A Kernel Based Bootstrap Method for Dependent processes

Paulo M.D.C. Parente
Department of Economics
University of Exeter Business School

Richard J. Smith
c*e*mmap
U.C.L and I.F.S.
and
Faculty of Economics
University of Cambridge

First Draft: August 2009
This Draft: October 2011

Abstract

A novel bootstrap method for stationary strong mixing processes is proposed in this article. The method consists in transforming the original data in an appropriate way using a kernel and applying standard m out of n bootstrap for independent and identically distributed observations. We investigate the first order asymptotic properties of the method in the case of the mean of the process and prove that the bootstrap distribution is consistent. Additionally, we show how the method can be applied to mean regression and quasi-maximum likelihood and demonstrate the first-order asymptotic validity of the bootstrap approximation in this context.

JEL Classification: C14, C15, C22

Keywords: Bootstrap; heteroskedastic and autocorrelation consistent inference; sample mean; least squares linear regression; quasi-maximum likelihood estimation.

1 Introduction

This article introduces a new bootstrap method for stationary and weakly dependent data that allows inferences to be made on features of the population of interest. The proposed method is based on results in Smith (2004) where the main idea is to replace the original sample of observations by a transformed sample of the same size, whose components are weighted moving averages of the original data points; see also Kitamura and Stutzer (1997) and Smith (1997). These weights are formed *via* a kernel function. Smith (2004) shows that the sample mean of the transformed data is consistent for the true population mean. More importantly, the standard formula of the sample variance for `i.i.d.` data applied to the transformed sample delivers a consistent estimator for the variance of the standardized mean of the original sample although the transformed sample points are not themselves `i.i.d.`

The validity of the bootstrap procedure method suggested here hinges on this latter property. Indeed it applies the standard (m out of n) nonparametric bootstrap for `i.i.d.` data, originally proposed in Bickel and Freedman (1981), to the transformed data, ignoring the serial dependence in these data. The proof for asymptotic validity is an immediate consequence of the theorems and lemmata in Smith (2004). However, since the results of Smith (2004) are appropriate for mean zero random variates, we provide two bootstrap methods. The first requires the mean of the stochastic process to be zero; we also show that this method remains appropriate in the non-zero mean case if the kernel weights are obtained from a class of truncated kernels. The second method allows a non-zero mean and for any type of kernel. The parameters of the linear regression model and population counterpart of the quasi-maximum likelihood (QML) objective function satisfy moment restrictions. Therefore the first bootstrap method can be used to make inference in both of these settings. The proofs for bootstrap asymptotic validity in the QML case are based on general results in Gonçalves and White (2004) on resampling methods for extremum estimators.

Since its introduction in the landmark article Efron (1979), the bootstrap has become extremely popular in empirical work. Its attractiveness lies in its being a computer intensive method that allows testing hypotheses or constructing confidence intervals without the necessity of the derivation of possibly complicated formulae for, e.g., asymptotic variances. Additionally, under some additional suitable assumptions it may provide a better approximation to the asymptotic distribution of estimators and test statistics than that obtained from first order theory; see, e.g., Beran (1988) and section 3 in Horowitz (2001).

The bootstrap methods described in Efron (1979) are only relevant for random samples. Indeed, Singh (1981) showed that these methods are not asymptotically valid if there is some dependency in the data; e.g., under stationarity, the bootstrap distribution of the sample mean converges uniformly to a normal distribution with mean zero but variance equal to that of the first observation. Therefore, to overcome this problem, myriad variants of Efron's bootstrap method have been proposed in the literature under different assumptions on the dependency of the data; e.g., 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 [Buhlmann (1997)]. For details on these methods, see the monographs of Shao and Tu (1995) and Lahiri (2003). Our method constitutes an alternative to these.

Our bootstrap method under a particular transformation of the data bears some similarities with MBB, although they differ in one fundamental aspect explained in section 4. Given a sample of size T, overlapping blocks of h consecutive observations are constructed in MBB. Each bootstrap sample is then obtained by drawing $l = T \times h$ blocks of data. Bootstrap statistics are computed using the resulting sample. In this method the blocks of data are highly correlated since they overlap, although the distribution of h-consecutive observations is preserved in each block. Similarly, each observation comprising the new bootstrap sample is a particular linear combination of all of the original data and contains all the relevant information on data dependency required for inference on the mean of the original sample. The bootstrap method proposed here also has some important common features with the transformation-based bootstrap of Lahiri (2003); this bootstrap applies a data transformation that yields asymptotically independent observations and then the `i.i.d.` bootstrap to the transformed data. We also apply a transformation to the original sample but, in contradistinction to Lahiri (2003), it does not yield asymptotically independent observations. Indeed, this dependence is irrelevant for asymptotic validity.

Importantly, the results proved here for the new resampling scheme only justify the use of the bootstrap percentile method. Indeed, the current paper only proves that the bootstrap distribution of particular estimators converge uniformly to their asymptotic distribution. The results do not guarantee that the bootstrap variance estimator is consistent. Although this point have been stressed in the bootstrap literature by Gonçalves and White (2004) and Shao and Tu (1995) for other bootstrap schemes, it has generally been ignored in empirical work. Nevertheless the method proposed here also implies that the alternative variance matrix estimator based on the bootstrap percentile method introduced by Parente and Machado (2005) is valid.

The article is organized as follows. Section 2.1 introduces the new bootstrap method for the zero mean case with the method appropriate for the non-zero mean case described in section 2.2. In section 4 the differences between the methods presented here and MBB are highlighted and discussed. Sections 5 and 6 demonstrate how our method can be applied in the mean regression and maximum quasi-likelihood frameworks respectively. Finally section 7 concludes. Proofs of the results in the main text are provided in an appendix.

2 A Bootstrap Method

Consider a sample of T observations, $(X_1, ..., X_T)$, on the zero mean finite dimensional stationary and strong mixing stochastic process $\{X_t\}_{t=1}^{\infty}$.[1] Let $\bar{X} = \sum_{t=1}^{T} X_t/T$. Define

[1]The validy of MBB has been shown under less stringent conditions than those imposed here; see Goncalves and White (2002). For the sake of simplicity, though, we assume stationarity.

the transformed variables

$$Y_{tT}=\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})X_{t-s},(t=1,...,T),$$

where S_T is a bandwidth parameter and $k(\cdot)$ is a kernel function standardized such that $\int_{-\infty}^{\infty}k(v)dv=1$; cf. Smith (2004). The new variates Y_{tT}, $(t=1,...,T)$, are weighted moving averages of the original observations. Also let $\bar{Y}=\sum_{t=1}^{T}Y_{tT}/T$ and $k_2=\int_{-\infty}^{\infty}k(v)^2dv$.

Define an additional transformation of the data

$$Z_{tT}=\bar{Y}+(S_T/k_2)^{1/2}(Y_{tT}-\bar{Y}),(t=1,...,T), \tag{2.1}$$

and the corresponding sample mean $\bar{Z}=\sum_{t=1}^{T}Z_{tT}/T$. Consider applying the standard non-parametric bootstrap for `i.i.d.` data method to the transformed sample $(Z_{1T},...,Z_{TT})$. Denote the resultant bootstrap sample by $(Z^*_{1T},...,Z^*_{TT})$, where each bootstrap observation is drawn from $(Z_{1T},...,Z_{TT})$ with equal probability $1/T$, and bootstrap sample mean $\bar{Z}^*=\sum_{t=1}^{T}Z^*_{tT}/T$. Also define the bootstrap observations Y^*_{tT} corresponding to Z^*_{tT}, $(t=1,...,T)$, obtained by inversion of (2.1) and associated sample mean $\bar{Y}^*=\sum_{t=1}^{T}Y^*_{tT}/T$.

Remark 1 As will be seen later the transformation (2.1) is not actually required for bootstrap validity since the sample $(Y_{1T},...,Y_{TT})$ may be used directly. However, for expositional reasons we prefer to use the transformed data $(Z_{1T},...,Z_{TT})$.

For ease of exposition we confine discussion to consideration of a scalar stochastic process $\{X_t\}_{t=1}^{\infty}$.

2.1 Zero Mean Stochastic Processes

Under conditions to be stated below Lemmata A.1 and A.2 in Smith (2004) imply that

$$\bar{Z}\xrightarrow{p}0,$$

and

$$T^{1/2}(\bar{Z}/\sigma_\infty)\xrightarrow{d}N(0,1),$$

where $\sigma_\infty^2 = \lim_{T\to\infty} var[T^{1/2}\bar{X}]$. Additionally, it follows from these lemmata that

$$\hat{\sigma}_n^2 = T^{-1}\sum_{t=1}^{T}(Z_t - \bar{Z})^2 \xrightarrow{p} \sigma_\infty^2;$$

see the proof of Theorem 2.

Write $E^*[\cdot]$ as conditional expectation given the sample $(Z_{1T}, ..., Z_{TT})$. Now

$$\begin{aligned} E^*[\bar{Z}^*] &= E[T^{-1}\sum_{t=1}^{T} Z_{tT}^*|(Z_{tT})_{t=1}^T] \\ &= T^{-1}\sum_{t=1}^{T} E[Z_{tT}^*|(Z_{tT})_{t=1}^T] \\ &= T^{-1}\sum_{t=1}^{T} \bar{Z} = \bar{Z}. \end{aligned}$$

Hence, from Smith (2004, Lemma A.1),

$$E^*[\bar{Z}^*] \xrightarrow{p} 0.$$

Smith (2004) required the following assumptions to hold.

Assumption 2.1 *The finite dimensional stochastic process $\{X_t\}_{t=1}^\infty$ is stationary and strong mixing with mixing coefficients α of size $-3v/(v-1)$ for some $v > 1$.*

Remark 2 The mixing coefficient condition in Assumption 2.1 guarantees that $\sum_{j=1}^\infty j^2\alpha(j)^{(v-1)/v} < \infty$ is satisfied, see Andrews (1991, p.824), a condition required for the results in Smith (2004).

Assumption 2.2 **(a)** *$S_T \to \infty$, $S_T/T^{1/2} \to 0$;* **(b)** *$k(.) : \mathbb{R} \to [-k_{\max}, k_{\max}].k_{\max} < \infty, k(0) \neq 0, k_1 \neq 0$ and is continuous at zero at almost everywhere;* **(c)** *$\int_{-\infty}^{\infty} \bar{k}(x)dx < \infty$ where $\bar{k}(x) = I(x \geq 0)\sup_{y\geq x}|k(y)| + I(x < 0)\sup_{y\leq x}|k(y)|$;* **(d)** *$|K(\lambda)| \geq 0$ for all $\lambda \in \mathbb{R}$, where $K(\lambda) = (2\pi)^{-1}\int k(x)\exp(-ix\lambda)dx$.*

Assumption 2.3 **(a)** *$E[|X_t|^{4v}] < \Delta < \infty$;* **(b)** *$\sigma_\infty^2$ is finite.*

Remark 3 Assumptions 2.1 and 2.3 guarantee than the central limit theorem of Wooldridge and White holds, see White (1999, Theorem 5.20). Assumption 2.3(a) may be relaxed to allow more heterogeneity by using the results in Hansen (1992) or Davidson and de Jong (2000), although there will be a resultant trade-off between the rate of divergence of S_T and the existence of moments.

Similarly to Gonçalves and White (2004) $\mathcal{P}$ denotes the probability measure of the original time series and $\mathcal{P}^*$ that induced by the bootstrap method. For a bootstrap statistic θ_T^* we write $\theta_T^* \rightarrow 0$ prob-$\mathcal{P}^*$, prob-$\mathcal{P}$ if for any $\varepsilon > 0$ and any $\delta > 0$, $\lim_{T\rightarrow\infty} \mathcal{P}\{\mathcal{P}^*\{|\theta_T^*| > \varepsilon\} > \delta\} = 0$.

The following lemma details a consistency result for the resampling mean and is similar to MBB-Lemma A.3, p.265, of Fitzenberger (1997) on MBB and Lemma A.5, p.213, of Gonçalves and White (2004).

Lemma 1 *Suppose Assumptions 2.1-2.3 hold. If* $E[X_t] = 0$*, then*

$$\bar{Z}^* - \bar{Z} \rightarrow 0, \textit{ prob-}\mathcal{P}^*, \textit{ prob-}\mathcal{P}; \bar{Y}^* - \bar{Y} \rightarrow 0, \textit{ prob-}\mathcal{P}^*, \textit{ prob-}\mathcal{P}.$$

The next result states the asymptotic validity of the proposed bootstrap.

Theorem 2 *Under Assumptions 2.1-2.3, if* $E[X_t] = 0$*,*

$$\lim_{T\rightarrow\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{T^{1/2}\left(\bar{Z}^* - \bar{Z}\right) \leq x\} - \mathcal{P}\{T^{1/2}\bar{X} \leq x\}\right| \geq \varepsilon\right\} = 0.$$

Remark 4 Since, from (2.1), $Z_{tT}^* = \bar{Y} + (S_T/k_2)^{1/2}(Y_{tT}^* - \bar{Y})$, where Y_{tT}^* are the bootstrap observations Y corresponding to Z_{tT}^*, $(t = 1, ..., T)$, $\bar{Z}^* = \bar{Y} + (S_T/k_2)^{1/2}(\bar{Y}^* - \bar{Y})$. Also $\bar{Z} = \bar{Y}$. Therefore, $\bar{Z}^* - \bar{Z} = (S_T/k_2)^{1/2}(\bar{Y}^* - \bar{Y})$. Consequently Theorem 2 becomes

$$\lim_{T\rightarrow\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{T^{1/2}(S_T/k_2)^{1/2}(\bar{Y}^* - \bar{Y}) \leq x\} - \mathcal{P}\{T^{1/2}\bar{X} \leq x\}\right| \geq \varepsilon\right\} = 0,$$

i.e., the bootstrap observations could have equivalently been drawn directly from $(Y_{1T}, ..., Y_{TT})$.

Hence the rate of convergence of $\bar{Y}^*$ to $\bar{Y}$ is "too fast", i.e., $(TS_T)^{-1/2}$ rather than the standard rate $T^{-1/2}$ for the convergence of $\bar{Y}$ to 0. Generally m-estimators are $T^{1/2}$-consistent. Consequently, this result cannot be used to approximate the distribution of such estimators. Our solution is to apply what is commonly referred to as the m out of n bootstrap, see Bickel and Freedman (1981); i.e., draw bootstrap samples of size m from of an original sample of size n. In our circumstance we draw a bootstrap sample of size $m_T = Tk_2/(S_T)$ from the transformed sample of size T with a consequent redefinition of the bootstrap sample means $\bar{Y}^* = \sum_{T=1}^{m_T} Y_{tT}^*/m_T$ and $\bar{Z}^* = \sum_{T=1}^{m_T} Z_{tT}^*/m_T$. However, we impose stronger assumptions on the rate of divergence of the bandwidth and on the existence of moments which coincide with those made in Smith (2004).

Assumption 2.4 **(a)** *$m_T = Tk_2/S_T$, $S_T \to \infty$, $S_T = O(T^{1/2-\eta})$ for some $\eta \in (0, 1/2)$;* **(b)** *$E[|X_t|^\alpha] < \Delta < \infty$, for some $\alpha > \max(4v, 1/\eta)$.*

Theorem 3 *Let Assumptions 2.1-2.3(b) and 2.4 be satisfied. If $E[X_t] = 0$, then*

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{m_T^{1/2}\left(\bar{Z}^* - \bar{Z}\right) \leq x\} - \mathcal{P}\{T^{1/2}\bar{X} \leq x\}\right| \geq \varepsilon\right\} = 0.$$

Remark 5 Theorem 2.2 is equivalent to

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{T^{1/2}(\bar{Y}^* - \bar{Y}) \leq x\} - \mathcal{P}\{T^{1/2}\bar{X} \leq x\}\right| \geq \varepsilon\right\} = 0.$$

Although a bootstrap sample of size m_T is now drawn, $\bar{Y}^*$ converges to $\bar{Y}$ at rate $T^{-1/2}$.

2.2 Non-Zero Mean Stochastic Processes

The method described in the previous section requires that $\{X_t\}_{t=1}^{\infty}$ is a zero mean stochastic process. Let $E[X_t] = \mu$ where now we allow for the possibility that μ is non-zero. In this case, the bootstrap proposed above may not work for general kernel functions. The source of difficulty is understood from the following Lemma.

Lemma 4 *Let the bootstrap sample be of size T. Under Assumptions 2.1-2.3,then*

$$Tvar^*[\bar{Z}^*] = \sigma_\infty^2 + B_T\mu^2 + o_p(1),$$

where $B_T = O(S_T)$ and $var^*[\cdot]$ denotes variance conditional on $(Z_{1T}, ..., Z_{TT})$.

Hence $Tvar^*[\bar{Z}^*]$ no longer converges to the correct variance σ_∞^2. In fact, it may be explosive if $\mu \neq 0$. The bias indicated in Lemma 4, however, vanishes for the class of kernels considered by Anatolyev (2005). This result is given by the following Lemma.

Lemma 5 *Let the assumptions of Lemma 4,be satisfied.If $k(\cdot) : [-b, b] \to [-k_{\max}, k_{\max}]$, $0 < k_{\max} < \infty$, $0 < b < \infty$, then $B_T = o(1)$.*

As noted by Anatolyev (2005), in addition to the truncated kernel, the Bartlett, Parzen and Tukey-Hanning kernels, see Andrews (1991), are also members of this class. Therefore, Lemma 5 demonstrates the bootstrap method of section 2.1 remains valid even if the mean of the process is non-zero if a member of this class of kernels is used.

The bootstrap weak law of large numbers of Lemma 1 still holds in the non-zero mean case for any kernel. Note, however, that this Lemma is very demanding in terms of existence of moments and dependency of the data. The following Lemma relaxes these assumptions but with an additional condition on the bootstrap sample size m_T.

Assumption 2.5 **(a)** *The finite dimensional stochastic process* $\{X_t\}_{t=1}^{\infty}$ *is stationary and strong mixing with mixing coefficients* α *of size* $-v/(v-1)$ *for some* $v > 1$; **(b)** $E[|X_t|^{v+\eta}] < \Delta < \infty$ *for some* $\eta > 0$; **(c)** $T^{1/(v+\eta)}/m_T = o(1)$.

Remark 6 Assumption 2.5(c) is satisfied if $m_T = T$.

Let $\bar{Y}^* = \sum_{T=1}^{m_T} Y_{tT}^*/m_T$.

Lemma 6 *Suppose that the bootstrap sample is of size* m_T. *Then, under Assumptions 2.2-2.5 and if* $m_T \to \infty$ *as* $T \to \infty$,

$$\bar{Y}^* - \bar{Y} \to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P}.$$

Remark 7 Note that m_T need not equal Tk_2/S_T.

Even if $\mu \neq 0$ it is possible to modify the method introduced in section 2.1 to provide the correct variance for any kernel. The transformation (2.1) is altered to

$$W_{tT} = \bar{X} + (k_2 S_T)^{-1/2} \sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})(X_{t-s} - \bar{X}), (t = 1, ..., T). \qquad (2.1)$$

The standard non-parametric bootstrap for `i.i.d.` data method is applied to $(W_{1T}, ..., W_{TT})$ from (2.1). Denote the bootstrap sample by $(W_{1T}^*, ..., W_{TT}^*)$, with each bootstrap observation drawn from $(W_{1T}, ..., W_{TT})$ with equal probability $1/T$. Write $\bar{W} = \sum_{t=1}^{T} W_{tT}/T$ and $\bar{W}^* = \sum_{t=1}^{T} W_{tT}^*/T$.

The validity of the resultant bootstrap method is stated in the following theorem.

Theorem 7 *If the bootstrap sample is of size T and Assumptions 2.1-2.3 are satisfied, then*

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{T^{1/2}\left(\bar{W}^* - \bar{W}\right) \le x\} - \mathcal{P}\{T^{1/2}(\bar{X}-\mu) \le x\}\right| \ge \varepsilon\right\} = 0.$$

Remark 8 This method has some similarities to the external bootstrap method of Shi and Shao (1988), although rather than using kernels their transformed observations depend on means of non-overlapping blocks. Additionally, they use a procedure similar to the Wild bootstrap rather than applying the standard non-parametric bootstrap to the transformed data.

3 Some Comparisons

3.1 MBB

If a truncated kernel is used, the method proposed in section 2.1 may be related to MBB. However, they do differ in one respect. To see this, following Kitamura and Stutzer (1997) and Example 2.1 of Smith (2004) and defining $S_T = (2q_T + 1)/2$,

$$Y_{tT} = \frac{1}{2q_T+1} \sum\nolimits_{s=\max[t-T,-q_T]}^{\min[t-1,q_T]} X_{t-s}, (t = 1, ..., T).$$

Suppose $q_T = 2$ and $T = 10$. Then $Y_{1T} = \sum_{t=1}^{3} X_t/5$, $Y_{2T} = \sum_{t=1}^{4} X_t/5$, $Y_{3T} = \sum_{t=1}^{5} X_t/5$, ..., $Y_{10T} = \sum_{t=8}^{10} X_t/5$. Hence, each transformed observation at the beginning or end of the sample depends on sums of a smaller number of terms than those in the middle of the sample. In contradistinction, for MBB, cf. the empirical likelihood estimator for dependent data discussed in Kitamura (1997), each block has the same size. Therefore, MBB differs from our bootstrap using a truncated kernel only in those data points at the beginning and end of the bootstrap sample, e.g., for the above example, Y_{1T}, Y_{2T}, Y_{9T} and Y_{10T} would be ignored.

Lemma 5 shows that the bias term in the variance $Tvar^*[\bar{Z}^*]$ vanishes, i.e., $B_T = o(1)$, for the kernel class defined there and thus provides an alternative justification for the use of MBB for the non-zero mean case.

3.2 HAC Estimation

Politis and Romano (1994) and Fitzenberger (1997) remark that MBB and the stationary bootstrap variance estimators are approximate equivalently to the Bartlett kernel variance estimator proposed by Newey and West (1987). The discussion of Smith (2004), section 2.6, implies that a similar conclusion holds for our bootstrap method too when the truncated kernel is used. Additionally, the results presented there also unveil that if the transformation is based on the Bartlett kernel, the bootstrap variance estimator obtained will be equivalent to the Parzen kernel variance estimator of Gallant (1987). The results in Smith (2004) also indicate that if the kernel used in the transformation is

$$k(x) = (\frac{5\pi}{8})^{1/2}\frac{1}{x}J_1(\frac{6\pi x}{5}), \tag{3.1}$$

where the Bessel function $J_1(\cdot)$ is given by

$$J_1(z) = \frac{z}{2}\sum\nolimits_{k=0}^{\infty}(-1)^k\frac{z^{2k}}{2^{2k}\Gamma(k+1)\Gamma(k+2)},$$

then a bootstrap variance estimator equivalent to the quadratic spectral variance estimator of Andrews (1991) is obtained. Andrews (1991) proved that this estimator is the best in terms of asymptotic mean square error in the class of kernels that satisfy Assumption 2.2. Hence it is not too unreasonable to conjecture a bootstrap method based on the kernel (3.1), might inherit similar desirable properties.

4 Mean Regression

The above results are useful for inference in the mean regression model. Consider the linear regression model

$$y_t = x_t'\beta_0 + \varepsilon_t,$$

where x_t is a random k-vector, $(t = 1, ..., T)$. The least squares (LS) estimator is defined as

$$\hat{\beta} = (\sum\nolimits_{t=1}^{T} x_t x_t'/T)^{-1}\sum\nolimits_{t=1}^{T} x_t y_t/T.$$

To introduce our bootstrap method define the following function

$$g_{tT}(b) = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})(y_{t-s} - x'_{t-s}b)^2$$

and consider the transformed sample $g_{tT}(b)$, $(t = 1, ..., T)$. Now draw a random sample of observations of size m_T from $g_{tT}(b)$, $(t = 1, ..., T)$, to obtain the bootstrap sample $g^*_{tT}(b)$, $(t = 1, ..., m_T)$.The bootstrap estimator is then defined by

$$\hat{\beta}^* = \arg\min_{b\in\mathcal{B}} \sum_{t=1}^{m_T} g^*_{tT}(b)/m_T,$$

where $\mathcal{B}$ is the parameter space.

To provide some intuition for the estimator $\hat{\beta}^*$, an alternative and equivalent manner for its definition may be given. To do so re-write the LS objective function in terms of

$$\begin{aligned} g_{tT}(b) \;\; = \;\; & (\frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})y^2_{t-s} + b'\frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})x_{t-s}x'_{t-s}b \\ & -2b'\frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})x_{t-s}y_{t-s}), \end{aligned}$$

$(t = 1, ..., T)$. Define

$$z^a_{tT} = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})x_{t-s}y_{t-s},\; z^b_{tT} = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})x_{t-s}x'_{t-s},$$

and construct the sample of pairs (z^a_{tT}, z^b_{tT}), $(t = 1, ..., T)$. Now draw a random sample of size m_T from this sample to obtain $(z^{a*}_{tT}, z^{b*}_{tT})$, $(t = 1, ..., T)$. The bootstrap estimator $\hat{\beta}^*$ is given by

$$\hat{\beta}^* = (\sum_{t=1}^{m_T} z^{b*}_{tT}/m_T)^{-1}\sum_{t=1}^{m_T} z^{a*}_{tT}/m_T.$$

The asymptotic properties of the bootstrap estimator can be studied using the general theorems of Gonçalves and White (2004), although for the sake of clarity we provide direct proofs.

To prove consistency of the bootstrap estimator we require the following assumptions.

Assumption 4.1 **(a)** *The finite dimensional stochastic process* $\{x'_t, \varepsilon_t\}_{t=1}^{\infty}$ *is stationary and strong mixing with mixing coefficients* α *of size* $-v/(v-1)$ *for some* $v > 1$; **(b)** $E[x_t\varepsilon_t] = 0$; **(c)** $E[\|x_t\varepsilon_t\|^{v+\gamma}] < \Delta < \infty$ *for some* $\gamma > 0$; **(d)** $E[\|x_t\|^{2v+\zeta}] < \Delta < \infty$ *for some* $\zeta > 0$; **(e)** $E[x_t x'_t]$ *is finite and positive definite;* **(e)** $T^{\psi}/m_T = o(1)$ *where* $\psi = \max\{1/(v+\eta), 1/(2v+\zeta)\}$.

Theorem 8 *Under Assumptions 2.2 and 4.1, if* $m_T \to \infty$ *and* $T \to \infty$*, then* **(a)** $\hat{\beta} \xrightarrow{p} \beta_0$; **(b)** $\hat{\beta}^* - \beta_0 \to 0$*, prob-*$\mathcal{P}^*$*, prob-*$\mathcal{P}$*.*

To show that the bootstrap distribution is close uniformly to its asymptotic counterpart we require the following additional conditions.

Assumption 4.2 **(a)** *The finite dimensional stochastic process* $\{x_t', \varepsilon_t\}_{t=1}^{\infty}$ *is stationary and strong mixing with mixing coefficients* α *of size* $-3v/(v-1)$ *for some* $v > 1$; **(b)** $m_T = Tk_2/(S_T)$, $S_T \to \infty$, $S_T = O(T^{1/2-\eta})$; **(c)** $J = \lim_{n\to\infty} Var((1/\sqrt{T})\sum_{t=1}^{T} x_t\varepsilon_t)$ *is positive definite;* **(d)** $E[\|x_t\varepsilon_t\|^{\alpha}] < \Delta < \infty$ *for some* $\alpha > \max(4v, 1/\eta)$.

Theorem 9 *Under Assumptions 2.2, 4.1 and 4.2,*

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{T^{1/2}(\hat{\beta}^* - \hat{\beta}) \leq x\} - \mathcal{P}\{T^{1/2}(\hat{\beta} - \beta_0) \leq x\}\right| \geq \varepsilon\right\} = 0.$$

Remark 9 Here Assumption 4.1(e) becomes $T^{\psi^*}/m_T = o(1)$ with $\psi^* = 1/(2v+\zeta)$. This condition is automatically satisfied under the remaining assumptions as $T^{\psi^*}/m_T = O(T^{-1/2-\eta+\psi^*}) = o(1)$ since $\psi^* < 1/2$.

5 Quasi-Maximum Likelihood

In this section we show that under some regularity conditions our bootstrap method may be used to test hypotheses and to construct confidence intervals in a quasi-maximum likelihood (QML) setting. The proofs of the results basically rely on verifying that the conditions of several general lemmata proven by Gonçalves and White (2004) are satisfied. Indeed, notice that although the paper of Gonçalves and White (2004) focus on MBB their results also apply to other bootstrap schemes.

Let us first describe the set-up. The QML estimator $\hat{\theta}$ is the optimiser

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \mathcal{L}_T(\theta), (T = 1, 2, ...),$$

where $\mathcal{L}_T(\theta) = \sum_{t=1}^{T} \log f(x_t, \theta)/T$, x_t is a vector of observations at time t taken from the stationary stochastic process $\{x_t\}_{t=1}^{\infty}$, $(t = 1, ..., T)$, $\theta \in \Theta$ with the parameter space

Θ a compact subset of $\mathbb{R}^p$. Denote

$$\theta_0 = \arg\max_{\theta\in\Theta} E[\log f(x_t, \theta)],$$

which we assume to be finite for simplicity.

Henceforth for any function $g : \Omega \times \Theta \to \mathbb{R}$ we denote $\bigtriangledown g(.,\theta) = \partial g(.,\theta)/\partial\theta$ and $\bigtriangledown^2 g(.,\theta) = \partial^2 g(.,\theta)/\partial\theta\partial\theta'$.

Define $A_0 = E[\partial^2 \log f(x_t, \theta_0)/\partial\theta\partial\theta']$ and $B_0 = \sum_{s=-\infty}^{\infty} \Gamma(s)$, where

$$\Gamma(s) = E[(\partial \log f(x_{t+s}, \theta_0)/\partial\theta)(\partial \log(f(x_t, \theta_0)/\theta')];$$

note that $\Gamma(s) = \Gamma(-s)'$. Under certain regulatory assumptions to be stated below, it follows from Gallant and White (1988, Theorem 5.7) that

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

To describe our bootstrap method for QML denote

$$h_{tT}(\theta) = \frac{1}{S_T} \sum_{s=t-T}^{t-1} k(\frac{s}{S_T}) \log f(x_{t-s}, \theta), (t = 1, ..., T),$$

and consider the sample $h_{tT}(\theta)$, $(t = 1, ..., T)$. Draw a random sample of size m_T with replacement from $h_{tT}(\theta)$, $(t = 1, ..., T)$, to obtain the bootstrap sample $h^*_{tT}(\theta)$, $(t = 1, ..., m_T)$. The bootstrap estimator is then defined as

$$\hat{\theta}^* = \arg\max_{\theta\in\Theta} \sum_{t=1}^{m_T} h^*_{tT}(\theta)/m_T.$$

We invoke the following assumption to prove consistency for the bootstrap estimator.

Assumption 5.1 **(a)** $(\Omega, \mathcal{F}, P)$ *is a complete probability space;* **(b)** *the finite dimensional stochastic process* $x_t : \Omega \to \mathbb{R}^l$, $(t = 1, 2, ...)$, *is stationary and strong mixing with mixing coefficients* α *of size* $-v/(v-1)$ *for some* $v > 1$ *and is measurable for all* t*;* **(c)** $f : \mathbb{R}^l \times \Theta \to \mathbb{R}^+$ *is measurable for each* $\theta \in \Theta$, Θ *a compact subset of* $\mathbb{R}^p$, *and* $f(x_t, \cdot)$ *is continuous a.s.-*P, $(t = 1, 2, ...)$*;* **(d)** θ_0 *is the unique maximizer of* $E[\log f(x_t, \theta)]$*;* **(e)** $\log f(x_t, \theta)$ *is Lipschitz continuous on* Θ, *i.e.,* $|\log f(x_t, \theta) - \log f(x_t, \theta_0)| \leq L_t \|\theta - \theta_0\|$ *a.s.-*P *for all* $\theta, \theta^0 \in \Theta$, *where* $\sup_T E[\sum_{t=1}^T L_t/T^{-1}] < \infty$*;* **(f)** $E[\sup_{\theta\in\Theta} |\log f(x_t, \theta)|^{v+\gamma}]$ $< \Delta < \infty$ *for some* $\gamma > 0$*;* **(g)** $T^{1/(v+\gamma)}/m_T = o(1)$.

Theorem 10 *Let Assumptions 2.2 and 5.1 hold. Then, if* $m_T \to \infty$ *and* $T \to \infty$, *(a)* $\hat{\theta} \xrightarrow{p} \theta_0$; **(b)** $\hat{\theta}^* - \hat{\theta}_0 \to 0$, *prob-*$\mathcal{P}^*$, *prob-*$\mathcal{P}$.

For consistency of the bootstrap distribution we make use of the following additional assumption.

Assumption 5.2 **(a)** *The finite dimensional stochastic process* $\{x_t\}_{t=1}^{\infty}$ *is stationary and strong mixing with mixing coefficients* α *of size* $-3v/(v-1)$ *for some* $v > 1$; **(b)** $m_T = Tk_2/(S_T)$, $S_T \to \infty$, $S_T = O(T^{1/2-\eta})$ *for some* $\eta \in (0, 1/2)$ *and* $1/2 + \eta > \psi$ *where* $\psi \equiv \max\{1/(v+\gamma), 1/(v+\xi)\}$; **(c)** $\partial^2 \log f(x_t, \theta)/\partial\theta\partial\theta'$ *is Lipschitz continuous on* Θ; **(d)** $E[\sup_{\theta\in\Theta} \|\partial \log f(x_t, \theta)/\partial\theta\|^{\alpha}] < \Delta < \infty$ *for some* $\alpha > \max(4v, 1/\eta)$, $E[\sup_{\theta\in\Theta} \|\partial^2 \log f(x_t, \theta)/\partial\theta\partial\theta'\|^{v+\xi}] < \Delta < \infty$ *for some* $\xi > 0$; **(e)** A^0 *is non-singular and* B^0 *is positive definite..*

Theorem 11 *Under Assumptions 2.2, 5.1 and 5.2,*

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{T^{1/2}(\hat{\theta}^* - \hat{\theta}) \le x\} - \mathcal{P}\{T^{1/2}(\hat{\theta} - \theta_0) \le x\}\right| \ge \varepsilon\right\} = 0.$$

Remark 10 Assumptions 5.2(b) and (d) indicate there is a trade off between existence of moments and the rate of divergence of S_T. To interpret the conditions suppose that v is slightly above 1. If η is close to zero, we would require higher moments to exist. For instance if η were 0.1 we would require $E[\sup_{\theta\in\Theta} \|\partial \log f(x_t, \theta)/\partial\theta\|^{10}]$ to be finite. Additionally, as $\psi < 3/5$, we would also need $E[\sup_{\theta\in\Theta} \|\partial^2 \log f(x_t, \theta)/\partial\theta\partial\theta'\|^{5/3+\delta}]$ and $E[\sup_{\theta\in\Theta} |\log f(x_t, \theta)|^{5/3+\delta}]$ to be finite for some $\delta > 0$. On the other hand, if η were close to 1/2, say 0.4, then $E[\sup_{\theta\in\Theta} \|\partial \log f(x_t, \theta)/\partial\theta\|^{4+\delta}]$ would need to be finite. Since $\psi < 9/10$, we would then require $E[\sup_{\theta\in\Theta} \|\partial^2 \log f(x_t, \theta)/\partial\theta\partial\theta'\|^{10/9+\delta}] < \infty$ and $E[\sup_{\theta\in\Theta} |\log f(x_t, \theta)|^{10/9+\delta}] < \infty$.

6 Conclusion

In this article we introduce a new bootstrap method for weakly dependent processes that requires two steps. First, we transform the original data using a kernel function. In the second step we apply the m out of n bootstrap to the transformed data. In the case of the sample mean, we prove that its asymptotic distribution is uniformly close to the bootstrap distribution of the proposed method.

The new method encompasses a variant of the well-known MBB method that consists in drawing both complete and incomplete blocks of consecutive observations. Drawing also these incomplete blocks will not affect in the asymptotic results.

We propose two versions of the new method, one that applies to zero mean stochastic processes and a second that allows this mean to be different from zero. Additionally, we show that, provided that a truncated kernel is used, the first method is also valid for processes with a non-zero mean.

These result allow us to show how the method can be applied to mean regression and quasi-maximum likelihood estimation in order to make inferences on the parameters of interest. We prove the first-order asymptotic validity of the new bootstrap method in these cases.

Appendix: Proofs of Results

Throughout the Appendix, 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. A similar notation is adopted to that in Gonçalves and White (2004). For a bootstrap statistic $W_T^*(\cdot, \omega)$ we write $W_T^*(\cdot, \omega) \to 0$ prob-$\mathcal{P}^*$, prob-$\mathcal{P}$ if, for any $\varepsilon > 0$ and any $\delta > 0$, $\lim_{T\to\infty} \mathcal{P}\{\omega : \mathcal{P}^*\{\lambda : |W_T^*(\lambda, \omega)| > \varepsilon\} > \delta\} = 0$. For ease of exposition we deal with the scalar case.

Proof of Lemma 1: The proof is similar to that of Lemma A.5 of Gonçalves and White (2004).

First

$$\begin{aligned} E^*[\bar{Z}^*] &= \bar{Z} \\ &= \bar{Y} = o_p(1) \end{aligned}$$

by Lemma A.1 of Smith (2004). Thus by C

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{Z}^* - \bar{Z}\right| &> \varepsilon\} \leq var^*[\bar{Z}^*] \\ &= \frac{S_T}{T^2}\sum\nolimits_{t=1}^{T}(Y_{tT} - \bar{Y})^2 = O_p(T^{-1})\end{aligned}$$

by Lemma A.3 of Smith (2004). It therefore follows from M and the Lebesgue Theorem that

$$\mathcal{P}\{\mathcal{P}^*\{\left|\bar{Z}^* - \bar{Z}\right| > \varepsilon\} > \delta\} = O(T^{-1}).$$

Secondly, similarly

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{Y}^* - \bar{Y}\right| &> \varepsilon\} \leq var^*[\bar{Y}^*] \\ &= T^{-2}\sum\nolimits_{t=1}^{T}(Y_{tT} - \bar{Y})^2 \\ &= O_p((S_T T)^{-1})\end{aligned}$$

also by Lemma A.3 of Smith (2004). The result then follows similarly to above. ■

Proof of Theorem 2: The result is proven if we are able to show the following steps; cf. Politis and Romano (1992b, Proof of Theorem 2). Step 1: $\bar{X} \xrightarrow{p} 0$. Step 2: $T^{1/2}\bar{X}/\sigma_\infty \xrightarrow{d} N(0,1)$. Step 3: $\sup_{x\in\mathbb{R}}\left|\mathcal{P}\{T^{1/2}\bar{X} \leq x\} - \Phi(x/\sigma_\infty)\right| \to 0$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Step 4: $Tvar^*[\bar{Z}^*] \xrightarrow{p} \sigma^2_\infty$. Step 5:

$$\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{\frac{\bar{Z}^* - E^*[\bar{Z}^*]}{var^*[\bar{Z}^*]^{1/2}} \leq x\} - \Phi(x)\right| \geq \varepsilon\right\} = 0.$$

Step 1: Follows from Theorem 3.47 of White (1999).

Step 2: By White (1999, Theorem 5.20).

Step 3: From Step 2 and the Polya Theorem, Serfling (2002, p.18), as $\Phi(\cdot)$ is a continuous c.d.f.

Step 4: Now $E^*[\bar{Z}^*] = \bar{Z}$. Thus

$$\begin{aligned}Tvar^*[\bar{Z}^*] &= T^{-1}\sum\nolimits_{t=1}^{T}(Z_t - \bar{Z})^2 \\ &= \frac{S_T}{T}\sum\nolimits_{t=1}^{T}(Y_t - \bar{Y})^2 \\ &= \frac{S_T}{T}\sum\nolimits_{t=1}^{T}Y_t^2 - S_T\bar{Y}^2.\end{aligned}$$

By Lemma A.2 of Smith (2004) $S_T\bar{Y}^2 = O_p(S_T/T) = o_p(1)$ and by Lemma A.3 of Smith (2004) $(S_T/T)\sum_{t=1}^{T} Y_t^2 \xrightarrow{p} \sigma_\infty^2$. Thus the result follows.

STEP 5: Since the bootstrap sample observations are independent, we can apply Berry-Esséen inequality. Thus

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{T^{1/2}\frac{\bar{Z}^*-\bar{Z}}{var^*[Z_{1T}^*]^{1/2}} \le x\} - \Phi(x)\right| &\le \frac{C}{T^{1/2}}E^*[(\frac{\left|Z_{1T}^*-\bar{Z}\right|}{var^*[Z_{1T}^*]^{1/2}})^3] \\ &= \frac{C}{T^{1/2}}var^*[Z_{1T}^*]^{-3/2}E^*[\left|Z_{1T}^*-\bar{Z}\right|^3].\end{aligned}$$

Note that $var^*[Z_{1T}^*] = T^{-1}\sum_{t=1}^{T}\left(Z_{tT}-\bar{Z}\right)^2$ and $E^*[\left|Z_{1T}^*-\bar{Z}\right|^3] = T^{-1}\sum_{t=1}^{T}\left|Z_{tT}-\bar{Z}\right|^3$. Thus, cf. the Proof of Step 4, $var^*[Z_{1T}^*] \xrightarrow{p} \sigma_\infty^2 > 0$. Also

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left|Z_{tT}-\bar{Z}\right|^3 &\le \frac{1}{T}\sum_{i=1}^{T}\left|Z_{tT}-\bar{Z}\right|^2\max_t\left|Z_{tT}-\bar{Z}\right| \\ &= O_p(1)O_p(S_T^{1/2}T^{1/4v})\end{aligned}$$

since $\max_t\left|Z_{tT}-\bar{Z}\right| = O(S_T^{1/2})\max_t\left|Y_t-\bar{Y}\right|$ and $\max_t\left|Y_t-\bar{Y}\right| \le \max_t|Y_t-\mu| + \left|\bar{Y}-\mu\right| = O_p(T^{1/4v}) + o_p(1)$ by Lemma A.2 of Smith (2004) and M. Thus

$$\begin{aligned}\frac{C}{T^{1/2}}var^*[Z_{1T}^*]^{-3/2}E^*[\left|Z_{1T}^*-\bar{Z}\right|^3] &= T^{-1/2}O_p(S_T^{1/2}T^{1/4v}) \\ &= o(T^{-1/4})O_p(T^{1/4v}) = o_p(1)\end{aligned}$$

since $S_T = o(T^{1/2})$ and $v > 1$. ■

Proof of Theorem 3: The proof is similar to that for Theorem 2 above. The only step that changes is the proof of Step 5. Given the sample, the bootstrap observations are independent. Hence we can apply Berry-Esséen inequality. Thus

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{m_T^{1/2}\frac{\bar{Z}^*-\bar{Z}}{var^*[Z_{1T}^*]^{1/2}} \le x] - \Phi(x)\right| &\le \frac{C}{m_T^{1/2}}E^*[(\frac{\left|Z_{1T}^*-\bar{Z}\right|}{var^*[Z_{1T}^*]^{1/2})})^3] \\ &= \frac{C}{m_T^{1/2}}var^*[Z_{1T}^*]^{-3/2}E^*[\left|Z_{1T}^*-\bar{Z}\right|^3].\end{aligned}$$

As above $var^*[Z_{1T}^*] \xrightarrow{p} \sigma_\infty^2 > 0$. Similarly $E^*[(\left|Z_{1T}^*-\bar{Z}\right|^3] = O_p(S_T^{1/2}T^{1/\alpha})$ since $\max_t\left|Z_{tT}-\bar{Z}\right| = O(S_T^{1/2})\max_t\left|Y_t-\bar{Y}\right|$ and $\max_t\left|Y_t-\bar{Y}\right| \le O_p(T^{1/\alpha})$. Thus,

$$\begin{aligned}\frac{C}{m_T^{1/2}}var^*[Z_{1T}^*]^{-3/2}E^*[\left|Z_{1T}^*-\bar{Z}\right|^3] &= m_T^{-1/2}O_p(S_T^{1/2}T^{1/\alpha}) \\ &= O(S_T/T^{1/2})O_p(T^{1/\alpha}) \\ &= O_p(T^{1/\alpha-\eta}) = o_p(1)\end{aligned}$$

since $m_T = Tk_2/S_T$, $S_T = O(T^{1/2-\eta})$ and $\alpha > \max(4v, 1/\eta)$. ■

Proof of Lemma 4: First

$$Tvar^*[\bar{Z}^*] = \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(Y_{tT} - \bar{Y})^2.$$

Write

$$\mu_{tT} = \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})\mu, (t = 1, ..., T).$$

Thus

$$\begin{aligned} Tvar^*[\bar{Z}^*] &= \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(Y_{tT} - \mu_{tT} + \mu_{tT} - \bar{Y})^2 \\ &= \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(Y_{tT} - \mu_{tT})^2 + \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(\mu_{tT} - \bar{Y})^2 \\ &\quad +2\frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(Y_{tT} - \mu_{tT})(\mu_{tT} - \bar{Y}). \end{aligned}$$

Now, from Lemma A.3 of Smith (2004),

$$\frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(Y_{tT} - \mu_{tT})^2 = \sigma_\infty^2 + o_p(1).$$

Also

$$\frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(Y_{tT} - \mu_{tT})(\mu_{tT} - \bar{Y}) = O_p((S_T^2/T)^{1/2}) = o_p(1)$$

as $\bar{Y} - \mu_{tT} = O_p(1)$ uniformly t and $T^{-1}\sum\nolimits_{t=1}^{T}(Y_{tT} - \mu_{tT}) = O_p(T^{-1/2})$ by Lemma A.2 of Smith (2004).

It remains to study the behaviour of $(S_T/k_2 T)\sum\nolimits_{t=1}^{T}(\mu_{tT} - \bar{Y})^2$. Write $\bar{\mu} = \sum\nolimits_{t=1}^{T}\mu_{tT}/T$. Notice that

$$\begin{aligned} \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(\mu_{tT} - \bar{Y})^2 &= \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(\mu_{tT} - \bar{\mu} + \bar{\mu} - \bar{Y})^2 \\ &= \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(\mu_{tT} - \bar{\mu})^2 + \frac{S_T}{k_2}(\bar{\mu} - \bar{Y})^2 \\ &= B_T\mu^2 + o_p(1) \end{aligned}$$

since $\sum\nolimits_{t=1}^{T}(\mu_{tT} - \bar{\mu}) = 0$, $S_T(\bar{\mu} - \bar{Y})^2 = O_p(S_T/T)$ and

$$B_T \equiv \frac{S_T}{k_2 T}\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T}) - \frac{1}{T}\sum\nolimits_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T}))^2. \tag{A.1}$$

As $S_T^{-1}\sum_{s=t-T}^{t-1} k(s/S_T) = O(1)$ uniformly t it follows that $B_T = O(S_T)$. ■

Proof of Lemma 5: First, recall the expression for B_T in eq. (A.1). Secondly, since $k(\cdot)$ is a member of the class of truncated kernels B_T may be written as

$$B_T \equiv \frac{S_T}{k_2 T}\sum_{t=1}^{T}(\frac{1}{S_T}\sum_{s=\max[t-T,-r_T]}^{\min[t-1,r_T]} k(\frac{s}{S_T}))^2 - \frac{S_T}{k_2}(\frac{1}{TS_T}\sum_{t=1}^{T}\sum_{s=\max[t-T,-r_T]}^{\min[t-1,r_T]} k(\frac{s}{S_T}))^2$$

where $r_T = \lfloor S_T b \rfloor$.

Now

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=\max[t-T,-r_T]}^{\min[t-1,r_T]} k(\frac{s}{S_T}) &= \sum_{t=1}^{r_T}\sum_{s=-r_T}^{t-1,} k(\frac{s}{S_T}) \\ &+ \sum_{t=r_T+1}^{T-r_T}\sum_{s=-r_T}^{r_T} k(\frac{s}{S_T}) \\ &+ \sum_{t=T-r_T+1}^{T}\sum_{s=t-T}^{r_T} k(\frac{s}{S_T}).\end{aligned}$$

Denote $C_T = S_T^{-1}\sum_{s=-r_T}^{r_T} k(s/S_T)$. Since $C_T = O(1)$ by Smith (2004, eq. (A.9)),

$$\begin{aligned}\frac{1}{S_T}\sum_{s=-r_T}^{t-1,} k(\frac{s}{S_T}) &\leq C_T = O(1) \text{ for } t \leq r_t, \\ \frac{1}{S_T}\sum_{s=t-T}^{r_T} k(\frac{s}{S_T}) &\leq C_T = O(1) \text{ for } t \geq T - r_T + 1.\end{aligned} \tag{A.2}$$

It then follows that

$$\begin{aligned}\sum_{t=1}^{r_T}\sum_{s=-r_T}^{t-1,} k(\frac{s}{S_T}) &= O(S_T^2), \\ \sum_{t=r_T+1}^{T-r_T}\sum_{s=-r_T}^{r_T} k(\frac{s}{S_T}) &= C_T(T-2r_T)S_T, \\ \sum_{t=T-r_T+1}^{T}\sum_{s=t-T}^{r_T} k(\frac{s}{S_T}) &= O(S_T^2).\end{aligned}$$

Consequently

$$\begin{aligned}\frac{1}{TS_T}\sum_{t=1}^{T}\sum_{s=\max[t-T,-r_T]}^{\min[t-1,r_T]} k(\frac{s}{S_T}) &= \frac{T-2r_T}{T}C_T + O(\frac{S_T}{T}) \\ &= C_T + O(\frac{S_T}{T}).\end{aligned} \tag{A.3}$$

Consider now

$$\begin{aligned}\sum_{t=1}^{T}(\frac{1}{S_T}\sum_{s=\max[t-T,-r_T]}^{\min[t-1,r_T]} k(\frac{s}{S_T}))^2 &= \sum_{t=1}^{r_T}(\frac{1}{S_T}\sum_{s=-r_T}^{t-1,} k(\frac{s}{S_T}))^2 \\ &+ \sum_{t=r_T+1}^{T-r_T}(\frac{1}{S_T}\sum_{s=-r_T}^{r_T} k(\frac{s}{S_T}))^2 \\ &+ \sum_{t=T-r_T+1}^{T}(\frac{1}{S_T}\sum_{s=t-T}^{r_T} k(\frac{s}{S_T}))^2.\end{aligned}$$

From eq. (A.2)

$$\begin{aligned}
\sum_{t=1}^{r_T}(\frac{1}{S_T}\sum_{s=-r_T}^{t-1,} k(\frac{s}{S_T}))^2 &= O(S_T), \\
\sum_{t=r_T+1}^{T-r_T}(\frac{1}{S_T}\sum_{s=-r_T}^{r_T} k(\frac{s}{S_T}))^2 &= C_T^2(T-2r_T), \\
\sum_{t=T-r_T+1}^{T}(\frac{1}{S_T}\sum_{s=t-T}^{r_T} k(\frac{s}{S_T}))^2 &= O(S_T).
\end{aligned}$$

Thus

$$\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{S_T}\sum_{s=\max[t-T,-S_T]}^{\min[t-1,S_T]} k(\frac{s}{S_T}))^2 = C_T^2 + O(\frac{S_T}{T}) \tag{A.4}$$

and therefore, from eqs. (A.3) and (A.4),

$$B_T = \frac{S_T}{k_2}[C_T^2 + O(\frac{S_T}{T})] - \frac{S_T}{k_2}[C_T^2 + O(\frac{S_T}{T})] = O(\frac{S_T^2}{T}) = o(1)$$

since $S_T = o(T^{-1/2})$. ■

Before proving Lemma 6, we demonstrate the following auxiliary Lemma.

Lemma 12 *Under Assumptions 2.5 and 2.2,*

$$\frac{1}{T}\sum_{T=1}^{T} Y_{tT} \xrightarrow{p} E[X_t]$$

Proof: The proof follows that for Lemma A.1 of Smith (2004) except the UWL required there is replaced by Corollary 3.48 of White (1999). ■

Proof of Lemma 6: First notice that

$$\begin{aligned}
E^*[|Y_{tT}^*|] &= \frac{1}{T}\sum_{t=1}^{T}|Y_{tT}| = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})X_{t-s}\right| \\
&\leq O(1)\frac{1}{T}\sum_{t=1}^{T}|X_t| = O_p(1)
\end{aligned}$$

Thus $E^*[|Y_{tT}^*|] = O_p(1)$. In addition

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}| - \frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|\, I(|Y_{tT}| &< \delta m_T) = \frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|\, I(|Y_{tT}| \geq \delta m_T) \\
&\leq \frac{1}{T}\sum_{t=1}^{T}|Y_{tT}| \max_t I(|Y_{tT}| \geq \delta m_T)
\end{aligned}$$

Now by M

$$\max_t |Y_{tT}| = O(1)\max_t |X_t| = O_p(T^{1/(v+\eta)}).$$

Since $T^{1/(v+\eta)}/m_T = o(1)$ it follows that $\max_t I(|Y_{tT}| \geq \delta m_T) = o_p(1)$. Thus

$$\frac{1}{T}\sum\nolimits_{t=1}^{T} |Y_{tT}|\, I(|Y_{tT}| \geq \delta m_T) = o_p(1).$$

The remaining part of the proof is similar to the proof of Khinchine's weak law of large numbers given in Rao (2002). Define a pair of new random variables for each T, $(t = 1, ..., m_T)$,

$$\begin{aligned} W_{tT} &= Y_{tT}^*, Z_{tT} = 0 \text{ if } |Y_{tT}^*| < \delta m_T, \\ W_{tT} &= 0, Z_{tT} = Y_{tT}^* \text{ if } |Y_{tT}^*| \geq \delta m_T. \end{aligned}$$

Hence $Y_{tT}^* = W_{tT} + Z_{tT}$. Define

$$\begin{aligned} \mu_T &= E^*[W_{tT}] \\ &= \frac{1}{T}\sum\nolimits_{t=1}^{T} Y_{tT} I[|Y_{tT}| < \delta m_T]. \end{aligned}$$

Note that $E^*[Y_{tT}^*] = \bar{Y}$ and $\left|\bar{Y} - \mu_T\right| < \varepsilon$ for any $\varepsilon > 0$ and T large enough. The latter claim holds since by T

$$\begin{aligned} \left|\frac{1}{T}\sum\nolimits_{t=1}^{T} Y_{tT} I[|Y_{tT}| < \delta m_T] - \frac{1}{T}\sum\nolimits_{t=1}^{T} Y_{tT}\right| &\leq \frac{1}{T}\sum\nolimits_{t=1}^{T} |Y_{tT}|\, I(|Y_{tT}| \geq \delta m_T) \\ &= o_p(1). \end{aligned}$$

Now

$$var^*[W_{tT}^*] = E^*[W_{tT}^2] - \mu_T^2 \leq E^*[W_{tT}^2] \leq \delta m_T E^*[|W_{tT}|].$$

Thus, writing $\bar{W} = \sum\nolimits_{t=1}^{m_T} W_{tT}/m_T$, using C,

$$\begin{aligned} \mathcal{P}^*\{\left|\bar{W} - \mu_T\right| &\geq \varepsilon\} \leq \frac{var^*[W_{tT}^*]}{\varepsilon^2 m_T} \\ &\leq \frac{\delta E^*[|W_{tT}|]}{\varepsilon^2}. \end{aligned}$$

Hence, since $\left|\bar{Y} - \mu_T\right| < \varepsilon$ for any $\varepsilon > 0$ and T large enough,

$$\mathcal{P}^*\{\left|\bar{W} - \bar{Y}\right| \geq 2\varepsilon\} \leq \frac{\delta E^*[|W_{tT}|]}{\varepsilon^2}. \tag{A.5}$$

Now by M it follows that

$$\begin{aligned}\mathcal{P}^*\{Z_{tT} &\neq 0\} = \mathcal{P}^*\{|Y_{tT}^*| \geq \delta m_T\} \\ &\leq \frac{1}{\delta m_T} E^*[|Y_{tT}^*| \, I[|Y_{tT}^*| \geq \delta m_T]] \leq \frac{\delta}{m_T}.\end{aligned}$$

To see this, as $E^*[|Y_{tT}^*|] = O_p(1)$, it follows that $E^*[|Y_{tT}^*| \, I[|Y_{tT}^*| \geq \delta m_T]] = o_p(1)$. Thus, we can always choose a constant δ^2 such that for T large enough $E^*[|Y_{tT}^*| \, I[|Y_{tT}^*| \geq \delta m_T]] \leq \delta^2$ w.p.a.1. Write $\bar{Z} = \sum_{t=1}^{m_T} Z_{tT}/m_T$. In addition

$$\begin{aligned}\mathcal{P}^*\{\bar{Z} &\neq 0\} \leq \mathcal{P}^*\{\max_t Z_{tT} \neq 0\} \\ &\leq \sum_{t=1}^{m_T} \mathcal{P}^*\{Z_{tT} \neq 0\} \leq \delta.\end{aligned} \tag{A.6}$$

Write $\bar{Y} = \sum_{t=1}^{T} Y_{tT}/m_T$ and $\bar{Y}^* = \sum_{t=1}^{m_T} Y_{tT}^*/m_T$. Therefore, from eqs. (A.5) and (A.6)

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{Y}^* - \bar{Y}\right| &\geq 4\varepsilon\} = \mathcal{P}^*\{\left|\bar{W} - \bar{Y} + \bar{Z}\right| \geq 4\varepsilon\} \\ &\leq \mathcal{P}^*\{\left|\bar{W} - \bar{Y}\right| + \left|\bar{Z}\right| \geq 4\varepsilon\} \\ &\leq \mathcal{P}^*\{\left|\bar{W} - \bar{Y}\right| \geq 2\varepsilon\} + \mathcal{P}^*\{\left|\bar{Z}\right| \geq 2\varepsilon\} \\ &\leq \frac{\delta E^*[|W_{tT}|]}{\varepsilon^2} + \mathcal{P}^*\{\left|\bar{Z}\right| \neq 0\} = \frac{\delta E^*[|W_{tT}|]}{\varepsilon^2} + \delta.\end{aligned}$$

Now choose δ small enough. As $E^*[|W_{tT}|] \leq E^*[|Y_{tT}^*|] = O_p(1)$, the result follows from M. ■

Proof of Theorem 7: Given the subsidiary results already shown in the proof of Theorem 2 this result follows if we are able to show the following steps. Step 1: $Tvar^*[\bar{W}^*] \xrightarrow{p} \sigma_\infty^2$. Step 2:

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}^k}\left|\mathcal{P}^*\{T^{1/2}\frac{\left(\bar{W}^* - E^*(\bar{W}^*)\right)}{var^*[\bar{W}^*]^{1/2}} \leq x\} - \Phi(x)\right| \geq \varepsilon\right\} = 0.$$

STEP 1: Notice that

$$\begin{aligned}E^*[\bar{W}^*] &= \bar{W} \\ &= \bar{X} + \frac{1}{T}\sum_{t=1}^{T}(k_2 S_T)^{-1/2}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})(X_{t-s} - \bar{X}).\end{aligned}$$

Now

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\bar{X}) \tag{A.7}$$

$$\begin{aligned} &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})[(X_{t-s}-\mu)-(\bar{X}-\mu)] \\ &= O_p(T^{-1/2}) \end{aligned}$$

by Lemma A.2 of Smith (2004), $T^{-1}\sum_{t=1}^{T}S_T^{-1}\sum\nolimits_{s=t-T}^{t-1}k(s/S_T) = O(1)$ and Theorem 5.20 of White (1999). Hence, from (A.7),

$$\begin{aligned} Tvar^*[\bar{W}^*] &= \frac{1}{T}\sum\nolimits_{t=1}^{T}(W_{tT}-\bar{W})^2 \\ &= \frac{S_T}{Tk_2}\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\bar{X}))^2 + O_p(\frac{S_T}{T}) \\ &= \frac{S_T}{Tk_2}\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\bar{X}))^2 + o_p(1). \end{aligned}$$

Also

$$\begin{aligned} &\frac{S_T}{Tk_2}\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\bar{X}))^2 \\ =\ &\frac{S_T}{Tk_2}\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})[(X_{t-s}-\mu)-(\bar{X}-\mu)])^2 \\ =\ &\frac{S_T}{Tk_2}\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\mu))^2 \\ &-\frac{2S_T}{Tk_2}(\bar{X}-\mu)\sum\nolimits_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\mu)\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T}) \\ &+\frac{S_T}{Tk_2}(\bar{X}-\mu)^2\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T}))^2 \\ =\ &\sigma_\infty^2 + o_p(1) \\ &-\frac{2S_T}{Tk_2}O(1)(\bar{X}-\mu)\sum\nolimits_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\mu) \\ &+\frac{S_T}{Tk_2}(\bar{X}-\mu)^2\sum\nolimits_{t=1}^{T}(\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T}))^2 \\ =\ &\sigma_\infty^2 + O_p(\frac{S_T}{T}) + o_p(1) = \sigma_\infty^2 + o_p(1) \end{aligned}$$

by Lemma A.2 of Smith (2004), $S_T^{-1}\sum\nolimits_{s=t-T}^{t-1}k(s/S_T) = O(1)$ uniformly t and Theorem 5.20 of White (1999).

STEP 2: The proof follows systematically that for Step 5 in Theorem 2. The bootstrap sample observations are independent. Hence, we again apply the Berry-Esséen inequality. Thus

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{T^{1/2}\frac{\bar{W}^*-W}{var^*[W^*_{1T}]^{1/2}}\le x\}-\Phi(x)\right| &\le \frac{C}{T^{1/2}}E^*[(\frac{\left|W^*_{1T}-\bar{W}\right|}{var^*[W^*_{1T}]^{1/2}})^3] \\ &= \frac{C}{T^{1/2}}var^*[W^*_{1T}]^{-3/2}E^*[\left|W^*_{1T}-\bar{W}\right|^3]\end{aligned}$$

From Step 1

$$var^*[W^*_{1T}]=\frac{1}{T}\sum\nolimits_{t=1}^{T}\left(W_{tT}-\bar{W}\right)^2\xrightarrow{p}\sigma^2_\infty>0.$$

Also

$$\begin{aligned}E^*[\left|W^*_{1T}-\bar{W}\right|^3] &= \frac{1}{T}\sum\nolimits_{t=1}^{T}\left|W_{tT}-\bar{W}\right|^3 \\ &\le \frac{1}{T}\sum\nolimits_{t=1}^{T}\left|W_{tT}-\bar{W}\right|^2\max_t\left|W_{tT}-\bar{W}\right| \\ &= O_p(1)\max_t\left|W_{tT}-\bar{W}\right|.\end{aligned}$$

By T and M, cf. eq. (A.7),

$$\begin{aligned}\max_t\left|W_t-\bar{W}\right| &= O(S_T^{1/2})\max_t\left|\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(X_{t-s}-\bar{X})+O_p(T^{-1/2})\right| \\ &\le O(S_T^{1/2})\max_t|X_t-\mu|+\left|\bar{X}-\mu\right|++O_p((S_T/T)^{1/2})=O_p(T^{1/4v})+o_p(1).\end{aligned}$$

Thus

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{T^{1/2}\frac{\bar{W}^*-W}{var^*[W^*_{1T}]^{1/2}}\le x\}-\Phi(x)\right| &\le O_p((S_T/T)^{1/2})O_p(T^{1/4v}) \\ &= o_p(1);\end{aligned}$$

cf. the proof of Theorem 2. ■

Proof of Theorem 8: First

$$\begin{aligned}z^a_{tT} &= \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})x_{t-s}y_{t-s} \\ &= \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})(x_{t-s}x'_{t-s}\beta_0+x_{t-s}\varepsilon_{t-s}) \\ &= z^b_{tT}\beta_0+z^c_{tT}\end{aligned}$$

Next by Lemma 6

$$\begin{aligned}\frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{b*} - \frac{1}{T}\sum_{t=1}^{T} z_{tT}^{b} &\to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P}, \\ \frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{c*} - \frac{1}{T}\sum_{t=1}^{T} z_{tT}^{c} &\to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P}.\end{aligned}$$

Also by Lemma 12

$$\frac{1}{T}\sum_{t=1}^{T} z_{tT}^{b} = E[x_t x_t'] + o_p(1), \frac{1}{T}\sum_{t=1}^{T} z_{tT}^{c} = E[x_t \varepsilon_t] + o_p(1).$$

Now

$$\hat{\beta} - \beta_0 = (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t.$$

By Corollary 3.48 of White (1999)

$$\frac{1}{T}\sum_{t=1}^{T} x_t x_t' = E[x_t x_t'] + o_p(1), \frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t = E[x_t \varepsilon_t] + o_p(1).$$

Hence

$$\frac{1}{T}\sum_{t=1}^{T} z_{tT}^{b} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t' + o_p(1), \frac{1}{T}\sum_{t=1}^{T} z_{tT}^{c} = \frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t + o_p(1).$$

Also, as $\sum_{t=1}^{T} x_t x_t'/T$ is positive definite for large enough T,

$$(\frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{b*})^{-1} - (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1} \to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P}.$$

Now

$$\begin{aligned}\hat{\beta}^* - \hat{\beta} &= (\frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{b*})^{-1}[\frac{1}{T}\sum_{t=1}^{T} z_{tT}^{c} - \frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t] \\ &\quad +(\frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{b*})^{-1}(\frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{*c} - \frac{1}{T}\sum_{t=1}^{T} z_{tT}^{c}) \\ &\quad +[(\frac{1}{m_T}\sum_{t=1}^{m_T} z_{tT}^{b*})^{-1} - (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}]\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t.\end{aligned} \tag{A.8}$$

Each term in eq. (A.8) converges to zero prob-$\mathcal{P}^*$, prob-$\mathcal{P}$ by earlier results. The result then follows by the conditional Slutsky Theorem, see Lemma 4.1 of Lahiri (2003). Alternatively a subsequence argument as in the proof of Lemma 3.2 of Gonçalves and White (2000) could also be used directly to demonstrate this result.

Proof of Theorem 9: With the same notation as in the proof of Theorem 8

$$\begin{aligned}\sqrt{T}(\hat{\beta}^* - \hat{\beta}) &= (\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} z_{tT}^{b*})^{-1}\sqrt{T}[\frac{1}{T}\sum\nolimits_{t=1}^{T} z_{tT}^{c} - \frac{1}{T}\sum\nolimits_{t=1}^{T} x_t\varepsilon_t] \\ &+(\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} z_{tT}^{b*})^{-1}\sqrt{T}(\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} z_{tT}^{*c} - \frac{1}{T}\sum\nolimits_{t=1}^{T} z_{tT}^{c}) \\ &+[(\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} z_{tT}^{b*})^{-1} - (\frac{1}{T}\sum\nolimits_{t=1}^{T} x_t x_t')^{-1}]\sqrt{T}\frac{1}{T}\sum\nolimits_{t=1}^{T} x_t\varepsilon_t.\end{aligned} \tag{A.9}$$

By Lemmata 12 and 6 and Corollary 3.48 of White (1999) $(\sum\nolimits_{t=1}^{m_T} z_{tT}^{b*}/m_T)^{-1} - (\sum\nolimits_{t=1}^{T} x_t x_t'/T)^{-1} \to 0$, prob-$\mathcal{P}^*$, prob-$\mathcal{P}$. Also the proof of Lemma A.2 of Smith (2004) demonstrates that

$$\frac{1}{T^{1/2}}\sum\nolimits_{t=1}^{T} z_{tT}^{c} = \frac{1}{T^{1/2}}\sum\nolimits_{t=1}^{T} x_t\varepsilon_t + O_p(T^{-1/2}).$$

The result then follows from arguments as in the proof of Theorem 8, the Cramér-Wold device, Theorem 3 and the conditional Slutsky Theorem [Lahiri (2003, Lemma 4.1)] [cf. proof of Theorem 3.2 of Fizenberger (1997)]. We could also have used a subsequence argument as in the proof of Lemma 3.3 of Gonçalves and White (2000). ■

To prove that our bootstrap method is applicable in the ML framework we make use of the following bootstrap UWL.

Lemma 13 *Let*

$$q_{tT}(\theta) = \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})g(x_{t-s}, \theta), \tag{A.10}$$

and consider the sample $q_{tT}(\theta)$, $(t = 1, ..., T)$. Draw a random sample of size m_T with replacement from $q_{tT}(\theta)$, $(t = 1, ..., T)$, to obtain the bootstrap sample $q_{tT}^*(\theta)$, $(t = 1, ..., m_T)$. Assume: **(a)** *Bootstrap Pointwise Weak Law of Large Numbers. for each $\theta \in \Theta \subset \mathbb{R}^p$, Θ a compact set,*

$$\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} q(x_t, \theta) \to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P};$$

(b) *Uniform Convergence:*

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} g(x_t, \theta)\right| \xrightarrow{p} 0;$$

(c) *Global Lipschitz: for all* $\theta, \theta^0 \in \Theta$ $|g(X_t, \theta) - g(X_t, \theta^0)| \leq L_t \|\theta - \theta^0\|$ *a.s.*$\mathcal{P}$ *and* $\sup_T E[T^{-1} \sum_{t=1}^{T} L_t] < \infty$. *Then, as* $m_T \to \infty$ *and* $S_T = o_p(T^{1/2})$, *for any* $\delta > 0$ *and* $\xi > 0$

$$\lim_{T\to\infty} \mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\Theta} \left| \frac{1}{m_T} \sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T} \sum\nolimits_{t=1}^{T} g(x_t, \theta) \right| > \delta\} > \xi\} = 0.$$

Proof: First $q_{tT}(\theta)$ is also Global Lipshitz for T large enough. As $S_T^{-1} \sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T}) = O(1)$, for large enough T,

$$\begin{aligned} \left|q_{tT}(\theta) - q_{tT}(\theta^0)\right| &= \left| \frac{1}{S_T} \sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})(g(X_{t-s}, \theta) - g(X_{t-s}, \theta^0)) \right| \\ &\leq C \left|g(X_{t-s}, \theta) - g(X_{t-s}, \theta^0)\right| \\ &\leq CL_t \left\|\theta - \theta^0\right\|. \end{aligned}$$

Similarly $|q_{tT}^*(\theta) - q_{tT}^*(\theta^0)| \leq C^* L_t^* \|\theta - \theta^0\|$.

From (b) the result is proven if

$$\lim_{T\to\infty} \mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\Theta} \left| \frac{1}{m_T} \sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T} \sum\nolimits_{t=1}^{T} q_{tT}(\theta) \right| > \delta\} > \xi\} = 0.$$

The remaining part of the the proof follows the steps of the proof of Lemma 8 of Hall and Horowitz (1996) and is identical to the proof of Lemma A.2 of Gonçalves and White (2000). Given $\varepsilon > 0$, let $\{\eta(\theta_i, \varepsilon) : i = 1, ..., I\}$ be a finite subcover of Θ where $\eta(\theta_i, \varepsilon) = \{\theta \in \Theta : \|\theta - \theta_i\| < \varepsilon\}$. Now

$$\sup_{\theta\in\Theta} \left| \frac{1}{m_T} \sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T} \sum\nolimits_{t=1}^{T} q_{tT}(\theta) \right| = \max_{i=1,\dots,I} \sup_{\theta\in\eta(\theta_i,\varepsilon)} \left| \frac{1}{m_T} \sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T} \sum\nolimits_{t=1}^{T} q_{tT}(\theta) \right|.$$

It then follows that for any $\delta > 0$ and any fixed ω

$$\begin{aligned} \mathcal{P}^*\{\sup_{\theta\in\Theta} \left| \frac{1}{m_T} \sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T} \sum\nolimits_{t=1}^{T} q_{tT}(\theta) \right| &> \delta\} \leq \\ \sum\nolimits_{i=1}^{I} \mathcal{P}^*\{\sup_{\theta\in\eta(\theta_i,\varepsilon)} \left| \frac{1}{m_T} \sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T} \sum\nolimits_{t=1}^{T} q_{tT}(\theta) \right| &> \delta\}. \end{aligned}$$

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

$$\begin{aligned}
\left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta)\right| &\leq \left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta_i) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta_i)\right| \\
&+ \left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta_i)\right| \\
&+ \left|\frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta_i)\right| \\
&\leq \left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta_i) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta_i)\right| \\
&+ C^*\varepsilon\frac{1}{m_T}\sum_{t=1}^{m_T} L_t^* + C\varepsilon\frac{1}{T}\sum_{t=1}^{T} L_t
\end{aligned}$$

for T large enough. Now $T^{-1}E[\sum_{t=1}^{T} L_t] = O(1)$. Thus, for any fixed $\delta > 0$ and $\xi > 0$, by M

$$\mathcal{P}\{C\varepsilon\sum_{t=1}^{T} L_t/T > \delta/3\} \leq \frac{3\varepsilon\Delta}{\delta} < \frac{\xi}{3}$$

with the choice $\varepsilon < \delta\xi/9C\Delta$ for some sufficiently large but finite constant Δ such that $\sup_T E[T^{-1}\sum_{t=1}^{T} L_t] < \Delta$ and large enough T. Hence

$$\begin{aligned}
&\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\eta(\theta_i,\varepsilon)}\left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta)\right| > \delta\} > \xi\} \\
&\leq \mathcal{P}\{\mathcal{P}^*\{\left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta_i) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta_i)\right| > \frac{\delta}{3}\} > \frac{\xi}{3}\} \\
&+\mathcal{P}\{\mathcal{P}^*\{C^*\varepsilon\frac{1}{m_T}\sum_{t=1}^{m_T} L_t^* > \frac{\delta}{3}\} > \frac{\xi}{3}\} + \\
&+\mathcal{P}\{C\varepsilon\frac{1}{T}\sum_{t=1}^{T} L_t > \frac{\delta}{3}\}.
\end{aligned}$$

By (a)

$$\mathcal{P}\{\mathcal{P}^*\{\left|\frac{1}{m_T}\sum_{t=1}^{m_T} q_{tT}^*(\theta_i) - \frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta_i)\right| > \frac{\delta}{3}\} > \frac{\xi}{3}\} < \frac{\xi}{3}$$

for T large enough. Also for fixed ω by M

$$\begin{aligned}
\mathcal{P}^*\{C^*\sum_{t=1}^{m_T} L_t^*/m_T &> \frac{\delta}{3\varepsilon}\} \leq \frac{3\varepsilon C^*}{\delta}\frac{1}{m_T}\sum_{t=1}^{m_T} E^*[L_t^*] \\
&= \frac{3\varepsilon C^*}{\delta}\frac{1}{T}\sum_{t=1}^{T} L_t
\end{aligned}$$

as $T^{-1}\sum_{t=1}^{T} L_t$ satisfies a LLN under the conditions of the theorem. Hence by M

$$\begin{aligned}
\mathcal{P}\{\mathcal{P}^*\{C^*\sum_{t=1}^{m_T} L_t^*/m_T &> \frac{\delta}{3\varepsilon}\} > \frac{\xi}{3}\} \leq \mathcal{P}\{\frac{3\varepsilon C^*}{\delta}\frac{1}{T}\sum_{t=1}^{T} L_t > \frac{\xi}{3}\} \\
&= \mathcal{P}\{\frac{1}{T}\sum_{t=1}^{T} L_t > \frac{\delta\xi}{9\varepsilon C^*}\} \\
&\leq \frac{9\varepsilon C^*}{\delta\xi}E[\frac{1}{T}\sum_{t=1}^{T} L_t] \\
&\leq \frac{9\varepsilon C^*\Delta}{\delta\xi} < \frac{\xi}{3}
\end{aligned}$$

with the choice $\varepsilon < \delta\xi^2/27C^*\Delta$.

Therefore, if ε is chosen such that

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

the result follows. ■

We also require the following Lemma.

Lemma 14 *Let $g(x_t, \theta)$ be a Lipschitz continuous function on Θ, i.e., for all $\theta, \theta^0 \in \Theta$ $|g(X_t, \theta) - g(X_t, \theta^0)| \leq L_t \|\theta - \theta^0\|$ a.s.$\mathcal{P}$ and $\sup_T E[T^{-1}\sum_{t=1}^{T} L_t] < \infty$. Assume additionally that the process $\{x_t\}_{t=1}^{\infty}$ is a finite dimensional stationary and strong mixing process with mixing coefficients α of size $-v/(v-1)$ for some $v > 1$ and $E[\sup_{\theta\in\Theta} |g(x_t, \theta\}|^{v+\eta}]$ $< \Delta < \infty$ for some $\eta > 0$. Then*

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta) - \frac{1}{T}\sum_{t=1}^{T} g(x_t, \theta)\right| \xrightarrow{p} 0,$$

where $q_{tT}(\theta)$, $(t = 1, ..., T)$, is defined in eq. (A.10).

Proof: By Lemma A.1 of Smith (2004)

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta) - E[g(x_t, \theta)]\right| = o_p(1).$$

Also by a standard UWL for global Lipschitz functions, e.g., Corollary 3.31 of Newey (1991), combined with Corollary 3.48 of White (1999),

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T} g(x_t, \theta) - E[g(x_t, \theta)]\right| = o_p(1).$$

Thus by T

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)-\frac{1}{T}\sum\nolimits_{t=1}^{T} g(x_t,\theta)\right| = o_p(1).$$

■

Proof of Theorem 10: Apply Lemma A.2 of Gonçalves and White (2004), with n replaced by T, $Q_T(\cdot,\theta)=\mathcal{L}_T(\theta)$ and $Q_T^*(\cdot,\omega,\theta)=\sum_{t=1}^{m_T} h_{tT}^*(\theta,\omega)/m_T$. Conditions a1-a3 of that theorem hold under Assumption 5.1.

Conditions b1 and b2 follow by Assumptions 5.1(a) and (b). Condition b3 is obtained from the bootstrap uniform weak law of large numbers, i.e., Lemmata 6, 13 and 14 which are implied by Assumption 5.1. ■

Proof of Theorem 11: The proof is identical to the proof of Theorem 2.2 of Gonçalves and White (2004) for MBB replacing Theorem 2.2 of Gonçalves and White (2002) by Theorem 3. Theorem 3 can be applied in this context since $E[\partial\log f(x_t,\theta_0)/\partial\theta]=0$. Additionally, the bootstrap uniform weak law of large numbers, Lemmata 6, 13 and 14, is used rather that the analogous results based on MBB in Gonçalves and White (2004). ■

A References

Anatolyev, S. (2005): "GMM, GEL, serial correlation, and asymptotic bias," *Econometrica,* 73, 983–1002

Andrews, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica,* 59, 817–858.

Beran, R. (1988): "Prepivoting test statistics: a bootstrap view of aymptotic refinements," *Journal of the American Statistical Association*, 83, 687-697.

Bickel, P. and Freedman, D. (1981): "Some asymptotic theory of the bootstrap," *Annals of Statistics*, 9, 1196-1217;

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

Davidson, J. and de Jong, R. (2000): "Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices," *Econometrica*, 68, 407-424

Efron, B. (1979): "Bootstrap Methods: Another look at the jackknife," *The Annals of Statistics*, 7, 1-26;

Fitzenberger, B. (1997): "The moving blocks bootstrap and robust inference for linear least squares and quantile regressions," *Journal of Econometrics*, 82, 235–287.

Gallant, R. (1987): *Nonlinear Statistical Models*, John Wiley and Sons, New York,

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*, vol. 18, 1367-1384.

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

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

Hansen, B. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60, 967-972.

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

Horowitz, J. (2001): "The bootstrap," In: Heckman, J.J., Leamer, E. (Eds.), *Handbook of Econometrics*, vol. 5. Elsevier, pp. 3159–3228. Ch52.

Hurvich, C. and Zeger, S. (1987): "Frequency domain bootstrap methods for time series," *Statistics and Operation Research Working Paper, New York University*, New York.

Kitamura, Y. (1997): “Empirical Likelihood Methods with Weakly Dependent Processes,” *The Annals of Statistics,* 25, 2084-2102.

Kitamura, Y., and Stutzer, M. (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation”, *Econometrica*, 65, 861-874.

Künsch, H. (1989): “The jackknife and the bootstrap for general stationary observations,” *The Annals of Statistics* 17, 1217–1241.

Lahiri (2003): *Resampling Methods for Dependent Data*, Springer-Verlag, New York, Inc.

Liu, R. and Singh, K. (1992): “Moving blocks jackknife and bootstrap capture weak dependence,” in *Exploring the Limits of Bootstrap*, (Raoul LePage and Lynne Billard, eds.), John Wiley, 224–248 New York: Wiley

Machado, J. and Parente, P. (2005): “Bootstrap estimation of covariance matrices via the percentile method,” *Econometrics Journal*, 8, 70-78.

Newey, W. (1991): “Uniform Convergence in Probability and Stochastic Equicontinuity,” *Econometrica,* 59, 1161-1167.

Newey, W. and West, K., (1987): “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703-08.

Politis, D. and Romano, J. (1992a): “A Circular Block-Resampling Procedure for Stationary Data,” in *Exploring the Limits of Bootstrap*, (Raoul LePage and Lynne Billard, eds.), John Wiley, 263-270.

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

Rao, C. (2002): *Linear Statistical Inference and its Applications*, Wiley.

Serfling, R. (2002): *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

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

Shi, X. and Shao, J. (1988): "Resampling estimation when the observations are m-dependent," *Communications in Statistics*, A, 17, 3923-3934.

Singh, K. (1981): "On the asymptotic accuracy of Efron's bootstrap," *The Annals of Statistics,* 9, 1187–1195.

Smith, R.J., (1997): "Alternative Semi-parametric Likelihood Approaches to Generalised Method of Moments Estimation," *Economic Journal*, vol. 107(441), 503-19.

Smith. R.J. (2004): "GEL criteria for moment condition models," *cemmap working paper CWP19/04*.

White, H., (1999): *Asymptotic theory for econometricians.* Academic Press.