Testing time series linearity: traditional and bootstrap methods

Arthur Berg* Timothy McMurry† Dimitris N. Politis‡

Abstract

We review the notion of time series linearity and describe recent advances in linearity and Gaussianity testing via data-resampling methodologies. Many advances have been made since the first published tests of linearity and Gaussianity by Subba Rao and Gabr in 1980, including several resampling-based proposals. This article is intended to be instructive in explaining and motivating linearity testing. Recent results on the validity of the AR–sieve bootstrap for linearity testing are reviewed. In addition, a subsampling-based linearity and Gaussianity test is proposed where asymptotic consistency of the testing procedure is justified.

*Department of Biostatistics, Penn State University, Hershey, PA 17033 email: berg@psu.edu.

†Department of Mathematics, DePaul University Chicago, IL 60614 email: tmcmurry@depaul.edu.

‡Department of Mathematics, University of California at San Diego, La Jolla, CA 92093-0112; email: dpolitis@ucsd.edu.

1 Introduction

Ever since the fundamental recognition of the potential role of the computer in modern statistics [23, 24], the bootstrap and other resampling methods have been extensively developed for inference in independent data settings; see e.g. [22, 25, 29, 59]. Such methods are even more important in the context of dependent data where the distribution theory for estimators and test statistics may be difficult to obtain even asymptotically.

In the time series context, different resampling and subsampling methods have been proposed, and are currently receiving the attention of the statistical community. Reviews of the impact of bootstrap methods on time series analysis may be found in books [46, 54], and the papers [17, 51] and the review by J.-P. Kreiss and S.N. Lahiri in this volume of the Handbook.

In the paper at hand, we revisit the problem of assessing whether a given time series is linear vs. nonlinear, or Gaussian vs. non-Gaussian. In practice, a Gaussian classification would indicate an Auto-Regressive Moving Average (ARMA) model with Gaussian innovations is appropriate while a linear classification would indicate that an ARMA model with independent but possibly non-Gaussian innovations can still be considered. However, the rejection of linearity typically requires the practitioner to carefully select an appropriate nonlinear model for the underlying time series, or even to proceed in a model-free, nonparametric manner.

We review the traditional linearity and Gaussianity tests that are based on the normalized bispectrum. The critical regions of these tests have been traditionally determined via asymptotic methods. As an alternative, we describe how these critical regions can be determined via resampling (e.g., the AR–sieve bootstrap) and/or subsampling. One of the advantages of subsampling methodology is the generality under which it is valid. There are a number of examples where subsampling yields consistent estimation but the bootstrap fails [54]. Although subsampling is more widely applicable, it is noted that when the bootstrap is indeed valid it may possess second-order asymptotic properties [29] giving the bootstrap an advantage.

The literature on linearity and Gaussianity tests is reviewed in the next section. The concept of ime series linearity is thoroughly described in Section 3. Sections 4 and 5 focus on the AR–sieve bootstrap and subsampling tests respectively.

2 A Brief Survey of Linearity and Gaussianity Tests

Several parametric and semiparametric tests of linearity designed with a specific nonlinear model as an alternative hypothesis have been proposed, including [3, 4, 18–20, 31, 32, 42, 47, 49, 50, 62, 63, 67]. Some tests have model-based assumptions on the null hypothesis (e.g. assuming the null to be AR(p) where p may or may not be assumed known) and some test induce model-based assumptions on the alternative hypothesis (e.g. assuming the specific GARCH nonlinear alternative hypothesis). Such model-based assumptions may help to increase the power of the various tests, but only when the respective assumptions are satisfied.

Many nonparametric or model-free tests, including the first published linearity test due to Subba Rao and Gabr [60], are based on nonparametric estimates of the normalized bispectrum,and thus involve much less restrictive assumptions under the null and alternative hypothesis; the normalized bispectrum will be defined and discussed in our Section 3. Further bispectrum-based tests include [5, 12, 15, 33, 40, 61, 69]. Tests based on the normalized bispectrum are frequently used in practice when data are available in abundance, for example, when analyzing financial time series; see e.g. [1, 2, 35, 36, 39]. Note that there are other nonparametric or model-free tests of linearity that are not based on the normalized bispectrum; see e.g., [38, 64, 65]. An overview of some of these tests is provided in [21].

Because of the nonparametric nature of the bispectrum-based tests, their critical regions have traditionally been determined via asymptotic approximations. However, considerably large sample sizes can be necessary in order to accurately estimate the two-dimensional bispectral density. As such, a number of resampling-based methods have been proposed in the recent literature to overcome this limitation in a finite sample size setting.

There are many published reports, especially in recent years, that utilize some form of resampled data in linearity testing [9, 12, 34, 37, 45]. Many of these methods involve bootstraping residuals obtained from fitting a parametric model which is equivalent to resampling the data obtained after a prewhitening step that has removed (to large extent) the presence of autocorrelation. If the prewhitening is performed by fitting an AutoRegressive AR(p) model to the data, then typically practitioners would choose the order p in a data-dependent manner, say by minimizing an information criterion such as AIC, BIC, etc. In practice, it is extremely rare that a finite-order AR(p) would explain the data perfectly; more often than not, the practitioner would use an order p that would be an increasing function of the sample size n, thereby creating an approximating *sieve* of AR models. This is the essence of the AR–sieve bootstrap that is reviewed in detail in the paper by J.-P. Kreiss and S.N. Lahiri in this volume of the Handbook; the application of the AR–sieve bootstrap to linearity testing is discussed in our Section 4.

Another popular approach for linearity testing is the *surrogate data* approach of Theiler et al. [65]. The idea of the surrogate data [1] method is to apply the bootstrap on the phases of the Discrete Fourier Transform (DFT) of the data while keeping the magnitudes of the DFT unchanged. With an inverse DFT, bootstrap pseudo-series can then be created. It is immediate that these pseudo-series have identical second order structure as the original series, since the second order structure is coded in the periodogram which remains unchanged in this process.

Alternative uses of the bootstrap in the literature of linearity and Gaussianity testing include a phase scrambling bootstrap [6], the use of bootstrapped residuals to obtain the correct false alarm rate [12, 34], and the Time Frequency Toggle (TFT)-bootstrap [43]. The TFT-bootstrap can actually be seen as a generalization of the surrogate data method since it involves resampling both the phases and the magnitudes of the Fourier coefficients. Several surrogate and bootstrap tests for linearity in time series were compared in [45]. Finally, a different test has recently been proposed that combines an entropy measure of (non)linearity with bootstrap critical regions [27].

In this article, we chose to highlight two resampling-based tests for time series linearity and Gaussianity. The first is the aforementioned AR-sieve method that bootstraps the residuals obtained from an appropriate AR(p) fit. The AR-sieve methodology has been popular for quite some time but its validity for testing Gaussianity or linearity has only recently been proven [9]; it is discussed in Section 4. In Section 5, we also describe in detail a novel subsampling-based approach to Gaussianity and linearity testing. The next section defines and discusses the notion of linearity in time series.

3 Linear and nonlinear time series

Consider data $X_1, \ldots, X_n$ arising from a strictly stationary time series $\{X_t\}$ that—for ease of notation—is assumed to have mean zero.[2] The most basic tool for quantifying the inherent strength of dependence is given by the autocovariance function $\gamma(k) = EX_tX_{t+k}$ and the corresponding Fourier series $f(w) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty}\gamma(k)e^{-iwk}$; the latter function is termed the *spectral density*, and is well-defined (and continuous) when $\sum_k |\gamma(k)| < \infty$. We can also define the autocorrelation function (ACF) as $\rho(k) = \gamma(k)/\gamma(0)$. If $\rho(k) = 0$ for all $k > 0$, then the series $\{X_t\}$ is said to be a *white noise*, i.e., an uncorrelated sequence; the reason

[1] In this paper, we reserve the term *surrogate data* for the method of Theiler et al. [65]; however, the reader should be warned that other authors use the term as a generic way of refering to bootstrap data including even the AR–sieve bootstrap [34, 66].

[2] Centering the data at their sample mean (instead of the true mean) is perfectly acceptable for the subsequent discussion as the resulting error is negligible.

for the term 'white' is the constancy of the associated spectral density function.

The function $\gamma(k)$ represents the second order moments of the time series $\{X_t\}$; more technically, it represents the second order *cumulants* [14, 57]. The third order cumulants are encapsulated by the function $\Gamma(j,k) = EX_tX_{t+j}X_{t+k}$ and the resulting two-dimensional Fourier series

$$f(w_1, w_2) = (2\pi)^{-2} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \Gamma(j,k) e^{-iw_1 j - iw_2 k}$$

is termed the *bispectral density*. For reasons to be apparent soon, we also define the *normalized bispectrum* as

$$K(w_1, w_2) = \frac{|f(w_1, w_2)|^2}{f(w_1) f(w_2) f(w_1 + w_2)}.$$

We can similarly define the cumulants of higher order whose corresponding multi-dimensional Fourier series are termed higher order spectral densities or *polyspectra*; see Section 5 for details. The set of cumulant functions of *all* orders, or equivalently the set of *all* higher order spectral density functions, is a complete description of the dependence structure of the general time series $\{X_t\}$. Of course, working with an infinity of functions is intractable; a welcome short-cut is offered by the notion of linearity.

A time series $\{X_t\}$ is called *linear* if it satisfies an equation of the type:

$$X_t = \sum_{k=-\infty}^{\infty} \beta_k Z_{t-k} \tag{1}$$

where the coefficients β_k are (at least) square-summable, and the series $\{Z_t\}$ is independent, identically distributed (i.i.d.) with mean zero and variance $\sigma^2 > 0$. To avoid the confounding of the β's with the scale parameter σ, it is helpful to assume that $\beta_0 = 1$.

A linear time series $\{X_t\}$ is called *causal* if $\beta_k = 0$ for $k < 0$, i.e., if

$$X_t = \sum_{k=0}^{\infty} \beta_k Z_{t-k}. \tag{2}$$

Eq. (2) should not be confused with the Wold decomposition that *all* purely nondeterministic time series possess [30]. In the Wold decomposition the 'error' series $\{Z_t\}$ is only assumed to be a white noise and not i.i.d.; the latter assumption is much stronger. The causality assumption has been used successfully in the context of nonlinear time series as well; see e.g. [28] and [68].

Linear time series are easy objects to work with since their dependence structure is perfectly captured by the sequence of coefficients $\{\beta_k\}$. To elaborate, if $\{X_t\}$ satisfies eq. (1),

then its autocovariance and spectral density functions are given by $\gamma(k) = \sigma^2 \sum_{s=-\infty}^{\infty} \beta_s \beta_{s+k}$ and $f(w) = (2\pi)^{-1}\sigma^2|\beta(w)|^2$ respectively; here $\beta(w)$ is the Fourier series of the β_k coefficients, i.e., $\beta(w) = \sum_{k=-\infty}^{\infty} \beta_k e^{iwk}$. In addition, the bispectral density is simply given by

$$f(w_1, w_2) = (2\pi)^{-2}\mu_3 \; \beta(-w_1)\beta(-w_2)\beta(w_1 + w_2) \tag{3}$$

where $\mu_3 = EZ_t^3$ is the 3rd moment of the errors. Similarly, all higher order spectra can be calculated in terms of $\beta(w)$.

It is now apparent that the normalized bispectrum $K(w_1, w_2)$ satisfies:

$$K(w_1, w_2) = \frac{|f(w_1, w_2)|^2}{f(w_1)f(w_2)f(w_1 + w_2)} \overset{\text{linearity}}{=} \frac{(\mu_3)^2}{(2\pi)^2\sigma^6} \overset{\text{Gaussianity}}{=} 0.$$

As indicated by the right-hand-side of the above equation, when the time series is in fact linear, the normalized bispectrum will be constant. Furthermore, if the time series is Gaussian (and therefore also linear), the normalized bispectrum will be constantly equal to zero. These two observations form the basis for a host of test of linearity and/or Gaussianity starting with the original paper of Subba Rao and Gabr [60]. Note, however, that although linearity implies the normalized bispectrum is constant, the converse is not necessarily true. Thus there is the implicit, though presumably unlikely, limitation in producing a falsely negative result in the presence of certain nonlinear or non-Gaussian processes.

A prime example of a linear time series is given by the Auto-Regressive (AR) family in which the time series $\{X_t\}$ satisfies a linear relationship with respect to its own lagged values, namely

$$X_t = \sum_{k=1}^{p} \theta_k X_{t-k} + Z_t \tag{4}$$

with the error process $\{Z_t\}$ being i.i.d. $(0, \sigma^2)$ as in eq. (1). AR modeling lends itself ideally to the problem of predicting future values of the time series; this is particularly true if the AR model is causal. Causality of an AR model is ensured if all roots of the characteristic polynomial $1 - \sum_{k=1}^{p} \theta_k z^k$ have modulus greater than one; see e.g. [16].

For example, let $\hat{X}_{n+1}$ denote the predictor of X_{n+1} on the basis of the observed data $X_1, \ldots, X_n$. It is well-known [11] that the optimal predictor with respect to Mean Squared Error is given by the the conditional expectation, i.e., $\hat{X}_{n+1} = E(X_{n+1}|X_1, \ldots, X_n)$. Thus, $\hat{X}_{n+1} = g_n(X_1, \ldots, X_n)$ where $g_n(\cdot)$ is a (generally nonlinear) function of the data $X_1, \ldots, X_n$. In the case of a *causal* AR model, however, it is easy to show that the function $g_n(\cdot)$ is actually *linear*, and that $\hat{X}_{n+1} = \sum_{k=1}^{p} \theta_k X_{n+1-k}$. Note also the property of 'finite memory' in that the prediction function $g_n(\cdot)$ is only sensitive to its last p arguments. Although the finite memory property is specific to finite-order causal AR (and Markov) models, the linearity of the optimal prediction function $g_n(\cdot)$ is a property shared by all *causal* linear time series

satisfying eq. (2); this broad class includes all causal and invertible, i.e. "minimum-phase" [58], ARMA models with i.i.d. innovations.

However, the property of linearity of the optimal prediction function $g_n(\cdot)$ is shared by a larger class of processes. To define this class, consider a weaker form of (2) that amounts to relaxing the i.i.d. assumption on the errors to the assumption of a martingale difference, i.e., to assume that

$$X_t = \sum_{i=0}^{\infty} \beta_i \nu_{t-i} \tag{5}$$

where $\{\nu_t\}$ is a stationary martingale difference adapted to $\mathcal{F}_t$, the σ-field generated by $\{X_s, s \leq t\}$, i.e., that

$$E[\nu_t|\mathcal{F}_{t-1}] = 0 \;\text{ and }\; E[\nu_t^2|\mathcal{F}_{t-1}] = 1 \;\text{ for all }\; t. \tag{6}$$

As in [44], we will use the term *weakly linear* for a time series $\{X_t\}$ that satisfies eq. (5) and (6). As it turns out, the linearity of the optimal prediction function $g_n(\cdot)$ is shared by *all* members of the family of weakly linear time series;[3] see e.g. Theorem 1.4.2 of [30].

The family of Gaussian sequences is an interesting subset of the class of linear time series. Gaussian series occur when the series $\{Z_t\}$ of eq. (1) is i.i.d. $N(0,1)$, and they too exhibit the useful linearity of the optimal prediction function $g_n(\cdot)$. To see this, recall that the conditional expectation $E(X_{n+1}|X_1,\ldots,X_n)$ turns out to be a linear function of $X_1,\ldots,X_n$ when the variables $X_1,\ldots,X_{n+1}$ are jointly normal [16].

Furthermore, in the Gaussian case all spectra of order higher than two are identically zero. It follows that all dependence information is concentrated in the spectral density $f(w)$. Thus, the investigation of the dependence structure of a Gaussian series can focus on the simple study of second order properties, namely the ACF $\rho(k)$ and/or the spectral density $f(w)$. For example, an uncorrelated Gaussian series, i.e., one satisfying $\rho(k) = 0$ for all k, necessarily consists of independent random variables. Note that to check/test whether an estimated ACF, denoted by $\hat{\rho}(k)$, is significantly different from zero, the Bartlett confidence limits are typically used. Bartlett's formula, however, is only valid for linear or weakly linear time series [26, 30, 56]. In the (potentially) nonlinear case, even testing the simple null hypothesis $\rho(1) = 0$ becomes highly nontrivial, and is greatly facilitated by a computer-intensive methods such as resampling or subsampling [51, 56].

[3]Nonetheless, the class of time series for which the best predictor is linear is larger than the family of weakly linear series. A prime example of a non-weakly linear time series that actually admits a linear optimal predictor can be given by a series of *squared* financial returns, i.e.,when the series $\{X_t\}$ satisfies $X_t = r_t^2$ for all t, and $\{r_t\}$ is modeled by an ARCH/GARCH model; see [44] for details.

4 AR-Sieve Bootstrap Tests of Linearity

The popular AR-sieve bootstrap method has also been recently shown to be an effective and robust method for Gaussianity and linearity testing. The following gives the general AR-sieve bootstrap algorithm, including separate procedures for Gaussianity and linearity testing as well as third possibility that sits between Gaussianity and linearity–a linear process with symmetric (though possibly non-Gaussian) innovations. The proof of asymptotic consistency of this procedure—under both the null and the alternative hypotheses—can be found in [9] along with simulations demonstrating its finite-sample effectiveness.

AR-sieve bootstrap Algorithm

Step 0: According to some criterion (AIC, BIC, etc.), choose the order p of the AR(p) model to fit to the data $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$.

Step 1: Fit an AR(p) model to $\{X_t\}$ with estimated coefficients $\hat{\boldsymbol{\theta}}_p = (\hat{\theta}_{1,p}, \hat{\theta}_{2,p}, \ldots, \hat{\theta}_{p,p})$; i.e., $\hat{\boldsymbol{\theta}}_p$ is an estimator for $\boldsymbol{\theta}_p$ where

$$\boldsymbol{\theta}_p = (\theta_{1,p}, \theta_{2,p}, \ldots, \theta_{p,p}) = \underset{(c_1,\ldots,c_p)}{\arg\min}\ \mathrm{E}\left[(X_t - \sum_{j=1}^{p} c_j X_{t-j})^2\right].$$

Step 2: Let $\boldsymbol{X}^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ be a series of n pseudo-observations generated by

$$X_t^* = \sum_{j=1}^{p} \hat{\theta}_{j,p} X_{t-j}^* + u_t^* \quad (t = -b, -b+1, \ldots, 0, 1, \ldots, n) \tag{7}$$

where $X_t^* := 0$ for $t < -b$; the positive number b denotes the so-called 'burn-in' period to ensure (approximate) stationarity of the bootstrap series.

In (7), the u_t^*'s are iid random variables having mean zero and distribution function F_n which is selected based on the purpose of the analysis. One of three distribution functions can be selected depending on the null hypothesis under consideration:

Linear null ($H_0^{(1)}$): If the null hypothesis states the time series is linear, then set $F_n = F_n^{(1)}$ to be the empirical distribution function of the centered residuals $\hat{u}_t - \overline{u}_n$, where

$$\hat{u}_t = X_t - \sum_{j=1}^{p} \hat{\theta}_{j,p} X_{t-j} \quad (t = p, p+1, \ldots, n)$$

and

$$\overline{u}_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{u}_t.$$

Linear symmetric null ($H_0^{(2)}$): If the null hypothesis states the time series is linear with a symmetric distribution of errors, then set $F_n = F_n^{(2)}$ to be a symmetrized version of $F_n^{(1)}$ obtained by setting $u_t^* = S_t u_t^+$ with $S_t \overset{\text{iid}}{\sim} \text{unif}\{-1, 1\}$ (the discrete uniform distribution on -1 and 1) and $u_t^+ \sim F_n^{(1)}$.

Gaussian null ($H_0^{(3)}$): If the null hypothesis states the time series is linear with Gaussian errors, then set $F_n = F_n^{(3)} = N(0, \widehat{\sigma}_p^2)$, where

$$\widehat{\sigma}_p^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} (\widehat{u}_t - \overline{u}_n)^2.$$

Step 3: Compute $T(\boldsymbol{X}^*)$ from the bootstrap series $\boldsymbol{X}^*$ where $T(\cdot)$ is the chosen statistic for the null hypothesis of interest. In the next section, examples of such statistics are provided for testing Gaussianity and linearity.

Repeat: Steps 2 and 3 are repeated a large number (say B) of times. The empirical distribution of the B bootstrap pseudo-statistics can then be used to approximate the true distribution of $T(\boldsymbol{X})$ under the null hypothesis thus making the test feasible.

For example, consider the aforementioned tests based on nonparametric estimates of the normalized bispectrum. In testing for linearity, the normalized bispectral estimator is evaluated over a grid of points and the variability of the estimates are quantified by the interquartile range. If the time series is in fact nonlinear, then the normalized bispectrum should exhibit great variability yielding an interquartile range larger than what would have been expected under linearity. Therefore, linearity is rejected for large values of the estimated interquartile range; see Section 5.2 for more details. Traditionally, the threshold of such a test has been determined from the asymptotic distribution of the test statistic under the null; the AR–sieve bootstrap offers us a non-asymptotic alternative critical value—see [9] for details.

In closing, note that a new bootstrap method for time series, the Linear Process Bootstrap (LPB), has been recently introduced [48]. The LPB generates linear time series in the bootstrap world whether the true model is linear or not, i.e., under the null of linearity but also under the alternative. As in the AR–sieve bootstrap case, this property makes the LPB bootstrap a promising alternative in connection with bootstrapping the test of linearity.

5 Subsampling Tests of Linearity

The general subsampling methodology for time series approximates the distribution of a statistic by evaluating the statistic on subsampled blocks or contiguous subsets of the original

time series. As with any resampling procedure, there are certain assumptions required on the data and the statistic to guarantee convergence; however, the assumptions needed to achieve consistency of subsampling are generally weaker or easier to verify than the assumptions required for bootstrap procedures [54].

To fix ideas, we consider in detail two statistics: a linearity test statistic, t_n^L, and a Gaussianity test statistic, t_n^G. These test statistics are derived from estimates of the normalized bispectrum, and they are based on the statistics originally proposed by Hinich in [33]. Whereas Hinich utilized asymptotic theory to determine the distribution of the statistics under their respective null hypotheses, the approach described here uses subsampling to approximate the distributions of the statistics.

The test statistics t_n^L and t_n^G are described and the asymptotic conditions needed to justify the subsampling tests based on these statistics are provided. These test statistics are based on estimates of the spectral density and the bispectrum. Therefore we first present some theory for polyspectral inference followed by the bispectrum-based method of linearity and Gaussianity testing.

5.1 Kernel-based polyspectral estimation

Let $X_1, X_2, \ldots, X_n$ be a realization of an s^{th}-order stationary time series with (possibly nonzero) mean μ. The s^{th}-order joint cumulant is defined as

$$C(\tau_1, \ldots, \tau_{s-1}) = \sum_{(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)!\, \mu_{\nu_1} \cdots \mu_{\nu_p}, \tag{8}$$

where the sum is over all partitions $(\nu_1, \ldots, \nu_p)$ of $\{0, \ldots, \tau_{s-1}\}$ and $\mu_{\nu_j} = E\left[\prod_{\tau_i \in \nu_j} X_{\tau_i}\right]$; refer to [41] for another expression of the joint cumulant. The s^{th}-order spectral density is defined as

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{s-1}} \sum_{\boldsymbol{\tau} \in \mathbb{Z}^{s-1}} C(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau}\cdot\boldsymbol{\omega}} \tag{9}$$

where the bold-face notation $\boldsymbol{\omega}$ denotes an $(s-1)$–dimensional, vector argument, i.e., $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{s-1})$. We adopt the usual assumption on $C(\boldsymbol{\tau})$ that it be absolutely summable, thus guaranteeing the existence and continuity of the spectral density. A natural estimator of $C(\boldsymbol{\tau})$ is given by

$$\widehat{C}(\tau_1, \ldots, \tau_{s-1}) = \sum_{(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)!\, \hat{\mu}_{\nu_1} \cdots \hat{\mu}_{\nu_p}, \tag{10}$$

where the sum is over all partitions of $(\nu_1, \ldots, \nu_p)$ of $\{0, \ldots, \tau_{s-1}\}$ and

$$\hat{\mu}_{\nu_j} = \frac{1}{n - \max(\nu_j) + \min(\nu_j)} \sum_{k=-\min(\nu_j)}^{n-\max(\nu_j)} \prod_{t \in \nu_j} X_{t+k}.$$

The previously discussed second– and third–order cumulant functions, as given by $s = 2$ and $s = 3$ in (8), simplify to the following centered expectations:

$$C(\tau_1) = \mathrm{E}\left[(X_t - \mu)(X_{t+\tau_1} - \mu)\right]$$
$$C(\tau_1, \tau_2) = \mathrm{E}\left[(X_t - \mu)(X_{t+\tau_1} - \mu)(X_{t+\tau_2} - \mu)\right].$$

In these cases, the corresponding estimator in (10) simplifies to

$$\widehat{C}(\boldsymbol{\tau}) = \frac{1}{n} \sum_{t=1}^{n-\gamma} \prod_{j=1}^{s} (X_{t-\alpha+\tau_j} - \bar{X}), \tag{11}$$

where $\alpha = \min(0, \tau_1, \ldots, \tau_{s-1})$ and $\gamma = \max(0, \tau_1, \ldots, \tau_{s-1}) - \alpha$, and $\bar{X}$ represents the sample mean of the data. We extend the domain of $\widehat{C}$ to all of $\mathbb{Z}^s$ by defining $\widehat{C}(\boldsymbol{\tau}) = 0$ when the sum in (10) or (11) is empty.

Consistent estimation of the polyspectra (9) is obtained by taking the Fourier transform of the sample cumulant function, $\widehat{C}(\boldsymbol{\tau})$, multiplied by a smoothing kernel κ_m with bandwidth $m = m(n)$ that grows asymptotically with n but with $m/n \to 0$; in other words, let

$$\hat{f}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| < n} \kappa_m(\boldsymbol{\tau}) \widehat{C}(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}. \tag{12}$$

Typically, the kernel κ_m is obtained by 'dilation' of a fixed underlying kernel κ, i.e., letting $\kappa_m(\boldsymbol{\tau}) = \kappa(\boldsymbol{\tau}/m)$. Several different shapes for κ have been proposed in the literature, particularly for second-order spectral density estimation; cf. [55]. In particular, utilizing a "flat-top" lag-window function, such as the trapezoidal function [53] or the conical frustum [52], will yield a (poly)spectral density estimate with optimal mean square error properties.

Asymptotic theory of the kernel-based polyspectral density estimators (12) is detailed in [10, 13, 57]. Two assumptions are generally required:

Assumption I: The cumulant function $C(\tau_1, \ldots, \tau_{s-1})$ satisfies

$$\sum_{(t_1, \ldots, t_{s-1}) \in \mathbb{Z}^{s-1}} t_j C(t_1, \ldots, t_{s-1}) \qquad \text{for each } j = 1, \ldots, s-1.$$

This assumption implies the existence of a continuously differentiable polyspectral density.

Assumption II: The kernel $\kappa(\boldsymbol{\tau})$ is continuously differentiable and satisfies

$$\max\left(|\tau_j\kappa(\boldsymbol{\tau})|, \left|\frac{\partial}{\partial\tau_j}\kappa(\boldsymbol{\tau})\right|\right) \leq M(1+\|\boldsymbol{\tau}\|)^{-(s-1)-\epsilon} \qquad \text{for each } j=1,\ldots,s-1,$$

where $\|\boldsymbol{\tau}\| = \left(\sum_{j=1}^{s-1}\tau_j^2\right)^{1/2}$, $M>0$, and $\epsilon>0$.

If $\{X_t\}$ is a strictly stationary process, Assumptions I and II can be used to show that

$$\sqrt{n/m^{s-1}}\left(\hat{f}(\boldsymbol{\omega}) - \mathrm{E}\left[\hat{f}(\boldsymbol{\omega})\right]\right) \longrightarrow_d \mathcal{N}\left(0,\sigma^2\right) \tag{13}$$

when $n\to\infty$ but $n/m^{s-1}\to\infty$; here σ^2 is a complex-valued functional of f and κ.

Remark 1. If the bias of $\hat{f}(\boldsymbol{\omega})$ is of smaller order than $\sqrt{n/m^{s-1}}$, then $\mathrm{E}[\hat{f}(\boldsymbol{\omega})]$ in (13) can be replaced with $f(\boldsymbol{\omega})$. This minimal bias property can be achieved in two ways: 1) by selecting a bandwidth m that is (slightly) bigger than the optimal one resulting in a certain *under*smoothing, or 2) by using an infinite-order kernel κ which possesses reduced-bias properties [52]. Selecting an optimal bandwidth in finite samples is an unavoidable issue in nonparametric function estimation; a practical and effective method for selecting an appropriate bandwith for polyspectral estimation is given in [10].

5.2 The test statistics t_n^G and t_n^L

Due to the symmetries inherent to polyspectra [7], the normalized bispectrum, $K(\omega_1,\omega_2)$ is uniquely defined by its values on Ω given by

$$\Omega := \{(\omega_1,\omega_2) : 0<\omega_1<\pi,\ 0<\omega_2<\min(\omega_1, 2(\pi-\omega_1)\}.$$

Utilizing estimates of the polyspectra in (12) yields $\hat{K}(\omega_1,\omega_2)$, the estimator of the normalized bispectrum. The Subba Rao and Gabr [61] Gaussianity test statistic is then defined as

$$t_n^G = \sum_{j=1}^{k}\hat{K}(\omega_j^1,\omega_j^2) \tag{14}$$

where (ω_j^1,ω_j^2) $(j=1,\ldots,k)$ constitutes a grid of k points inside Ω; the number of points k increases with n to ensure consistency of the test. The null hypothesis of Gaussianity is rejected if t_n^G is too large.

Hinich [33] proposed an improved and more robust version of the original bispectrum-based linearity test proposed by Subba Rao and Gabr. The Hinich linearity test statistic is

given as

$$t_n^L = IQR\left(\left\{\hat{K}(\omega_j^1, \omega_j^2)\right\}_{j=1}^{k}\right) \tag{15}$$

where IQR stands for the interquartile range. The null hypothesis of linearity is rejected if t_n^L is too large.

In either case, t_n^G or t_n^L, the practitioner must determine the threshold of the critical region, i.e., decide what constitutes "too large" a value of the test statistic. This has been traditionally accomplished via asymptotic arguments [33, 61]. However, as discussed in Section 4, we can alternatively determine the threshold by a resampling approximation offered by the AR-sieve bootstrap. The following Section describes how to obtain a subsampling approximation to such a critical region.

5.3 Subsampling for t_n^G and t_n^L

In order to establish the consistency of subsampling for the test statistics t_n^G and t_n^L, it must be shown that their sampling distribution converges to a continuous limit law under their respective null hypothesis. The asymptotics of the t_n^G and t_n^L have been established in the literature as presented below.

If the time series is Gaussian, then [60, 61]

$$\left(\frac{n}{m^2} \cdot \frac{2\pi}{\zeta_2}\right) t_n^G \longrightarrow_d \chi_{2k}^2 \tag{16}$$

where $m = m(n)$ is the bandwidth used for the estimator (12) and $\zeta_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \kappa^2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2$.

If the time series is linear, then [9, 33]

$$\left(\frac{n}{m^2} \cdot \frac{2\pi}{\zeta_2}\right) t_n^L \longrightarrow_d N\left(\xi_{3/4} - \xi_{1/4}, \frac{1}{16k}\left(\frac{3}{g^2(\xi_{1/4})} + \frac{3}{g^2(\xi_{3/4})} - \frac{2}{g^2(\xi_{1/4})g^2(\xi_{3/4})}\right)\right) \tag{17}$$

where $\xi_\cdot$ and $g(\cdot)$ are the quantile and density functions, respectively, of the χ_{2k}^2 distribution.

Let t_n denote either t_n^G or t_n^L as appropriate. It is easy to see that t_n satisfies the property $t_n \to 0$ under its respective null and $t_n \to t > 0$ under the alternative; convergence of t_n under the alternative is investigated in [33]. Define $t_{n,b,t}$ to be the statistic defined by (14) or (15), whichever appropriate, calculated using only the subsample $\{X_t, X_{t+1}, \ldots, X_{t+b-1}\}$ for $t \in \{1, 2, \ldots, n-b+1\}$.

We now consider two candidate subsampling distributions for subsampling the hypothesis test of Gaussianity or linearity. First we define the uncentered subsampling distribution as presented in [54],

$$S_{n,b}^{U}(x) := \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1\{\tau_b t_{n,b,t} \leq x\}. \tag{18}$$

where[4] $\tau_b = b/m(b)^2$.

Alternatively, a centered version of the above subsampling distribution has been shown to possess improved power in many contexts [8]. The centered subsampling distribution is given by

$$S_{n,b}^{C}(x) := \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1\{\tau_b (t_{n,b,t} - t_n) \leq x\}. \tag{19}$$

It follows from (16) and (17) that the sampling distribution of $\tau_n t_n$ converges, under the respective null hypothesis, to a continuous limit law with cumulative distribution function denoted by $H(x)$. The consistency of the subsampling method as applied to linearity and Gaussianity testing is now stated; the following theorem follows directly from Theorem 3.5.1 in [54].

Theorem [Validity of subsampling for t_n^G and t_n^L]. *Let $H_{n,b}(x)$ denote either $S_{n,b}^{U}(x)$ or $S_{n,b}^{C}(x)$. Assume either* (16) *or* (17) *according to whether T_n denotes t_n^G or t_n^L. Assume $b \to \infty$, $b/n \to 0$ and $\tau_b/\tau_n \to 0$ as $n \to \infty$. Assume the bandwidth m for the polyspectra estimates used in the construction of the test statistics obeys the undersmoothing condition outlined in Remark 1. Further assume the time series $\{X_t\}$ is strictly stationary, and strong mixing. For $\alpha \in (0,1)$, define the two quantities*

$$h_{n,b}(1-\alpha) = \inf\{x : H_{n,b}(x) \geq 1-\alpha\}$$
$$h(1-\alpha) = \inf\{x : H(x) \geq 1-\alpha\}$$

Then under the null hypothesis

i. $h_{n,b}(1-\alpha) \longrightarrow g(1-\alpha)$ *in probability;*

ii. $Prob\{\tau_n t_n > h_{n,b}(1-\alpha)\} \longrightarrow \alpha \quad \text{as } n \to \infty.$

And under the alternative hypothesis,

iii. $Prob\{\tau_n t_n > h_{n,b}(1-\alpha)\} \longrightarrow 1 \quad \text{as } n \to \infty.$

[4]Recall that $m = m(n)$; for example, if $m(n) = n^\delta$ for some $\delta \in (0, 1/2)$, then $\tau_b = b/[b^\delta]^2 = b^{1-2\delta}$.

The above theorem shows that both subsampling distributions $S^U_{n,b}(x)$ or $S^C_{n,b}(x)$ yield consistent α—level tests. However, by analogy to other simpler examples [8], we expect that the test based on the centered subsampling distribution $S^C_{n,b}(x)$ would be more powerful than the one based on $S^U_{n,b}(x)$, i.e., that the convergence in part (iii) of the Theorem would be faster when $H_{n,b}(x) = S^C_{n,b}(x)$. By the same token, the convergence in part (ii) of the Theorem is expected to be faster when $H_{n,b}(x) = S^U_{n,b}(x)$, i.e., the level of the test would be more accurately achieved with the uncentered subsampling distribution $S^U_{n,b}(x)$.

References

[1] A. Abhyankar, L.S. Copeland, and W. Wong. Nonlinear dynamics in real-time equity market indices: evidence from the United Kingdom. *The Economic Journal*, 105(431):864–880, 1995.

[2] A. Abhyankar, L.S. Copeland, and W. Wong. Uncovering nonlinear structure in real-time stock-market indexes: the S&P 500, the DAX, the Nikkei 225, and the FTSE-100. *Journal of Business & Economic Statistics*, 15(1):1–14, 1997.

[3] H.Z. An, L.X. Zhu, and R.Z. Li. A mixed-type test for linearity in time series. *Journal of Statistical Planning Inference*, 88:339–353, 2000.

[4] R.A. Ashley and D.M. Patterson. A Test of the GARCH(1,1) Specification For Daily Stock Returns. Working paper presented at the 17th Society for Nonlinear Dynamics and Econometrics on April 16, 2009), 2009.

[5] R.A. Ashley, D.M. Patterson, and M.J. Hinich. A diagnostic test for nonlinear serial dependence in time series fitting errors. *Journal of Time Series Analysis*, 7(3):165–178, 1986.

[6] AG Barnett and RC Wolff. A time-domain test for some types of nonlinearity. *IEEE Transactions on Signal Processing*, 53(1):26–33, 2005.

[7] A. Berg. Multivariate lag-windows and group representations. *Journal of Multivariate Analysis*, 99(10):2479–2496, 2008.

[8] A. Berg, T.L. McMurry, and D.N. Politis. Subsampling p-values. *Statistics & Probability Letters*, 80(17-18):1358–1364, 2010.

[9] A. Berg, E. Paparoditis, and D.N. Politis. A bootstrap test for time series linearity. *Journal of Statistical Planning and Inference*, 140(12):3841–3857, 2010.

[10] A. Berg and D.N. Politis. Higher-order accurate polyspectral estimation with flat-top lag-windows. *Annals of the Institute of Statistical Mathematics*, 61:1–22, 2009.

[11] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, 1995.

[12] Y. Birkelund and A. Hanssen. Improved bispectrum based tests for gaussianity and linearity. *Signal Processing*, 89(12):2537–2546, 2009.

[13] D. Brillinger and M. Rosenblatt. *Spectral Analysis of Time Series*, chapter Asymptotic theory of kth order spectra in spectral analysis of time series. Wiley, New York, 1967.

[14] D.R. Brillinger. *Time Series: Data Analysis and Theory*. Society for Industrial Mathematics, 2001.

[15] P.L. Brockett, M.J. Hinich, and D. Patterson. Bispectral-based tests for the detection of gaussianity and linearity in time series. *Journal of the American Statistical Association*, 83(403):657–664, 1988.

[16] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer Verlag, 2009.

[17] P. Bühlmann. Bootstraps for time series. *Statististical Science*, 17:52–72, 2002.

[18] K.S. Chan. Testing for threshold autoregression. *The Annals of Statistics*, 18(4):1886–1894, 1990.

[19] K.S. Chan and H. Tong. On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B*, 52(3):469–476, 1990.

[20] W.S. Chan and H. Tong. On tests for non-linearity in time series analysis. *Journal of Forecasting*, 5(4):217–228, 1986.

[21] M. Corduas. Nonlinearity tests in time series analysis. *Statistical Methods and Applications*, 3(3):291–313, 1994.

[22] A.C. Davison and D.V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.

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

[24] B. Efron. Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, 21(4):460–480, 1979.

[25] B. Efron and R. Tibshirani. *An introduction to the bootstrap*, volume 57. Chapman & Hall/CRC, 1993.

[26] C. Francq and J.M. Zakoïan. Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis*, 30(4):449–465, 2009.

[27] S. Giannerini, E. Maasoumi, and E. Bee Dagum. Entropy testing for nonlinearity in time series. Technical report, Università di Bologna, 2011.

[28] C. Gourieroux and J. Jasiak. Nonlinear innovations and impulse responses with application to VaR sensitivity. *Annales d'Economie et de Statistique*, (78):1–31, 2005.

[29] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer Verlag, 1997.

[30] E.J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1988.

[31] B.E. Hansen. Testing for linearity. *Journal of Economic Surveys*, 13(5):551–76, 1999.

[32] D.I. Harvey and S.J. Leybourne. Testing for time series linearity. *Econometrics Journal*, 10(1):149–165, 2007.

[33] M.J. Hinich. Testing for gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3(3):169–176, 1982.

[34] M.J. Hinich, E.M. Mendes, and L. Stone. Detecting nonlinearity in time series: Surrogate and bootstrap approaches. *Studies in Nonlinear Dynamics & Econometrics*, 9(4):3, 2005.

[35] M.J. Hinich and D.M. Patterson. Evidence of nonlinearity in daily stock returns. *Journal of Business & Economic Statistics*, pages 69–77, 1985.

[36] M.J. Hinich and D.M. Patterson. Evidence of nonlinearity in the trade-by-trade stock market return generating process. In W.A. Barnett, J. Geweke, and K. Shell, editors, *Economic complexity: chaos, sunspots, bubbles and nonlinearity-international symposium in economic theory and econometrics*, pages 383–409. Cambridge University Press, Cambridge, UK, 1989.

[37] V. Hjellvik and D. Tjostheim. Nonparametric tests of linearity for time series. *Biometrika*, 82(2):351–368, 1995.

[38] A. Hong-Zhi and C. Bing. A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *International Statistical Review/Revue Internationale de Statistique*, pages 287–307, 1991.

[39] D.A. Hsieh. Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business*, pages 339–368, 1989.

[40] N. Jahan and J.L. Harvill. Bispectral-based goodness-of-fit tests of gaussianity and linearity of stationary time series. *Communications in Statistics-Theory and Methods*, 37(20):3216–3227, 2008.

[41] S.R. Jammalamadaka, T.S. Rao, and G. Terdik. Higher order cumulants of random vectors and applications to statistical inference and time series. *Sankhyā: The Indian Journal of Statistics*, 68(2):326–356, 2006.

[42] D.M.R. Keenan. A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, 72(1):39–44, 1985.

[43] C. Kirch and D.N. Politis. TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Annals of Statistics*, (in press), 2011.

[44] P.S. Kokoszka and D.N. Politis. Nonlinearity of ARCH and stochastic volatility models and Bartlett's formula. *Probability and Mathematical Statistics*, 31 (in press), 2011.

[45] D. Kugiumtzis. Evaluation of surrogate and bootstrap tests for nonlinearity in time series. *Studies in Nonlinear Dynamics & Econometrics*, 12(1):4, 2008.

[46] S.N. Lahiri. *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer, New York, 2003.

[47] R. Luukkonen, P. Saikkonen, and T. Terasvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499, 1988.

[48] T. McMurry and D.N. Politis. Banded and tapered estimates of autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31:471–482, 2010.

[49] J. Petruccelli and N. Davies. A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika*, 73(3):687–694, 1986.

[50] J.D. Petruccelli. A comparison of tests for setar-type non-linearity in time series. *Journal of Forecasting*, 9(1):25–36, 1990.

[51] D.N. Politis. The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219–230, 2003.

[52] D.N. Politis. Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, (in press), 2011.

[53] D.N. Politis and J.P. Romano. Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis*, 16(1):67–103, 1995.

[54] D.N. Politis, J.P. Romano, and M. Wolf. *Subsampling*. Springer Verlag, 1999.

[55] Maurice Priestley. *Spectral Analysis and Time Series, Vol 1 and II*. Academic Press, 1983.

[56] J.P. Romano and L.A. Thombs. Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, 91(434):590–600, 1996.

[57] M. Rosenblatt. *Stationary Sequences and Random Fields*. Springer, 1985.

[58] M. Rosenblatt. *Gaussian and non-Gaussian linear time series and random fields*. Springer Verlag, 2000.

[59] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer, 1995.

[60] T. Subba Rao and MM Gabr. A test for linearity of stationary time series. *Journal of Time Series Analysis*, 1(2):145–158, 1980.

[61] T. Subba Rao and MM Gabr. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, volume 24 of *Lecture Notes in Statistics*. Springer, New York, 1984.

[62] T. Terasvirta. Testing linearity and modelling nonlinear time series. *Kybernetika*, 30(3):319–330, 1994.

[63] T. Terasvirta, C.F. Lin, and C.W.J. Granger. Power of the neural network linearity test. *Journal of Time Series Analysis*, 14(2):209–220, 1993.

[64] G. Terdik and J. Math. A new test of linearity of time series based on the bispectrum. *Journal of Time Series Analysis*, 19(6):737–753, 1998.

[65] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J.D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.

[66] J. Theiler and D. Prichard. Using 'surrogate surrogate data'to calibrate the actual rate of false positives in tests for nonlinearity in time series. *Fields Inst. Comm.*, 11:99–113, 1997.

[67] R.S. Tsay. Nonlinearity tests for time series. *Biometrika*, 73(2):461–466, 1986.

[68] W.B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150, 2005.

[69] J. Yuan. Testing linearity for stationary time series using the sample interquartile range. *Journal of Time Series Analysis*, 21(6):713–722, 2000.