*T}\right|]}{\varepsilon^2}. \tag{A.6}$$

By M,

$$\begin{aligned}\mathcal{P}^*\{W_{t_s^*T} &\neq 0\} = \mathcal{P}^*\{\left|\mathcal{K}_{t_s^*T}\right| \geq m_T\delta\} \\ &\leq \frac{1}{m_T\delta}\mathrm{E}^*[\left|\mathcal{K}_{t_s^*T}\right|\mathbb{I}(\left|\mathcal{K}_{t_s^*T}\right| \geq m_T\delta)] \leq \frac{\delta}{m_T}.\end{aligned} \tag{A.7}$$

To see this, $\mathrm{E}^*[\left|\mathcal{K}_{t_s^*T}\right|\mathbb{I}(\left|\mathcal{K}_{t_s^*T}\right| \geq m_T\delta)] = o_p(1)$ from eq. (A.4). Thus, for T large enough, $\mathrm{E}^*[\left|\mathcal{K}_{t_s^*T}\right|\mathbb{I}(\left|\mathcal{K}_{t_s^*T}\right| \geq m_T\delta)] \leq \delta^2$ w.p.a.1. Write $\bar{W}^*_{m_T} = \sum_{s=1}^{m_T} W_{t_s^*T}/m_T$. Thus, from eq. (A.7),

$$\mathcal{P}^*\{\bar{W}^*_{m_T} \neq 0\} \leq \sum_{s=1}^{m_T} \mathcal{P}^*\{W_{t_s^*T} \neq 0\} \leq \delta. \tag{A.8}$$

Therefore,

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{\mathcal{K}}^*_{m_T} - \bar{\mathcal{K}}_T\right| &\geq 4\varepsilon\} \leq \mathcal{P}^*\{\left|\bar{V}^*_{m_T} - \bar{\mathcal{K}}_T\right| + \left|\bar{W}^*_{m_T}\right| \geq 4\varepsilon\} \\ &\leq \mathcal{P}^*\{\left|\bar{V}^*_{m_T} - \bar{\mathcal{K}}_T\right| \geq 2\varepsilon\} + \mathcal{P}^*\{\left|\bar{W}^*_{m_T}\right| \geq 2\varepsilon\} \\ &\leq \frac{\delta E^*[\left|\bar{V}_{t_s^*T}\right|]}{\varepsilon^2} + \mathcal{P}^*\{\left|\bar{W}^*_{m_T}\right| \neq 0\} \leq \frac{\delta E^*[\left|V_{t_s^*T}\right|]}{\varepsilon^2} + \delta.\end{aligned}$$

where the first inequality follows from T, the third from eq. (A.6) and the final inequality from eq. (A.8). Since δ may be chosen arbitrarily small enough and $\mathrm{E}^*[\left|V_{t_s^*T}\right|] \leq \mathrm{E}^*[\left|\mathcal{K}_{t_s^*T}\right|] = O_p(1)$, the result follows by M.■

LEMMA A.3. (Bootstrap CLT.) Let Assumptions 3.2(a)(b), 3.3, 3.4 and 3.6(b)(c) hold. Then, if $S_T \to \infty$ and $S_T = O(T^{\frac{1}{2}-\eta})$ with $0 < \eta < \frac{1}{2}$,

$$\sup_{x\in\mathcal{R}}\left|\mathcal{P}^*_\omega\{m_T^{1/2}(\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{\partial\theta} - \frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta}) \leq x\} - \mathcal{P}\{T^{1/2}\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta} \leq x\}\right| \to 0,\ \text{prob-}\mathcal{P}.$$

Proof. The result is proven in Steps 1-5 below; cf. Politis and Romano (1992b, Proof of Theorem 2, pp. 1994-1995). To ease exposition, let $m_T = T/S_T$ be integer and $d_\theta = 1$.

Step 1. $d\bar{\mathcal{L}}(\theta_0)/d\theta \to 0$ prob-$\mathcal{P}$. Follows by White (1984, Theorem 3.47, p.46) and $\mathrm{E}[\partial \mathcal{L}_t(\theta_0)/\partial\theta] = 0$.

Step 2. $\mathcal{P}\{B_0^{-1/2}T^{1/2}d\bar{\mathcal{L}}(\theta_0)/d\theta \le x\} \to \Phi(x)$, where $\Phi(\cdot)$ is the standard normal distribution function. Follows by White (1984, Theorem 5.19, p.124).

Step 3. $\sup_x \left|\mathcal{P}\{B_0^{-1/2}T^{1/2}d\bar{\mathcal{L}}(\theta_0)/d\theta \le x\} - \Phi(x)\right| \to 0$. Follows by Pólya's Theorem (Serfling, 1980, Theorem 1.5.3, p.18) from Step 2 and the continuity of $\Phi(\cdot)$.

Step 4. $\mathrm{var}^*[m_T^{1/2}d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta] \to B_0$ prob-$\mathcal{P}$. Note $\mathrm{E}^*[d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta] = d\bar{\mathcal{L}}_T(\theta_0)/d\theta$. Thus,

$$\begin{aligned}
\mathrm{var}^*[m_T^{1/2}\frac{d\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{d\theta}] &= \mathrm{var}^*[\frac{d\mathcal{L}_{t^*T}(\theta_0)}{d\theta}] \\
&= \frac{1}{T}\sum_{t=1}^{T}(\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta} - \frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}(\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta})^2 - (\frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta})^2.
\end{aligned}$$

the result follows since $(d\bar{\mathcal{L}}_T(\theta_0)/d\theta)^2 = O_p(S_T/T)$ (Smith, 2011, Lemma A.2, p.1219), $S_T = o(T^{1/2})$ by hypothesis and $T^{-1}\sum_{t=1}^{T}(d\mathcal{L}_{tT}(\theta_0)/d\theta)^2 \to B_0$ prob-$\mathcal{P}$ (Smith, 2011, Lemma A.3, p.1219).

Step 5.

$$\lim_{T\to\infty}\mathcal{P}\left\{\sup_x\left|\mathcal{P}^*_\omega\{\frac{d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta - \mathrm{E}^*[d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta]}{\mathrm{var}^*[d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta]^{1/2}} \le x\} - \Phi(x)\right| \ge \varepsilon\right\} = 0.$$

Applying the Berry-Esséen inequality, Serfling (1980, Theorem 1.9.5, p.33), noting the bootstrap sample observations $\{d\mathcal{L}_{t_s^*T}(\theta_0)/d\theta\}_{s=1}^{m_T}$ are independent and identically distributed,

$$\begin{aligned}
\sup_x\left|\mathcal{P}^*_\omega\{\frac{m_T^{1/2}(d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta - d\bar{\mathcal{L}}_T(\theta_0)/d\theta)}{\mathrm{var}^*[m_T^{1/2}d\bar{\mathcal{L}}^*_{m_T}(\theta_0)/d\theta]^{1/2}} \le x\} - \Phi(x)\right| &\le \frac{C}{m_T^{1/2}}\mathrm{var}^*[\frac{d\mathcal{L}_{t^*T}(\theta_0)}{d\theta}]^{-3/2} \\
&\quad \times \mathrm{E}^*[\left|\frac{d\mathcal{L}_{t^*T}(\theta_0)}{d\theta} - \frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta}\right|^3].
\end{aligned}$$

Now $\mathrm{var}^*[d\mathcal{L}_{t^*T}(\theta_0)/d\theta] \to B_0 > 0$ prob-$\mathcal{P}$; see the Proof of Step 4 above. Furthermore,

$\mathrm{E}^*[\left|d\mathcal{L}_{t^*T}(\theta_0)/d\theta - d\bar{\mathcal{L}}_T(\theta_0)/d\theta\right|^3] = T^{-1}\sum_{t=1}^{T}\left|d\mathcal{L}_{tT}(\theta_0)/d\theta - d\bar{\mathcal{L}}_T(\theta_0)/d\theta\right|^3$ and

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta}-\frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta}\right|^3 &\leq \max_t\left|\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta}-\frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta}\right|\frac{1}{T}\sum_{t=1}^{T}(\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta}-\frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta})^2\\ &= O_p(S_T^{1/2}T^{1/\alpha}).\end{aligned}$$

The equality follows since

$$\begin{aligned}\max_t\left|\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta}-\frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta}\right| &\leq \max_t\left|\frac{d\mathcal{L}_{tT}(\theta_0)}{d\theta}\right|+\left|\frac{d\bar{\mathcal{L}}_T(\theta_0)}{d\theta}\right|\\ &= O_p(S_T^{1/2}T^{1/\alpha})+O_p((S_T/T)^{1/2})=O_p(S_T^{1/2}T^{1/\alpha})\end{aligned}$$

by M and Assumption 3.6(b), cf. Newey and Smith (2004, Proof of Lemma A1, p.239), and $\sum_{t=1}^{T}(d\mathcal{L}_{tT}(\theta_0)/d\theta - d\bar{\mathcal{L}}_T(\theta_0)/d\theta)^2/T = O_p(1)$, see the Proof of Step 4 above. Therefore

$$\begin{aligned}\sup_x\left|\mathcal{P}_\omega^*\{\frac{(T/S_T)^{1/2}(d\bar{\mathcal{L}}_{m_T}^*(\theta_0)/d\theta - d\bar{\mathcal{L}}_T(\theta_0)/d\theta)}{\mathrm{var}^*[(T/S_T)^{1/2}d\bar{\mathcal{L}}_{m_T}^*(\theta_0)/d\theta]^{1/2}}\leq x\}-\Phi(x)\right| &\leq \frac{1}{m_T^{1/2}}O_p(1)O_p(S_T^{1/2}T^{1/\alpha})\\ &= \frac{S_T^{1/2}}{m_T^{1/2}}O_p(T^{1/\alpha})=o_p(1),\end{aligned}$$

by hypothesis, yielding the required conclusion.■

Lemma A.4. Suppose that Assumptions 3.2(a)(b), 3.3, 3.4 and 3.6(b)(c) hold. Then, if $S_T\to\infty$ and $S_T=O(T^{\frac{1}{2}-\eta})$ with $0<\eta<\frac{1}{2}$,

$$(k_2/S_T)^{1/2}\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta}=k_1\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta}+o_p(T^{-1/2}).$$

Proof. Cf. Smith (2011, Proof of Lemma A.2, p.1219). Recall

$$(k_2/S_T)^{1/2}\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta}=\frac{1}{S_T}\sum_{r=1-T}^{T-1}k\left(\frac{r}{S_T}\right)\frac{1}{T}\sum_{t=\max[1,1-r]}^{\min[T,T-r]}\frac{\partial\mathcal{L}_t(\theta_0)}{\partial\theta}.$$

The difference between $\sum_{t=\max[1,1-r]}^{\min[T,T-r]}\partial\mathcal{L}_t(\theta_0)/\partial\theta$ and $\sum_{t=1}^{T}\partial\mathcal{L}_t(\theta_0)/\partial\theta$ consists of $|r|$ terms. By C, using White (1984, Lemma 6.19, p.153),

$$\begin{aligned}\mathcal{P}\{\frac{1}{T}\left|\sum_{t=1}^{|r|}\frac{\partial\mathcal{L}_t(\theta_0)}{\partial\theta}\right| &\geq \varepsilon\}\leq\frac{1}{(T\varepsilon)^2}\mathrm{E}[\left|\sum_{t=1}^{|r|}\frac{\partial\mathcal{L}_t(\theta_0)}{\partial\theta}\right|^2]\\ &= |r|\,O(T^{-2})\end{aligned}$$

where the $O(T^{-2})$ term is independent of r. Therefore, using Smith (2011, Lemma C.1, p.1231),

$$\begin{aligned}(k_2/S_T)^{1/2}\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta} &= \frac{1}{S_T}\sum_{r=1-T}^{T-1}k\left(\frac{r}{S_T}\right)(\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta}+|r|\,O_p(T^{-2})) \\ &= \frac{1}{S_T}\sum_{s=1-T}^{T-1}k\left(\frac{s}{S_T}\right)\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta}+O_p(T^{-1}) \\ &= (k_1+o(1))\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta}+O_p(T^{-1}) \\ &= k_1\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta}+o_p(T^{-1/2}).\blacksquare\end{aligned}$$

Appendix B: Proofs of Results

Proof of Theorem 3.1. Theorem 3.1 follows from a verification of the hypotheses of Gonçalves and White (2004, Lemma A.2, p.212). To do so, replace n by T, $Q_T(\cdot,\theta)$ by $\bar{\mathcal{L}}(\theta)$ and $Q_T^*(\cdot,\omega,\theta)$ by $\bar{\mathcal{L}}_{m_T}^*(\omega,\theta)$. Conditions (a1)-(a3), which ensure $\hat{\theta}-\theta_0\to 0$, prob-$\mathcal{P}$, hold under Assumptions 3.1 and 3.2. To establish $\hat{\theta}^*-\hat{\theta}\to 0$, prob-$\mathcal{P}^*$, prob-$\mathcal{P}$, Conditions (b1) and (b2) follow from Assumption 3.1 whereas Condition (b3) is the bootstrap UWL Lemma A.1 which requires Assumption 3.3.■

Proof of Theorem 3.2. The structure of the proof is identical to that of Gonçalves and White (2004, Theorem 2.2, pp.213-214) for MBB requiring the verification of the hypotheses of Gonçalves and White (2004, Lemma A.3, p.212) which together with Pólya's Theorem, Serfling (1980, Theorem 1.5.3, p.18), and the continuity of $\Phi(\cdot)$ gives the result.

Assumptions 3.2-3.4 ensure Theorem 3.1, i.e., $\hat{\theta}^*-\hat{\theta}\to 0$, prob-$\mathcal{P}^*$, prob-$\mathcal{P}$, and $\hat{\theta}-\theta_0\to 0$, prob-$\mathcal{P}$. The assumptions of the complete probability space $(\Omega,\mathcal{F},\mathcal{P})$ and compactness of Θ are stated in Assumptions 3.4(a) and 3.5(a). Conditions (a1) and (a2) follow from Assumptions 3.5(a)(b). Condition (a3) $B_0^{-1/2}T^{1/2}\partial\bar{\mathcal{L}}(\theta_0)/\partial\theta\xrightarrow{d}N(0,I_{d_\theta})$ is satisfied under Assumptions 3.4, 3.5(a)(b) and 3.6(b)(c) using the CLT White (1984, Theorem 5.19, p.124); cf. Step 4 in the Proof of Lemma A.3 above. The continuity of $A(\theta)$ and the UWL Condition (a4) $\sup_{\theta\in\Theta}\left\|\partial^2\bar{\mathcal{L}}(\theta)/\partial\theta\partial\theta'-A(\theta)\right\|\to 0$, prob-$\mathcal{P}$, follow since the hypotheses of the UWL Newey and McFadden (1994, Lemma 2.4, p.2129) for stationary and mixing (and, thus, ergodic) processes are satisfied under Assumptions 3.4-3.6. Hence, invoking Assumption 3.6(c), from a mean value expansion of $\partial\bar{\mathcal{L}}(\hat{\theta})/\partial\theta=0$ around $\theta=\theta_0$ with $\theta_0\in\mathrm{int}(\Theta)$ from Assumption 3.5(c), $T^{1/2}(\hat{\theta}-\theta_0)\xrightarrow{d}N(0,A_0^{-1}B_0A_0^{-1})$.

Conditions (b1) and (b2) are satisfied under Assumptions 3.5(a)(b) as above. To verify Condition (b3),

$$
\begin{aligned}
m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta} &= m_T^{1/2}(\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{\partial\theta}-\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta}) \\
&\quad +m_T^{1/2}\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta}+m_T^{1/2}(\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}-\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{\partial\theta}).
\end{aligned}
$$

With Lemma A.3 replacing Gonçalves and White (2002, Theorem 2.2(ii), p.1375), the first term converges in distribution to $N(0, B_0)$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$. The sum of the second and third terms converges to 0, prob-$\mathcal{P}^*$, prob-$\mathcal{P}$. To see this, first, using the mean value theorem for the third term, i.e.,

$$
m_T^{1/2}(\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}-\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{\partial\theta})=\frac{1}{S_T^{1/2}}\frac{\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})}{\partial\theta\partial\theta'}T^{1/2}(\hat{\theta}-\theta_0),
$$

where $\dot{\theta}$ lies on the line segment joining $\hat{\theta}$ and θ_0. Secondly, $(k_2/S_T)^{1/2}\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})/\partial\theta\partial\theta' \to k_1A_0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$, using the bootstrap UWL $\sup_{\theta\in\Theta}(k_2/S_T)^{1/2}\left\|\partial^2\bar{\mathcal{L}}^*_{m_T}(\theta)/\partial\theta\partial\theta'-\partial^2\bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta'\right\| \to 0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$, cf. Lemma A.1, and the UWL $\sup_{\theta\in\Theta}\left\|(k_2/S_T)^{1/2}\partial^2\bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta'-k_1A(\theta)\right\| \to 0$, prob-$\mathcal{P}$, cf. Remark A.2. Condition (b3) then follows since $T^{1/2}(\hat{\theta}-\theta_0)+A_0^{-1}T^{1/2}\partial\bar{\mathcal{L}}(\theta_0)/\partial\theta \to 0$, prob-$\mathcal{P}$, and $m_T^{1/2}\partial\bar{\mathcal{L}}_T(\theta_0)/\partial\theta-(k_1/k_2^{1/2})T^{1/2}\partial\bar{\mathcal{L}}(\theta_0)/\partial\theta \to 0$, prob-$\mathcal{P}$, cf. Lemma A.4. Finally, Condition (b4) $\sup_{\theta\in\Theta}\left\|(k_2/S_T)^{1/2}[\partial^2\bar{\mathcal{L}}^*_{m_T}(\theta)/\partial\theta\partial\theta'-\partial^2\bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta']\right\| \to 0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$, is the bootstrap UWL Lemma A.1 appropriately revised using Assumption 3.6.

Because $\hat{\theta}\in\text{int}(\Theta)$ from Assumption 3.5(c), from a mean value expansion of the first order condition $\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta}^*)/\partial\theta=0$ around $\theta=\hat{\theta}$,

$$
T^{1/2}(\hat{\theta}^*-\hat{\theta})=[\frac{\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})}{\partial\theta\partial\theta'}/S_T^{1/2}]^{-1}m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}, \tag{B.1}
$$

where $\dot{\theta}$ lies on the line segment joining $\hat{\theta}^*$ and $\hat{\theta}$. Noting $\hat{\theta}^*-\hat{\theta}\to 0$, prob-$\mathcal{P}^*$, prob-$\mathcal{P}$, and $\hat{\theta}-\theta_0\to 0$, prob-$\mathcal{P}$, $(k_2/S_T)^{1/2}\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})/\partial\theta\partial\theta'\to k_1A_0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$. Therefore, $T^{1/2}(\hat{\theta}^*-\hat{\theta})$ converges in distribution to $N(0,(k_2/k_1^2)A_0^{-1}B_0A_0^{-1})$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$.■

Proof of Corollary 3.1. It follows immediately from Lemma A.3 that

$$
m_T^{1/2}(\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{\partial\theta}-\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta})\xrightarrow{d}N(0,B_0),\ \text{prob-}\mathcal{P}^*_\omega,\ \text{prob-}\mathcal{P}.
$$

Moreover, from the Proof of Theorem 3.2,

$$m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}-m_T^{1/2}(\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\theta_0)}{\partial\theta}-\frac{\partial\bar{\mathcal{L}}_T(\theta_0)}{\partial\theta})\to 0,\ \text{prob-}\mathcal{P}^*_\omega,\ \text{prob-}\mathcal{P}.$$

Therefore,

$$\sup_{x\in\mathcal{R}^{d_\theta}}\left|\mathcal{P}^*_\omega\{m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}\le x\}-\mathcal{P}\{T^{1/2}\frac{\partial\bar{\mathcal{L}}(\theta_0)}{\partial\theta}\le x\}\right|\to 0,\ \text{prob-}\mathcal{P},$$

follows by Pólya's Theorem (Serfling, 1980, Theorem 1.5.3, p.18) and the continuity of the normal c.d.f. $\Phi(\cdot)$ recalling $\sup_x\left|\mathcal{P}\{B_0^{-1/2}T^{1/2}d\bar{\mathcal{L}}(\theta_0)/d\theta\le x\}-\Phi(x)\right|\to 0$ from Step 3 of the Proof of Lemma A.3.

Recall from eq. (B.1) in the Proof of Theorem 3.2 that, because $\hat{\theta}\in\text{int}(\Theta)$, from a mean value expansion of the first order condition $\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta}^*)/\partial\theta=0$ around $\theta=\hat{\theta}$,

$$T^{1/2}(\hat{\theta}^*-\hat{\theta})=[\frac{\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})}{\partial\theta\partial\theta'}/S_T^{1/2}]^{-1}m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}.$$

First, $(k_2/S_T)^{1/2}\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})/\partial\theta\partial\theta'\to k_1A_0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$, using the bootstrap UWL $\sup_{\theta\in\Theta}(k_2/S_T)^{1/2}\left\|\partial^2\bar{\mathcal{L}}^*_{m_T}(\theta)/\partial\theta\partial\theta'-\partial^2\bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta'\right\|\to 0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$, cf. Lemma A.1, and the UWL $\sup_{\theta\in\Theta}\left\|(k_2/S_T)^{1/2}\partial^2\bar{\mathcal{L}}_T(\theta)/\partial\theta\partial\theta'-k_1A(\theta)\right\|\to 0$, prob-$\mathcal{P}$, cf. Remark A.2. Secondly, similarly, $(k_2/S_T)^{1/2}\left\|\partial^2\bar{\mathcal{L}}^*_{m_T}(\dot{\theta})/\partial\theta\partial\theta'-\partial^2\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})/\partial\theta\partial\theta'\right\|\to 0$, prob-$\mathcal{P}^*_\omega$, prob-$\mathcal{P}$. Hence,

$$T^{1/2}(\hat{\theta}^*-\hat{\theta})-[\frac{\partial^2\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta\partial\theta'}/S_T^{1/2}]^{-1}m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}\to 0,\ \text{prob-}\mathcal{P}^*_\omega,\ \text{prob-}\mathcal{P}. \tag{B.2}$$

Therefore, from Theorem 3.2, after substitution of (B.2),

$$\sup_{x\in\mathcal{R}^{d_\theta}}\left|\mathcal{P}^*_\omega\{[\frac{\partial^2\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta\partial\theta'}/S_T^{1/2}]^{-1}m_T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}/k^{1/2}\le x\}-\mathcal{P}\{T^{1/2}(\hat{\theta}-\theta_0)\le x\}\right|\to 0,\ \text{prob-}\mathcal{P},$$

or

$$\sup_{x\in\mathcal{R}^{d_\theta}}\left|\mathcal{P}^*_\omega\{[\frac{\partial^2\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta\partial\theta'}]^{-1}T^{1/2}\frac{\partial\bar{\mathcal{L}}^*_{m_T}(\hat{\theta})}{\partial\theta}/k^{1/2}\le x\}-\mathcal{P}\{T^{1/2}(\hat{\theta}-\theta_0)\le x\}\right|\to 0,\ \text{prob-}\mathcal{P}.\blacksquare$$

References

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59, 817-858.

Andrews, D.W.K., and Monahan, J.C. (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica*, 60, 1992, 953-966.

Bickel, P., and Freedman, D. (1981). Some Asymptotic Theory of the Bootstrap. *Annals of Statistics*, 9, 1196-1217.

Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory.* San Francisco: Holden Day.

Bühlmann, P. (1997). Sieve Bootstrap for Time Series. *Bernoulli*, 3, 123-148.

Davidson, R. and J. G. MacKinnon (1999). Bootstrap testing in Nonlinear Models, *International Economic Review*, 40, 487-508.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, 7, 1-26.

Epanechnikov, V.A. (1969). Non-parametric Estimation of a Multivariate Probability Density. *Theory of Probability and Its Applications*, 14, 153-158.

Fitzenberger, B. (1997). The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions. *Journal of Econometrics*, 82, 235-287.

Gonçalves, S., and White, H. (2000). Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models. University of California at San Diego, Economics Working Paper Series 2000-32, Department of Economics, UC San Diego.

Gonçalves, S., and White, H. (2002). The Bootstrap of the Mean for Dependent Heterogeneous Arrays. *Econometric Theory*, 18, 1367-1384.

Gonçalves, S., and White, H. (2004). Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models. *Journal of Econometrics*, 119, 199-219.

Gradshteyn, I.S., and Ryzhik, I.M. (1980). *Tables of Integrals, Series and Products.* New York: Academic Press.

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

Hall, P., and Horowitz, J. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, 64, 891-916.

HIDALGO, J. (2003). An Alternative Bootstrap to Moving Blocks for Time Series Regression Models. *Journal of Econometrics*, 117, 369-399.

HURVICH, C., AND ZEGER, S. (1987). Frequency Domain Bootstrap Methods for Time Series. Statistics and Operation Research Working Paper, New York University.

IBRAGIMOV, I.A., AND LINNIK, Y.V. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen : Wolters-Noordhoff.

KITAMURA, Y., AND STUTZER, M. (1997). An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica*, 65, 861-874.

KÜNSCH, H.R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics*, 17, 1217–1241.

LAHIRI, S.N. (2003): *Resampling Methods for Dependent Data.* New York: Springer.

LIU, R., AND SINGH, K. (1992). Moving Blocks Jackknife and Bootstrap Capture Weak Dependence. In LePage, R., and Billard L. (Eds.), *Exploring the Limits of Bootstrap*, 224-248. New York: Wiley.

MAGNUS, J. R. AND NEUDECKER, H. (1979). The commutation matrix: some properties and applications. *Annals of Statistics*, **7**, 381-394.

NEWEY, W.K., AND MCFADDEN, D.L. (1994). Large Sample Estimation and Hypothesis Testing. Chapter 36 in *Handbook of Econometrics,* Vol. 4, eds. R.F. Engle and D.L. McFadden, 2111-2245. New York: North Holland.

NEWEY, W.K. AND SMITH, R.J. (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, 72, 219-255.

PAPARODITIS, E., AND POLITIS, D.N. (2001). Tapered Block Bootstrap. *Biometrika*, 88, 1105-1119.

PAPARODITIS, E., AND POLITIS, D.N. (2002). The Tapered Block Bootstrap for General Statistics from Stationary Sequences. *Econometrics Journal*, 5, 131-148.

PARENTE, P.M.D.C., AND SMITH, R.J. (2019). Kernel Block Bootstrap. Revision of CWP 48/18, Centre for Microdata Methods and Practice, U.C.L and I.F.S.

PARZEN, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Annals.of Mathematical Statistics*, 28, 329-348.

Politis, D.N., and Romano, J. (1992a). A Circular Block-Resampling Procedure for Stationary Data. In LePage, R., and Billard L. (Eds.), *Exploring the Limits of Bootstrap*, 263-270. New York: Wiley.

Politis, D.N., and Romano, J. (1992b). A General Resampling Scheme for Triangular Arrays of α-Mixing Random Variables with Application to the Problem of Spectral Density Estimation. *Annals of Statistics*, 20, 1985-2007.

Politis, D.N., and Romano, J. (1994). The Stationary Bootstrap. *Journal of the American Statististical Association*, 89, 1303-1313.

Politis, D.N., and Romano, J. (1995). Bias-Corrected Nonparametric Spectral Estimation. *Journal of Time Series Analysis*, 16, 68-103.

Politis, D.N., and White, H. (2004). Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, 23, 53-70.

Priestley, M.B. (1962). Basic Considerations in the Estimation of Spectra. *Technometrics*, 4, 551-5154,

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. New York: Academic Press.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. Second Edition. New York: Wiley.

Sacks, J., and Ylvisacker, D. (1981). Asymptotically Optimum Kernels for Density Estimation at a Point. *Annals of Statistics*, 9, 334-346.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.

Shi, X., and Shao, J. (1988). Resampling Estimation when the Observations are m-Dependent. *Communications in Statistics*, A, 17, 3923-3934.

Smith, R.J. (1997). Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation. *Economic Journal*, 107, 503-19.

Smith, R.J. (2005). Automatic Positive Semidefinite HAC Covariance Matrix and GMM Estimation. *Econometric Theory*, 21, 158-170.

Smith. R.J. (2011). GEL Criteria for Moment Condition Models. *Econometric Theory*, 27, 1192-1235.

White, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press: New York.

Table 1. Empirical Coverage Rates: Nominal 95% Confidence Intervals. Homoskedastic Innovations.

T	64					128					256				
ρ	0	0.2	0.5	0.7	0.9	0	0.2	0.5	0.7	0.9	0	0.2	0.5	0.7	0.9
KBB_{BT}	93.94	93.70	88.50	79.70	65.38	95.06	95.06	88.94	86.50	76.88	95.48	94.88	92.54	88.94	83.84
KBB^a_{BT}	94.06	93.68	88.76	79.84	65.96	95.02	94.98	88.76	86.60	77.30	95.50	94.80	92.74	89.06	84.06
KBB^b_{BT}	88.18	86.82	82.02	79.54	59.24	89.12	88.16	90.38	82.42	69.08	90.08	88.96	89.22	88.08	79.10
KBB^c_{BT}	93.18	92.72	86.74	75.64	58.16	94.62	94.48	87.82	84.38	70.40	95.42	94.62	92.14	87.56	79.76
KBB_{PZ}	92.80	92.64	89.84	85.50	73.02	93.88	93.90	91.72	87.90	81.86	95.00	94.20	92.24	90.22	86.50
KBB^a_{PZ}	92.82	92.56	90.32	85.46	73.36	93.86	93.82	91.68	88.02	81.80	95.00	94.14	92.36	90.36	86.72
KBB^b_{PZ}	93.40	92.60	88.20	79.70	60.42	94.82	94.76	91.06	85.74	71.28	96.58	95.26	92.56	88.86	80.74
KBB^c_{PZ}	92.18	91.32	87.04	79.46	60.98	93.62	93.22	90.32	85.16	72.22	94.90	94.12	91.70	88.82	81.32
KBB_{QS}	93.08	92.96	89.92	85.44	74.30	93.94	93.52	91.46	88.22	82.78	94.42	93.62	92.22	90.18	87.38
KBB^a_{QS}	93.20	92.76	90.10	85.62	74.40	93.98	93.52	91.54	88.24	82.64	94.44	93.62	92.32	90.34	87.52
KBB^b_{QS}	91.54	90.66	85.54	77.90	59.26	93.16	92.68	89.60	84.66	70.92	94.12	93.30	91.30	88.24	80.76
KBB^c_{QS}	91.72	90.84	85.92	78.48	59.94	93.26	92.74	89.76	84.94	71.80	94.18	93.38	91.44	88.48	81.24
KBB_{PP}	93.80	93.38	91.42	88.76	79.54	93.82	93.82	92.38	89.28	86.22	94.50	93.88	92.46	90.76	89.26
KBB^a_{PP}	93.66	93.32	91.46	88.86	80.02	93.90	93.82	92.44	89.56	86.50	94.52	93.96	92.50	90.86	89.32
KBB^b_{PP}	91.52	90.88	86.06	77.74	59.38	93.30	92.84	89.44	84.46	71.34	94.46	93.56	91.22	88.44	80.82
KBB^c_{PP}	91.70	90.92	86.58	79.32	60.96	93.08	92.66	90.18	84.80	72.54	94.08	93.54	91.38	88.72	81.38
MBB	93.08	92.72	89.42	85.40	73.90	93.90	93.38	91.52	87.62	81.06	94.54	93.76	91.94	89.68	85.80
MBB^a	93.50	92.66	90.08	85.96	74.76	94.08	93.56	91.50	88.10	82.00	94.62	93.84	91.98	89.66	86.40
MBB^b	91.26	90.24	84.92	77.08	56.48	93.26	92.32	89.06	83.44	68.28	94.38	93.42	91.08	87.68	78.50
MBB^c	91.26	90.34	85.08	77.16	57.58	93.28	92.24	89.00	83.66	69.00	94.42	93.44	91.10	87.62	79.02
TBB	91.70	91.56	89.36	86.70	78.82	92.78	92.86	91.40	88.70	85.00	93.60	93.40	92.12	90.40	87.98
TBB^a	92.24	92.00	89.94	87.92	80.24	93.00	92.98	91.60	89.56	86.14	93.68	93.34	92.30	90.72	89.36
TBB^b	90.46	90.20	85.16	78.00	58.06	92.86	92.52	89.70	84.32	70.26	93.86	93.44	91.46	88.76	80.32
TBB^c	90.18	89.56	84.68	77.92	59.92	92.28	91.88	89.40	84.42	71.94	93.36	92.92	91.42	88.42	80.88
TR	93.02	92.06	85.18	73.62	50.72	94.16	93.10	86.98	76.14	68.14	94.56	93.56	86.76	84.80	80.20
BT	92.88	91.82	85.98	78.20	58.82	93.86	93.14	89.44	84.12	70.38	94.48	93.54	91.12	87.82	80.00
PZ	90.54	90.38	86.36	78.72	60.52	92.62	92.78	90.02	85.44	71.38	94.18	93.56	91.96	88.82	80.80
TH	91.90	91.30	87.70	80.56	62.04	93.40	93.20	90.72	86.18	72.52	94.46	93.72	92.34	89.20	81.76
QS	92.44	91.68	87.50	79.80	61.20	93.72	93.20	90.38	85.38	72.36	94.60	93.58	91.90	88.74	81.62
S_{BT}	92.88	92.94	86.00	72.64	55.76	94.18	94.58	85.34	84.66	69.62	94.72	93.70	93.20	87.00	79.24
S_{PZ}	91.60	90.66	87.42	77.04	58.78	93.42	93.10	88.84	84.52	71.10	94.78	94.36	91.68	88.86	81.08
S_{QS}	91.18	90.68	85.62	77.86	59.32	93.16	92.78	89.78	84.56	71.04	94.06	93.40	91.48	88.26	81.02
S_{PP}	91.96	90.98	83.80	73.66	54.36	93.54	92.68	87.42	80.54	66.94	94.38	93.26	89.32	86.20	77.76

Table 2. Empirical Coverage Rates: Nominal 95% Confidence Intervals. Heteroskedastic Innovations.

T	64					128					256				
ρ	0	0.2	0.5	0.7	0.9	0	0.2	0.5	0.7	0.9	0	0.2	0.5	0.7	0.9
KBB_{BT}	91.78	91.10	88.60	80.74	63.50	93.36	92.94	88.94	84.26	73.72	94.78	94.16	90.20	88.04	80.66
KBB^a_{BT}	91.86	91.24	88.66	80.58	63.80	93.32	92.98	88.98	84.30	74.22	94.76	94.18	90.20	88.20	80.94
KBB^b_{BT}	85.66	83.90	77.22	73.94	57.88	88.76	87.46	85.06	81.04	64.88	90.38	89.08	88.68	86.32	75.14
KBB^c_{BT}	89.58	88.70	85.34	75.28	55.96	92.68	91.96	87.28	81.68	66.48	94.44	93.98	89.32	86.58	76.14
KBB_{PZ}	91.12	90.30	88.16	83.40	69.74	92.68	92.18	90.00	85.68	77.24	93.98	93.44	90.82	88.98	83.00
KBB^a_{PZ}	91.06	90.36	88.42	83.56	69.72	92.76	92.20	89.98	85.82	77.50	94.08	93.48	91.04	89.04	83.20
KBB^b_{PZ}	90.76	89.52	85.68	78.22	58.78	93.52	92.94	89.14	83.40	67.38	95.32	94.76	90.72	87.72	76.96
KBB^c_{PZ}	88.72	87.50	83.54	77.10	58.40	91.80	91.38	88.02	82.52	67.72	93.64	93.24	89.80	87.54	77.36
KBB_{QS}	91.40	90.82	88.68	83.88	70.50	92.70	92.20	89.68	85.66	78.36	93.94	93.38	90.66	88.92	83.86
KBB^a_{QS}	91.48	90.76	88.72	83.68	70.44	92.70	92.30	89.62	85.92	78.36	93.86	93.38	90.74	89.00	84.08
KBB^b_{QS}	88.06	86.80	82.60	75.18	56.58	91.16	90.94	87.20	81.86	66.68	93.24	92.80	89.12	86.92	76.80
KBB^c_{QS}	88.32	86.94	82.98	75.62	57.30	91.24	91.10	87.24	82.12	67.28	93.32	92.84	89.18	87.04	77.24
KBB_{PP}	91.22	90.78	88.68	84.72	72.96	92.38	92.04	89.94	86.14	79.30	93.62	93.26	90.84	88.82	83.88
KBB^a_{PP}	91.24	90.74	88.86	84.82	73.10	92.28	91.96	89.88	86.24	79.40	93.64	93.26	90.80	88.88	83.98
KBB^b_{PP}	87.90	86.80	82.42	74.64	56.22	91.22	90.84	86.94	81.46	66.64	93.48	92.70	88.90	86.56	76.54
KBB^c_{PP}	87.94	86.80	82.54	75.96	57.82	91.14	90.74	87.20	81.88	67.24	93.20	92.48	89.40	86.78	77.16
MBB	91.34	90.54	88.06	83.30	69.28	92.56	92.10	89.66	85.40	76.02	93.86	93.24	90.36	88.52	81.48
MBB^a	91.52	90.72	88.52	83.50	70.12	92.56	92.02	89.52	85.38	76.64	93.88	93.20	90.54	88.60	82.28
MBB^b	87.44	86.36	81.74	74.04	54.78	91.34	90.68	86.30	80.90	64.62	93.26	92.50	88.82	86.14	73.96
MBB^c	87.36	86.30	81.90	74.18	55.30	91.18	90.68	86.34	81.14	65.00	93.20	92.44	88.80	86.22	74.18
TBB	89.56	88.66	86.30	82.78	72.44	91.02	90.58	88.84	85.72	78.12	92.68	92.14	90.24	88.64	82.16
TBB^a	89.50	88.92	87.60	83.72	73.66	91.12	90.98	88.94	86.04	79.28	92.74	92.20	90.40	88.96	83.42
TBB^b	86.52	85.62	81.46	74.96	56.02	90.92	90.22	86.54	81.58	66.02	92.68	92.14	89.18	86.92	75.70
TBB^c	85.74	84.68	80.68	74.44	56.58	89.94	89.32	86.06	81.50	66.66	92.16	91.44	89.12	86.52	75.64
TR	91.08	89.34	84.32	72.82	49.82	92.76	91.84	85.84	76.06	62.20	93.88	93.08	86.40	80.34	75.64
BT	90.76	89.32	85.06	76.56	57.08	92.62	91.92	87.70	82.60	67.64	93.96	93.12	89.74	87.26	76.92
PZ	88.54	87.00	84.58	78.30	59.82	91.82	91.14	88.80	84.40	69.20	93.38	92.98	90.52	88.36	78.56
TH	90.00	88.60	86.02	79.58	61.34	92.42	91.72	89.54	84.64	70.38	93.88	93.34	90.96	88.84	79.16
QS	90.44	88.98	86.06	78.82	59.90	92.36	91.78	88.98	83.84	69.52	93.98	93.38	90.54	88.26	78.84
S_{BT}	89.52	89.18	87.10	75.52	54.44	92.48	92.18	87.02	82.40	67.66	94.46	93.40	89.28	86.58	76.56
S_{PZ}	88.28	87.00	84.40	77.42	57.56	91.62	90.80	88.58	82.84	68.68	93.52	92.76	89.92	88.80	78.32
S_{QS}	88.86	87.64	83.88	76.62	57.72	91.80	91.06	88.18	83.08	68.46	93.70	93.10	90.08	87.90	78.26
S_{PP}	89.92	88.42	82.54	72.06	52.42	92.48	91.34	85.96	79.68	63.64	93.82	92.80	88.10	84.88	73.34