Volterra bootstrap: Resampling higher-order statistics for strictly stationary univariate time series

Natalia Sirotko-Sibirskaya, Matthias O. Franz, and Thorsten Dickhaus

October 21, 2020

Abstract

We are concerned with nonparametric hypothesis testing of time series functionals. It is known that the popular autoregressive sieve bootstrap is, in general, not valid for statistics whose (asymptotic) distribution depends on moments of order higher than two, irrespective of whether the data come from a linear time series or a nonlinear one. Inspired by nonlinear system theory we circumvent this non-validity by introducing a higher-order bootstrap scheme based on the Volterra series representation of the process. In order to estimate coefficients of such a representation efficiently, we rely on the alternative formulation of Volterra operators in reproducing kernel Hilbert space. We perform polynomial kernel regression which scales linearly with the input dimensionality and is independent of the degree of nonlinearity. We illustrate the applicability of the suggested Volterra-representation-based bootstrap procedure in a simulation study where we consider strictly stationary linear and nonlinear processes.

MSC 2020 classification numbers: Primary 62M15, 62F40; secondary 62M10.

JEL Classification: C22, C12.

*University of Bremen, Institute for Statistics, Germany
†HTWG Konstanz, Germany
‡Corresponding Author. University of Bremen, Institute for Statistics, Bibliothekstr. 1, 28359 Bremen, Germany. E-mail: dickhaus@uni-bremen.de
Key words: autocorrelation, autoregressive processes, higher-order cumulants, hypothesis testing, nonlinearity, reproducing kernel Hilbert space.

1 Introduction

Over the recent years the bootstrap procedure initially introduced by Efron (1979a,b) for stochastically independent and identically distributed (iid) observables has been extended to cope with dependent data, see the overviews by Härdle et al. (2003), Kreiss and Paparoditis (2011) and Kreiss and Lahiri (2012) as well as the monographs by Politis et al. (1999) and Lahiri (2003), among others. Most of the suggested methods deal with linear processes and often the sample mean is the only statistic of interest. However, real data often exhibit nonlinear patterns and statistics of higher order such as autocovariances, autocorrelations and spectral densities are of considerable interest. This motivates us to introduce a bootstrap procedure which takes into account nonlinear features of the time series reflected in its higher-order moments and to consider scenarios where it is especially beneficial to take such nonlinear features into consideration.

Alongside with linear strictly stationary time series we consider nonlinear strictly stationary time series \( (X_t)_t \) as described by Wu (2005, 2011). They are of the form

\[
X_t = H(\ldots, \varepsilon_{t-1}, \varepsilon_t),
\]

where \( \{ \varepsilon_t, \ t \in \mathbb{N} \} \) are iid random variables and \( H \) is a measurable function such that \( X_t \) is well-defined. As Wu (2005) argues the representation in (1) can be viewed as a nonlinear analogue of the Wold representation. However, whereas for the Wold decomposition to hold one needs weakly stationary time series, asymptotic theory established in Wu (2005) under the representation as in (1) requires the time series to be strictly stationary.

Throughout this work, we consider the following representation of nonlinear time series:

\[
X_t = H(\ldots, \varepsilon_{t-1}, \varepsilon_t) = \sum_{p=0}^{\infty} \sum_{u_1, \ldots, u_p=0}^{\infty} h^{(p)}(u_1, \ldots, u_p) \varepsilon_{t-u_1} \ldots \varepsilon_{t-u_p},
\]

or, equivalently,

\[
X_t = h^{(0)} + \sum_{u=0}^{\infty} h^{(1)}(u) \varepsilon_{t-u} + \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} h^{(2)}(u,v) \varepsilon_{t-u} \varepsilon_{t-v} + \ldots
\]

\[
+ \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \ldots \sum_{w=0}^{\infty} h^{(p)}(u,v,\ldots,w) \varepsilon_{t-u} \varepsilon_{t-v} \ldots \varepsilon_{t-w} + \ldots,
\]
where $h^{(p)}$ is a Volterra operator of order $p \geq 0$ and $(\varepsilon_t)_t$ denotes a sequence of real-valued random variables. The representation in (2) is called (discrete time) Volterra series expansion, due to the Italian mathematician Vito Volterra who suggested a continuous-time analogue of this functional form in the 1880s. The Volterra representation can be thought of as a Taylor series type expansion, but unlike Taylor series Volterra series capture so-called memory effects of time series reflected in the lags of the $\varepsilon_t$'s.

Representation as in Equations (1) and (2) were studied by Wiener (1958), whose work plays an important role in the nonlinear system theory, see, e.g., Schetzen (2006), Mathews and Sicuranza (2000), and Rugh (1981), among others. Wiener (1958) conjectured that if the process is stationary and ergodic, then there exists a function $H$ and iid random variables $(\varepsilon_t)_t$, such that (1) holds. These conditions were, however, shown to be insufficient by Rosenblatt (2009). On the other hand, it has been proven by Nisio (1960) that every strictly stationary time series has a two-sided polynomial representation in terms of Gaussian iid random variables. However, according to our knowledge the sufficient conditions for the time series to have a one-sided representation as in (1) have not been established so far. In this work we consider the class of processes which admit the representation as in (1) without providing further conditions which completely characterize this class.

As targets of statistical inference we consider higher-order statistics contained in the following broad class of functions of generalized means as considered in Example 2.2 of Künsch (1989), Assumption C of Bühlmann (1997) and Assumption (A2) of Kreiss et al. (2011). Suppose we can observe univariate random variables $X_1, \ldots, X_n$ from some stationary process $\mathbf{X} = \{X_t : t \in \mathbb{Z}\}$. For functions $g : \mathbb{R}^m \to \mathbb{R}^d$ and $w : \mathbb{R}^d \to \mathbb{R}$ let

$$T_n = w \left( \frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t, \ldots, X_t+m-1) \right), \quad (3)$$

where $m \in \mathbb{N}$ and $d \in \mathbb{N}$ are given numbers and the functions $g$ and $w$ fulfill some smoothness assumptions like in Assumption C of Bühlmann (1997). The so-defined class of statistics is quite rich and contains, e.g., sample means, sample autocovariances, sample partial autocorrelations, sample partial autocorrelations, and Yule-Walker estimators.

Under appropriate mixing or weak dependence conditions, central limit theorems for $T_n$ can be established for sufficiently smooth functions $g$ and $w$; cf., for example, Künsch.
However, in finite samples a normal approximation of the distribution of $T_n$ is often inaccurate and/or the limiting variance $\tau^2$ (say) is difficult to estimate or cannot be derived analytically. Therefore, and in line with previous literature, we suggest to employ a bootstrap procedure to approximate the unknown finite sample distribution of $T_n - \theta$, where $\theta$ is a centering constant or a parameter value under a null hypothesis of a statistical test, respectively. In particular, we base our bootstrap procedure on the Volterra representation (2) of the process (or a truncated version thereof) which can mimic higher-order moments of the process $X = \{X_t : t \in \mathbb{Z}\}$; see Example 2.1 for a statistic which requires correctly mimicked fourth-order moments.

The remainder of the work is structured as follows. Section 2 outlines the proposed Volterra-based procedure, Section 3 analyzes theoretical properties of the suggested procedure, Section 4 explains how coefficients in the Volterra representation are estimated based on polynomial kernel regression, and Section 5 presents results of Monte-Carlo simulations highlighting the advantages of the suggested procedure over the autoregressive (AR) sieve bootstrap. Finally, Section 6 concludes.

2 Volterra bootstrap

Before we describe our proposed methodology, let us consider a motivating example, which we will get back to in our numerical examples in Section 5.

**Example 2.1** (Sample autocorrelations at lag 1). Consider the statistic $T_n$ from (3) for the special case of $m = d = 2$, $g(x, y) = (yx, x^2)^\top$, and $w(x, y) = x/y$. We obtain that

$$T_n = \frac{\sum_{t=1}^{n-1} X_{t+1} X_t}{\sum_{t=1}^{n-1} X_t^2}.$$ (4)

Up to (empirical) centering, this statistic is for large sample size $n$ essentially equivalent to the sample autocorrelation $\hat{\rho}(1) = \hat{\gamma}(1)/\hat{\gamma}(0)$, where $\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)$ and $\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t$.

For convenience and due to practical relevance, we present here results pertaining to $\hat{\rho}(1)$, but they would apply in an analogous manner to $T_n$ from (4). Namely, large sample properties of $\{\hat{\rho}(h)\}_{1 \leq h \leq k}$ for $k \in \mathbb{N}$ have been discussed by Romano and Thombs (1996) under weak assumptions. In particular, the authors provided the following result.
Proposition 2.1 (Thm. 3.2 in Romano and Thombs (1996)). Suppose \( X_1, \ldots, X_n \) is a sample from a stationary mean zero process such that \( \gamma(0) = \text{Var}(X_1) \in (0, \infty) \). Then, under appropriate moment and mixing conditions, the random vector \( \sqrt{n}(\hat{\rho}(1) - \rho(1), \ldots, \hat{\rho}(k) - \rho(k))^\top \) is asymptotically normal with mean vector zero. The asymptotic covariance \( \tau_{i,j} \) of \( \sqrt{n}(\hat{\rho}_{i+1,j+1} - \rho(i)c_{1,i+1} + \rho(i)\rho(j)c_{1,j+1}) \) is given by

\[
\tau_{i,j} = \lim_{n \to \infty} \{ n \text{Cov}(\hat{\gamma}_n(i), \hat{\gamma}_n(j)) \}
\]

\[
\equiv \sum_{h=-\infty}^{\infty} \left\{ \gamma(h)\gamma(h + j - i) + \gamma(h + j)\gamma(h - i) + \kappa(h, i, j - i) \right\}
\]

\[
= \sum_{h=-\infty}^{\infty} \text{Cov}(X_0X_i, X_hX_{h+j})
\]

and \( \kappa(h, i, j - i) \) denotes the fourth joint cumulant of the distribution of \( (X_0, X_i, X_h, X_{j+h})^\top \).

In the case that \( \kappa(h, i, j - i) \) vanishes for all \( (h, i, j) \), we arrive at Bartlett’s formula (see, e.g., Theorem 7.2.1. in Brockwell and Davis (1991)). However, this is only the case for restrictive special cases. For instance, Bartlett’s formula is valid in the case that \( X \) is a Gaussian process or if \( X \) can be represented as a linear process of the form

\[
X_t = \sum_{j=-\infty}^{\infty} b_j \varepsilon_{t-j}, \quad b_0 = 1, \ t \in \mathbb{Z}, \tag{5}
\]

where \( \{\varepsilon_t : t \in \mathbb{Z}\} \) are iid with zero mean and finite fourth moments, and the coefficients \( \{b_j\}_{j \in \mathbb{Z}} \) are absolutely summable. In general and for many processes of practical interest, however, fourth-order moments appear in the limiting (co-)variances \( (\tau_{i,j})_{i,j} \).

Remark 2.1. The AR sieve bootstrap can only mimic first and second order moments of \( X \) correctly; see, e.g., Section 2.2.6. of Kreiss and Paparoditis (2011) and Section 3.1 of Jentsch and Politis (2013).

Our proposed methodology relies on a truncated version of the Volterra representation of the time series as in Equation (2). For given order \( p \in \mathbb{N} \) and degree \( m \in \mathbb{N} \), it is given
by
\[ h^{(0)} + \sum_{u=0}^{m} h^{(1)}_{t-u} + \sum_{u=0}^{m} \sum_{v=0}^{m} h^{(2)}_{t-u} \varepsilon_{t-v} + \ldots + \sum_{u=0}^{m} \sum_{v=0}^{m} \ldots \sum_{w=0}^{m} h^{(p)}_{t-u} \varepsilon_{t-v} \ldots \varepsilon_{t-w}. \] (6)

Appropriate truncation is essential in the case of a finite sample size \( n \). A natural explanation provided by Volterra himself is that the memory [of a process] "gradually fades out", see Volterra (1959). It is further formalized by Boyd and Chua (1985) and Sandberg (2002) among others. We provide details on an automated choice of \( p \) and \( m \) in Section 5.

Let \( T_n \) be as in (3), and suppose that for some appropriately increasing sequence of real numbers \( \{c_n : n \in \mathbb{N}\} \) and a given real parameter \( \theta \), the distribution \( L_n \equiv L(c_n(T_n - \theta)) \) has a nondegenerate limit. The Volterra bootstrap procedure to estimate the distribution \( L_n \) is then performed as follows.

**Algorithm 2.1** (Volterra bootstrap procedure).

1. Select an appropriate order \( p \ll n \), and an appropriate degree \( m \ll \infty \), and fit a \( p \)-th order \( m \)-th degree Volterra series to \( X_1, \ldots, X_n \). The fitted process is denoted by \( \hat{X}_t \) and is given as follows:

\[ \hat{X}_t = \sum_{u=0}^{m} \hat{h}^{(1)}_{t-u} + \sum_{u=0}^{m} \sum_{v=0}^{m} \hat{h}^{(2)}_{t-u} \varepsilon_{t-v} + \ldots + \sum_{u=0}^{m} \sum_{v=0}^{m} \ldots \sum_{w=0}^{m} \hat{h}^{(p)}_{t-u} \varepsilon_{t-v} \ldots \varepsilon_{t-w}, \] (7)

where the \( \varepsilon_t \)'s are iid, and \( \hat{h}^{(c)} \) is an estimated Volterra kernel of a corresponding order. For example, in Section 5 we consider \( (\varepsilon_t)_{t \sim \mathcal{N}(0,1)} \).

2. Let \( X_1^*, \ldots, X_n^* \) be constructed as follows:

\[ X_t^* = \sum_{u=0}^{m} \hat{h}^{(1)}_{t-u} \varepsilon_{t-u}^* + \sum_{u=0}^{m} \sum_{v=0}^{m} \hat{h}^{(2)}_{t-u} \varepsilon_{t-v}^* + \ldots + \sum_{u=0}^{m} \sum_{v=0}^{m} \ldots \sum_{w=0}^{m} \hat{h}^{(p)}_{t-u} \varepsilon_{t-v}^* \ldots \varepsilon_{t-w}^*, \] (8)

where \( (\varepsilon_t^*) \) has the same (joint) distribution as \( (\varepsilon_t) \).

3. Let \( T_n^* = T_n(X_1^*, \ldots, X_n^*) \) be the statistic \( T_n \) applied to the pseudo-time series \( X_1^*, \ldots, X_n^* \), and denote by \( \theta^* \) the analogue of \( \theta \) associated with the bootstrap process \( X^* \). The Volterra bootstrap approximation of \( L_n \) is then given by \( L_n^* = L(c_n(T_n^* - \theta^*)) \), where \( L^* \) refers to the distribution of \( (X_t^*)_{1 \leq t \leq n} \). In practice, a Monte Carlo-variant of \( L^* \) will be applied.
In the next section we provide theoretical considerations regarding the consistency of the bootstrap procedure defined by Algorithm 2.1.

3 Theoretical considerations

In this section we present the key definitions and assumptions required in order to establish the consistency of the proposed Volterra bootstrap procedure. In this, the so-called cumulant matching approach as suggested by Kalouptsidis and Koukoulas (2005) plays an important role.

**Definition 3.1.** We call a given bootstrap procedure, which generates pseudo observables $X_1^*, \ldots, X_k^*(n)$, consistent for $T_n$, if $d(L_n, L_n^*) \to 0$ in probability for $n \to \infty$, where $d(\cdot, \cdot)$ is any distance that metrizes weak convergence, e. g., the Prohorov distance. Here, $\{k(n)\}_{n \in \mathbb{N}}$ is an increasing sequence of integers which denotes bootstrap pseudo sample sizes.

The following assertion on the consistency of the bootstrap procedure is well-known in the bootstrap literature and has been discussed extensively by Kreiss and Paparoditis (2011), among others.

**Proposition 3.1** (Conditions for bootstrap consistency). The consistency of a given bootstrap procedure for approximating $L_n$ depends on the following two conditions.

(a) The bootstrap procedure is such, that its resulting companion process (in the sense of Kreiss and Paparoditis (2011)) captures all distributional characteristics of $X$ which are relevant for the limiting distribution of $c_n(T_n - \theta)$.

(b) The functions $g$ and $w$ are sufficiently smooth, such that distributional closeness of $X$ and the companion process of the bootstrap procedure implies distributional closeness of $T_n$ and $T_n^*$.

The exact mathematical assumptions for smoothness of $g$ and $w$ can be found, e. g., under Assumption C of Bühlmann (1997) and in (A2) of Kreiss et al. (2011), respectively. The following more explicit corollary is tailored to the setting of Example 2.1, or a similar setting in which the limiting distribution of $c_n(T_n - \theta)$ is a normal distribution.
Corollary 3.1. Assume that the functions $g$ and $w$ as well as the process $X$ are such, that a central limit theorem holds for $c_n(T_n - \theta)$, where the limiting normal distribution is centered and its variance depends only on (joint) cumulants of finite order $\Xi \in \mathbb{N}$ of the distribution of $X$. Then, a given bootstrap procedure for approximating $\mathcal{L}_n$ is consistent, if all (joint) cumulants up to order $\Xi$ of the distribution of $X$ are correctly mimicked by the companion process of that bootstrap procedure.

Kalouptsidis and Koukoulos (2005) provide the following results on the relationship between input cumulants and output cumulants of a Volterra system.

Proposition 3.2 (cf. Sections II and III of Kalouptsidis and Koukoulos (2005)). Let $\Xi \in \mathbb{N}$ be a given integer and assume that the (random) input of a Volterra system is chosen to be stationary higher order white noise. Then there exist integers $m \in \mathbb{N}$ and $p \in \mathbb{N}$ as well as Volterra kernels $h^{(0)}, \ldots, h^{(p)}$, such that the (joint) cumulants up to order $\Xi$ of the finite Volterra series (6) match given target values.

Proposition 3.2 guarantees that a cumulant matching up to a given order is possible by appropriately chosen Volterra kernels. In particular, this implies that the (joint) cumulants of our original process $(X_t)_t$ and the (joint) cumulants of the approximation $(\hat{X}_t)_t$ can be made identical by applying a suitable estimation procedure for Volterra kernels. Furthermore, since under (8) in Step 2 of Algorithm 2.1 we use the same (joint) distribution for $(\epsilon^*_t)_t$ as for $(\epsilon_t)_t$ in Step 1, we can deduce that the companion process corresponding to the bootstrap procedure defined by Algorithm 2.1 can mimic the (joint) cumulants up to a required order $\Xi$ of the original process $(X_t)_t$. This argumentation implies the conceptual validity of the proposed Volterra bootstrap approach under the assumptions of Corollary 3.1.

It remains to describe an appropriate estimation and model selection procedure for $m$, $p$, and $h^{(0)}, \ldots, h^{(p)}$. In the following section we employ a technique which is based on the theory of reproducing kernel Hilbert space (RKHS) and polynomial kernel regression. The reason for this choice is that this estimation method scales linearly with the input dimensionality and is independent of the degree of nonlinearity. This avoids stability issues (cf., e. g., Franz and Schölkopf (2005, 2006)) of direct cumulant matching approaches, especially for larger values of $m$ and $p$. 
4 Estimation approach

Several methods to estimate Volterra kernels exist in the literature. Among others, there are the cross-correlation method by [Lee and Schetzen (1965)] and its extensions such as, e.g., in [Orcioni et al. (2018)], the exact orthogonal method as in [Korenberg and Hunter (1996)], the neural network-based method as in [Wray and Green (1994)] and the polynomial kernel regression method as in [Franz and Schölkopf (2005)]. The cross-correlation method is considered to be a traditional method to estimate the Volterra representation and is widely applied. However, as outlined by [Franz and Schölkopf (2005)], it suffers from several shortcomings: (1) It requires large sample sizes before sufficient convergence is reached. (2) Generally (and initially) it is developed under the assumption of Gaussian iid inputs. (3) The number of coefficients to be estimated for the finite-sample Volterra expansion is \((p + m - 1)!/(p!(m-1)!)\), which can be computationally prohibitive already in moderately scaled models. (4) Estimation is performed under the noise-free data assumption which is unrealistic as real data is likely to be noise-contaminated, see Section 2 in [Franz and Schölkopf (2005)].

For these reasons we adopt the estimation method suggested by [Franz and Schölkopf (2005)], which overcomes the disadvantages of the cross-correlation method as listed above and can provide estimates of the Volterra kernels in a much more (computationally) efficient way. The key idea of [Franz and Schölkopf (2005)] consists in reformulating the Volterra series as a polynomial kernel regression in a RKHS. In the remainder of this section we provide a summary on this estimation method. Further details can be found in [Franz and Schölkopf (2004); Franz and Schölkopf (2005, 2006)] and references therein.

We use bold letters to denote vectors and matrices, respectively.

It is convenient to explain polynomial kernel regression in RKHS by starting with the linear regression. Assume that the process \(X\) is approximated as a function of \(\epsilon\), meaning that the following representation holds:

\[
\hat{X}_t = f(\epsilon_t) = \sum_{j=0}^{M} \gamma_j \varphi_j(\epsilon_t), \quad 1 \leq t \leq n, \tag{9}
\]

where \(\epsilon_t = (\epsilon_t, \ldots, \epsilon_{t-m+1}) \in \mathbb{R}^m\), \(\gamma_j \in \mathbb{R}\), \(\varphi_j : \mathbb{R}^m \to \mathbb{R}\) and \(\varphi_0(\epsilon_t) = 1\), and where the \(\varphi_j\)'s contain all monomials of the elements of the vector \(\epsilon_t\) up to order \(j\) for the \(j\)-th order Volterra series. The coefficients \(\{\gamma_j\}_{0 \leq j \leq M}\) are found by minimizing the mean squared
Finally, we stack the maps \( \phi \) where \( \lambda \) follows:

\[
\eta_i \in \mathbb{R}^{m_i}
\]

such that \( \phi \) maps the input \( \varepsilon_t \in \mathbb{R}^m \) into a vector \( \phi(\varepsilon_t) \in \mathbb{R}^{m'} \). By stacking the coefficients of the \( i \)-th order Volterra operator into a single vector \( \eta_i = (h_{1,1,\ldots,1}, h_{1,2,\ldots,1}, \ldots) \in \mathbb{R}^{m'} \) we can rewrite it as a scalar product as follows:

\[
H_i(\varepsilon_t) = \eta_i^\top \phi_i(\varepsilon_t), \quad 0 \leq i \leq p.
\]

Finally, we stack the maps \( \phi_i \) with positive weights \( a_i \in \mathbb{R}_{>0} \) into a single map \( \phi^{(p)}(\varepsilon_t) = (a_0 \phi_0(\varepsilon_t), a_1 \phi_1(\varepsilon_t), \ldots, a_p \phi_p(\varepsilon_t)) \top \), where \( \phi^{(p)}(\varepsilon_t) : \mathbb{R}^m \to \mathbb{R} \times \mathbb{R}^{m \times m^2} \times \ldots \times \mathbb{R}^{m^p} = \mathbb{R}^M \) and \( M = (1 - m^{p+1})/(1 - m) \). It follows that Equation (11) can be rewritten as a scalar product as follows

\[
\hat{X}_t = f(\varepsilon_t) = \sum_{i=0}^{p} H_i(\varepsilon_t) = (\eta^{(p)})^\top \phi^{(p)}(\varepsilon_t), \quad 1 \leq t \leq n,
\]

where \( \eta^{(p)} \in \mathbb{R}^M \). Similar to Equation (10) the optimal solution can be expressed as follows:

\[
\hat{\eta}^{(p)} = \arg \min_{\eta^{(p)}} n^{-1} \sum_{t=1}^{n} (f(\varepsilon_t) - X_t)^2 + \lambda(\eta^{(p)})^\top \eta^{(p)},
\]

where \( \lambda \) is additionally introduced as a regularizing penalty, which accounts for the noise in the real data and can be determined in practice, e. g., via cross-validation. This
solution is not yet based on kernels and is computationally no more efficient than the solution to Equation (9). However, by reformulating (6) as in (12) one can employ the fact that the space of functions \( \phi_i(\varepsilon_t), i = 0, \ldots, p \), has the structure of a RKHS, see Schölkopf and Smola (2001). Namely, it can be shown that

\[
\phi_i(\varepsilon_t) \phi_i(\varepsilon_{t'}) = (\varepsilon_t^\top \varepsilon_{t'})^i \equiv k_i(\varepsilon_t, \varepsilon_{t'}), \quad 1 \leq t, t' \leq n,
\]

where \( k_i(\varepsilon_t, \varepsilon_{t'}) \) is the \( i \)-th degree homogeneous polynomial kernel. Consequently, one can also write the scalar product of the maps \( \phi^{(p)}(\varepsilon_t) \) as follows:

\[
\phi^{(p)}(\varepsilon_t) \phi^{(p)}(\varepsilon_{t'}) = \sum_{i=0}^{p} a_{i}^{2}(\varepsilon_t^\top \varepsilon_{t'})^i \equiv k^{(p)}(\varepsilon_t, \varepsilon_{t'}), \quad 1 \leq t, t' \leq n.
\]

Due to the RKHS structure of the space of the functions \( \phi_i(\varepsilon_t), i = 0, \ldots, p \), it follows from the representer theorem that the optimal solution to Equation (12) can be expressed in terms of kernels as follows:

\[
\hat{X}_t = f(\varepsilon_t) = \sum_{i=0}^{p} H_i(\varepsilon_t) = X^\top (K_p + \lambda I_n)^{-1} k^{(p)}(\varepsilon_t), \quad 1 \leq t \leq n, \quad (14)
\]

where \( X = (X_1, \ldots, X_n) \) denotes a \( n \times 1 \) vector, \( K_p \) is the (positive definite) \( n \times n \) Gram matrix with entries \( k^{(p)}(\varepsilon_t, \varepsilon_{t'}), \quad 1 \leq t, t' \leq n \), and \( k^{(p)}(\varepsilon_t) \in \mathbb{R}^{n \times 1} \) denotes the \( t \)-th column of \( K_p \), \( 1 \leq t \leq n \).

To recover the coefficients of each Volterra kernel individually, note that the coefficient vector \( \eta_i = (h^{(i)}_{1,1,\ldots,1}, h^{(i)}_{1,2,\ldots,1}, \ldots)^\top \) of the \( i \)-th order Volterra operator can equivalently written as follows:

\[
\eta_i = a_i \Phi_i^\top (K_p + \lambda I_n)^{-1} X, \quad 1 \leq i \leq p,
\]

where \( \Phi_i = (\phi_1(\varepsilon_1), \ldots, \phi_i(\varepsilon_n))^\top \) is the matrix containing all monomials corresponding to the \( i \)-th order Volterra operator.

The choice of an appropriate penalty \( \lambda \) as well as the choice of an order \( p \) and a degree \( m \) of the finite (truncated) Volterra representation can be performed either by minimizing the in-sample MSE of a corresponding fit for each possible \( \lambda, p \), or by cross-validation in the frequency-domain as suggested by Hurvich and Zeger (1990). The latter approach is computationally much more intensive, however, less biased, whereas the former is faster, but often leads to overfitting due to the fact that the crucial assumption of cross-validation on independence of test and training sets is not valid for time series.
5 Simulation studies

In this section we summarize results of our simulation studies. We consider a linear process and several nonlinear processes to illustrate the performance of the suggested Volterra bootstrap procedure for the case of testing for autocorrelation at lag 1 based on the estimator \( \hat{\rho}(1) \); see Example 2.1. To highlight the usefulness of the suggested procedure we also perform the AR sieve bootstrap for the processes under consideration.

In particular, we consider two-sided test problems of the form

\[ H_0 : \rho(1) = c_0 \quad \text{versus} \quad H_1 : \rho(1) \neq c_0 \]

for a given value \( c_0 \in [-1, 1] \). The accuracy of the approximation of the null distribution of \( \hat{\rho}(1) \) by means of the Volterra bootstrap is assessed by reporting empirical type I error rates (i.e., relative rejection frequencies) of the hypothesis test which is given by the following scheme.

Algorithm 5.1.

1. Fix the significance level \( \alpha \) of the test, and fix a number \( B \) of bootstrap repetitions.

2. Let a Studentized version of the absolute difference between \( \hat{\rho}(1) \) and \( c_0 \) be given by

\[ D_n = \sqrt{n} \left| \frac{\hat{\rho}(1) - c_0}{\sqrt{\hat{\operatorname{Var}}(\hat{\rho}(1))}} \right|, \]

and let \( D_n^*,b \) denote the analogue of \( D_n \) based on the Volterra bootstrap process \( X^* \) according to Section 2 in the \( b \)-th bootstrap repetition.

3. Let a bootstrap \( p \)-value for testing \( H_0 \) versus \( H_1 \) be given by

\[ p_{\text{boot}} = \frac{|\{b : D_n^*,b > D_n\}| + 1}{B + 1}. \]

4. Reject \( H_0 \) in favor of \( H_1 \) iff \( p_{\text{boot}} < \alpha \).

In our simulations, we have set \( c_0 = \rho(1) \), meaning that the null hypothesis \( H_0 \) is true, and we have set \( \alpha = 5\% \). In analogy to Jentsch and Politis (2013), the true autocorrelation has been approximated by means of 20,000 Monte-Carlo simulations for each of the processes under consideration. The variance of \( \hat{\rho}(1) \) has been estimated using the formulas in
Proposition 2.1 based on 10 lead and 10 lags of the simulated process. For the AR sieve bootstrap we used the Akaike information criterion to fit the model and $p_{\text{max}} = 20$.

We consider the following processes:

P1 \textbf{AR}, $X_t = 0.75X_{t-1} + \varepsilon_t,$ $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)$.

P2 \textbf{GARCH}, $X_t = \sigma_t \varepsilon_t, \sigma_t^2 = 1 + 0.2\sigma_{t-1}^2 + 0.65\varepsilon_{t-1}^2, \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)$.

P3 \textbf{Bilinear}, $X_t = 0.6X_{t-1} + \varepsilon_t + 0.75X_{t-1}\varepsilon_{t-1}, \varepsilon_t \overset{iid}{\sim} \text{UNI}(-\sqrt{3}, \sqrt{3})$.

P4 \textbf{EXPAR}, $X_t = (0.45 + 0.48\exp(-0.96X_{t-1}^2))X_{t-1} + \varepsilon_t, \varepsilon_t \overset{iid}{\sim} \text{UNI}(-\sqrt{3}, \sqrt{3})$.

Each of the processes is generated under stationarity assumptions as stated, e. g., in Wu (2011). When simulating the corresponding time series as well as in the bootstrap procedure we have skipped the first $N$ values until the process achieves stationarity, where $N = 100$ is typically sufficient. In expressions as, for instance, the right-hand side of (13), a corresponding shift of the time index has to be considered.

For each model P1-P4 we consider time series of length $n = 100$ and 200 simulation runs with 250 bootstrap repetitions within each simulation run to assess the empirical type I error rate of the proposed bootstrap test. For all processes, we used $(\varepsilon_t)_t \overset{iid}{\sim} \mathcal{N}(0, 1)$ as well as $(\varepsilon^*_t)_t \overset{iid}{\sim} \mathcal{N}(0, 1)$ in estimation and bootstrapping based on the Volterra representation. The order $p$ and the degree $m$ of the Volterra representation have been chosen either based on minimizing the in-sample MSE or based on the procedure as in Hurvich and Zeger (1990). For regularized estimation as explained in Section 4 we used the following penalties: $\lambda \in \{10e - 7, 10e - 6, 10e - 5\}$. The maximum degree in the Volterra representation is set equal to 30, and the maximum order is set equal to $p_{\text{max}} = 10$. We report our simulation results in Table 1.

Our simulation results for P1 for the AR sieve procedure are as expected from the theoretical point of view, i. e., given that the innovations are Gaussian, the AR sieve bootstrap performs well. However, for the processes P3 - P4 this is not the case anymore as either the innovations are not Gaussian, or the process under consideration is no longer representable as in Equation (5). Interestingly, also for P2 when the process is nonlinear, but the innovations are Gaussian, the test based on AR sieve seems to violate the type I error rate in finite samples. On the other hand, the Volterra bootstrap is able to keep the type I
error rate approximately at the pre-specified significance level for nonlinear processes due to its ability to replicate the higher-order structure of the underlying process. However, for the linear processes unless the order of the Volterra representation is restricted to one, the type I error rate is slightly larger than $\alpha$, because the values of $p$ and $m$ chosen both by cross-validation and by in-sample MSE minimization result in overfitting. It is therefore necessary before deciding on the appropriate bootstrapping scheme to test the time series under consideration for nonlinearity.

Table 1: Type I errors (significance level 5%) for AR sieve and Volterra bootstrap procedures, $n = 100$, the number of Monte-Carlo repetitions is 200 and the number of bootstrap repetitions is 250. The symbol * indicates that the order of Volterra representation was restricted to one.

| Model | AR sieve | Average $p$ | Volterra | Average $(p, m)$ |
|-------|----------|-------------|----------|-----------------|
| P1    | 0.048    | 1.0         | 0.042    | $(1.0^*, 30.0)$ |
| P2    | 0.084    | 1.4         | 0.057    | $(2.8, 29.2)$   |
| P3    | 0.068    | 1.2         | 0.042    | $(2.1, 22.4)$   |
| P4    | 0.059    | 1.0         | 0.052    | $(2.4, 28.2)$   |

6 Discussion

In the present work we focus on the bootstrap procedure based on the Volterra series representation. In particular, we estimate and mimic the original process based on iid random variables. An alternative procedure can be constructed based on the lags of the original process similarly as in the AR sieve bootstrap method. For example, consider the following representation:

$$X_t = f(X_{t-1}, \varepsilon_t),$$
where $f$ is some measurable function such that $X_t$ is well-defined. This type of approach has been indicated by Barahona and Poon (1996). To our knowledge the necessary and sufficient conditions for the existence of such a representation have not been established so far. In Wiener (1958) the problem of finding a so-called (infinite) nonlinear moving average representation is dealt with in Lecture 11 ("Coding"), whereas an (infinite) autoregressive representation is addressed in Lecture 12 ("Decoding"). A somewhat more detailed discussion of these ideas is available in Kallianpur (1981) and Wiener (1964). However, neither Wiener (1958) nor Wiener (1964) established nonlinear AR-type filtering theory in full detail. A so called “coefficient matching” approach with the goal of generalizing linear AR results is attempted in Hunt et al. (1995). Furthermore, a nonlinear autoregressive representation for the case when the process under consideration is a Markov chain is worked out in Rosenblatt (1971), see also Tong (1990).

Another direction for formulating bootstrap procedures for nonlinear processes might be to consider its frequency domain representation employing higher-order spectra as in Brillinger (1970), Brillinger (1994), Shiryaev (1960), Shiryaev (1963), and Priestley (1988).

We reserve these ideas as well as practical applications of the suggested bootstrap procedure for future research.

References

Barahona, M. and Poon, C. (1996). Detection of nonlinear dynamics in short, noisy time series. *Nature*, 381:215–217.

Boyd, S. and Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Transactions on circuits and systems*, 32(11):1150–1161.

Brillinger, D. (1970). The identification of polynomial systems by means of higher order spectra. *Journal of Sound and Vibration*, 12(3):301–313.

Brillinger, D. R. (1994). Some basic aspects and uses of higher-order spectra. *Signal Processing*, 36(3):239–249.
Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148.

Efron, B. (1979a). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26.

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

Franz, M. O. and Schölkopf, B. (2004). Implicit estimation of Wiener series. In *Proceedings of the 14th IEEE Signal Processing Society Workshop Machine Learning for Signal Processing, 2004*, pages 735–744.

Franz, M. O. and Schölkopf, B. (2005). Implicit Wiener Series for Higher-Order Image Analysis. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 465–472. MIT Press.

Franz, M. O. and Schölkopf, B. (2006). A unifying view of Wiener and Volterra theory and polynomial kernel regression. *Neural Comput.*, 18(12):3097–3118.

Härdle, W., Horowitz, J., and Kreiss, J.-P. (2003). Bootstrap methods for time series. *International Statistical Review*, 71(2):435–459.

Hunt, L., DeGroat, R., and Linebarger, D. (1995). Nonlinear AR modeling. *Circuits, Systems and Signal Processing*, 14(5):689–705.

Hurvich, C. M. and Zeger, S. L. (1990). A frequency domain selection criterion for regression with autocorrelated errors. *Journal of the American Statistical Association*, 85(411):705–714.

Jentsch, C. and Politis, D. N. (2013). Valid resampling of higher-order statistics using the linear process bootstrap and autoregressive sieve bootstrap. *Comm. Statist. Theory Methods*, 42(7):1277–1293.

Kallianpur, G. (1981). Some ramifications of Wiener’s ideas on nonlinear prediction. In P. Masani (Ed.): *Norbert Wiener, Collected Works III with Commentaries*, pages 402–424. MIT Press Cambridge, MA.
Kalouptsidis, N. and Koukoulas, P. (2005). Blind identification of Volterra-Hammerstein systems. *IEEE Trans. Signal Process.*, 53(8, part 1):2777–2787.

Korenberg, M. J. and Hunter, I. W. (1996). The identification of nonlinear biological systems: Volterra kernel approaches. *Annals of biomedical engineering*, 24(2):250–268.

Kreiss, J.-P. and Lahiri, S. N. (2012). Bootstrap methods for time series. In *Handbook of statistics*, volume 30, pages 3–26. Elsevier.

Kreiss, J.-P. and Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40(4):357–378.

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

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer Series in Statistics. Springer, New York.

Lee, Y. and Schetzen, M. (1965). Measurement of the Wiener kernels of a non-linear system by cross-correlation. *International Journal of Control*, 2(3):237–254.

Mathews, V. and Sicuranza, G. (2000). *Polynomial signal processing*. Wiley Series in Telecommunications and Signal Processing. Wiley.

Nisio, M. (1960). On polynomial approximation for strictly stationary processes. *Journal of the Mathematical Society of Japan*, 12(2):207–226.

Orcioni, S., Terenzi, A., Cecchi, S., Piazza, F., and Carini, A. (2018). Identification of volterra models of tube audio devices using multiple-variance method. *Journal of the Audio Engineering Society*, 66(10):823–838.

Politis, D., Romano, J., and Wolf, M. (1999). *Subsampling*. Springer Series in Statistics. Springer, New York.

Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis*. Academic Press.
Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, 91(434):590–600.

Rosenblatt, M. (1971). *Markov Processes. Structure and asymptotic behavior*. Springer, New York, Heidelberg, Berlin.

Rosenblatt, M. (2009). A comment on a conjecture of N. Wiener. *Statistics & Probability Letters*, 79(3):347–348.

Rugh, W. (1981). *Nonlinear system theory*. Johns Hopkins University Press, Baltimore, Maryland.

Sandberg, I. W. (2002). Fading memory and extensions of input-output maps. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 49(11):1586–1591.

Schetzen, M. (2006). *The Volterra and Wiener theories of nonlinear systems. Revised edition*. Krieger Publishing Company.

Schölkopf, B. and Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press Cambridge, MA.

Shiryaev, A. N. (1960). Some problems in the spectral theory of higher-order moments. *Theory of Probability & Its Applications*, 5(3):265–284.

Shiryaev, A. N. (1963). On conditions for ergodicity of stationary processes in terms of higher order moments. *Theory of Probability & Its Applications*, 8(4):436–439.

Tong, H. (1990). *Nonlinear time series: a dynamical system approach*. Clarendon Press, Oxford, England.

Volterra, V. (1959). *Theory of functionals and of integral and integro-differential equations*. Dover Publications.

Wiener, N. (1958). *Nonlinear problems in random theory*. MIT Press Cambridge, MA.

Wiener, N. (1964). *Selected Papers of Norbert Wiener. Edited by Y. M. Lee, Norman Levinson and W. T. Martin*. MIT Press Cambridge, MA.

Wray, J. and Green, G. G. (1994). Calculation of the Volterra kernels of non-linear dynamic systems using an artificial neural network. *Biological cybernetics*, 71(3):187–195.
Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface*, 4(2):207–226.