A frequency domain wild bootstrap for dependent data

James Davidson* *University of Exeter*
Rami Chehab *Aston University*

8th July 2019

Abstract

We propose a resampling method for stationary dependent time series, based on Rademacher wild bootstrap draws from the Fourier transform of the data. The main distinguishing feature of our method is that the bootstrap draws share their periodogram identically with the sample, implying good properties under autocorrelation of arbitrary form. A drawback of the basic procedure, that the bootstrap distribution of the mean is degenerate, is overcome by a simple Gaussian augmentation with variance estimated by a response surface fitted to preliminary simulations. Extensive Monte Carlo evidence is reported comparing alternative bootstrap methods in tests of significance and location in a regression model with autocorrelated shocks, and also of unit roots.

1 Introduction

In econometric applications, bootstrap inference in dependent data is typically motivated as a remedy for an incorrect or incomplete model specification. The bootstrap principle can be viewed in this case as a vehicle for estimating the missing model components. Since the object is to model not merely the marginal distribution of the sample data but its joint distribution, this is an exercise entailing difficult compromises. Unless the distribution possesses some basic regularities, at a minimum stationarity, the sample reduces in effect to a single observation. Even in a nonparametric context, it is inevitably necessary to assume the joint distribution has sufficient structure to allow it to be reconstructed from the sample data.

The various forms of block bootstrap represent one approach to performing this reconstruction. They are nonparametric in character, and mimic the sample dependence of the whole by the joint distributions of short segments. However, since these segments must be independently drawn with replacement and concatenated, the 'joins' problem places a limitation on the effectiveness of the mimicry. A leading alternative to the blocks method is the sieve autoregressive bootstrap, which is surely well adapted to capturing certain types of dependence but equally must fail to capture other types, as discussed in Kreiss, Paparoditis and Politis (2011).

A third approach is to use an estimate of the spectral density of the series to model the dependence. A number of such methods are proposed in Kirch and Politis (2011), defining what they call the 'time-frequency toggle' (TFT) class. One of their methods involves randomly resampling the discrete Fourier transform (DFT) of the regression residuals after standardizing the DFT points to equal variance, using a kernel estimator of the spectral density. The resampled points, with real and imaginary parts treated equivalently, are recoloured by multiplication by the spectral weights before inversion back to the time domain. Another variant uses standard

*Corresponding author: james.davidson@exeter.ac.uk.

Gaussian drawings in place of the residual DFT, and a third uses normalized periodogram weights as probabilities to resample DFT points, at random but with a higher probability of being drawn from the same frequency range as the sample points they replace. A related approach is that of Theiler et al. (1992), who resample the phase components of the data distribution while holding the frequency magnitudes fixed, estimated by the periodogram.

A feature of these methods is that implementation requires a choice of kernel function and bandwidth. By contrast, Hidalgo (2003) proposes two methods that do not depend on estimating the spectral density. In one proposal, the residuals' DFT coordinates divided by their moduli are randomly resampled before being renormalized. In the second, the standardized time domain residuals are randomly resampled and the Fourier transform of this draw is constructed to form coordinates multiplying the absolute DFT coordinates of the original sample. These latter methods require the test regressions to be run in the frequency domain.

The bootstrap proposed in this paper has features in common with all these, specifically in resampling the DFT, and is similar to Hidalgo's methods in not being dependent, in its basic form, on estimation of the periodogram. The bootstrap draw is made using the Rademacher wild bootstrap to switch signs of the DFT points, but the bootstrap regressions are run conventionally after applying the inverse transform to the DFT. While the bootstrap draws are conditionally independent, they have the special property that their periodogram is numerically identical to that of the sample data. Hence, if the data dependence is fully embodied in second moments the bootstrap distribution mimics the sample distribution as closely as is possible. By the same token, the method is not applicable to inference problems relating to the periodogram points themselves or functions thereof, autocovariances and autoregressive parameters in particular.[1] All the above-cited procedures are aimed at the basic problem of inference in regression models with autocorrelated disturbances. They stand in contrast to the extensive literature on inference on the periodogram and so-called ratio statistics, as surveyed in Lahiri (2003) Chapter 9, and Kreiss and Paparoditis (2011), who focus chiefly on resampling the periodogram. This class of procedures have a different motivation to our own.

The paper is organized as follows. Section 2 describes the proposed bootstrap algorithm and derives key statistical properties. The proofs for these and subsequent results are gathered in Appendix A.1. Among these properties is the peculiarity that the bootstrap distributions of the series mean and variance are degenerate. This feature poses no problem for significance tests on slope coefficients in regressions with autocorrelated disturbances, and Section 3 shows such tests to be consistent. The basic method is unsuitable for tests of location and of unit roots, although an augmented version of the algorithm is shown to be consistent for tests of this type given a consistent long-run variance estimator. We propose implementing these latter tests by fitting a response surface to Monte Carlo-estimated performance measures, the details of which are given in Appendix A.3.

Section 4 summarizes the results of a large Monte Carlo study that compares our method with leading bootstrap alternatives in a variety of forms for the neglected autocorrelation. The full results are available in an online supplement. The wide range of experiments is made possible, given limited computing resources, by the 'warp-speed' Monte Carlo method of Giacomini et al. (2013). Appendix A.2 provides a brief description and motivation for this approach. Section 5 considers the multivariate case and shows that the cross-periodogram and cross-autocovariance properties extend the univariate case as expected. Simulation of a bivariate location test when the data are generated by a VAR(1) with correlated shocks indicates that our simple augmentation formula can render good size properties. Section 6 concludes.

[1] See Proposition 2.1 of Kirch and Politis (2011) for a related result.

2 The Fourier wild bootstrap

2.1 The FWB algorithm

Let $\boldsymbol{x} = (x_0, \ldots, x_{n-1})'$ denote the random n-vector whose joint distribution is to be modelled by the bootstrap. We assume this to be a finite realization of a covariance stationary process with population mean of zero. In the sequel, to establish the properties of the algorithm we shall assume that $\boldsymbol{x}$ satisfies conditions sufficient to satisfy a central limit theorem for dependent processes.

The procedure, which we refer to by the acronym FWB, consists of three basic steps:

1. Compute the discrete Fourier transform (DFT)

$$\boldsymbol{z} = \boldsymbol{Q}\boldsymbol{x} \tag{2.1}$$

where $\boldsymbol{Q}$ $(n \times n)$ denotes the unitary symmetric Fourier matrix with elements

$$q_{jk} = n^{-1/2} \mathrm{e}^{-2\pi \mathrm{i} jk/n}$$

for and $j, k = 0, \ldots, n-1$, where $\mathrm{i} = \sqrt{-1}$.

Then, for replications $b = 1, ..., B$:

2. Apply the Rademacher wild bootstrap to $\boldsymbol{z}$, switching signs (of the real and imaginary parts together) with probability $\frac{1}{2}$. Symbolically,

$$\boldsymbol{z}_b^* = \boldsymbol{W}_b \boldsymbol{z} \tag{2.2}$$

where $\boldsymbol{W}_b = \mathrm{diag}(\boldsymbol{w}_b)$ and $\boldsymbol{w}_b = (w_{b0}, \ldots, w_{b,n-1})'$ is a n-vector of independent random draws with

$$P(w_{bj} = 1) = P(w_{bj} = -1) = \tfrac{1}{2}.$$

3. Create the bootstrap draw in the time domain by

$$\boldsymbol{x}_b^* = \mathrm{Re}(\boldsymbol{Q}^\dagger \boldsymbol{z}_b^*) + \mathrm{Im}(\boldsymbol{Q}^\dagger \boldsymbol{z}_b^*) \tag{2.3}$$

where $\boldsymbol{Q}^\dagger$ denotes the conjugate transpose of $\boldsymbol{Q}$.

This algorithm has some features requiring careful explanation. While the rationale for summing the real and imaginary components at Step 3 is explained in the discussion following Theorem 2.3 below, it may be helpful to motivate this choice of algorithm by contrasting it with another.

Because $\boldsymbol{x}$ is real-valued, a well-known property of the DFT $\boldsymbol{z}$ is that $\mathrm{Re}(z_{n-j}) = \mathrm{Re}(z_j)$ and $\mathrm{Im}(z_{n-j}) = -\mathrm{Im}(z_j)$ for $j = 0, \ldots, n/2 - 1$ (n even) or $j = 0, \ldots, (n-1)/2$ (n odd). Likewise, for any $\boldsymbol{z}$ with this property $\boldsymbol{Q}^\dagger \boldsymbol{z}$ is real valued. Therefore, consider replacing (2.2) at Step 2 by

$$\widetilde{\boldsymbol{z}}_b^* = \widetilde{\boldsymbol{W}}_b \boldsymbol{z} \tag{2.4}$$

where $\widetilde{\boldsymbol{W}}_b$ is a special case of $\boldsymbol{W}_b$ constructed so that $\tilde{w}_{bj}$ is drawn like w_{bj} for j in the indicated ranges, but $\tilde{w}_{b,n-j} = \tilde{w}_{bj}$. Hence, only half as many independent random drawings are taken. Since the sign pattern is preserved, and hence $\widetilde{\boldsymbol{z}}_b^*$ shares the indicated property with $\boldsymbol{z}$, Step 3 must then yield a bootstrap draw with the form

$$\widetilde{\boldsymbol{x}}_b^* = \boldsymbol{Q}^\dagger \widetilde{\boldsymbol{z}}_b^* \tag{2.5}$$

which is real-valued, as the special case of (2.3) when the imaginary component is zero. It will be noted that (2.5) is just a special case of (2.3). Imposing the restriction on $\boldsymbol{W}_b$ must imply a different conditional distribution of the bootstrap since fewer Rademacher draws are utilized, and additional bootstrap replications must implicitly be needed to estimate the bootstrap distribution with equivalent precision. However, apart from this fact, the draws (2.5) share all the relevant properties of (2.3) to be derived in Section 2.2. The restricted form of the algorithm confers no apparent advantages over the general form. It is the latter form that is used in our simulation experiments reported in Section 4.

2.2 Properties of the Algorithm

Begin by writing

$$\boldsymbol{Q} = \boldsymbol{A} + \mathrm{i}\boldsymbol{B}$$
$$\boldsymbol{Q}^\dagger = \boldsymbol{A} - \mathrm{i}\boldsymbol{B}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ $(n \times n)$ are the symmetric matrices defined elementwise by

$$\{\boldsymbol{A}\}_{ij} = n^{-1/2}\cos(2\pi ij/n)$$
$$\{\boldsymbol{B}\}_{ij} = n^{-1/2}\sin(2\pi ij/n)$$

for $i, j = 0, \ldots, n-1$. $\boldsymbol{A}$ and $\boldsymbol{B}$ are mutually orthogonal, with $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{0}$, since for each pair $i, j = 0, \ldots, n-1$,

$$\begin{aligned}\{\boldsymbol{AB}\}_{ij} &= \frac{1}{n}\sum_{k=0}^{n-1}\cos(2\pi ik/n)\sin(2\pi jk/n) \\ &= \frac{1}{2n}\sum_{k=0}^{n-1}\sin(2\pi(i+j)k/n) + \frac{1}{2n}\sum_{k=0}^{n-1}\sin(2\pi(i-j)k/n) \\ &= 0.\end{aligned}$$

As in a number of similar identities to be determined subsequently, the sums vanish here since the terms are either zero, or equal and opposite in pairs. Note that the unitary property of $\boldsymbol{Q}$ implies

$$\boldsymbol{QQ}^\dagger = \boldsymbol{Q}^\dagger\boldsymbol{Q} = \boldsymbol{AA} + \boldsymbol{BB} = \boldsymbol{I}. \tag{2.6}$$

Consider the structure of these matrices in more detail. The row (column) of $\boldsymbol{A}$ labelled 0, and also that labelled $n/2$ for the case n even, is a unit vector times $n^{-1/2}$. Otherwise, the pairs of rows (columns) of $\boldsymbol{A}$ labelled j and $n-j$, respectively, are equal. Accordingly $\boldsymbol{A}$ has rank $(n+1)/2$ when n is odd or $n/2+1$ when n is even. These facts, and standard trigonometric identities, imply in particular that

$$\{\boldsymbol{AA}\}_{ij} = \begin{cases} 1, & i = j = 0, \text{ and } i = j = n/2 \ (n \text{ even}) \\ 0.5, & i = j \text{ and } i = n-j, \text{ for } j > 0 \text{ and } j \neq n/2 \ (n \text{ even}) \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

Note that $\boldsymbol{AAA} = \boldsymbol{A}$. $\boldsymbol{B}$ has a structure comparable to $\boldsymbol{A}$ except that the rows (columns) labelled 0 and $n/2$ (n even) are zero, and otherwise the pairs of rows (columns) labelled j and $n-j$ are equal in magnitude but opposite in sign. The rank of $\boldsymbol{B}$ is $(n-1)/2$ (n odd) or $n/2-1$ (n even),

$$\{\boldsymbol{BB}\}_{ij} = \begin{cases} 0, & i = j = 0, \text{ and } i = j = n/2 \ (n \text{ even}) \\ 0.5, & i = j > 0, \text{ and } j \neq n/2 \ (n \text{ even}) \\ -0.5, & i = n-j,\ j > 0, \text{ and } j \neq n/2 \ (n \text{ even}) \\ 0, & \text{otherwise,} \end{cases} \tag{2.8}$$

and $\boldsymbol{BBB} = \boldsymbol{B}$. Also note the direct confirmation from (2.7) and (2.8) of (2.6)

In the light of these facts, consider the steps of the proposed bootstrap draw. For ease of notation let the subscript b denoting the draw be henceforth implicit. The first Fourier transform yields

$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \mathrm{i}\boldsymbol{B}\boldsymbol{x} \tag{2.9}$$

and

$$\boldsymbol{z}^* = \boldsymbol{W}\boldsymbol{A}\boldsymbol{x} + \mathrm{i}\boldsymbol{W}\boldsymbol{B}\boldsymbol{x}. \tag{2.10}$$

The inverse Fourier transform leads to

$$\begin{aligned} \boldsymbol{Q}^\dagger \boldsymbol{z}^* &= \boldsymbol{A}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{x} + \mathrm{i}\boldsymbol{W}\boldsymbol{B}\boldsymbol{x}) - \mathrm{i}\boldsymbol{B}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{x} + \mathrm{i}\boldsymbol{W}\boldsymbol{B}\boldsymbol{x}) \\ &= (\boldsymbol{U} + \mathrm{i}\boldsymbol{V})\boldsymbol{x} \end{aligned} \tag{2.11}$$

(say) where

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{W}\boldsymbol{A} + \boldsymbol{B}\boldsymbol{W}\boldsymbol{B} \tag{2.12}$$

$$\boldsymbol{V} = \boldsymbol{A}\boldsymbol{W}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{W}\boldsymbol{A}. \tag{2.13}$$

Defining

$$\boldsymbol{R} = \boldsymbol{U} + \boldsymbol{V} \tag{2.14}$$

the FWB draw has the form

$$\boldsymbol{x}^* = \boldsymbol{R}\boldsymbol{x}. \tag{2.15}$$

We may compare (2.15) with other bootstrap formulae, having this form with different randomly drawn matrices. In the standard Efron (1979) bootstrap, $\boldsymbol{R}$ is constructed with n columns drawn randomly from the identity matrix with replacement. In the wild bootstrap $\boldsymbol{R}$ is diagonal, with randomly drawn diagonal elements. The various block bootstrap schemes construct $\boldsymbol{R}$ from random blocks of consecutive columns of the identity matrix, while the sieve autoregressive method forms $\boldsymbol{R}$ as an Efron matrix (as defined above) postmultiplied by an upper triangular matrix of moving average weights.

For $i, j = 0, \ldots, n-1$ the elements of $\boldsymbol{U}$ and $\boldsymbol{V}$ take the form

$$\{\boldsymbol{U}\}_{ij} = \sum_{k=0}^{n-1} w_k \Phi_{ijk}$$

and

$$\{\boldsymbol{V}\}_{ij} = \sum_{k=0}^{n-1} w_k \Psi_{ijk}$$

where

$$\begin{aligned} \Phi_{ijk} &= n^{-1}\left[\cos(2\pi ik/n)\cos(2\pi jk/n) + \sin(2\pi ik/n)\sin(2\pi jk/n)\right] \\ &= n^{-1}\cos(2\pi(i-j)k/n) \end{aligned} \tag{2.16}$$

and

$$\begin{aligned} \Psi_{ijk} &= n^{-1}\left[\cos(2\pi ik/n)\sin(2\pi jk/n) - \sin(2\pi ik/n)\cos(2\pi jk/n)\right] \\ &= -n^{-1}\sin(2\pi(i-j)k/n). \end{aligned} \tag{2.17}$$

Setting $\Upsilon_{ijk} = \Phi_{ijk} + \Psi_{ijk}$, the bootstrap series therefore have coordinates

$$x_i^* = \sum_{j=0}^{n-1} \left(\sum_{k=0}^{n-1} w_k \Upsilon_{ijk} \right) x_j \tag{2.18}$$

where the parenthesized sums of terms are the elements r_{ij} of the matrix $\boldsymbol{R}$ in (2.15).

For $0 \leq i, j < n-1$ and for each $k = 0, \ldots, n-1$, note that

$$\Upsilon_{i+1,j+1,k} = \Upsilon_{ijk} \tag{2.19}$$

and since $\cos(2\pi mk/n) = \cos(2\pi(m-n)k/n)$ and $\sin(2\pi mk/n) = \sin(2\pi(m-n)k/n)$ for $0 \leq m \leq n$ we also find that

$$\Upsilon_{i+1,0,k} = \Upsilon_{i,n-1,k} \tag{2.20}$$

for $0 \leq i < n-1$. Thus, $\boldsymbol{U}$ and $\boldsymbol{V}$ are both Toeplitz matrices having the circulant property, that each row reproduces the one above with a shift of one place to the right, with the last column entry wrapping around to the first position. The x_i^* in (2.18) are linear combinations of the sample series having random weights with this circulant form, which makes it easy to see how the bootstrap draws inherit the autocorrelation structure of the sample.

Letting $\boldsymbol{\iota}$ $(n \times 1)$ denote the column of ones, $\boldsymbol{U}$ and $\boldsymbol{V}$ have the following additional properties.

Theorem 2.1

(i) $\boldsymbol{U\iota} = w_0 \boldsymbol{\iota}$.

(ii) $\boldsymbol{V\iota} = \boldsymbol{0}$.

(iii) $\boldsymbol{U'V} = \boldsymbol{0}$.

(iv) $\boldsymbol{U'U} + \boldsymbol{V'V} = \boldsymbol{I}_n$.

(All proofs are given in Appendix A.1). Hence, (2.15) has the following special properties. First,

$$\boldsymbol{\iota' x}^* = \boldsymbol{\iota' R x} = w_0 \boldsymbol{\iota' x} \tag{2.21}$$

which says that the sum of the bootstrap series matches the sum of either $\boldsymbol{x}$ or $-\boldsymbol{x}$, with equal probability $\frac{1}{2}$, depending on the value of w_0. Second, since

$$\boldsymbol{R'R} = \boldsymbol{I}_n \tag{2.22}$$

by (2.14) and Theorem (2.1) (ii) and (iv),

$$\boldsymbol{x}^{*\prime} \boldsymbol{x}^* = \boldsymbol{x' x}. \tag{2.23}$$

If the sample data are expressed in mean deviation form such that $\boldsymbol{\iota' x} = 0$, as is usual in applications, the conditional distributions of the mean and variance of the bootstrap draws are therefore both degenerate, being equal with probability one to the sample mean (0) and the sample variance, respectively. We explore the implications of these facts in the sequel.

Consider the alternative algorithm obtained by replacing $\boldsymbol{W}$ by the special case $\widetilde{\boldsymbol{W}}$, as defined following equation (2.4), and so define $\widetilde{\boldsymbol{U}}$ and $\widetilde{\boldsymbol{V}}$ as the obvious variants of (2.12) and (2.13). It is easy to verify from the forms of $\boldsymbol{A}$ and $\boldsymbol{B}$ that $\boldsymbol{A}\widetilde{\boldsymbol{W}}\boldsymbol{B} = \boldsymbol{B}\widetilde{\boldsymbol{W}}\boldsymbol{A} = \boldsymbol{0}$ and hence $\widetilde{\boldsymbol{V}} = \boldsymbol{0}$. This verifies the earlier assertion that the imaginary component of the inverse DFT vanishes, and $\widetilde{\boldsymbol{x}}^* = \widetilde{\boldsymbol{U}}\boldsymbol{x}$. The following bootstrap properties depend on the forms of $\boldsymbol{A}$ and $\boldsymbol{B}$ and hold for arbitrary choices of $\boldsymbol{W}$, including $\widetilde{\boldsymbol{W}}$ in particular. Since $\widetilde{\boldsymbol{V}} = \boldsymbol{0}$ is merely a special case, they hold for $\widetilde{\boldsymbol{x}}^*$ just as for $\boldsymbol{x}^*$, and therefore do not require separate statements. For compactness we leave the alternative cases implicit and refer only to $\boldsymbol{W}$ in formulae, noting that the result of substituting $\widetilde{\boldsymbol{W}}$ is to eliminate the imaginary terms.

Theorem 2.2 *The periodogram of a FWB draw is identical with that of the sample series.*

Corollary 2.1 *The periodogram of $\boldsymbol{x}^*$ in (2.15) matches that of the inverse transform in (2.11).*

Next, let $\mathrm{E}^*(\cdot)$ denote the expected value under the bootstrap distribution of the random elements $w_0, \ldots, w_{n-1}$, conditional on the sample. Let x_j for $j = 0, \ldots, n-1$ denote the elements of $\boldsymbol{x}$ and let x_i^* denote the corresponding elements of the FWB draw.

Theorem 2.3 $\mathrm{E}^*(x_i^*) = 0$, *and for $m \geq 0$ and $i \geq m$,*

$$\gamma_{nm}^* := \mathrm{E}^*(x_i^* x_{i-m}^*) = \frac{1}{n} \sum_{j=m}^{n-1} x_j x_{j-m}. \tag{2.24}$$

Note how the bootstrap draw must be defined by (2.14) to achieve this result. The real and imaginary parts in (2.11) both have the circulant property producing the correct autocorrelation structure of a draw, but the associated scale parameters are individually too small. The one exception is the case $\boldsymbol{W} = \widetilde{\boldsymbol{W}}$ so that the imaginary part is zero. If the sample data are expressed in mean deviation form, Theorem 2.3 gives the autocovariance function of the bootstrap.

These results show that the bootstrap distribution is closely affiliated with the sample distribution and, for some purposes, too closely affiliated to be useful. As has been noted by Kreiss and Paparoditis (2011), it is clearly not feasible to use this bootstrap to analyse the distribution of functionals of the periodogram points. However, the implications of (2.21) and (2.23) that the bootstrap distribution has restricted rank should not be confused with dependence under that distribution. The following is shown by direct calculation.

Theorem 2.4 *Let $\boldsymbol{W}_a$ and $\boldsymbol{W}_b$ denote independent Rademacher draws defining $\boldsymbol{R}_a$ and $\boldsymbol{R}_b$ in (2.14) and let $\boldsymbol{x}_a^* = \boldsymbol{R}_a \boldsymbol{x}$ and $\boldsymbol{x}_b^* = \boldsymbol{R}_b \boldsymbol{x}$. Letting $\boldsymbol{x}_a^* = (x_{a1}^*, \ldots, x_{an}^*)'$ and $\boldsymbol{x}_b^* = (x_{b1}^*, \ldots, x_{bn}^*)'$,*

$$\mathrm{E}^* \left(x_{ai}^{*r} x_{bj}^{*s} \right) = \mathrm{E}^*(x_{ai}^{*r}) \mathrm{E}^*(x_{bj}^{*s}).$$

for all $i, j = 0, \ldots, n-1$ and positive integers r and s. The same holds for pairs $\widetilde{\boldsymbol{x}}_a^*$ and $\widetilde{\boldsymbol{x}}_b^*$ defined in (2.5).

Next, consider the following remarkable implication of expression (2.15) noting that the elements of $\boldsymbol{R} = \{r_{ij}\}$ are the parenthesized expressions in (2.18). Defining a *Gaussian sequence* as a random sequence (finite or infinite) whose finite dimensional distributions are multivariate Gaussian, the r_{ij} have the following property.

Theorem 2.5 *(i) For each fixed pair i, j, $\sqrt{n} r_{ij} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$. (ii) For each i the sequences $\{\sqrt{n} r_{ij}\}_{j=0}^{n-1}$ are asymptotically Gaussian.*

Expression (2.18) can be viewed in two ways. From the standpoint of the bootstrap distribution conditional on the data, $\boldsymbol{x}_i^* = \boldsymbol{r}_i' \boldsymbol{x}$, where $\boldsymbol{r}_i'$ is a row of $\boldsymbol{R}$, is a linear combination with fixed weights $x_1, \ldots x_n$, of increments that are individually Gaussian in the limit. The further implication is that $\boldsymbol{x}_i^*$ itself is asymptotically Gaussian. On the other hand, from the standpoint of the data distribution in the context of a given bootstrap draw, $\boldsymbol{x}_i^*$ is the weighted sum of $x_1, \ldots, x_n$ with fixed weights $\{r_{ij}\}_{j=0}^{n-1}$, the parenthesized terms in (2.18). This implies the following.

Theorem 2.6 *If the process generating the sample $\boldsymbol{x}$ satisfies the conditions of the central limit theorem, the distribution of the FWB series in a given bootstrap draw is asymptotically Gaussian with probability 1.*

We do not spell out explicit sufficient conditions for this latter convergence in view of the extensive existing literature on this topic, but we may cite de Jong (1997) whose Theorem 2 specifies dependence conditions formulated as near-epoch dependence on a mixing process. These are among the least restrictive known and this framework admits both linear and nonlinear dependence structures. As is appropriate to the present case, de Jong's theorem is for triangular arrays where the weighting of observations depends on sample size in the manner of the elements of (2.15).

We thus obtain the curious result that the bootstrap draws form an asymptotically Gaussian sequence with respect to both data and bootstrap distributions. This asymptotic property extends to statistics that are linear functions of the data. Should higher moments of the data as well as the autocorrelation structure influence the distribution of sample statistics, Theorem 2.6 implies that the method is most appropriate to samples that are stationary with Gaussian characteristics. However, in linear regression models where only second moments are involved, the method should be robust to the form of the shock distribution, subject only to existence of the variance.

Finally, recalling the assumption that $E(x_j) = 0$, define the autocovariance sequence of the observed series by $\gamma_m = \mathrm{E}(x_j x_{j-m})$ for $-\infty < m < \infty$. Let

$$\hat{\gamma}^*_{nm} := \frac{1}{n} \sum_{i=m}^{n-1} x^*_i x^*_{i-m} \tag{2.25}$$

such that $\mathrm{E}^*(\hat{\gamma}^*_{nm}) = \gamma^*_{nm}$ as defined in (2.24).

Theorem 2.7 *For fixed $m \geq 0$,*

$$\gamma^*_{nm} \xrightarrow{\text{pr}} \gamma_m \tag{2.26}$$

and

$$\mathrm{E}^*(\hat{\gamma}^*_{nm} - \gamma^*_{nm})^2 \xrightarrow{\text{pr}} 0. \tag{2.27}$$

These convergences in probability relate to expectations under the bootstrap distribution and are defined with respect to the data distribution. Taken together, (2.26) and (2.27) imply that the bootstrap estimator is consistent for the autocovariances of the underlying data.

3 Implications for Testing

Consider how this bootstrap might operate in the context of inference on $\boldsymbol{\beta}$ in a linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \quad (n \times 1)$$

where $\boldsymbol{u}$ is autocorrelated. The formal assumptions are as follows.

Assumption 1 *The series $\boldsymbol{u}$ and $\boldsymbol{X}$ $(n \times k)$ are weakly dependent processes having summable autocovariances.*

Assumption 2 $n^{-1}\boldsymbol{X}'\,\boldsymbol{X} \to_{\text{pr}} \boldsymbol{M}_{XX} < \infty$, *positive definite.*

Assumption 3 $\boldsymbol{X}$ *is strongly exogenous with respect to $\boldsymbol{u}$, with zero dependence at all leads and lags.*

Since the autocorrelation is to be modelled nonparametrically, the final assumption is required to rule out cases where $\boldsymbol{X}$ coordinates might depend on lags of $\boldsymbol{u}$ coordinates. Assumption 1 is unavoidable since we shall need to construct autocorrelation consistent variance estimators to ensure the test statistics are asymptotically pivotal.

3.1 Significance Tests

A significance test of β_1, the first (without loss of generality) element of $\boldsymbol{\beta}$, may be based on the asymptotically pivotal statistic

$$t = \frac{\hat{\beta}_1}{\text{s.e.}(\hat{\beta}_1)}$$

where the standard error is computed using a suitable robust variance estimator. Partition the regressors into first column and remainder, as $\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{X}_2]$. Under the null hypothesis $\beta_1 = 0$ the statistic has the form

$$t = \frac{\sqrt{n}\sum_{t=1}^{n} l_{nt}u_t}{\sqrt{\hat{V}_n}} \tag{3.1}$$

where $\boldsymbol{u} = (u_1, \ldots, u_n)'$ and l_{nt} is the t^{th} element of the vector

$$\boldsymbol{l} = \frac{\boldsymbol{M}_2\boldsymbol{x}_1}{\boldsymbol{x}_1'\boldsymbol{M}_2\boldsymbol{x}_1} \quad (n \times 1) \tag{3.2}$$

with $\boldsymbol{M}_2 = \boldsymbol{I} - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'$. An autocorrelation-consistent (AC) variance estimator appropriate to appear in (3.1) has the form

$$\hat{V}_n = \sum_{m=-M_n}^{M_n} w_{nm}\hat{\gamma}_{nm} n \sum_{t=|m|+1}^{n} l_{nt}l_{n,t-|m|} \tag{3.3}$$

where $M_n = o(n)$ and

$$\hat{\gamma}_{nm} = \frac{1}{n}\sum_{t=|m|+1}^{n} \hat{u}_t\hat{u}_{t-|m|} = \frac{1}{n}\sum_{t=|m|+1}^{n} u_t u_{t-|m|} + O_p(n^{-1}).$$

The w_{nm} are kernel weights, from a formula ensuring $\hat{V}_n > 0$ and with $w_{nm} = w_{n,-m}$ and $w_{nm} \to 1$ as $n \to \infty$ for fixed m.[2]

The numerator of (3.1) has variance

$$\begin{aligned} V_n &= n\text{E}\left(\text{E}\left(\left(\textstyle\sum_{t=1}^{n} l_{nt}u_t\right)^2 \middle| \boldsymbol{X}\right)\right) \\ &= \sum\nolimits_{m=1-n}^{n-1} \gamma_m n \sum\nolimits_{t=|m|+1}^{n} \text{E}(l_{nt}l_{n,t-|m|}) \end{aligned} \tag{3.4}$$

where $\gamma_m = \gamma_{-m} = \text{E}(u_t u_{t-m})$. Under the regularity conditions specified in Theorem 2.6, which must include $\{\gamma_m,\ m \geq 0\}$ being an absolutely summable sequence, $\sqrt{n}\sum_{t=1}^{n} l_{nt}u_t$ is asymptotically normal with variance

$$V = \text{plim}\, V_n = \sum\nolimits_{m=-\infty}^{\infty} \gamma_m \,\text{plim}\, n \sum\nolimits_{t=|m|+1}^{n} l_{nt}l_{n,t-|m|}. \tag{3.5}$$

The convergence in probability in the right-hand member of (3.5) is specified with respect to the sampling distribution of the regressors, and holds under our assumptions which specify in particular that the probability limit appearing in the last member is finite. Regularity conditions for consistency of the kernel estimator are close to those sufficient for the cited CLT, as detailed in de Jong and Davidson (2000) *inter alia*, and $\hat{V}_n \to_{\text{pr}} V$ leads to the usual conclusion that $t \to_{\text{d}} \text{N}(0,1)$ under the null hypothesis.[3]

---

[2] The Bartlett and Parzen kernels are valid choices, see e.g. Andrews (1991).

[3] "$\to_{\text{pr}}$" and "$\to_{\text{d}}$" denote convergence in probability and distribution respectively.

Now consider the bootstrap statistic. The bootstrap regressand is $y^*_{nt} = \boldsymbol{x}'_t\hat{\boldsymbol{\beta}} + u^*_{nt}$ for $t = 1, \ldots, n$ where $u^*_{n1}, \ldots, u^*_{nn}$ are resampled by the FWB from the least squares residuals $\hat{u}_t$. Observe the use of array notation in this context, which is appropriate because in view of (2.21) the series u^*_{nt} for $t = 1, \ldots, n$ has the special property

$$\sum_{t=1}^{n} u^*_{nt} = 0 \tag{3.6}$$

for each n. Letting $\tilde{\gamma}^*_{nm} = n^{-1}\sum_{t=|m|+1}^{n} \hat{u}^*_{nt}\hat{u}^*_{n,t-|m|}$ where $\hat{u}^*_{n1}, \ldots, \hat{u}^*_{nn}$ are the residuals from the bootstrap regression, the bootstrapped test statistic is

$$t^* = \frac{\hat{\beta}^*_1 - \hat{\beta}_1}{\text{s.e.}(\hat{\beta}^*_1)} = \frac{\sum_{t=1}^{n} l_{nt}u^*_{nt}}{\sqrt{\hat{V}^*_n}} \tag{3.7}$$

where

$$\hat{V}^*_n = \sum_{m=-M_n}^{M_n} w_{nm}\tilde{\gamma}^*_{nm} n \sum_{t=|m|+1}^{n} l_{nt}l_{n,t-|m|}. \tag{3.8}$$

Note that $\tilde{\gamma}^*_{nm} = \hat{\gamma}^*_{nm} + O_p(n^{-1})$ where $\hat{\gamma}^*_{nm}$ is defined in (2.25).

Theorem 3.1 *If the conditions of Theorem 2.6 hold and $\boldsymbol{x}_1 \neq \boldsymbol{\iota}$ (the unit column), then $t^* \to_{\mathrm{d}} \mathrm{N}(0, 1)$.*

This convergence is with respect to the distribution of the sample data with the bootstrap draw conditionally fixed. It occurs with probability 1 under the bootstrap distribution, and so tells us what happens under the distribution of multiple independent bootstrap draws in large samples. It might be noted that even if the data series were wholly arbitrary and the CLT conditions failed, the limit distribution in question would still be Gaussian, as a corollary of Theorem 2.5. The conditions of Theorem 2.6 are required only to ensure correct calculation of the limiting variance.

The bootstrap disturbances have the special property (3.6) not shared with the sample disturbances but the distribution of the statistic would be the same without this restriction imposed, provided the $\boldsymbol{X}_2$ variables include the intercept (unit column). Then, the residuals sum to zero by the familiar property of least squares and $\sum_{t=1}^{n} l_{nt}u_t$ has the same value whether u_t is in mean-deviation form or otherwise. The same is true of $\sum_{t=1}^{n} l_{nt}u^*_{nt}$, notwithstanding that the mean-deviation form of u^*_{nt} is identical with the original. The distribution of the bootstrap statistic is therefore invariant to restriction (3.6). Even in a regression with intercept term excluded, provided the disturbances have population mean of zero the asymptotic variance is still given by (3.5), and the limit distribution is unchanged. An additional feature of (3.8) is that

$$\hat{\gamma}^*_{n0} = \frac{1}{n}\sum_{t=1}^{n} \hat{u}^2_t + O_p(n^{-1}) \tag{3.9}$$

where the right-hand side sum does not vary over the bootstrap replications thanks to (2.23) and the remainder reflects the differences $\hat{u}^*_{nt} - u^*_{nt}$. This is however a small-order effect on the bootstrap dispersion of $\hat{V}^*_n$ that does not affect its limit in probability.

An alternative to (3.3) is the heteroscedasticity and autocorrelation consistent (HAC) formulation

$$\hat{V}^H_n = \sum_{m=-M_n}^{M_n} w_{nm} n \sum_{t=|m|+1}^{n} l_{nt}l_{n,t-|m|}\hat{u}_t\hat{u}_{t-|m|}. \tag{3.10}$$

A minor advantage of (3.10) is that property (2.23) plays no part. Closely analogous arguments can be given for the validity of the bootstrap test based on (3.10) and its bootstrap counterpart. The key steps in linking the two formulations are to show that under the present assumptions, $\mathrm{E}(\hat{V}_n^H) = V_n$ as defined in (3.4), and that if $\hat{V}_n^{H*}$ denotes the bootstrap counterpart of (3.10) with $\hat{u}_{nt}^*$ replacing $\hat{u}_t$, then $\mathrm{E}^*(\hat{V}_n^{H*}) = \mathrm{E}^*(\hat{V}_n^*)$. The proof of the counterpart of Theorem 3.1 with $\hat{V}_n^H$ replacing $\hat{V}_n$ therefore follows very similar lines.

3.2 Tests of Location and Unit Roots

The case $\boldsymbol{x}_1 = \boldsymbol{\iota}$, as when β_1 is the intercept of the regression, leads to a different result. In this case,

$$\sqrt{n}\sum_{t=1}^{n} l_{nt}u_{nt}^* = \frac{\sqrt{n}}{d_n}\sum_{t=1}^{n} u_{nt}^* + \sqrt{n}\sum_{t=1}^{n}\left(l_{nt} - \frac{1}{d_n}\right)u_{nt}^* \tag{3.11}$$

where $d_n = \boldsymbol{\iota}'\boldsymbol{M}_2\boldsymbol{\iota}$. The first right-hand side sum in (3.11) vanishes identically by (3.6). The variance of (3.11) under the bootstrap distribution is different from that of the numerator in (3.1) and, in particular, is not estimated consistently by (3.8). The conclusion is that the FWB fails in tests of location.

This is in fact a common limitation of frequency domain methods. Hidalgo's (2003) method explicitly casts the data into mean deviation form, and while the bootstrap means in the TFT procedures of Kirsch and Politis (2011) do possess a distribution, the authors are careful to point out (page 3, third paragraph), that this *cannot* be identified with the distribution of the sample mean. Testing the significance of the intercept in regression models is sometimes viewed as of minor importance in applied econometric work, but tests of location are nonetheless an important testbed in methodological studies; see for example McElroy and Politis (2002).

There is also the problematic case of the unit root test. Under the null hypothesis of a unit root the observed process has the form $y_t = \sum_{s=1}^{t} u_s$, and Dickey-Fuller-type statistics feature (with $y_0 = 0$) the term

$$\sum\nolimits_{t=1}^{n} y_{t-1}u_t = \tfrac{1}{2}\left(y_n^2 - \sum\nolimits_{t=1}^{n} u_t^2\right). \tag{3.12}$$

Since bootstrap series $y_{n1}^*, \ldots, y_{nn}^*$ must be constructed by cumulating bootstrap draws $u_{n1}^*, \ldots, u_{nn}^*$, it is clear from (3.6) that the FWB bootstrap distribution of y_{nn}^* is degenerate, so that bootstrap unit root tests must fail. Simulated null distributions will in fact converge on limits involving a Brownian bridge in place of a Brownian motion.

We propose a simple device to overcome these limitations. If the bootstrap entails resampling a series $\hat{u}_t$ expressed in mean deviations, what is required is that draws based on this series should exhibit randomly distributed sample means having central tendency zero. This effect can be achieved by adding to u_{nt}^* for each t an independently drawn scalar random variable. Letting Z^* denote a $\mathrm{N}(0,1)$ drawing and

$$\omega^2 = \sum_{m=-\infty}^{\infty} \gamma_m \tag{3.13}$$

the coordinates

$$u_{nt}^{**} = u_{nt}^* + \omega Z^*/\sqrt{n} \tag{3.14}$$

for $t = 1, \ldots, n$ have a sample mean that is distributed normally with mean zero and variance ω^2/n. We call the additional term in (3.14) the 'surrogate mean' of the bootstrap draw. Provisionally assuming ω^2 is known, replacing u_{nt}^* with u_{nt}^{**} replaces the first right-hand side term in (3.11) with $\omega Z^*/d_n$. Since the augmentation is Gaussian and independently drawn, under

the assumptions of Theorem 2.6 the large-sample distribution of $n^{-1/2}\sum_{t=1}^{n} u_t^{**}$ matches the corresponding distribution for u_t. We note in view of (2.23) that

$$\sum_{t=1}^{n} u_t^{**2} = \sum_{t=1}^{n} \hat{u}_t^2 + \omega^2 Z^{*2}$$

similarly to (3.9). There therefore exists an upward bias in a variance estimate based on the augmented bootstrap, although this is of $O_p(n^{-1})$ as $n \to \infty$.

We refer to the FWB procedure applied to the augmented bootstrap disturbances as the "augmented Fourier wild bootstrap" (AFB). Be careful to note that the augmentation vanishes from formulae in every case where the residuals appear in mean-deviation form. Hence, tests of slope coefficients are identical in the AFB and the unaugmented FWB.

In a unit root test based on the regression $\Delta y_t = \phi y_{t-1} + u_t$, suppose u_t is a zero-mean I(0) process whose distribution validates the usual Dickey-Fuller result

$$n\hat{\phi} = \frac{n^{-1}\sum_{t=2}^{n} y_{t-1}\Delta y_t}{n^{-2}\sum_{t=2}^{n} y_{t-1}^2} \xrightarrow{d} \frac{B(1)^2 - \sigma^2/\omega^2}{2\int_0^1 B^2 dr} \tag{3.15}$$

where $\sigma^2 = \mathrm{E}(u_t^2)$, ω^2 is defined in (3.13) and B is standard Brownian motion on $[0,1]$. Following the substitution in (3.12), $B(1)$ is the limit in distribution of $y_n/\omega\sqrt{n}$ and is standard normal. This is the property that the FWB bootstrap draw fails to deliver. The key result for understanding the role of the augmentation is the following, which validates the application both to unit root tests and to tests of location of stationary series.

Theorem 3.2 *Let the array $\{u_{nt}^*, t = 1, \ldots, n\}$ represent the FWB draw subject to the restriction $\sum_{t=1}^{n} u_{nt}^* = 0$. Also define u_{nt}^{**} as in (3.14). Letting $y_{nt}^{**} = \sum_{s=1}^{t} u_{ns}^{**}$ and $Y_n^{**}(r) = n^{-1/2} y_{n,[nr]}^{**}$ for $0 \le r \le 1$, $Y_n^{**} \to_d Y^{**} = \omega B$ as $n \to \infty$.* [4]

Some remarks about the nature of this convergence are in order. The stationary array $\{u_{nt}^{**}\}$ is not ergodic but the strong law of large numbers nonetheless applies in the sense that, thanks to (2.21),

$$\frac{1}{n}\sum_{t=1}^{n} u_{nt}^{**} = n^{-1/2}\omega Z^* \to 0 \text{ a.s.}$$

as $n \to \infty$. Similarly, if $\{u_t\}$ is stationary and ergodic then thanks to (2.23),

$$\frac{1}{n}\sum_{t=1}^{n} u_{nt}^{**2} = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2 + \frac{\omega^2 Z^{*2}}{n} \to \mathrm{E}(u_1^2) \text{ a.s.}$$

It may also be useful in certain contexts such as cointegration testing (see Phillips and Ouliaris 1990, inter alia) to remark that this convergence is mixing in the sense of Rényi (compare Hall and Heyde 1980, page 57). Let $\{\mathcal{F}_{nt}, 1 \le t \le n, n \ge 1\}$ denote the filtration defined by the bootstrap draw $\{u_{nt}^{**}\}$ conditional on the data, and $\mathcal{F} = \bigvee_{n,t} \mathcal{F}_{nt}$. That the events $\{Y^{**} \le y\}$ for $y \in \mathbb{R}$ are independent of every $E \in \mathcal{F}$ follows from continuity of the probability measure and the facts that $|u_{nt}^{**} - u_{nt}^*| \to 0$ a.s. for each t as $n \to \infty$, and that Z^* is an independent drawing.

From these considerations, it is straightforward to establish that when the null hypothesis is true, the various unit root test statistics under the bootstrap distribution converge to the

[4] The indicated convergence in distribution is with respect to the Skorokhod topology on $[0,1]$.

corresponding functionals of Brownian motion specified for the sample data. In particular, Theorem 3.2 is sufficient to establish the limit distribution indicated in (3.15) for $n\hat{\phi}^*$, the formula corresponding to $n\hat{\phi}$ in which u_t is replaced by u_{nt}^{**} throughout.

Of course, the critical assumption underlying this claim is that ω is known. In practice this has to be estimated and the effectiveness of the augmentation depends on the properties of this estimator. A kernel-based HAC variance estimator computed for the sample data is the natural expedient, and Kirsch and Politis (2011) adopt a similar device to implement their unit root test. However, we find experimentally that the standard kernel estimator does not behave too well, with an excessive error in rejection probability (ERP) remaining, appearing greater both as the sample is small and the degree of autocorrelation is large. Estimators computed from fitted residuals are of course biased downwards in finite samples. This effect is compensated by the usual factor of $n/(n-k)$ when the disturbances are independent but in correlated cases it is harder to control. The solution adopted in the present study is to construct a response surface correction depending on both sample size and a measure of the degree of autocorrelation in the sample, based on preliminary calibration experiments. The details of this method are given in Appendix A.3.

4 Monte Carlo Evidence

Simulation experiments were performed using the 'warp speed' technique proposed by Giacomini et al. (2013). In this approach, the bootstrap distribution is estimated once for the whole experiment, rather than repeatedly for each replication, permitting an order-of-magnitude reduction in computational cost. By the use of this technique, it has been feasible to explore the performance of our proposed methods in a wide range of models. The method is explained and motivated in Appendix A.2. The number of Monte Carlo replications is set to $K = 50,000$, a feasible choice for the warp speed method.

One way in which warp-speed experiments differ from conventional simulations is that they do not capture the approximation due to a feasibly small number of bootstrap draws. By contrast, the results approach the asymptote in measuring bootstrap rejection rates. There do exist circumstances where warp speed could fail, as when the bootstrap distribution is not well defined, but there are no grounds to suppose such problems exist in our case. As a check, a conventional experiment was run with 1000 replications and 399 bootstrap draws in each. No difference in the results was found that could not be accounted for by the usual margin of experimental error.

The full set of experimental results are reported in the accompanying supplement, available for download.[5] Systematic comparisons entail looking at a range of cases and the burden of both reporting and absorbing the resulting plethora of results threatens to become excessive. Therefore the tables below show only summaries of the full results, the average performance over fixed sets of disturbance dependence structures and sample sizes. Since the same cases are simulated for each procedure these averages provide rankings for comparison. Table entries are quoted to the nearest two digits for values below 99, noting that additional digits are essentially noise.

Tables 1, 3 and 5 show averages of the percentage absolute size distortions in tests of nominal significance level 5%. For example, an average experimental rejection rate of either 6% or 4% would result in a table entry of 20, while an average rejection rate of either 10% or 0% would appear as 100. Tables 2, 4 and 6 show the averages of estimated percentage powers under specified

[5] The online supplement is at http://people.ex.ac.uk/jehd201/AFB_Supplement.pdf. In addition to the 5% test size distortions and test powers summarized here, the supplement contains tables of Cramer-von Mises statistics, showing the differences between the distribution of the bootstrap null p-values and the uniform distribution.

alternatives. True powers are being compared here, since the tests are in each case correctly sized making use of the empirical tabulations obtained from simulating the null hypothesis.

The sample sizes compared are $n = 50$, 200 and 800. The table columns show the comparisons for different *types* of autocorrelation, as well as for alternative i.i.d. distributions for the shocks, denoted E_t. These are either standard normal or standardized Student's $t(3)$. In the columns headed AR in the tables, the disturbances are generated as

$$U_t = \rho U_{t-1} + E_t \tag{4.1}$$

with $\rho = 0.3$, 0.6 and 0.9. Thus, in these entries a total of nine cases are averaged, three values of ρ and three of n. The columns headed MA show the performance over moving average schemes of the form

$$U_t = E_t + E_{t-m} \tag{4.2}$$

for cases $m = 1, 2$ and 4, and the three sample sizes, again 9 cases in total. These latter models provide a contrast to the smooth and monotone AR spectral densities, and are of interest in particular since the AFB and SAR methods depend on fitted autoregressive approximations. The columns headed Fractional show the averages over fractionally integrated disturbance processes having the form

$$U_t = (1 - L)_+^{-d} E_t \tag{4.3}$$

where the fractional operator $(1 - L)_+^{-d}$ truncates lags to positive values of t. Here, six cases are averaged, $d = 0.1$ and $d = 0.3$ with the three sample sizes. These models are in a different category from the AR and MA because under strong dependence the HAC test statistics are not asymptotically pivotal.

Since the cases are arbitrarily chosen, take care to note that there is *no* comparability across columns in the tables. The full sets of results in the online supplement should be consulted to see the individual results that go to make up these averages.

4.1 Significance tests

Tables 1–4 show the results of experiments from a regression model with intercept and two exogenous regressors,

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t,\ t = 1, \ldots, n$$

Regressor X_1 is a serially independent standard normal (N(0,1)) sequence, and regressor X_2 is a AR(1) process with coefficient equal to 0.5, driven by N(0,1) shocks. The regressors are sampled afresh at each replication, so the experimental results are not contingent on a fixed regressor set but on the specified distributions. With one exception, the test statistics are robust t ratios with standard errors computed by the HAC estimator with Parzen kernel and Newey-West (1994) plug-in bandwidth. The bootstrap significance tests use the signed statistics with equal-tailed rejection regions.

The experiments compare the AFB test[6] with a number of bootstrap alternatives, with results in the rows of the tables. In the first of the Kirsch and Politis (2011) TFT methods cited in the introduction, the spectral density is estimated with the Parzen window and a bandwidth of $[n^{1/3}]$. HB1 denotes the first of the Hidalgo (2003) algorithms, in which the regressions are performed in the frequency domain with robust test statistics in this case calculated using formula (10) of Hidalgo's paper. Also compared are the moving blocks method of Künsch (1989)

[6] In the case of slope coefficient tests the AFB and FWB procedures are identical, and the table rows are labelled accordingly.

	AR		MA		Fractional	
	Normal	t(3)	Normal	t(3)	Normal	t(3)
FWB	4.0	5.3	7.4	6.9	4.0	6.0
TFT	9.8	8.2	7.9	7.2	4.8	9.1
HB1	4.4	5.1	3.5	3.0	3.0	4.0
MBB	7.4	6.8	7.6	7.2	5.5	5.6
SAR	3.4	4.3	6.0	9.1	4.8	6.1
Asy.	63	54	52	42	51	42

Table 1: Average size distortion: regression slope coefficients

	AR		MA		Fractional	
	Normal	$t(3)$	Normal	$t(3)$	Normal	$t(3)$
FWB	54	58	55	60	76	54
TFT	54	58	54	59	73	75
HB1	64	67	56	61	79	82
MBB	54	59	54	61	73	78
SAR	53	58	53	59	72	75
Asy.	66	68	66	70	86	85

Table 2: Average powers: regression slope coefficients

with block length $[n^{1/3}]$, denoted MBB in the tables,[7] and the sieve-autoregressive bootstrap of Bühlmann (1997), denoted SAR, with lag length chosen by the Akaike criterion up to a maximum of $[0.6n^{1/3}]$. Asymptotic tests using the standard tabulations were also run to provide a baseline for comparison. The bootstrap methods were applied subject to a pretest for autocorrelation, the standard Efron (1979) bootstrap being used if the absolute value of the first-order residual autocorrelation coefficient is less than $2/\sqrt{n}$.[8] The averages in Tables 1 and 2 are taken over the two slope coefficients, as well as over the sample sizes and autocorrelation parameters. Tables 3 and 4 shows the averages for the test for significance of the intercept, a case for which the TFT and HB1 bootstraps are unavailable. Tables 2 and 4 show average power performances for alternatives $\beta_j = 3/\sqrt{n}$, for $j = 0, 1$, and 2. These cases were selected to yield rejection rates exceeding the null case, but at the same time not too close to unity.

For the intercept experiments, we recorded the means of the generated samples and also the surrogate means of the AFB draws. To check on the match of distributions, the variances of these two quantities were calculated for a representative experiment. These are tabulated on page 19 of the online supplement. The average of these ratios (variance of the bootstrap means over variance of the sample means) over the 12 models compared is 1.067. While the discrepancies appear linked to sample size, they are scarcely larger overall than what experimental error might produce, increasing confidence in our response surface variance estimate.

[7] The stationary blocks bootstrap method of Politis and Romano (1994) was also investigated, but the results were generally similar to moving blocks. The experimental results were also relatively insensitive to block length, with $[2n^{1/3}]$ also being tried.

[8] In preliminary experiments this strategy appeared to improve performance in several contexts. There are many ways to implement such a pre-test, and in a practical setting, a Q test (Box and Pierce 1970) might also be considered for this purpose. The HB1 tests are based on frequency-domain regression and do not use a pre-test.

	AR		MA		Fractional	
	Normal	t(3)	Normal	t(3)	Normal	t(3)
AFB	19	17	60	58	430	449
MBB	109	101	60	57	425	417
SAR	26	27	58	54	403	392
Asy.	255	257	126	123	550	560

Table 3: Average size distortion: regression intercept

	AR		MA		Fractional	
	Normal	$t(3)$	Normal	$t(3)$	Normal	$t(3)$
AFB	42	33	34	45	77	70
MBB	42	43	44	45	77	76
SAR	30	30	40	43	75	77
Asy.	60	62	55	58	83	84

Table 4: Average powers: regression intercept

4.2 Unit Root Tests

Alternative bootstrap tests of the unit root hypothesis have been studied by Ferretti and Romo (1996), Park (2003) and Paparoditis and Politis (2005), and in the TFT framework by Kirsch and Politis (2011 Section 6.3). See also Park (2002) and Kreiss and Paparoditis (2003) for related research on these methods. Tables 5 and 6 summarize the results obtained with the augmented Dickey-Fuller test using the AFB and alternatives, where the lag length is chosen to optimize the Akaike criterion up to a maximum of $[2n^{1/3}]$. In their experiments Kirsch and Politis (2011) apply an augmentation comparable to our own, employing a kernel estimator (see their Section.7.2). We have not attempted to replicate this version of the technique but results obtained with the basic TFT algorithm are reported in the online supplement.

The data sets generated in each replication under the null hypothesis of a unit root have the form

$$Y_t = Y_{t-1} + U_t, \quad t = 1, \ldots, n$$

where $Y_0 = 0$ and U_t is generated by one of (4.1), (4.2) or (4.3) with E_t i.i.d. drawings from either the Gaussian or standardized Student $t(3)$, as indicated in the table. To create the bootstrap draws, the data set is differenced and the differences regressed unrestrictedly on the lagged level, to form the residual series for resampling. This unrestricted filtering step is necessary to ensure the tests have power by undoing the effect of over-differencing under the alternative, although this must increase small-sample size distortion in some degree. The entries of Table 5 and 6 are the averages of the same combinations of autocorrelation and sample size as in Section 4.1, the full results being reported in the online supplement. Cases of the alternative hypothesis are generated by setting $Y_t = U_t$ so that the rejection rates are related in the obvious way to the degrees of induced autocorrelation in the series. These features of the experiment can be clearly seen in the tables of individual model results in the online supplement.

Unlike the regression tests of the previous sections which rely only on a nonparametric HAC variance estimator, the ADF test incorporates a parametric autocorrelation correction whose advantages are dramatically illustrated in Table 5. Even when the dependence is not autoregressive, the asymptotic tests are competitive with the various bootstraps in the average comparison.

	AR		MA		Fractional	
	Normal	$t(3)$	Normal	$t(3)$	Normal	$t(3)$
AFB	10	8	18	19	21	18
MBB	13	10	17	16	27	26
SAR	3.4	5.4	22	20	18	24
Asy	13	8.1	19	17	26	25

Table 5: Average size distortion: ADF tests.

	AR		MA		Fractional	
	Normal	$t(3)$	Normal	$t(3)$	Normal	$t(3)$
AFB	83	84	88	88	96	96
MBB	83	83	88	88	96	96
SAR	84	85	87	88	96	96
Asy	85	85	88	88	96	96

Table 6: Average powers: ADF tests.

5 The Multivariate Case

For simplicity's sake the analysis has thus far dealt with the case of a single autocorrelated series. In the obvious extension to multiple time series that may be cross-autocorrelated, the essential feature of the basic multivariate draw, to preserve the required characteristics, is that the same Rademacher variates must be used for each element of the process. If $\boldsymbol{X}$ $(n \times r)$ denotes the matrix whose columns are the time series of each of r variables, the FWB draw takes the form

$$\boldsymbol{X}^* = \boldsymbol{R}\boldsymbol{X} \tag{5.1}$$

where $\boldsymbol{R}$ is defined as before by (2.14).

Without loss of generality it suffices to consider the bivariate case, with $r = 2$. Two additional theorems suffice to establish the requisite properties, as follows. The cross-periodogram of $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ is in general a complex-valued process.

Theorem 5.1 *The cross-periodogram of* $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ *is identical with that of* $\boldsymbol{X}$*.*

We can calculate the means under the bootstrap distribution of the cross-autocovariances via the following corollary of Theorem 2.3.

Theorem 5.2 *Let* x_{pi} *denote the* i*th element of vector* $\boldsymbol{x}_p$*, for* $i = 0, \ldots, n-1$*, and* $p = 1, 2$*, and let* x_{pi}^* *denote the corresponding element of the bootstrap draw. For* $m \geq 0$ *and* $i \geq m$*,*

$$\mathrm{E}^*(x_{1i}^* x_{2,i-m}^*) = \frac{1}{n} \sum_{j=m}^{n-1} x_{1j} x_{2,j-m}$$

and

$$\mathrm{E}^*(x_{2i}^* x_{1,i-m}^*) = \frac{1}{n} \sum_{j=m}^{n-1} x_{2j} x_{1,j-m}.$$

These results establish the theoretical properties of the basic multivariate bootstrap, on the same basis as the univariate case. It remains to consider how best to implement the AFB procedure of Section 3. Let $\hat{\boldsymbol{\Omega}}$ denote an empirical HAC covariance matrix of $\boldsymbol{X}$ and let lower triangular

	Y_{1t}	Y_{2t}	$Y_{1,t-1}$	$Y_{2,t-1}$
Y_{1t}	1.124	0.424	0.309	0.379
Y_{2t}	0.424	1.138	0.312	0.241

Table 7: VAR covariances

Nominal significance level	0.1	0.05	0.025	0.01
Rejection rates, $\chi^2(2)$ criterion	0.1256	0.0692	0.0400	0.0194
Rejection rates, AFB criterion	0.1024	0.0511	0.0264	0.0109

Table 8: Wald test of joint significance of series means - results from 50,000 replications

$\hat{\boldsymbol{P}}$ denote its Cholesky decomposition, such that $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{P}}'$. Then, letting $\boldsymbol{Z}^*$ $(r \times 1)$ denote a standard Gaussian vector, the vector of surrogate means might be computed as $n^{-1/2}\hat{\boldsymbol{P}}\hat{\boldsymbol{D}}^{1/2}\boldsymbol{Z}^*$ where $\hat{\boldsymbol{D}}$ denotes an $r \times r$ diagonal matrix of suitable weights. A low-cost implementation of this method would be to use the univariate augmentation described in Appendix A.3, applied equation by equation, to generate the diagonal elements of $\hat{\boldsymbol{D}}$.

Systematic comparisons are not attempted in this instance, but we report evidence on the adequacy of this augmentation scheme for testing locations in a bivariate model. Experimental series of length $n = 500$ were generated by the VAR(1) model

$$Y_{1t} = 0.2Y_{1,t-1} + 0.2Y_{2,t-1} + E_{1t}$$
$$Y_{2t} = 0.3Y_{1,t-1} + 0.1Y_{2,t-1} + E_{2t}$$

where (E_{1t}, E_{2t}) are serially independent Gaussian with unit variances and contemporaneous covariance $\gamma_{12} = 0.3$. Table 7 shows the variances and zero-order and first-order covariances of the resulting series, solved from the Yule-Walker equations.

The estimation stage of the experiment tests the hypothesis that the series means are jointly zero using the asymptotically pivotal Wald statistic

$$\mathrm{W} = n\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\mu}}$$

where $\hat{\boldsymbol{\mu}}$ is the vector of sample means and $\hat{\boldsymbol{\Omega}}$ the HAC variance matrix implemented with the Parzen kernel and Newey-West (1994) plug-in bandwidth. Table 8 compares null rejection rates at various nominal significance levels, in 50,000 replications of the experiment, for two cases; the asymptotic test using the $\chi^2(2)$ critical values, and the AFB procedure as described. As before, the warp-speed Monte Carlo method has been used for the bootstrap case.

6 Concluding Remarks

We investigate a simple bootstrap for dependent processes based on a Rademacher wild bootstrap draw from the discrete Fourier transform. While similar methods have been proposed, our variant has the feature of exactly reproducing the sample periodogram in the bootstrap draws. The basic FWB algorithm does not require a choice of bandwidth or other arbitrary parameterization, a feature shared with the Hidalgo (2003) methods.

In the tests of significance of slope coefficients, the simulations show that the FWB, Hidalgo 1 and sieve autoregressive algorithm all perform comparably with autoregressive disturbances. The poorer performance of the TFT algorithm is, we conjecture, due to its dependence on a kernel periodogram estimator. Hidalgo's method does well under non-autoregressive dependence,

and possibly this is due to the fact that all the time-domain methods rely on an HAC kernel estimator. These findings suggest possible directions for refinement of the techniques.

In the location and unit root tests, the AFB and sieve autoregression again behave relatively well under autoregressive dependence. The implementation of the AFB can doubtless be refined, although the method achieves a reasonable match of sample mean distributions. As an alternative to the response surface approach to bias correction, we considered pre-whitening the kernel estimator using an AR(1) filter as suggested by Andrews and Monahan (1992). In trials this method did appear to reduce the ERP, although not as effectively as the response surface. The response surface is calibrated using autoregressive dependence, and there is certainly evidence in the tables that it performs better in this framework that when faced with the moving average cases. An improved calibration method is another possible avenue for further research.

A Appendices

A.1 Proofs

Proof of Theorem 2.1. Parts (i) and (ii) follow directly from the facts that $\boldsymbol{A\iota} = (\sqrt{n}, 0, \ldots, 0)'$ and $\boldsymbol{B\iota} = \boldsymbol{0}$. For Part (iii), using the facts that $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric and orthogonal we obtain

$$\boldsymbol{U'V} = \boldsymbol{AWAAWB} - \boldsymbol{BWBBWA}.$$

Observe that $\boldsymbol{WA}$ is a matrix obtained from $\boldsymbol{A}$ by changing the signs of certain rows, and $\boldsymbol{AW}$ is its transpose. $\boldsymbol{WAAW}$ is therefore a matrix defined in the same way as $\boldsymbol{AA}$ in (2.7), except that a number of the non-diagonal elements may have changed signs. Specifically, since $w_i^2 = 1$,

$$\{\boldsymbol{WAAW}\}_{ij} = \begin{cases} 1, & i = j = 0 \text{ and } i = j = n/2 \ (n \text{ even}) \\ 0.5, & i = j,\ j \neq 0 \text{ and } j \neq n/2 \ (n \text{ even}) \\ 0.5 w_i w_j, & i = n - j \\ 0, & \text{otherwise} \end{cases}$$

It follows that the matrix $\boldsymbol{AWAAW}$ is defined in the same manner as $\boldsymbol{A}$, except that certain pairs of columns may be replaced by zero columns. Thus, for $j \neq 0$ and $j \neq n/2$ (n even), $\{\boldsymbol{A}\}_{ij} = \{\boldsymbol{A}\}_{i,n-j}$ and hence

$$\begin{aligned} \{\boldsymbol{AWAAW}\}_{ij} &= \sum_{k=0}^{n-1} \{\boldsymbol{A}\}_{ik} \{\boldsymbol{WAAW}\}_{kj} \\ &= \{\boldsymbol{A}\}_{ij} \{\boldsymbol{WAAW}\}_{jj} + \{\boldsymbol{A}\}_{i,n-j} \{\boldsymbol{WAAW}\}_{n-j,j} \\ &= \{\boldsymbol{A}\}_{ij} \left[\{\boldsymbol{WAAW}\}_{jj} + \{\boldsymbol{WAAW}\}_{n-j,j} \right] \\ &= \tfrac{1}{2} (1 + w_{n-j} w_j) \{\boldsymbol{A}\}_{ij}. \end{aligned}$$

The rule is that the replacement by zeros of the jth and $(n-j)$th columns occurs if $w_j w_{n-j} = -1$. Further,

$$\begin{aligned} \{\boldsymbol{AWAAWB}\}_{ij} &= \frac{1}{2} \sum_{k=0}^{n-1} (1 + w_k w_{n-k}) \{\boldsymbol{A}\}_{ik} \{\boldsymbol{B}\}_{kj} \\ &= \frac{1}{2n} \sum_{k=0}^{n-1} (1 + w_k w_{n-k}) \cos(2\pi ik/n) \sin(2\pi jk/n) \\ &= \frac{1}{4n} \sum_{k=0}^{n-1} \left[\sin\left(2\pi k\left(i+j\right)/n\right) + \sin\left(2\pi k\left(i-j\right)/n\right) \right] \\ &\quad + \frac{1}{4n} \sum_{k=0}^{n-1} w_k w_{n-k} \left[\sin\left(2\pi k\left(i+j\right)/n\right) + \sin\left(2\pi k\left(i-j\right)/n\right) \right]. \end{aligned} \tag{A-1}$$

Note that the first sum of terms in the last member of (A-1) vanishes since the summands are equal and opposite in pairs, with the kth term cancelling the $n-k$th for $k \neq 0$ (n odd) and for $k \neq 0$, $n/2$ (n even). However, the sign of $w_k w_{n-k}$ is invariant under replacement of k by $n-k$. Hence, the second set of summands also cancel in pairs, in the identical manner. We conclude that $\boldsymbol{AWAAWB} = \boldsymbol{0}$. The same form of argument shows that $\boldsymbol{BWBBW}$ matches $\boldsymbol{B}$ apart from some column replacements by zero columns, according to the same rule, and that $\boldsymbol{BWBBWA} = \boldsymbol{0}$. This concludes the proof of (iii).

To prove part (iv), note that the identities $\boldsymbol{AB} = \boldsymbol{0}$, $\boldsymbol{AA} + \boldsymbol{BB} = \boldsymbol{I}$ and $\boldsymbol{WW} = \boldsymbol{I}$ imply respectively

$$\begin{aligned}\boldsymbol{U}'\boldsymbol{U} &= \boldsymbol{AWAAWA} + \boldsymbol{BWBBWB} \\ &= \boldsymbol{I} - \boldsymbol{AWBBWA} - \boldsymbol{BWAAWB}\end{aligned}$$

and

$$\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{AWBBWA} + \boldsymbol{BWAAWB}.$$

∎

Proof of Theorem 2.2. For any vector $\boldsymbol{a}$, let the notation $\boldsymbol{a}^{\circ 2}$ denote the element-wise square, the Hadamard product with itself. This is a vector with the same dimension as $\boldsymbol{a}$. Then the periodogram points of $\boldsymbol{x}$ can be written as

$$\frac{|\boldsymbol{z}|^{\circ 2}}{2\pi} = \frac{(\boldsymbol{Ax})^{\circ 2} + (\boldsymbol{Bx})^{\circ 2}}{2\pi}$$

The periodogram points of the bootstrap draw (2.14) take the corresponding form

$$\frac{(\boldsymbol{Ax}^*)^{\circ 2} + (\boldsymbol{Bx}^*)^{\circ 2}}{2\pi}$$

where

$$\boldsymbol{Ax}^* = \boldsymbol{AAW}(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{x} = \boldsymbol{P}_A\boldsymbol{x} \tag{A-2a}$$

$$\boldsymbol{Bx}^* = \boldsymbol{BBW}(\boldsymbol{B} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{P}_B\boldsymbol{x}, \tag{A-2b}$$

the second equalities defining matrices $\boldsymbol{P}_A$ and $\boldsymbol{P}_B$. Replacing $\boldsymbol{W}$ by $\tilde{\boldsymbol{W}}$ where $\tilde{\boldsymbol{W}}$ is defined following (2.4), note that these formulae become $\tilde{\boldsymbol{P}}_A = \boldsymbol{AA}\tilde{\boldsymbol{W}}\boldsymbol{A}$ and $\tilde{\boldsymbol{P}}_B = \boldsymbol{BB}\tilde{\boldsymbol{W}}\boldsymbol{B}$. Referring to the formula in (2.7), we deduce that, except for the cases $j = 0$ and $j = n/2$ (n even), the rows of the matrix $\boldsymbol{AAWA}$ have three possible forms. If $w_j = w_{n-j} = 1$, rows j and $n-j$ match rows j and $n-j$ of $\boldsymbol{A}$. If $w_j = w_{n-j} = -1$, rows j and $n-j$ match the corresponding rows of $-\boldsymbol{A}$. And, if the signs of w_j and w_{n-j} are different, the corresponding pairs of rows are zero. Rows 0 and $n/2$ (n even) match the corresponding rows of $\boldsymbol{A}$ or $-\boldsymbol{A}$, with the signs taken from w_0 and $w_{n/2}$ (n even), respectively.

On the other hand, the matrix $\boldsymbol{AAWB}$ has zeros for rows j and $n-j$ if $w_j = w_{n-j} = \pm 1$. It has jth and $n-j$th rows matching those of $\boldsymbol{B}$, if $w_j = 1$ and $w_{n-j} = -1$, and matching those of $-\boldsymbol{B}$ if $w_j = -1$ and $w_{n-j} = 1$. Rows 0 and $n/2$ (n even) are zero, matching those of $\boldsymbol{B}$. Notice how the zero rows arise in complementary positions in the two matrices.

Putting these results together, we can set out the following table where the notation identifies the jth rows of the matrices in question, for $j \neq 0$ and $j \neq n/2$ (n even), depending on the signs of the Rademacher pairs:

$$\{\boldsymbol{P}_A\}_{j\cdot} = \begin{cases} \{\boldsymbol{A}\}_{j\cdot} & \text{if } w_j = w_{n-j} = 1, \\ \{-\boldsymbol{A}\}_{j\cdot} & \text{if } w_j = w_{n-j} = -1, \\ \{\boldsymbol{B}\}_{j\cdot} & \text{if } w_j = 1 \text{ and } w_{n-j} = -1, \\ \{-\boldsymbol{B}\}_{j\cdot} & \text{if } w_j = -1 \text{ and } w_{n-j} = 1. \end{cases} \tag{A-3}$$

The last two of these cases do not arise when w_j is replaced by $\tilde{w}_j$. Also, $\{\boldsymbol{P}_A\}_{0\cdot} = w_0\{\boldsymbol{A}\}_{0\cdot}$, and $\{\boldsymbol{P}_A\}_{n/2\cdot} = w_{n/2}\{\boldsymbol{A}\}_{n/2\cdot}$ (n even).

The corresponding analysis of the matrices $\boldsymbol{BBWA}$ and $\boldsymbol{BBWB}$, making use of (2.8), yields

$$\{\boldsymbol{P}_B\}_{j\cdot} = \begin{cases} \{\boldsymbol{B}\}_{j\cdot} & \text{if } w_j = w_{n-j} = 1, \\ \{-\boldsymbol{B}\}_{j\cdot} & \text{if } w_j = w_{n-j} = -1, \\ \{-\boldsymbol{A}\}_{j\cdot} & \text{if } w_j = 1 \text{ and } w_{n-j} = -1, \\ \{\boldsymbol{A}\}_{j\cdot} & \text{if } w_j = -1 \text{ and } w_{n-j} = 1 \end{cases} \tag{A-4}$$

and also $\{\boldsymbol{P}_B\}_{0\cdot} = \boldsymbol{0}'$, and $\{\boldsymbol{P}_B\}_{n/2\cdot} = \boldsymbol{0}'$ (n even).

After squaring has eliminated the negative signs, the implication of (A-3) is that the elements of $(\boldsymbol{P}_A\boldsymbol{x})^{\circ 2}$ match either the corresponding elements of $(\boldsymbol{A}\boldsymbol{x})^{\circ 2}$ or the corresponding elements of $(\boldsymbol{B}\boldsymbol{x})^{\circ 2}$, depending on whether the Rademacher pairs match or differ. (They of course cannot differ for $\tilde{\boldsymbol{P}}_A$ and $\tilde{\boldsymbol{P}}_B$.) The elements of $(\boldsymbol{P}_B\boldsymbol{x})^{\circ 2}$ on the other hand, according to (A-4), are the complementary cases from $(\boldsymbol{B}\boldsymbol{x})^{\circ 2}$ and $(\boldsymbol{A}\boldsymbol{x})^{\circ 2}$. In other words, if an element of $(\boldsymbol{P}_A\boldsymbol{x})^{\circ 2}$ matches that of $(\boldsymbol{A}\boldsymbol{x})^{\circ 2}$, the corresponding element of $(\boldsymbol{P}_B\boldsymbol{x})^{\circ 2}$ matches that of $(\boldsymbol{B}\boldsymbol{x})^{\circ 2}$. If an element of $(\boldsymbol{P}_A\boldsymbol{x})^{\circ 2}$ matches that of $(\boldsymbol{B}\boldsymbol{x})^{\circ 2}$, the corresponding element of $(\boldsymbol{P}_B\boldsymbol{x})^{\circ 2}$ matches that of $(\boldsymbol{A}\boldsymbol{x})^{\circ 2}$. It follows that

$$(\boldsymbol{P}_A\boldsymbol{x})^{\circ 2} + (\boldsymbol{P}_B\boldsymbol{x})^{\circ 2} = (\boldsymbol{A}\boldsymbol{x})^{\circ 2} + (\boldsymbol{B}\boldsymbol{x})^{\circ 2}$$

and the theorem follows directly. ∎

Proof of Corollary 2.1. The DFT of (2.11) is identical with $\boldsymbol{z}^*$ in (2.10), noting that

$$\begin{aligned} (\boldsymbol{A} + \mathrm{i}\boldsymbol{B})\,[(\boldsymbol{AWA} + \boldsymbol{BWB}) + \mathrm{i}(\boldsymbol{AWB} - \boldsymbol{BWA})] \\ = (\boldsymbol{AA} + \boldsymbol{BB})\boldsymbol{WA} + \mathrm{i}(\boldsymbol{AA} + \boldsymbol{BB})\boldsymbol{WB} \\ = \boldsymbol{W}(\boldsymbol{A} + \mathrm{i}\boldsymbol{B}). \end{aligned}$$ ∎

Proof of Theorem 2.3. The result $\mathrm{E}^*(x_i^*) = 0$ is immediate since $\mathrm{E}^*(w_k) = 0$. Using formula (2.18), note that for $i \geq m$

$$x_i^* x_{i-m}^* = \sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\sum_{j'=0}^{n-1}\sum_{k'=0}^{n-1} w_k w_{k'} \Upsilon_{ijk}\Upsilon_{i-m,j'k'} x_j x_{j'}. \tag{A-5}$$

Since $\mathrm{E}^*(w_k w_{k'}) = \begin{cases} 1, & k = k' \\ 0, & k \neq k' \end{cases}$, we have

$$\begin{aligned} \mathrm{E}^*(x_i^* x_{i-m}^*) &= \sum_{j=0}^{n-1}\sum_{j'=0}^{n-1} x_j x_{j'} \sum_{k=0}^{n-1} \Upsilon_{ijk}\Upsilon_{i-m,j',k} \\ &= \sum_{j=m}^{n-1} x_j x_{j-m} \sum_{k=0}^{n-1} \Upsilon_{ijk}\Upsilon_{i-m,j-m,k} + \sum_{j=0}^{n-1}\sum_{j'\neq j-m} x_j x_{j'} \sum_{k=0}^{n-1} \Upsilon_{ijk}\Upsilon_{i-m,j',k} \\ &= \frac{1}{n}\sum_{j=m}^{n-1} x_j x_{j-m}. \end{aligned} \tag{A-6}$$

Here, the third equality makes use of the facts that

$$\sum_{k=0}^{n-1} \Upsilon_{ijk}\Upsilon_{i-m,j-m,k} = \frac{1}{n^2}\sum_{k=0}^{n-1} \left[\cos^2(2\pi(i-j)k/n) + \sin^2(2\pi(i-j)k/n)\right]$$

$$- 2\frac{1}{n^2}\sum_{k=0}^{n-1}\cos(2\pi(i-j)k/n)\sin(2\pi(i-j)k/n)$$
$$= \frac{1}{n} \tag{A-7}$$

in view of (2.19) and $\cos^2 + \sin^2 = 1$, and also that for each i and $j' \neq j - m$,

$$\begin{aligned}\sum_{k=0}^{n-1}\Upsilon_{ijk}\Upsilon_{i-m,j'k} &= \sum_{k=0}^{n-1}\cos(2\pi(i-j)k/n)\cos(2\pi(i-m-j')k/n)\\ &\quad + \sum_{k=0}^{n-1}\sin(2\pi(i-j)k/n)\sin(2\pi(i-m-j')k/n)\\ &\quad - \sum_{k=0}^{n-1}\cos(2\pi(i-j)k/n)\sin(2\pi(i-m-j')k/n)\\ &\quad - \sum_{k=0}^{n-1}\sin(2\pi(i-j)k/n)\cos(2\pi(i-m-j')k/n)\\ &= \sum_{k=0}^{n-1}\cos(2\pi(m-j+j')k/n) - \sum_{k=0}^{n-1}\sin(2\pi(2i-m-j-j')k/n)\\ &= 0, \end{aligned} \tag{A-8}$$

noting that $m - j + j'$ and $2i - m - j - j'$ are always nonzero integers. ∎

Proof of Theorem 2.4. Let $\boldsymbol{a}_0, \ldots, \boldsymbol{a}_{n-1}$ and $\boldsymbol{b}_0, \ldots, \boldsymbol{b}_{n-1}$ denote the columns of symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. Since $\boldsymbol{W}$ (omitting for clarity the subscript denoting the draw) is diagonal,

$$\begin{aligned}\boldsymbol{R} &= (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{W}\boldsymbol{A} + (\boldsymbol{A}+\boldsymbol{B})\boldsymbol{W}\boldsymbol{B}\\ &= \sum_{j=0}^{n-1} w_j \boldsymbol{D}_j\end{aligned}$$

where

$$\boldsymbol{D}_j = (\boldsymbol{a}_j - \boldsymbol{b}_j)\boldsymbol{a}_j' + (\boldsymbol{a}_j + \boldsymbol{b}_j)\boldsymbol{b}_j',\ j = 0, \ldots, n-1$$

are $n \times n$ matrices. The ith element of $\boldsymbol{x}^*$ is therefore

$$x_i^* = \sum_{j=0}^{n-1} w_j g_{ji} \tag{A-9}$$

where g_{ji} is the ith element of

$$\boldsymbol{g}_j = \boldsymbol{D}_j\boldsymbol{x}\ \ (n \times 1),$$

conditionally fixed under the bootstrap distribution. Since

$$w_j^r = \begin{cases} 1 & r \text{ even} \\ w_j & r \text{ odd,} \end{cases}$$

for r any positive integer, there exists the decomposition

$$x_i^{*r} = A_i^{(r)} + \sum_{j=0}^{n-1} w_j B_{ij}^{(r)} + \sum_{j=0}^{n-1}\sum_{k\neq j} w_j w_k C_{ijk}^{(r)} \tag{A-10}$$

where $A_i^{(r)} = \mathrm{E}(x_i^{*r})$ consists of the sum of those terms (if any) containing only even powers of $w_0, ..., w_{n-1}$, the $B_{ij}^{(r)}$ are the sums of the terms (if any) containing an odd power of either w_j or w_k, and $C_{ijk}^{(r)}$ is the sum of those terms (if any) containing odd powers of both w_j and w_k. Since $\mathrm{E}(w_j) = 0$, it follows that when $w_{a0}, \ldots, w_{a,n-1}, w_{b0}, \ldots, w_{b,n-1}$ are drawn independently,

$$\mathrm{E}^* \left(x_{ai}^{*r} x_{bk}^{*s} \right) = A_i^{(r)} A_k^{(s)}, \quad i, k = 0, \ldots, n-1$$

and the conclusion follows. For the case where w_j is replaced by $\tilde{w}_j$ as in (2.5), the argument is amended only by noting that the decomposition of (A-10) assigns the terms differently, and the sums run over $j = 0, ..., n/2 - 1$ (n even) and $j = 0, ..., (n-1)/2$ (n odd). ∎

Proof of Theorem 2.5 (i) $r_{ij} = \sum_{k=0}^{n-1} w_k \Upsilon_{ijk}$ where $\mathrm{E}^*(w_k) = 0$, $\mathrm{E}^*(w_k^2) = 1$ and $\mathrm{E}^*(w_k w_{k'}) = 0$ for $k \neq k'$. Hence, $\mathrm{E}^*(r_{ij}) = 0$ and $\mathrm{E}^*(r_{ij}^2) = 1/n$ as a consequence of (A-7) with $m = 0$. The theorem follows since the terms of the sum are independent with bounded variances. (ii) Consider a finite collection of row coordinates $j_1, \ldots, j_m$ with associated fixed finite weights $\alpha_1, \ldots, \alpha_m$. The random variable

$$\sqrt{n} \sum_{p=1}^{m} \alpha_p r_{ij_p} = \sqrt{n} \sum_{k=0}^{n-1} w_k \sum_{p=1}^{m} \alpha_p \Upsilon_{ij_p k},$$

has mean zero and variance $n \sum_{k=0}^{n-1} (\sum_{p=1}^{m} \alpha_p \Upsilon_{ij_p k})^2 = O(1)$, and is asymptotically Gaussian by the argument for part (i). The distribution of the m-fold collection $\sqrt{n} r_{ij_1}, \ldots, \sqrt{n} r_{ij_m}$ therefore becomes multivariate Gaussian as $n \to \infty$ by the Cramér-Wold theorem (Davidson 1994, Theorem 25.5). The Gaussianity of the limit sequence as a whole, subject to a mild consistency condition that certainly holds in the present case, is a consequence of the Kolmogorov consistency theorem (Davidson 1994, Theorem 12.4). ∎

Proof of Theorem 2.6. The terms Υ_{ijk} in (2.18) are bounded absolutely by $2/n$. Hence, since $\mathrm{E}^*(w_k) = 0$, $r_{ij} = O_p(n^{-1/2})$ as $n \to \infty$, where this rate of convergence holds almost surely with respect to the distribution of the bootstrap draws by the normal number theorem. With the bootstrap draw conditionally fixed, the marginal Gaussianity of the quantities x_i^* for each i follows under appropriate regularity conditions by a central limit theorem for dependent triangular arrays such as de Jong (1997) Theorem 2. The moment and dependence properties of the data are all that matter for this result. The bounded weights defined in (2.18), treated as conditionally fixed, are absorbed trivially by choice of the scaling constants specified in the cited theorem. A linear combination of any finite collection of points from a bootstrap draw, say $x_{i_p}^* = \boldsymbol{r}_{i_p}' \boldsymbol{x}$ for $p = 1, \ldots, m$ with finite weights $\{\alpha_1, \ldots, \alpha_m\}$, is also a linear combination $\sum_{p=1}^{m} \alpha_p \boldsymbol{r}_{i_p}' \boldsymbol{x}$ of $\boldsymbol{x}$ with compound weights summing to either $\sum_{p=1}^{m} \alpha_p$ or $-\sum_{p=1}^{k} \alpha_p$ by Theorem 2.1. The weights are bounded absolutely by $n^{-1/2} \sum_{p=1}^{m} |\alpha_p| < \infty$. The limiting Gaussianity of the finite dimensional distributions of the sequence, and hence the limiting Gaussianity of the sequence as a whole, follows by the same arguments as in Theorem 2.5. ∎

Proof of Theorem 2.7 Since $\mathrm{E}^*(x_i^*) = 0$, (2.24) implies

$$\mathrm{E}^*(\hat{\gamma}_{nm}^*) = \frac{1}{n} \sum_{i=m}^{n-1} \mathrm{E}^*(x_i^* x_{i-m}^*) = \frac{n-m}{n^2} \sum_{j=m}^{n-1} x_j x_{j-m} \xrightarrow{pr} \gamma_m$$

as in (2.26), where the convergence in probability is with respect to the distribution of the sample data as $n \to \infty$. This is a standard application of the law of large numbers (ergodic theorem) in view of the fact that $\mathrm{E}(x_j) = 0$ by assumption.

Next, consider the error of estimate under the bootstrap distribution. This takes the form

$$\begin{aligned}\hat{\gamma}^*_{nm} - \mathrm{E}^*(\hat{\gamma}^*_{nm}) &= \frac{1}{n}\sum_{i=m}^{n-1}\big(x^*_i x^*_{i-m} - \mathrm{E}^*(x^*_i x^*_{i-m})\big) \\ &= \frac{1}{n}\sum_{i=m}^{n-1}\sum_{j=0}^{n-1}\sum_{j'=0}^{n-1} x_j x_{j'} G_{im,j,j'} \qquad \text{(A-11)}\end{aligned}$$

where

$$G_{im,j,j'} = \sum_{k=0}^{n-1}\sum_{k'\neq k} w_k w_{k'} \Upsilon_{ijk}\Upsilon_{i-m,j',k'}.$$

Note how these terms complement (A-6) with the cases $k \neq k'$.

Since $\mathrm{E}^*(w_k w_{k'} w_l w_{l'}) = 1$ if either $k = l$ and $k' = l'$ or $k = l'$ and $k' = l$, and otherwise is zero, substituting the trigonometric identities corresponding to (A-7) and (A-8) yields

$$\begin{aligned}\mathrm{E}^*(G^2_{im,j,j'}) &= \sum_{k=0}^{n-1}\sum_{k'\neq k}\sum_{l=0}^{n-1}\sum_{l'\neq l} \mathrm{E}^*(w_k w_{k'} w_l w_{l'})\Upsilon_{ijk}\Upsilon_{i-m,j',k'}\Upsilon_{ijl}\Upsilon_{i-m,j',l'} \\ &= \sum_{k=0}^{n-1}\Upsilon^2_{ijk}\sum_{l=0}^{n-1}\Upsilon^2_{i-m,j',l} + \sum_{k=0}^{n-1}\Upsilon_{ijk}\Upsilon_{i-m,j',k}\sum_{l=0}^{n-1}\Upsilon_{ij',l}\Upsilon_{i-m,j,l} \\ &= \frac{1}{n^2}.\end{aligned}$$

On the other hand, for $p \neq i$ and/or $q \neq j$, $q' \neq j'$,

$$\begin{aligned}\mathrm{E}^*(G_{im,j,j'}G_{pm,q,q'}) &= \sum_{k=0}^{n-1}\sum_{k'\neq k}\sum_{l=0}^{n-1}\sum_{l'\neq l} \mathrm{E}^*(w_k w_{k'} w_l w_{l'})\Upsilon_{ijk}\Upsilon_{i-m,j',k'}\Upsilon_{pql}\Upsilon_{p-m,q',l'} \\ &= \sum_{k=0}^{n-1}\Upsilon_{ijk}\Upsilon_{pqk}\sum_{l=0}^{n-1}\Upsilon_{i-m,j',l}\Upsilon_{p-m,q',l} + \sum_{k=0}^{n-1}\Upsilon_{ijk}\Upsilon_{p-m,q',k}\sum_{l=0}^{n-1}\Upsilon_{pql}\Upsilon_{i-m,j',l} \\ &= 0\end{aligned}$$

holds similarly to (A-8), in each case. The variance of (A-11) under the bootstrap distribution therefore takes the form

$$\mathrm{E}^*\big(\hat{\gamma}^*_{nm} - \mathrm{E}^*(\hat{\gamma}^*_{nm})\big)^2 = \frac{n-m}{n^2}\left(\frac{1}{n}\sum_{j=0}^{n-1} x_j^2\right)^2 = O_p(n^{-1})$$

corresponding to (2.27), where the stochastic order of magnitude is defined with respect to the distribution of the sample data, and follows from the assumption that the series is stationary and ergodic with finite second moment. ∎

Proof of Theorem 3.1 It is sufficient for the specified weak limit to hold that (a) the term $\sqrt{n}\sum_{t=1}^{n} l_{nt}u^*_{nt}$ from (3.7) converges to a normal limit with zero mean and variance V, and (b) $\hat{V}^*_n \to_{\mathrm{pr}} V$.

Consider for any $0 \leq i \leq n-1$ u^*_i, the i^{th} bootstrap disturbance obtained from the sample least squares residuals $\hat{\boldsymbol{u}}$. Letting $\boldsymbol{r}'_i$ denote the i^{th} row of $\boldsymbol{R}$,

$$u^*_i = \boldsymbol{r}'_i\hat{\boldsymbol{u}} = \boldsymbol{r}'_i\boldsymbol{u} - \boldsymbol{r}'_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$$

$$= \boldsymbol{r}_i'\boldsymbol{u} - \frac{\boldsymbol{r}_i'\boldsymbol{X}}{\sqrt{n}}\Big(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\Big)^{-1}\frac{\boldsymbol{X}'\boldsymbol{u}}{\sqrt{n}} \tag{A-12}$$

Noting that the elements of $\boldsymbol{r}_i$ are $O_p(n^{-1/2})$ as shown in Theorem 2.6, the right-hand side terms in (A-12) are $O_p(1)$ and asymptotically jointly Gaussian under our assumptions. According to Theorems 2.3 and 2.7 and the assumptions,

$$\begin{aligned}\gamma^*_{nm} = \mathrm{E}^*(u^*_i u^*_{i-|m|}) &= \frac{1}{n}\sum_{j=|m|}^{n-1}\hat{u}_j\hat{u}_{j-|m|} \\ &= \frac{1}{n}\sum_{j=|m|}^{n-1}u_j u_{j-|m|} + O_p(n^{-1}) \\ &\xrightarrow{\text{pr}} \gamma_m. \end{aligned} \tag{A-13}$$

Let $\boldsymbol{\Gamma}$ $(n\times n)$ denote the Toeplitz matrix with elements γ_m on the m^{th} diagonals and let $\boldsymbol{l}$ $(n\times 1)$ be defined by (3.2). It follows from (A-13) that

$$n\mathrm{E}^*(\boldsymbol{l}'\boldsymbol{u}^*\boldsymbol{u}^{*\prime}\boldsymbol{l}) = n\boldsymbol{l}'\boldsymbol{\Gamma}\boldsymbol{l} + O_p(n^{-1}).$$

and $\operatorname{plim} n\boldsymbol{l}'\boldsymbol{\Gamma}\boldsymbol{l} = V$, as defined in (3.5). It follows from Theorem 2.6 with the Cramér and Slutsky Theorems (Davidson 1994, Th.18.10 and Th.22.14) that $\sqrt{n}\boldsymbol{l}'\boldsymbol{u}^* \xrightarrow{\text{d}} \mathrm{N}(0,V)$, the convergence holding with probability 1 under the bootstrap distribution.

Next, consider (b). The variance of $\sqrt{n}\sum_{t=1}^n l_{nt}u^*_{nt}$ under the bootstrap distribution, conditional on the sample, is

$$V^*_n = \sum\nolimits_{m=1-n}^{n-1}\gamma^*_{nm} n \sum\nolimits_{t=|m|+1}^{n} l_{nt}l_{n,t-|m|}$$

where γ^*_{nm} is the quantity defined in (A-13) so that $\operatorname{plim}\gamma^*_{nm} = \gamma_m$ for fixed $m \geq 0$. We show that $\operatorname{plim} V^*_n = V$ defined by (3.5), where the probability limit of the conditional mean is taken with respect to the sampling distribution of the data as $n\to\infty$.

For clarity of notation write $Q_{nm} = n\sum_{t=|m|+1}^n l_{nt}l_{n,t-|m|}$ and $\bar{Q}_m = \operatorname{plim} Q_{nm}$, which is finite by the assumption of weak dependence (Assumption 1). Then let $V^*_n = T_1 + 2T_2$ where for some fixed $M < \infty$, $T_1 = \sum_{-M}^{M}\gamma^*_{nm}Q_{nm}$ and $T_2 = \sum_{m=M+1}^{n-1}\gamma^*_{nm}Q_{nm}$. It is easily seen that $\operatorname{plim} T_1 = \sum_{-M}^{M}\gamma_m\bar{Q}_m$. Next, let $\gamma^*_{nm} = \gamma_m + u_{mn}$ and $Q_{nm} = \bar{Q}_m + v_{nm}$ where $\gamma_m = O(m^{-1-\delta_\gamma})$ by Assumption 1, with $\delta_\gamma > 0$ and $\bar{Q}_m = O(m^{-1-\delta_Q})$ with $\delta_Q > 0$. $u_{mn} = O(n^{-1/2})$ and $v_{mn} = O(n^{-1/2})$ are the errors of estimate of the two components, and are independently distributed, by Assumption 3, with means of zero. Then, we find

$$T_2 = \sum\nolimits_{m=M+1}^{n-1} u_{nm}v_{nm} + O(M^{-\min\{\delta_\gamma,\delta_Q\}}) \tag{A-14}$$

where the products $u_{nm}v_{nm}$ have means of zero and are $O(n^{-1})$, which implies that the first term of (A-14) is $o(1)$ as $n\to\infty$. Hence, T_2 can be made as small as desired by taking M and n large enough.

By contrast the expectation of (3.8) under the bootstrap distribution conditional on $\boldsymbol{X}$ is

$$\mathrm{E}^*(\hat{V}^*_n) = \sum_{m=-M_n}^{M_n} w_{mn}\gamma^*_{nm} n \sum_{t=|m|+1}^{n} l_{nt}l_{n,t-|m|}.$$

Given the probability limits of the $\hat{\gamma}^*_{nm}$ according to Theorem 2.7, an appropriate choice of kernel weights such that $w_{nm}\to 1$ for fixed m ensures that $\operatorname{plim}\hat{V}^*_n = V$ as required. ∎

Proof of Theorem 3.2 The assumptions of Theorem 2.6 are sufficient for the operation of the requisite functional central limit theorem for dependent processes to hold for the FWB process $\{u_t^*\}$. The finite dimensional distributions of the limit process Y^{**} are Gaussian, in view of Theorem 2.6 and the fact that Z^* is an independent Gaussian drawing, unrelated to sample size. Tightness of the limit distribution follows from the assumptions, since the bootstrap series are stationary and have autocorrelation structure matching that of the original data in the limit by Theorem 2.7, Therefore, Y^{**} is continuous with probability 1. The augmentation by the independent shift variable does not affect these features of the process.

It remains to establish the covariance properties of the augmented limit process. It is required that (a) for $0 \le r < s \le 1$, $\mathrm{E}(Y^{**}(s) - Y^{**}(r))^2 = \omega^2(s-r)$, and (b) non-overlapping increments are uncorrelated, and hence independent. We show that these properties hold for the process Y_n^{**} apart from components that are of small order as $n \to \infty$. The main step is to note that an arbitrary random variable added to each coordinate $u_{n1}^*, \ldots, u_{nn}^*$ automatically defines the sample mean of the series. Hence there always exists a companion process $\{v_1, \ldots, v_n\}$, having mean zero, such that $u_{nt}^* = v_t - \bar{v}_n$ where $\bar{v}_n = n^{-1}\sum_{t=1}^n v_t$. The series $\{v_t\}$ can be constructed to share all the properties of $\{u_{nt}^*\}$ apart from the centering, in particular the asymptotic Gaussianity and the autocorrelation structure. In particular,

$$\frac{1}{n}\mathrm{E}\Big(\sum_{t=[nr]+1}^{[ns]} v_t\Big)^2 = \omega^2(s-r) + O(n^{-1}) \tag{A-15}$$

where the order of magnitude of the remainder reflects the fact that the tail autocovariances are summable. Thus,

$$\begin{aligned}
\frac{1}{n}\mathrm{E}\big(y_{[ns]}^{**} - y_{[nr]}^{**}\big)^2 &= \frac{1}{n}\mathrm{E}\Big(\sum_{t=[nr]+1}^{[ns]} (v_t - \bar{v}_n + n^{-1/2}\omega Z^*)\Big)^2 \\
&= \frac{1}{n}\mathrm{E}\Big(\sum_{t=[nr]+1}^{[ns]} v_t - ([ns]-[nr])\,\bar{v}_n + n^{-1/2}\,([ns]-[nr])\,\omega Z^*\Big)^2 \\
&= \omega^2(s-r) + O(n^{-1}).
\end{aligned} \tag{A-16}$$

To show the last equality of (A-16), note that of the nine terms represented by the second member, the first square term is (A-15) and the other two squares are

$$\frac{([ns]-[nr])^2}{n^2}\mathrm{E}(\bar{v}_n^2) = \omega^2(s-r)^2 + O(n^{-1})$$

and

$$\frac{([ns]-[nr])^2}{n^2}\omega^2\mathrm{E}(Z^{*2}) = \omega^2(s-r)^2,$$

while two of the cross-products have the form

$$-\frac{[ns]-[nr]}{n^2}\mathrm{E}\Big(\sum_{t=1}^{n} v_t \sum_{t=[nr]+1}^{[ns]} v_t\Big) = -\omega^2(s-r)^2 + O(n^{-1})$$

and the other four are zero, since Z^* and $v_1, \ldots, v_n$ are mutually independent.

To show that property (b) holds for Y_n^{**}, the covariance of non-overlapping intervals with $r_1 < s_1 \le r_2 < s_2$ is

$$n^{-1}\mathrm{E}(y_{[ns_1]}^{**} - y_{[nr_1]}^{**})(y_{[ns_2]}^{**} - y_{[nr_2]}^{**})^2 = O(n^{-1}),$$

noting the cancellation of the terms $2\omega^2(s_1 - r_1)(s_2 - r_2)$ and $-2\omega^2(s_1 - r_1)(s_2 - r_2)$ similarly to (A-16). This completes the proof. ∎

Proof of Theorem 5.1. Similarly to the proof of Theorem 2.2, the cross-periodogram points of $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ can be written as the pair of vectors $(2\pi)^{-1}\boldsymbol{z}_1 \circ \boldsymbol{z}_2^\dagger$ and $(2\pi)^{-1}\boldsymbol{z}_2 \circ \boldsymbol{z}_1^\dagger$ where the $\boldsymbol{z}_j$ are defined by (2.9) with appropriate substitutions, '†' denotes the complex conjugate and '∘' denotes the element-wise Hadamard product of the two vectors. Thus,

$$\boldsymbol{z}_1 \circ \boldsymbol{z}_2^\dagger = \boldsymbol{A}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{B}\boldsymbol{x}_1 \circ \boldsymbol{B}\boldsymbol{x}_2 + \mathrm{i}(\boldsymbol{B}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2 - \boldsymbol{A}\boldsymbol{x}_1 \circ \boldsymbol{B}\boldsymbol{x}_2) \tag{A-17}$$

and $\boldsymbol{z}_2 \circ \boldsymbol{z}_1^\dagger$ is the corresponding expression with reversed sign of the imaginary component. Then, from (5.1) using (A-2a) and (A-2b), the cross-periodogram points of the bootstrap draw take the forms of

$$\boldsymbol{P}_A\boldsymbol{x}_1 \circ \boldsymbol{P}_A\boldsymbol{x}_2 + \boldsymbol{P}_B\boldsymbol{x}_1 \circ \boldsymbol{P}_B\boldsymbol{x}_2 + \mathrm{i}(\boldsymbol{P}_B\boldsymbol{x}_1 \circ \boldsymbol{P}_A\boldsymbol{x}_2 - \boldsymbol{P}_A\boldsymbol{x}_1 \circ \boldsymbol{P}_B\boldsymbol{x}_2) \tag{A-18}$$

and of its complex conjugate.

We now show, reprising the arguments in the proof of Theorem 2.2, that both the real and the imaginary parts of (A-18) are identical with those of (A-17). Consider the real terms first. Recall that the rows of $\boldsymbol{P}_A$ follow the scheme in (A-3) and the rows of $\boldsymbol{P}_B$ follow (A-4) similarly. The sign changes disappear in the Hadamard products, similarly to the cases of the squares except that here the signs are not always positive. What matters is that the signs of the elements of $\boldsymbol{A}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2$ must match those of $(-\boldsymbol{A})\boldsymbol{x}_1 \circ (-\boldsymbol{A})\boldsymbol{x}_2$, and similarly for $\boldsymbol{B}$. We therefore conclude that

$$\boldsymbol{P}_A\boldsymbol{x}_1 \circ \boldsymbol{P}_A\boldsymbol{x}_2 + \boldsymbol{P}_B\boldsymbol{x}_1 \circ \boldsymbol{P}_B\boldsymbol{x}_2 = \boldsymbol{A}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{B}\boldsymbol{x}_1 \circ \boldsymbol{B}\boldsymbol{x}_2$$

This result holds whether or not $\boldsymbol{x}_1 = \boldsymbol{x}_2$, which was the case shown previously.

Now consider the imaginary parts in the light of tables (A-3) and (A-4). We find that the elements of the vectors $\boldsymbol{P}_B\boldsymbol{x}_1 \circ \boldsymbol{P}_A\boldsymbol{x}_2$ and $\boldsymbol{P}_A\boldsymbol{x}_1 \circ \boldsymbol{P}_B\boldsymbol{x}_2$ match those of the respective vectors $\boldsymbol{B}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2$ and $\boldsymbol{A}\boldsymbol{x}_1 \circ \boldsymbol{B}\boldsymbol{x}_2$ when the signs of the Rademacher pairs agree. When the Rademacher pairs take opposite signs, the corresponding elements of $\boldsymbol{P}_B\boldsymbol{x}_1 \circ \boldsymbol{P}_A\boldsymbol{x}_2$ and $\boldsymbol{P}_A\boldsymbol{x}_1 \circ \boldsymbol{P}_B\boldsymbol{x}_2$ match, respectively, those of $-(\boldsymbol{A}\boldsymbol{x}_2 \circ \boldsymbol{B}\boldsymbol{x}_1)$ and $-(\boldsymbol{B}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2)$. It follows that

$$\boldsymbol{P}_B\boldsymbol{x}_1 \circ \boldsymbol{P}_A\boldsymbol{x}_2 - \boldsymbol{P}_A\boldsymbol{x}_1 \circ \boldsymbol{P}_B\boldsymbol{x}_2 = \boldsymbol{B}\boldsymbol{x}_1 \circ \boldsymbol{A}\boldsymbol{x}_2 - \boldsymbol{A}\boldsymbol{x}_1 \circ \boldsymbol{B}\boldsymbol{x}_2. \tag{A-19}$$

Since the same equalities holds for the complex conjugate, which merely changes the sign on both sides of (A-19), the proof is complete. ∎

Proof of Theorem 5.2. This follows immediately by the arguments of Theorem 2.3. Simply note that

$$x^*_{pi} = \sum_{j=0}^{n-1}\sum_{k=0}^{n-1} w_k \Upsilon_{ijk} x_{pj}$$

for $p = 1$ and $p = 2$. Accordingly, replace $x_j x_{j'}$ by $x_{1j}x_{2j'}$ in all the expressions where the latter product appears, and likewise replace $x^*_j x^*_{j-m}$ by $x^*_{1j}x^*_{2,j-m}$ and $x_j x_{j-m}$ by $x_{1j}x_{2,j-m}$. With these amendments, the derivation proceeds unchanged. Then, interchange subscripts 1 and 2 for the second part of the theorem. ∎

A.2 Warp-speed Monte Carlo

For each of K Monte Carlo replications, the procedure is to generate a sample data set, calculate and store the test statistic, and then take a *single* bootstrap draw, compute the matching test statistic from the bootstrap sample and also store this. At the termination of the experiment,

sort the latter set and for $k = 1, \ldots, K$ estimate the p-value by the position of the k^{th} sample statistic in the bootstrap distribution. The empirical rejection rate in an α-level test is estimated by

$$\hat{P}(\alpha) = K^{-1} \sum_{k=1}^{K} I(p_k \leq \alpha) \tag{A-20}$$

where $I(.)$ denotes the indicator function and p_k the warp-speed estimator of the p-value in the kth Monte Carlo replication.

The validity of the warp speed experiment depends on the fact that the conditional bootstrap draws have a common unconditional distribution. Formally, let ξ represent the Monte Carlo replicate and let ζ represent the random drawing that conditionally generates the bootstrap data. Let $t = t(\xi)$ denote the sample statistic with distribution $F(x) = P(t \leq x)$ and $t^* = t^*(\xi, \zeta)$ the bootstrap statistic. The conventional bootstrap estimates the random measure $F_\xi(x) = P(t^* \leq x|\xi)$ and locates t in this distribution to estimate the p-value, defined either as

$$g(t) = 1 - F_\xi(t) \tag{A-21}$$

in the case of a one-tailed test, or as

$$g(t) = 2\min(F_\xi(t), 1 - F_\xi(t)) \tag{A-22}$$

in the case of a two-tailed test. If F and F_ξ match and are continuous, then g is uniform on [0,1]. By contrast, the warp-speed Monte Carlo procedure estimates the distributions F and $F^* = \mathrm{E}(F_\xi)$. The warp-speed p-values are $g^*(t)$, defined by (A-21) or (A-22) with F^* replacing F_ξ in the formulae. If the bootstrap is valid and F_ξ does not depend on ξ, F^* and F_ξ match and hence F and F^* match. If $F_\xi = F$ with probability 1 and F is continuous then g^* is uniform on $[0, 1]$.

The one caveat is that it is not impossible to have F matching F^* in spite of bootstrap failure with positive probability, if the deviations of F_ξ from F average to zero. For example, we cannot predict how the warp-speed method might perform if the data have no variance, so that the distribution depends on ξ even in the limit (Athreya 1987). Excepting such cases, warp-speed experiments are a reliable technique of bootstrap evaluation.

A.3 Constructing the Surrogate Mean

Sets of Monte Carlo experiments were run, with different sample sizes and different patterns of autocorrelation, using data generated from the model

$$Y_t = \mu + V_t, \;\; V_t = \rho V_{t-1} + U_t, \;\; U_t \sim \mathrm{NI}(0, 1), \quad t = 1, \ldots, n \tag{A-23}$$

where the null hypothesis $\mu = 0$ is true with various values assigned to ρ and n, 24 cases in total represented by the combinations of $n = \{50, 100, 200, 400, 800, 1600\}$ and $\rho = \{0, 0.3, 0.6, 0.9\}$. For each of these cases, a sequence of warp speed Monte Carlo experiments with $K = 5000$ replications was run on AFB tests of the null hypothesis $\mu = 0$, as described in Section 3. For each replication an HAC variance estimator $\hat{\omega}^2$ was computed from the mean deviations $Y_t - \bar{Y}$ using the Bartlett kernel with bandwidth chosen automatically by the Newey and West (1994) plug-in method. The AFB test was then implemented with surrogate mean set to $R_i\hat{\omega}^2 Z^*/\sqrt{n}$ where R_i is a trial correction factor. A sequence of such experiments for $i = 0, 1, 2, \ldots$ was run with $R_i = 1.1R_{i-1}$, starting with $R_0 = 0.5$.

For each experiment the Cramer-von Mises goodness-of-fit criterion

$$\mathrm{CvM} = K^{-1} \sum_{k=1}^{K} \left(\hat{P}(k/K) - k/K\right)^2 \tag{A-24}$$

was computed for the p-value distribution where $\hat{P}(\alpha)$ is defined by (A-20). The sequence was terminated at the point where CvM was observed to increase at three successive grid points, and the value of R yielding the minimum of CvM over the grid was recorded. These calibration experiments yielded 24 triples, (R, n, ρ).

The problem is to construct a response surface making use of the data available in a practical testing situation. The residual variance of the regression, $\hat{\sigma}_V^2 = n^{-1}\sum_{t=1}^n \hat{V}_t^2$, would estimate $\sigma_U^2/(1-\rho^2)$ were the data actually generated by (A-23). Measuring scale by the residual variance $\hat{\sigma}_U^2$ from a sieve autoregression of $\hat{V}_t$,[9] a feasible measure of autocorrelation of $\hat{V}_t$ is [10]

$$\eta = \hat{\sigma}_V^2/\hat{\sigma}_U^2. \tag{A-25}$$

5000 replications of $\hat{\sigma}_V^2$ and $\hat{\sigma}_U^2$ are generated from the calibration runs, and their averages were used to compute η. A trans-log regression fitted to the data points (R, n, η) yields

$$\begin{aligned}\hat{R}(n,\eta) = \exp\{&4.33337 - 2.07486\log n + 0.3395(\log n)^2 - 0.01868(\log n)^3 \\ &+ 4.05463\log\eta + 0.56462(\log\eta)^2 - 0.30931(\log\eta)^3 \\ &- 1.08039\log n\log\eta + 0.01368\log n(\log\eta)^2 \\ &+ 0.07224(\log n)^2\log\eta\}.\end{aligned} \tag{A-26}$$

In experiments with the AFB the surrogate mean is generated as $\hat{R}(n,\eta)\hat{\omega}^2 Z^*/\sqrt{n}$ with η computed by (A-25) from the sample residuals and their sieve-AR filtered counterpart.

References

[1] Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59 (3) 817-858.

[2] Andrews, D. W. K. and J. C. Monahan (1992) An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953-66.

[3] Athreya, K. (1987) Bootstrap of the mean in the infinite variance case, *Annals of Statistics* 15, 724-731.

[4] Bühlmann, P. (1997) Sieve bootstrap for time series. *Bernoulli* 3, 123-148.

[5] Davidson, J. (1994) *Stochastic Limit Theory,* Oxford: Oxford University Press.

[6] de Jong, R. M. (1997) Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353-367.

[7] de Jong, R. M. and J. Davidson (2000) Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices, *Econometrica* 68, 407-423.

[8] Efron, B. (1979), 'Bootstrap methods: another look at the jackknife', *The Annals of Statistics* 7(1), 1-26.

[9] The same procedure as used to implement the Buhlmann (1997) bootstrap described in Section 4.

[10] Another approach considered was to fit the response surface to $\hat{\omega}_V^2/\hat{\sigma}_V^2$ where $\hat{\omega}_V^2$ is the HAC variance estimator, but this proved less effective in practice.

[9] Ferretti, N. and J. Romo (1996) Unit root bootstrap tests for AR(1) models, *Biometrika* 83(4) 849-860.

[10] Giacomini, R., D. Politis and H. White (2013) A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* 29, 567-589.

[11] Hall, P. and C. C. Heyde (1980) *Martingale Limit Theory and its Application*, London: Academic Press.

[12] Hidalgo, J. (2003) An alternative bootstrap to moving blocks for time series regression models. *Journal of Econometrics* 117, 369-399.

[13] Kirch, C. and D. N. Politis (2011), TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *The Annals of Statistics* 39 (3), 1427-1470.

[14] Kreiss, J-P. and E. Paparoditis (2003) Autoregressive-aided periodogram bootstrap for time series, *The Annals of Statistics* 31 (6), 1923-1955

[15] Kreiss, J-P. and E. Paparoditis (2011) Bootstrap methods for dependent data: A review, *Journal of the Korean Statistical Society* 40, 357-378.

[16] Kreiss, J-P., E. Paparoditis and D. N. Politis (2011) On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics*, 39 (4) 2103–2130.

[17] Künsch, H. R. (1989), The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics* 17, 1217-1261

[18] Lahiri S. N. (2003) *Resampling Methods for Dependent Data*, New York: Springer Verlag

[19] McElroy, T. and D. Politis (2002) Robust inference for the mean in the presence of serial correlation and heavy-tailed distributions, *Econometric Theory* 18, 1019-1039.

[20] Newey, W. K. and K. D. West (1994), Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631-54.

[21] Ng, S. and P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69 (6), 1519-1554.

[22] Paparoditis, E. and D. N. Politis (2005) Bootstrapping unit root tests for autoregressive time series, *Journal of the American Statistical Association* 100 (470), 545-553.

[23] Park, J. Y. (2002) An Invariance principle for sieve bootstrap in time series, *Econometric Theory* 18, (2) 469-490.

[24] Park, J. Y. (2003) Bootstrap unit root tests, *Econometrica* 71 (6), 1845–1895

[25] Phillips, P. C. B. and S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–93.

[26] Phillips, P.C.B. and P. Perron (1988) Testing for a Unit Root in Time Series Regression, *Biometrika* 75 (2), 335-346.

[27] Politis, D. N. and Romano, J. P. (1994) The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303-1313.

[28] Theiler, J., S. Eubank, A. Longtin, B. Galdrikan and J. D. Farmer (1992), Testing for nonlinearity in time series: The method of surrogate data. *Physica D* 58, 77-94.