Computational Statistics and Data Analysis 177 (2023) 107580

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Identification of the differencing operator of a non-stationary time series via testing for zeroes in the spectral density ☆

Tucker S. McElroy [a,*], Agnieszka Jach [b]

[a] U.S. Census Bureau, Research and Methodology Directorate, Washington, DC, USA
[b] Department of Finance and Economics, Hanken School of Economics, Helsinki, Finland

ARTICLE INFO

Article history:
Received 10 September 2021
Received in revised form 21 June 2022
Accepted 22 July 2022
Available online 29 July 2022

Keywords:
Seasonality
Subsampling
Unit roots

ABSTRACT

A nonparametric procedure for identifying the differencing operator of a non-stationary time series is presented and tested. Any proposed differencing operator is first applied to the time series, and the spectral density is tested for zeroes corresponding to the polynomial roots of the operator. A nonparametric tapered spectral density estimator is used, and the subsampling methodology is applied to obtain critical values. Simulations explore the effectiveness of the procedure under a variety of scenarios involving non-stationary processes.

Published by Elsevier B.V.

1. Introduction

The identification of time series models for forecasting, seasonal adjustment, and other applications is of central importance in federal statistical agencies (as well as in private industry) that process economic time series data. In such applications, trend and seasonality are important sources of non-stationarity (due to the nature of economic activity – see Findley et al. (1998)), and can be described through differencing polynomials (an approach summarized in Bell and Hillmer (1984)) or through deterministic functions of unknown structure – see McElroy and Politis (2020) for discussion of both approaches. This latter tactic can be approached through nonparametric techniques, such as filtering (e.g., based upon the reproducing kernel Hilbert space methodology of Dagum and Bianconcini (2008, 2016)), wavelets (Alexandrov et al., 2012), or regularized singular value decomposition (Lin et al., 2020). The former approach, which relies on identifying unit roots in a differencing polynomial, is more commonly used in official statistics (see McElroy (2021) or McElroy and Roy (2021) for discussion), and is the focus of this article.

The differencing polynomial $\delta(z)$ has all its roots on the unit circle of the complex plane (henceforth, "unit roots"), and is defined such that the time series $\{X_t\}$ is stationary after differencing, viz. $\{Y_t\}$ is stationary, where

$$Y_t = \delta(L)X_t \tag{1}$$

and L is the lag operator (McElroy and Politis, 2020). The approach to time series data analysis first advocated in Box and Jenkins (1970) involves determining marginal transformations of the data, followed by identification of $\delta(z)$, and finally the formulation of an appropriate stationary time series model, such as an Autoregressive Moving Average (ARMA). The early

☆ This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau.

* Corresponding author.
E-mail addresses: tucker.s.mcelroy@census.gov (T.S. McElroy), agnieszka.jach@hanken.fi (A. Jach).

https://doi.org/10.1016/j.csda.2022.107580
0167-9473/Published by Elsevier B.V.

exploratory approaches were later made rigorous through the unit-root testing literature – see Dickey et al. (1986), and the summary in Hamilton (1994). Although extensions were made in Elliott et al. (1996) and Ng and Perron (2001), difficulties with the practical performance have been remarked on by many authors (Schwert (1989), Pantula (1991), De Jong et al. (1992), Perron and Ng (1996)).

The asymptotic distributions for unit-root tests involve the stationary portions of the process. For instance (cf. Ling and Li (1998)), if we suppose that the data process is a pseudo-AR process with some unit roots and some stable roots (i.e., those with magnitude greater than one), then the asymptotic behavior of Ordinary Least Squares (OLS) parameter estimates depends on the stationary portions of the pseudo-AR polynomial. It is also presumed that one has identified the AR process correctly – but this identification is contingent upon having identified the unit roots, and carries its own uncertainty. McElroy (2021) studies the impact of AR order mis-specification and identification on test statistics related to unit root tests; identification is a serious issue. These difficulties persist if one considers ARMA models instead of just a simple AR.

The goal of this paper is to provide a procedure to identify $\delta(z)$ that avoids the circularity of the unit root testing literature as well as avoiding the uncertainty associated with ARMA order specification. We potentially over-difference the time series – beginning with a putative $\delta(z)$ determined by exploratory techniques – such that we are confident that $\{Y_t\}$ is stationary, and then test for whether there are zeroes in its spectral density at frequencies corresponding to the phase of the roots of $\delta(z)$. This approach of testing for over-specification has a rich history: Tanaka (1996), Davis and Dunsmuir (1996), and Tam and Reinsel (1997) studied first order moving average processes with a unit root, and this approach was generalized to higher orders in Chen et al. (2011), Davis and Song (2011), and Larsson (2014). In order to remove the nuisance of ARMA specification, nonparametric approaches to estimating the spectral density were considered in Lacroix (1999); our procedure is similar in spirit, but avoids the difficulties in implementation of Lacroix (1999) by utilizing a fixed-b bandwidth asymptotic theory (McElroy and Jach, 2019) together with subsampling methodology (Politis et al. (1999)), thereby obtaining a testing procedure that avoids post-model selection biases.

In over-specification testing, one supposes that an initial specification of $\delta(z)$ has been provided, perhaps obtained through exploratory analysis based on fitting an AR(p) model, or by using the identification procedure of Gómez (2013) – this latter technique is explained in detail in Section 2.3.1 below. Then one seeks to test whether this $\delta(z)$ is specified correctly. The main novelty of our contribution is to adopt a nonparametric approach to testing for zeroes in the spectral density using a more modern asymptotic theory (cf. the vanishing bandwidth fraction asymptotics of Lacroix (1999)), and to construct a feasible methodology through backwards deletion and subsampling. The pitfalls of specifying an ARMA model are avoided; although the user must choose taper and bandwidth for the spectral estimator, the impact of these choices upon critical values is automatically addressed through the subsampling methodology.

2. Methodology

2.1. Testing for zeroes in the spectral density

Suppose that a given series $\{X_t\}$ is thought to possibly have non-stationary behavior, such that there exists a unit root polynomial $\delta(z)$ with $\{Y_t\}$ given by (1) being weakly stationary and invertible. We can mis-identify $\delta(z)$ by omitting some roots (under-identification) or by including unnecessary roots (over-identification). Hence, it is possible to both under- and over-identify $\delta(z)$ at the same time. However, in this paper we will focus on the over-identification problem only, presuming that under-identification is not an issue. This approach is justified by a methodology that is liberal in assigning potential unit roots to $\delta(z)$.

For example, suppose that as a first stage we fit $\{X_t\}$ with an AR(p) via OLS for various orders p; we can use a p determined by some information criterion, such as AIC or BIC. By plotting the roots of the AR polynomial in the complex plane $\mathbb{C}$, we can identify potential unit roots by finding which roots are sufficiently close to the unit circle. In the approach of Gómez (2013) (which, however, focuses on the unit root $z = 1$) the real and imaginary parts of the roots are compared to sequences of numbers determined by the sample size T. In this way one can construct a prospective $\delta^{(0)}(z)$, and set $W_t = \delta^{(0)}(L)X_t$. We can attempt to ensure that there is no under-identification by examining a spectral density estimate of $\{W_t\}$ for large peaks.

Henceforth assuming that $\{W_t\}$ is weakly stationary, we wish to determine if $\{X_t\}$ has been over-differenced. Suppose that the specified $\delta^{(0)}(z)$ possibly over-differences; in such a case the superfluous unit roots correspond to a polynomial $\epsilon(z)$, and $\delta^{(0)}(z) = \delta(z)\epsilon(z)$. Then if f is the spectral density of $\{Y_t\}$, the spectral density of $\{W_t\}$ is

$$g(\lambda) = |\epsilon(e^{-i\lambda})|^2 f(\lambda).$$

Because $\{Y_t\}$ is invertible by assumption, $f > 0$ and hence any zeroes in g must correspond to unit roots of $\epsilon(z)$, i.e., $g(\lambda) = 0$ if and only if $e^{-i\lambda}$ is a root of $\epsilon(z)$. Clearly, it is only necessary to test all the roots of $\delta^{(0)}(z)$, whether or not they correspond to zeroes of g. If none of these are zeroes, then $\epsilon(z) = 1$ and $\delta^{(0)}(z) = \delta(z)$. Otherwise, any identified zeroes must correspond to $\epsilon(z)$, and then we can deduce the true differencing polynomial via $\delta(z) = \delta^{(0)}(z)/\epsilon(z)$.

Specifically, say $\delta^{(0)}(z)$ has degree d, with roots $\zeta_1, \ldots, \zeta_d$ that are assumed to be distinct. The case that the root $z = 1$ is double may arise in practice, but in such a case we apply our procedure to the trend-differenced data, which amounts to

testing whether the order of $z = 1$ is one or two. (The occurrence of a double root at other frequencies is unlikely to arise in practice – since it would correspond to seasonality of a highly non-stationary nature, which is seldom encountered in economic data – but could be handled in the same way as the case $z = 1$.) A root ζ_j may be complex, in which case there is always some other root equal to its conjugate.

Let $\omega_1, \ldots, \omega_d$ be the phase portions of the roots, such that $\zeta_j = \exp\{i\omega_j\}$, and consider the index set $\{1, 2, \ldots, d\}$ corresponding to the proposed roots. We then test for spectral zeroes using some test statistic; rejection indicates that $\delta(z) = \delta^{(0)}(z)$, but failure to reject indicates that $g(\omega_j) = 0$ for at least one ω_j. We can then examine various subsets of the frequencies to determine $\delta(z)$ corresponding to a null hypothesis of a sub-collection, i.e., an index set $J \subset \{1, 2, \ldots, d\}$ such that the minimum over $j \in J$ of the test statistic rejects, but the inclusion of any additional frequency does not reject.

2.2. Spectral density estimation and testing

Suppose that a given weakly stationary process $\{W_t\}$ is thought to be non-invertible, i.e., there are frequencies $\omega \in [-\pi, \pi]$ such that $g(\omega) = 0$, where $g(\lambda) = \sum_h \gamma_h e^{-i\lambda h}$ for any $\lambda \in [-\pi, \pi]$, and $\gamma_h = \mathrm{Cov}[W_{t+h}, W_t]$. If we thought that $\{W_t\}$ was q-dependent, we could use the estimator

$$\widehat{g}_q(\lambda) = \sum_{|h| \le q} \widehat{\gamma}_h e^{-i\lambda h}, \tag{2}$$

where $\widehat{\gamma}_h$ is the sample autocovariance based on a sample of size T from $\{W_t\}$. Note that whereas $\widehat{g}_q$ is real-valued, it need not be a non-negative function of frequency when $q < T - 1$. The asymptotic behavior of (2) is given by

$$\sqrt{T}\,\big(\widehat{g}_q(\lambda) - g(\lambda)\big) \stackrel{\mathcal{L}}{\Longrightarrow} \underline{Z}'\,\underline{\upsilon}(\lambda),$$

where $\underline{\upsilon}(\lambda)' = [1, 2\cos(\lambda), \ldots, 2\cos(\lambda q)]$ and $\underline{Z} \sim \mathcal{N}(0, V)$ such that V is the $q+1 \times q+1$ asymptotic covariance matrix of the sample autocovariances at lags 0 through q, i.e., the jkth entry (for $0 \le j, k \le q$) is given by

$$V_{jk} = \frac{2}{2\pi} \int_{-\pi}^{\pi} \cos(\lambda j)\, \cos(\lambda k)\, g(\lambda)^2\, d\lambda.$$

(This weak convergence follows directly from Theorem 3 of McElroy (2021), which is concerned with spectral density estimators for MA(q) processes.) The asymptotic variance can be estimated by plugging in the periodogram and dividing by 2, approximating the integral by a Riemann sum over Fourier frequencies – see Chiu (1988) and Deo and Chen (2000).

The above asymptotic result can be used as the basis for testing whether $g(\omega) = 0$ for some given ω. Such a test can also be rendered joint over a finite set of frequencies. In particular, suppose we wish to test

$$H_0 : g(\omega_j) = 0 \quad \text{for at least one } j \in \{1, \ldots, d_J\}, \tag{3}$$

for some subset J (with d_J elements) of the frequencies $\omega_1, \ldots, \omega_d$ (recall that these are the phase portions of the d roots of $\delta^{(0)}(z)$). Then our test statistic is

$$T \min_{1 \le j \le d_J} \big(\widehat{g}(\omega_j)\big)^2,$$

which under H_0 has asymptotic distribution $\min_{1 \le j \le d_J} \big(\underline{Z}'\,\underline{\upsilon}(\omega_j)\big)^2$. Hence this test statistic is large if and only if each $|\widehat{g}(\omega_j)|^2$ is large.

A problem with this approach is that size and power depend upon the correct specification of q, and incorporating the uncertainty due to estimation of q is something we wish to avoid. Instead, we can take a nonparametric approach that reflects our uncertainty about q. We can view (2) as a tapered spectral density estimator that uses a truncation taper at all lags h exceeding q; generalizing the taper yields the estimator

$$\widehat{g}_{b,T}(\lambda) = \sum_{|h| \le T} \Lambda_b(h/T)\, \widehat{\gamma}_h e^{-i\lambda h}, \tag{4}$$

where $\Lambda_b(x) = \Lambda(x/b)$ and Λ is the taper, an even function with domain $(-1, 1)$. Principal examples of tapers are the Parzen (p.391 of Taniguchi and Kakizawa (2000)) and the Bartlett (p. 292 of McElroy and Politis (2020)) tapers. The bandwidth is bT, and $b \in (0, 1)$ is known as the bandwidth fraction – see McElroy and Politis (2014) for fixed-b asymptotics for spectral density estimation. In the case that $g(\lambda) = 0$, it is shown in McElroy and Jach (2019) that

$$T\,\widehat{g}_{b,T}(\lambda) \stackrel{\mathcal{L}}{\Longrightarrow} S,$$

where S is an absolutely continuous random variable that depends on b, Λ, and parameters of the data process $\{W_t\}$. Note that in the case that Λ corresponds to the truncation taper, then the estimator (4) with $q = [bT]$ (where the bracket denotes the greatest integer function) reduces to (2); however, q is now a function of T, and grows proportionately to sample size.

Because the asymptotic distribution of the random variable S depends on unknown parameters of the data process, it is convenient to use the subsampling methodology to approximate the sampling distribution of $T\ \widehat{g}_{b,T}(\lambda)$. In general, a statistic $\widehat{\theta}_T$ of rate τ_T has subsampling distribution function (Politis et al. (1999))

$$L_{T,B}(x|\tau.) = \frac{1}{T-B+1} \sum_{t=1}^{T-B+1} 1_{\{\tau_B \widehat{\theta}_{T,B,t} \le x\}},$$

where B is the subsample size ($B \to \infty$, $B/T \to 0$), and $\tau_B = B$; also, $\widehat{\theta}_{T,B,t}$ is the evaluation of the estimator upon the tth subsample

$$\widehat{\theta}_{T,B,t} = \widehat{g}_{b,T,B,t}(\lambda) = \sum_{|h| \le B} \Lambda_b(h/B) \widehat{\gamma}_{h,B} e^{-i\lambda h}.$$

Here $\widehat{\gamma}_{h,B}$ is the sample autocovariance at lag h based on the tth subsample $W_t, W_{t+1}, \ldots, W_{t+B-1}$ of size B, for $t = 1, 2, \ldots, T-B+1$. For the bandwidth bB the same fraction b is used as for $\widehat{\theta}_T$. We reject the null hypothesis $g(\lambda) = 0$ at a significance level α if

$$\tau_T\, \widehat{\theta}_T \ge L_{T,B}(1-\alpha|\tau.)^{-1},$$

where $L_{T,B}(1-\alpha|\tau.)^{-1}$ is a lower $1-\alpha$ quantile of the subsampling distribution.

When the test is performed over several frequencies $H_0 : g(\omega_j) = 0$ for at least one $j \in \{1, \ldots, d_J\}$ (where J is a given subset of $\{1, 2, \ldots, d\}$ with d_J elements), for the tth subsample of size B we compute $B\ \widehat{g}_{b,T,B,t}(\omega_j)$ for several ω_j and take the minimum over $j \in \{1, \ldots, d_J\}$. We repeat this calculation for $t = 1, 2, \ldots, T-B+1$ to obtain the subsampling distribution function and $1-\alpha$ quantile. If the full-sample statistic $\min_{1 \le j \le d_J} T\ \widehat{g}_{b,T}(\omega_j)$ exceeds that quantile, we reject the (joint) null hypothesis.

2.3. Identification procedure

There are two facets to our proposed procedure. First, an initial specification of $\delta(z)$ is formulated, perhaps based on exploratory analysis. We call this *Stage 1*, and below we describe an adaptation of the method of Gómez (2013) to obtain the initial $\delta(z)$. Then in *Stage 2* we test for zeroes in $\delta(z)$ using the nonparametric estimator together with subsampling, and utilize a backwards deletion strategy. Stages 1 and 2 have been implemented in R software, which is available from the Github repository https://github.com/tuckermcelroy/RootId.

2.3.1. Stage 1: construction of $\delta^{(0)}(z)$

We explicitly describe the procedure of Gómez (2013), and its adaptation to unit roots $z \ne 1$. First, the original procedure of Gómez (2013) has as its objective to discriminate between I(1) (non-stationary with a unit root at frequency 0) and I(0) (stationary) processes.

Procedure of Gómez (2013):

- Step 1
  a) Fit a sample of length T from $\{X_t\}$ with an AR(p) process $X_t + \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} = Z_t$ via OLS for $p = 6$, denoting the estimates by $\widehat{\beta}_i$ for $i = 1, 2, \ldots, p$.
  b) Compute the inverse roots $\widehat{\lambda}_i$ of the polynomial $\widehat{\beta}(z) = 1 + \widehat{\beta}_1 z + \ldots + \widehat{\beta}_p z^p = \prod_{i=1}^{p} (1 - \widehat{\lambda}_i z)$.
  c) If $Re(\widehat{\lambda}_i) > 1 - h_T$ and $|Im(\widehat{\lambda}_i)| < h_T$ for some $1 \le i \le p$, where $h_T = 1/T^{\alpha_T}$ and $\alpha_T = 0.5 - 1/T$, declare $\{X_t\}$ to be an I(1) process and stop; otherwise, go to Step 2.
- Step 2
  a) Fit a sample of length T from $\{X_t\}$ with an ARMA(1,1) process $X_t + \phi X_{t-1} = Z_t + \theta Z_{t-1}$ via the Hannan and Rissanen (1982) method, denoting the estimates by $\widehat{\phi}$ and $\widehat{\theta}$, and setting $\widehat{\lambda} = -\widehat{\phi}$.
  b) If $\widehat{\lambda} > 1 - h_T$ and $|\widehat{\phi} - \widehat{\theta}| > c$, where $h_T = 1/T^{\beta_T}$, $\beta_T = 0.5 - 1/T^{0.55}$ and $c = 0.11$, declare $\{X_t\}$ to be an I(1) process; otherwise, as an I(0) process.

Modified Procedure of Gómez (2013): We propose a variant that allows for unit roots z away from frequency zero (so $z \ne 1$). We fit a sample of length T from $\{X_t\}$ with an AR(p) via OLS of order $p = 6 + 2 \cdot [s/2]$, where s is the number of seasons per year (e.g., $s = 12$ for monthly data), and $[s/2] = s/2 - 1$ if s is even and $(s-1)/2$ otherwise. We consider the trend frequency $\omega = 0$ as well as seasonal frequencies $\omega_j = 2\pi j/s$ ($j = 1, 2, \ldots, [s/2]$); we also examine the frequency π. For the trend frequency, we follow the procedure of Gómez (2013) exactly (with $p = 6 + 2 \cdot [s/2]$). For the seasonal frequencies, we include Step 1c) of Gómez (2013) as follows: for a given seasonal frequency ω_j, we identify a "large" (estimated) seasonal inverse root – together with its conjugate pair (at that frequency) – if there exist indices i and i', such that:

- $Mod(\widehat{\lambda}_i) > 1 - h_T$ and $|Arg(\widehat{\lambda}_i) - \omega_j| < h_T$, for $Arg(\widehat{\lambda}_i) > 0$, and
- $Mod(\widehat{\lambda}_{i'}) > 1 - h_T$ and $|\omega_j + Arg(\widehat{\lambda}_i)| < h_T$, for $Arg(\widehat{\lambda}_{i'}) < 0$, with $Mod(\widehat{\lambda}_i) = Mod(\widehat{\lambda}_{i'})$.

For frequency π we check the first of the above bullet points. In the case of a "large" inverse root at the trend frequency $\omega = 0$, the corresponding difference polynomial is $(1 - z)$. In the case of a "large" inverse root at the seasonal frequency ω_j, the corresponding difference polynomial is $(1 - 2\cos(\omega_j)z + z^2)$; at frequency π it is $(1 + z)$. Finally, the polynomial $\delta^{(0)}(z)$ is obtained as a product of the trend and seasonal polynomials.

This article focuses upon the over-identification problem, and hence we assume that $\delta^{(0)}(z)$ obtained at the end of Stage 1 contains all the true unit roots – although it may contain some spurious roots as well. The original procedure of Gómez (2013) is strongly consistent in its identification of the unit root at frequency 0, and the Modified Procedure is a straightforward adaptation of Gómez' technique. Nevertheless, decision errors can arise from such a methodology, and the more cautious investigator may simply construct $\delta^{(0)}(z)$ so as to include all seasonal (and trend) roots (i.e., roots of the form $\exp\{2\pi ik/s\}$ for $-[s/2] \le k \le [s/2]$). So long as s is integer, this means that $\delta^{(0}(z) = 1 - z^s$, corresponding to a seasonal difference.

2.3.2. Stage 2: identification of $\delta(z)$

Our input data is a sample of length T from $\{X_t\}$. We have d frequencies of interest ω_j (for $j = 1, 2, \ldots, d$), and unit root differencing polynomials $\delta^{(0j)}(z)$ associated with them, which after multiplication yield $\delta^{(0)}(z)$. In other words, $\delta^{(0)}(z) = \prod_{j=1}^{d} \delta^{(0j)}(z)$. We assume that $\delta^{(0j)}(z)$ for $j = 1, 2, \ldots, d$ have distinct roots, some of which can be complex, in which case we obtain a conjugate pair.

The overall strategy is to successively delete single factors from $\delta^{(0)}(z)$; we refer to this as our "backward deletion" approach. Initially, we consider deleting any of the d single factors, testing in each in turn; we combine these d tests by sorting the p-values and comparing the smallest p-value to our threshold α, or alternatively to α/d if we apply the Bonferroni procedure. Rejection means that we have found the true configuration, but otherwise we proceed to a subsequent step, where we consider removing any pair of factors. Now there are $\binom{d}{d-2}$ possibilities, and each of these is tested, again with the minimal p-value being compared to α (or alternatively to $\alpha/\binom{d}{d-2}$).

The reason we consider the minimal p-value in these comparisons, is that if the minimum exceeds the given threshold, then all the tests fail to reject at that level; then we know that some over-differencing has been done, although we don't know exactly which roots are responsible (since we are doing several joint tests). So we reduce the size of the differencing polynomial by one factor, considering all such subset polynomials, and proceed to test again. Proceeding in this stepwise fashion, rather than considering all 2^d polynomials at once, allows us to make deductions from the failure to reject the null hypothesis at each stage.

In summary, we proceed through the following steps:

- Perform differencing of $\{X_t\}$ with $\delta^{(0)}(L) = \prod_{j=1}^{d} \delta^{(0j)}(L)$ to obtain $W_t = \delta^{(0)}(L)X_t$, and jointly test $g(\omega_j) = 0$ for $j = 1, 2, \ldots, d$. Let the p-value be that obtained from the (full) joint test with null hypothesis (3).
- If the null hypothesis is rejected (p-value $\le \alpha$), set $J = \{1, 2, \ldots, d\}$ and stop. (This means that we did not over-difference $\{X_t\}$, i.e., there are no MA unit roots encountered in $\{W_t\}$ and the putative operator $\delta^{(0)}(L)$ coincides with $\delta(L)$.)
- If the null hypothesis is not rejected (p-value $> \alpha$), that means that we over-differenced $\{X_t\}$, and $\{W_t\}$ has at least one MA unit root. For $d = 1$, we set $J = \emptyset$ and terminate: since ω_1 is the only frequency considered, it is the one at which non-invertibility was found, and hence $\{X_t\}$ was generated by a stationary process that should not have been differenced. Otherwise, we employ backward deletion to identify the frequencies associated with non-invertibility (i.e., frequencies ω_j whose index j should *not* be in the index set J).
- **Backward deletion**: Begin by setting $j = d - 1$. While $j \ge 1$ and the p-value is "large", do the following:
  Consider all $K = \binom{d}{j}$ sets $J_1, J_2, \ldots, J_K$ obtained as subsets of the index set $\{1, 2, \ldots, d\}$ of cardinality j. For each $1 \le \ell \le K$, compute $\delta^{(J_\ell)}(z) = \prod_{j \in J_\ell} \delta^{(0j)}(z)$, obtain $W_t^{(J_\ell)} = \delta^{J_\ell}(L)X_t$, and perform the joint test $g(\omega_j) = 0$ for $j \in J_\ell$ (where g is the spectral density of $\{W_t^{(J_\ell)}\}$), denoting the corresponding p-value by p_ℓ. Sort all p-values from the smallest to the largest, expressed as $p_{(1)}, p_{(2)}, \ldots, p_{(K)}$, and set the p-value for the jth step to be $p_{(1)}$, the smallest, and let the test threshold be α (or alternatively, α/K).
  - If the p-value is greater than the test threshold, reduce j by 1 (and repeat this step).
  - If the p-value is less than or equal to the test threshold, we stop and conclude that the corresponding subset $J_{(1)}$ is the true configuration J.

  If we end up with $j = 1$ and a "large" p-value, that means that we over-differenced at each frequency. We return J equal to the empty set (this means that our data $\{X_t\}$ was generated by a stationary process).
- Final result of Stage 2: set $\delta(z) = \prod_{j \in J} \delta^{(0j)}(z)$ for non-empty J and $\delta(z) = 1$ otherwise.

An advantage of our nonparametric approach is that test statistics are quickly computed (no models need to be fitted), and the subsampling critical values are less expensive typically than other resampling techniques (such as time series

bootstraps). We remark that in the backward deletion step there are quite a few tests to be considered; in essence, there are 2^d configurations of unit roots in play. For any specified subset polynomial (described through a particular index set J), the joint test is asymptotically correctly sized due to the critical values being obtained from the subsampling procedure. However, multiple such tests are compared through the sorted p-values, and hence the overall Type I error rate is not α. To address this one can apply the Bonferroni procedure to control the family-wise error rate, and use α/K instead as the test threshold, although such procedures are predicated on having independent tests – which is not true in the context of multiple test statistics computed from a common time series sample. Comparing to α/K will make rejection of the null hypothesis more difficult, thereby favoring differencing polynomials with fewer unit roots. In our implementation we use the threshold α at each stage, which errs more on the side of including too many unit roots – in our view, a less serious error.

3. Simulations

We focus our simulations on Stage 2 of this article's procedure, taking as given the proposal of $\delta^{(0)}(z)$. This allows us to focus on the Type I and II error rates by isolating the novel contribution of our research, viz. Stage 2. In the first part of this section we assess the performance of Stage 2 of the procedure for data generated from AR(2d) processes with various parameterizations (corresponding to different specifications of seasonal frequencies), while in the second phase we consider AR(2d+1) processes (combining the seasonal specification with a trend frequency).

3.1. Specification of the simulations

The number of replicates is 500 and the sample size is $T \in \{120, 240\}$, which corresponds to 10 and 20 years of monthly observations, respectively; all figures for the $T = 120$, and most of the figures for the $T = 240$ case, are included in Appendix A of the Supplementary Material. Several values of the subsample size are used, which constitute about 3-20% of the sample size T; these are selected according to the formula $\lfloor T(0.75)^k \rfloor$ for integers k such that $k_{\min} \le k \le k_{\max}$, where $k_{\min} = 6$ and $k_{\max} = 12$. (We require at least 5 observations in a subsample.) In addition, three different fixed-b fractions, viz. $0.3, 0.5, 0.7$, are employed. Results for the Parzen taper are discussed below, but additional results with the Bartlett taper are included in Appendix A of the Supplementary Material for comparison (Figs. A.9, A.10, A.11 therein). The nominal level of the subsampling test is set to 0.01, with additional results obtained with 0.05 (and Parzen taper); these can be found in Appendix A of the Supplementary Material (Figs. A.15, A.16, A.17).

To begin, we suppose that $\{X_t\}$ is generated from an AR(2d) process for $d \ge 1$. Specifically, $\{X_t\}$ is defined such that $\prod_{j=1}^{d}(1 - 2\rho_j \cos(\omega_j)L + \rho_j^2 L^2)X_t = Z_t$, where $Z_t \sim$ i.i.d.$N(0,1)$ and $\omega_j = \frac{2\pi j}{12}$ for $j = 1, 2, \ldots, d$. In this formulation, ρ_j is the modulus of the inverse root at the frequency ω_j of the (atomic) seasonal polynomial $(1 - 2\rho_j \cos(\omega_j)z + \rho_j^2 z^2)$; if it is equal to 1, then there is a (seasonal) unit root at the corresponding frequency. For each value of $d \in \{1, 2, 3, 4\}$ we choose several sets of ρ_js, as described in Table A.1 of Appendix A of the Supplementary Material, with the first set (for each value of d) corresponding to a unit root at all seasonal frequencies of interest. For the prospective difference operator $\delta^{(0)}(z)$ (where $W_t = \delta^{(0)}(L)X_t$), we set

$$\delta^{(0)}(L) = \prod_{j=1}^{d} \overbrace{(1 - 2\cos(\omega_j)L + L^2)}^{\delta^{(0j)}(L)}.$$

For each given d, the first choice of the ρ_js (see Table A.1 of Appendix A of the Supplementary Material) is unit, and hence the corresponding $\delta^{(0)}(z)$ has the correct amount of differencing, whereas the other choices for the ρ_js – in subsequent rows of the table – all result in $\delta^{(0)}(z)$ being over-specified, such that there will be MA unit roots in the differenced series $\{W_t\}$ (at some, or even all, of the frequencies). The objective of Stage 2 of the procedure is to identify the largest index set $J \subset \{1, 2, \ldots, d\}$ for which the product of $\delta^{(0j)}(L)$, for $j \in J$, will *not* lead to over-differencing. This is the target operator $\delta(L)$.

In the second phase of our studies we add frequency zero to the seasonal frequencies, and suppose that $\{X_t\}$ is such that $\prod_{j=1}^{d}(1 - 2\rho_j \cos(\omega_j)L + \rho_j^2 L^2)(1 - \phi L)X_t = Z_t$, for $Z_t \sim$ i.i.d.$N(0,1)$, where $\omega_j = \frac{2\pi j}{12}$ for $j = 1, 2, \ldots, d$ and $\omega_{d+1} = 0$. In this formulation, ρ_j for $1 \le j \le d$ are as before while ϕ is the inverse root at the zero frequency (i.e., the trend frequency, $\omega_{d+1} = 0$). Specific values of the parameters can be found in Table A.2 of Appendix A of the Supplementary Material for $d \in \{2, 4\}$. We difference a sample from $\{X_t\}$ with $\delta^{(0)}(L) = \prod_{j=1}^{d}(1 - 2\cos(\omega_j)L + L^2)(1 - L)$, which will lead to over-differencing in all except for the first parametrization within a given AR(2d+1) Data Generating Process (DGP).

3.2. Summary of the results

The results are presented as a sequence of multi-panel figures (Figs. 1-2 and Figs. A.1-A.2 of Appendix A of the Supplementary Material for AR(2d) DGPs and Figs. 3-4 for AR(2d+1) DGPs), whose lay-out is as follows: subplot titles give the values of the ρ_js (and of ϕ, if applicable), x-axis shows the subsample size (for the sample size of $T = 240$), y-axis

gives the proportion of correct identifications of the set J, and the different plotting symbols correspond to different values of fixed-b fraction. Analogous figures for $T = 120$ can be found in Appendix A of the Supplementary Material. Also therein we consider a data-driven rule for choosing the subsampling size B outlined in Götze and Račkauskas (2001) and Bickel and Sakov (2008). Among the candidate subsample sizes B of the form described previously, we calculate the Kolmogorov-Smirnov distance between two consecutive subsampling distribution functions

$$\tilde{L}_{T,B}(x) = \frac{1}{T-B+1} \sum_{t=1}^{T-B+1} 1_{\{\min_{1\le j\le d} B\, \widehat{g}_{b,T,B,t}(\omega_j)\le x\}},$$

for B_1 and B_2, for B_2 and B_3, and so on. We focus on the pair that minimizes this distance, and choose the smaller of the two values of B as the optimal one. This procedure is applied to AR(2d) DGPs for $d \in \{1, 3\}$ and AR(2d+1) DGP for $d = 2$ (Figs. A.12, A.13, and A.14 of Appendix A of the Supplementary Material).

Overall, the performance of the methodology used in Stage 2 is quite good when the DGP is AR(2d) (Figs. 1-2 and Figs. A.1-A.2 of Appendix A of the Supplementary Material) and fairly good when the DGP is AR(2d+1) (Figs. 3-4) with $T = 240$; for $T = 120$ (Figs. A.3-A.8 of Appendix A of the Supplementary Material) the proportions of correct identifications of the set J are lower. Therefore we recommend to apply the procedure when sample size is at least 200. We note that there are some scenarios which prove to be more challenging for Stage 2 of the identification procedure. For example, the AR(2d) DGP for $d = 1$, when $\rho_1 = 0.9$ (column 2 of Figure A.1 of Appendix A of the Supplementary Material) is a difficult case, which is understandable given that ρ_1 is close to the inverse unit root. Likewise, a challenging identification arises for the AR(2d) DGPs with $d = 4$, when there are two seasonal unit roots at the two largest frequencies (row 2, column 4 of Fig. 1) or one seasonal unit root at the largest frequency (row 4, column 3 of Fig. 1). Another difficult case is the AR(2d+1) DGP for $d = 4$, when there is more unnecessary differencing conducted (row 1, columns 3-4 and row 2, column 3 of Fig. 4). Additionally, we note that the Parzen taper yields a somewhat higher proportion of correct identifications over the Bartlett taper (compare 'Parzen' Figs. A.3, A.5, A.7 and 'Bartlett' Figs. A.9, A.10, A.11 in Appendix A of the Supplementary Material). Also, large values of subsample size should not be used, as the performance of the method degrades (very small subsample sizes, however, should also be avoided). A practical remedy is to utilize a data-driven subsample size selection, which provides satisfactory results; we see that the subsample sizes that yield high proportions of correct identifications are matched by the data-driven subsample size with a similar proportion (compare Figs. A.3, A.5, A.7 with Figs. A.12, A.13, A.14 of Appendix A of the Supplementary Material). Finally, in most cases a fixed-b ratio of 0.7 provides inferior results, apparently being too large. (Also, the value of 0.3 for the AR(2) DGPs is too small.) We recommend using a moderate bandwidth fraction of 0.5.

4. Data application

We illustrate the method by applying it to two monthly time series published by the U.S. Census Bureau (see https://www.census.gov/construction/nrc/historical_data/index.html), from the New Residential Construction sector. Specifically, we study "Annual Rate for Housing Units Authorized in Permit-Issuing Places" (Series 1) and "Annual Rate for Housing Units Started" (Series 2), each of which is a monthly time series (Total Units, United States, Not Seasonally Adjusted) beginning on January 1959 (Series 1) and January 1968 (Series 2), respectively, and both ending on September 2019. (We downloaded the data on 6:51 EDT of November 5, 2019, and the data has not been updated in order to avoid Covid-19 effects.) The first time series is shown in the top left panel of Fig. 5 and the second in the top left of Fig. 6.

These are fairly typical economic time series, in that the strong trend and seasonal pattern are not unusual. Although macroeconomists may be interested in forecasting a time series, a statistical agency is often interested more simply in publishing seasonally adjusted data. Proper identification of the seasonal unit root structure facilitates identification and extraction of the seasonality (McElroy, 2008), and for this reason it is important to correctly identify the roots of the differencing operator; mis-identification of the unit roots can lead to inappropriate seasonal adjustment filters, whereby seasonality remains behind (McElroy and Roy, 2021). Whereas long-horizon forecast comparisons may be of interest in their own right, such comparisons have also long been used by the seasonal adjustment community to assess the appropriateness of stationary model specifications – see Findley (1990, 1991), Diebold and Mariano (1995), and Rivers and Vuong (2002) – and we use this technique below to compare specifications of the differencing polynomial.

With such an application in mind, we proceed to compare outcomes based on Stage 1 only (Section 2.3.1) with those obtained with Stage 1 and Stage 2 (Section 2.3.1 and Section 2.3.2) used together; we refer to these as Modified Gómez (MG) and Zero Testing (ZT) respectively, for short. We also compare these results to a unit root identification procedure based upon the root diagnostic of McElroy (2021), further discussed below. We begin by employing the TRAMO-SEATS procedure (U.S. Census Bureau, 2017, p. 70-72 of the documentation) implemented in the `R` package *seasonal* (Sax and Eddelbuettel, 2018) to each time series in order to identify a transformation, any outliers, and pertinent regression effects; these are identified based on a default (011)(011) SARIMA model, and the resulting output time series is referred to as "linearized." (Further details on this pre-processing are given in Appendix B of the Supplementary Material.) The sample autocorrelation function of the linearized data for Series 1 is shown in the top center panel of Fig. 5 and that of Series 2 in Fig. 6. Linearized data is inputted into Stage 1 to identify $\delta^{(0)}(z)$, and then is passed on to Stage 2 (in the ZT procedure), where a fixed-b ratio of 0.5 is used, the Parzen taper is employed, the nominal level is set to 0.05, and the fourth largest subsample size B from the data-driven rule is selected.

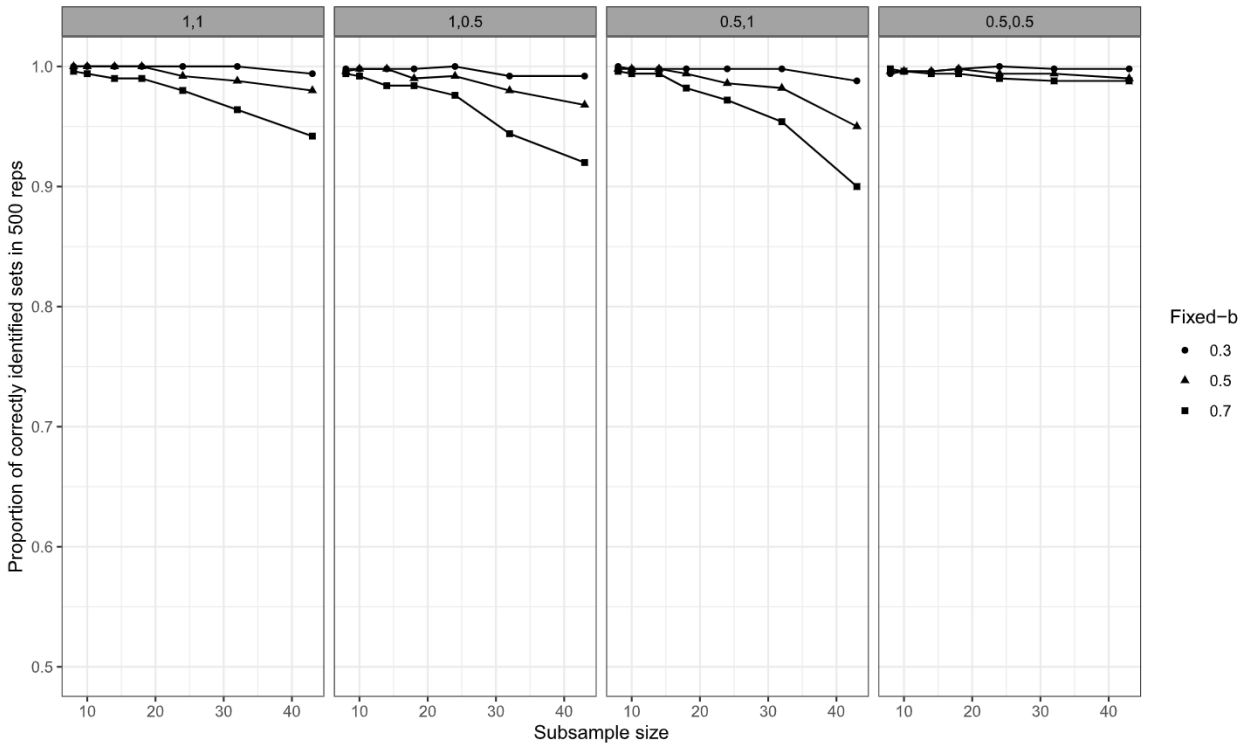

Fig. 1. Proportion of correct identification of the set J in Stage 2 of the procedure for AR(2d), $d = 2$, DGP of Table A.1 of Appendix A of the Supplementary Material and $\delta^{(0)}(L) = \prod_{j=1}^{d}(1 - 2\cos(\omega_j)L + L^2)$. Subplot titles give the ρ_j, $j = 1, 2, \ldots, d$. Sample size is 240, subsample size is on the x-axis, nominal size of the subsampling test is 0.01, fixed-b fraction is $0.3, 0.5, 0.7$ (legend). Parzen taper is employed.

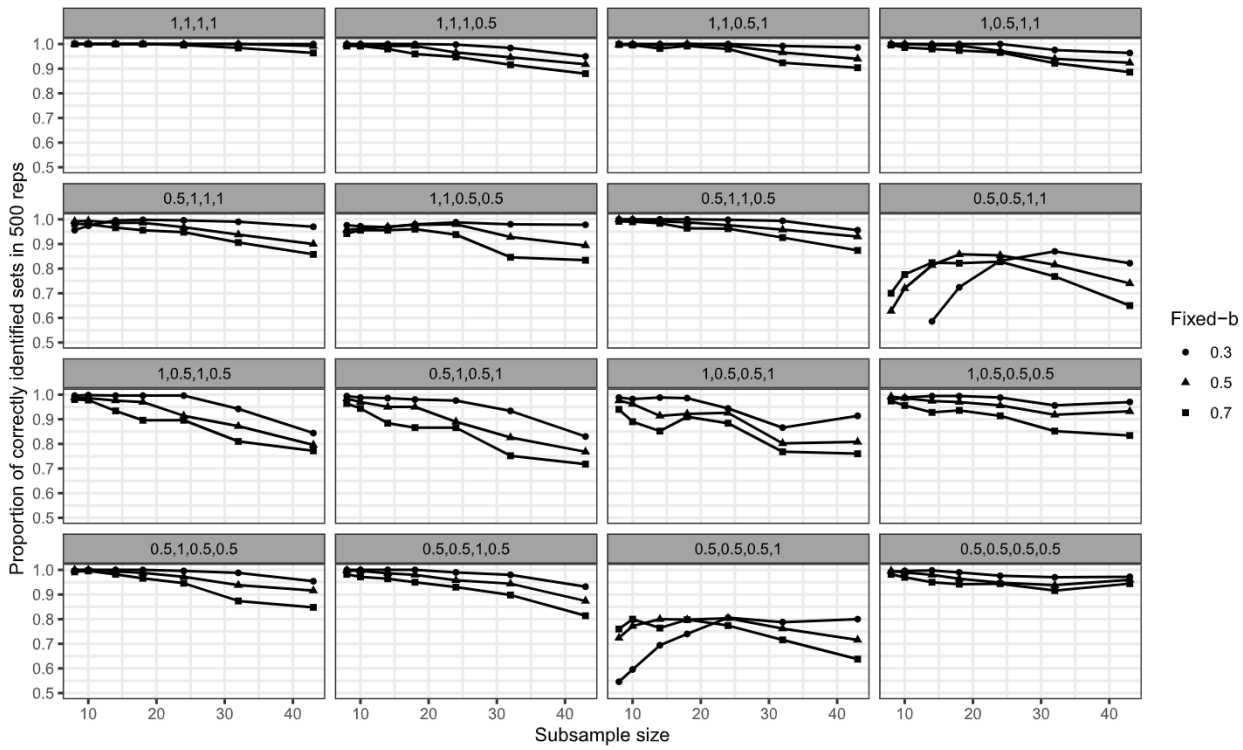

Fig. 2. Proportion of correct identification of the set J in Stage 2 of the procedure for AR(2d), $d = 4$, DGP of Table A.1 of Appendix A of the Supplementary Material and $\delta^{(0)}(L) = \prod_{j=1}^{d}(1 - 2\cos(\omega_j)L + L^2)$. Subplot titles give the ρ_j, $j = 1, 2, \ldots, d$. Sample size is 240, subsample size is on the x-axis, nominal size of the subsampling test is 0.01, fixed-b fraction is $0.3, 0.5, 0.7$ (legend). Parzen taper is employed.

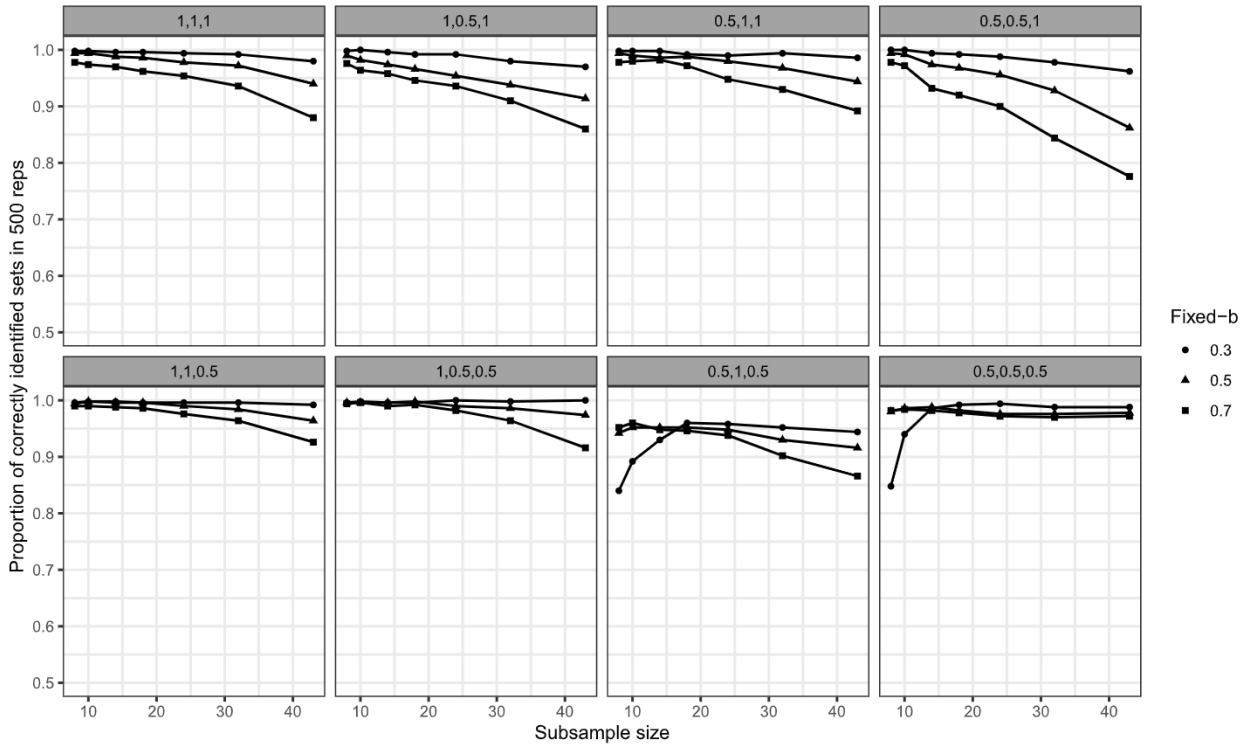

Fig. 3. Proportion of correct identification of the set J in Stage 2 of the procedure for AR(2d+1), $d = 2$, DGP of Table A.2 of Appendix A of the Supplementary Material and $\delta^{(0)}(L) = \prod_{j=1}^{d}(1 - 2\cos(\omega_j)L + L^2)(1 - L)$. Subplot titles give the ρ_j, $j = 1, 2, \ldots, d$ and ϕ. Sample size is 240, subsample size is on the x-axis, nominal size of the subsampling test is 0.01, fixed-b fraction is 0.3, 0.5, 0.7 (legend). Parzen taper is employed.

Table 1
Identification results for Series 1 and Series 2 based on the Modified Gómez (MG), AR root (ROOT), and Zero Testing (ZT) procedures. The tests are applied to linearizations of Series 1 and Series 2.

	Series 1			Series 2		
frequency	MG	ROOT	ZT	MG	ROOT	ZT
0	Yes	Yes	Yes	Yes	Yes	Yes
$2\pi \cdot 1/12$	Yes	Yes	Yes	Yes	Yes	Yes
$2\pi \cdot 2/12$	Yes	Yes	Yes	Yes	Yes	Yes
$2\pi \cdot 3/12$	Yes	Yes	No	No	Yes	No
$2\pi \cdot 4/12$	Yes	Yes	No	Yes	Yes	Yes
$2\pi \cdot 5/12$	Yes	Yes	No	Yes	Yes	No
π	Yes	Yes	No	Yes	Yes	Yes

Because in Stage 1 the estimated inverse roots of the prospective difference polynomial $\delta^{(0)}(z)$ at the frequencies of interest are examined, we display them in the complex plane in the top right panels of Figs. 5-6 for Series 1 and Series 2, respectively. Stage 1 suggests that for Series 1 all frequencies under consideration ($2\pi \cdot j/12$ for $j = 0, 1, 2, \ldots, 6$) are associated with a unit root (not an atypical result for monthly construction data); however, only the frequencies for $j = 0, 1, 2$ are selected in Stage 2. For Series 2, all except for frequency $2\pi \cdot 3/12$ are identified in Stage 1, but after Stage 2 the unit root at frequency $2\pi \cdot 5/12$ is excluded. The linearized data is differenced with $\delta^{(0)}(z)$ obtained from the MG procedure; the $\delta(z)$ obtained from the ZT procedure method is summarized in Table 1. The respective sample autocorrelation functions (ACF) are shown in the bottom panels of Fig. 5 for Series 1 and of Fig. 6 for Series 2.

As a comparison, we also implement the AR root diagnostic (as discussed on pages 376–377 of McElroy (2021)) on the linearized data to detect strong seasonal autoregressive roots. Based on the discussion in McElroy (2021), we formulate a criterion that the $1-\alpha$ confidence interval for ρ should contain all values of $\rho > .97$, or some other suitably high threshold. Since the test provides us with a p-value for each specified value of the null ρ, we can write this as $p(\rho)$, and the criterion states that the smallest $p(\rho)$ that is greater than α must exceed .97 in order for the corresponding root to be considered unit. For seasonal frequencies we use the threshold of $\rho > .97$, whereas for the trend frequency we set the threshold much lower, at .85; this is because under-specification of the trend unit root carries more substantial costs for forecasting and

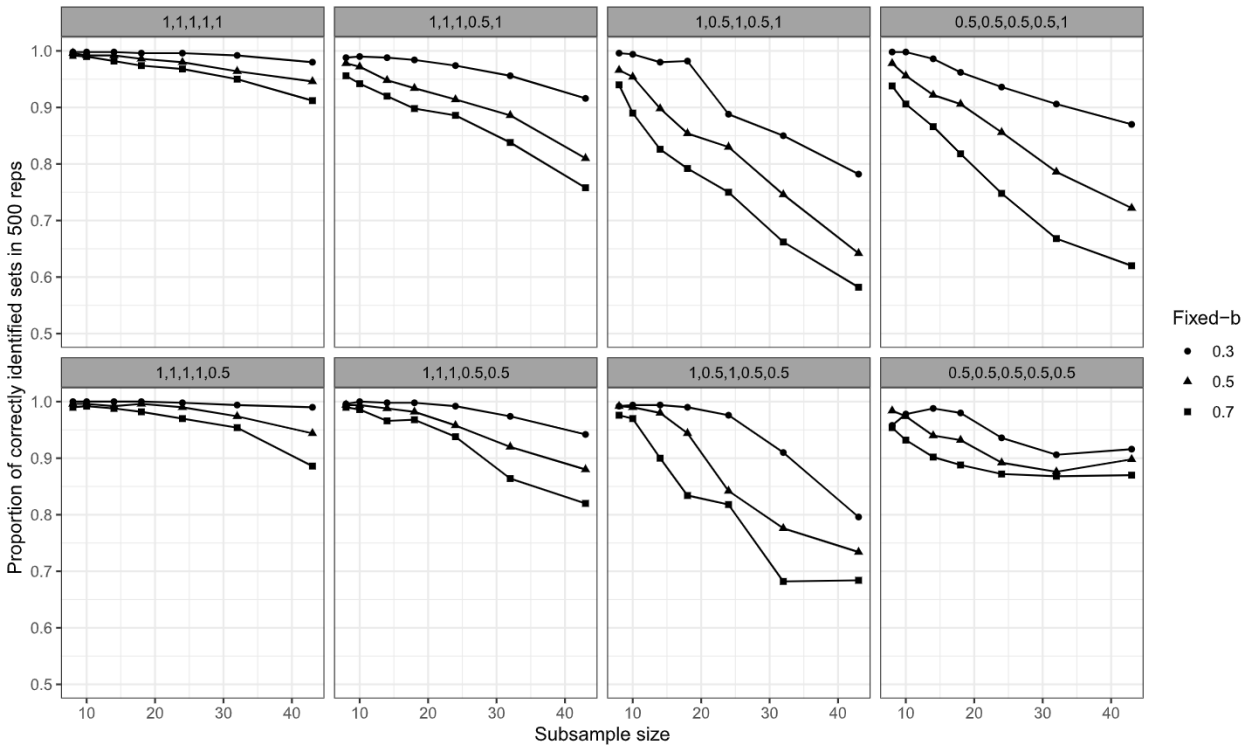

Fig. 4. Proportion of correct identification of the set J in Stage 2 of the procedure for AR(2d+1), $d = 4$, DGP of Table A.2 of Appendix A of the Supplementary Material and $\delta^{(0)}(L) = \prod_{j=1}^{d}(1 - 2\cos(\omega_j)L + L^2)(1 - L)$. Subplot titles give the ρ_j, $j = 1, 2, \ldots, d$ and ϕ. Sample size is 240, subsample size is on the x-axis, nominal size of the subsampling test is 0.01, fixed-b fraction is 0.3, 0.5, 0.7 (legend). Parzen taper is employed.

Table 2
The smallest value of ρ such that $p(\rho) > \alpha$, where $p(\rho)$ is the p-value for the test $H_0(\rho)$: $\pi(\rho^{-1}e^{i\omega_j}) = 0$, and $\omega_j = 2\pi \cdot j/12$ for $0 \le j \le 6$. The tests are applied to linearizations of Series 1 and Series 2.

	Series 1		Series 2	
frequency	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
0	0.923	0.924	0.896	0.897
$2\pi \cdot 1/12$	0.995	0.995	0.992	0.993
$2\pi \cdot 2/12$	0.994	0.995	0.988	0.989
$2\pi \cdot 3/12$	0.984	0.986	0.983	0.985
$2\pi \cdot 4/12$	0.994	0.994	0.991	0.992
$2\pi \cdot 5/12$	0.993	0.994	0.986	0.988
π	0.987	0.988	0.984	0.986

signal extraction. (We found that omitting $1 - L$ from $\delta(L)$ results in very poor forecast performance, in comparison to the MG and ZT specifications.) We refer to this as the ROOT procedure; for values of $\alpha = .01, .05$ we report the smallest ρ such that $p(\rho) > \alpha$ in Table 2, and the results are summarized in Table 1.

Because the true $\delta(z)$ for the data is unknown, to assess the outcomes we use out-of-sample forecast performance. We do this by using the identified differencing operator (either $\delta^{(0)}(L)$ for the MG procedure, or $\delta(L)$ from the ROOT procedure, or $\delta(L)$ for the ZT procedure), together with a fitted model for the differenced data, to forecast each time series up to 24 steps ahead (this corresponds to two years of monthly values). We use an AR model – fitted using OLS, based on an automatic order selection obtained via AIC – to capture the stationary portion of the differenced data; hence our forecast model amounts to

$$\delta(L)\phi(L)X_t \sim \text{i.i.d.}\,(0, \sigma^2),$$

where $\delta(z)$ is obtained from either the MG, ROOT, or ZT procedures, and $\phi(z)$ is the AR polynomial. A 1-step ahead forecast is obtained via $\widehat{X}_{t+1} = (1 - \delta(L)\phi(L))X_{t+1}$, i.e., simply multiply $\delta(z)$ with the fitted $\phi(z)$, subtract from one, and weight present and past values of the time series according to the coefficients. Multi-step ahead forecasts are obtained by iteratively

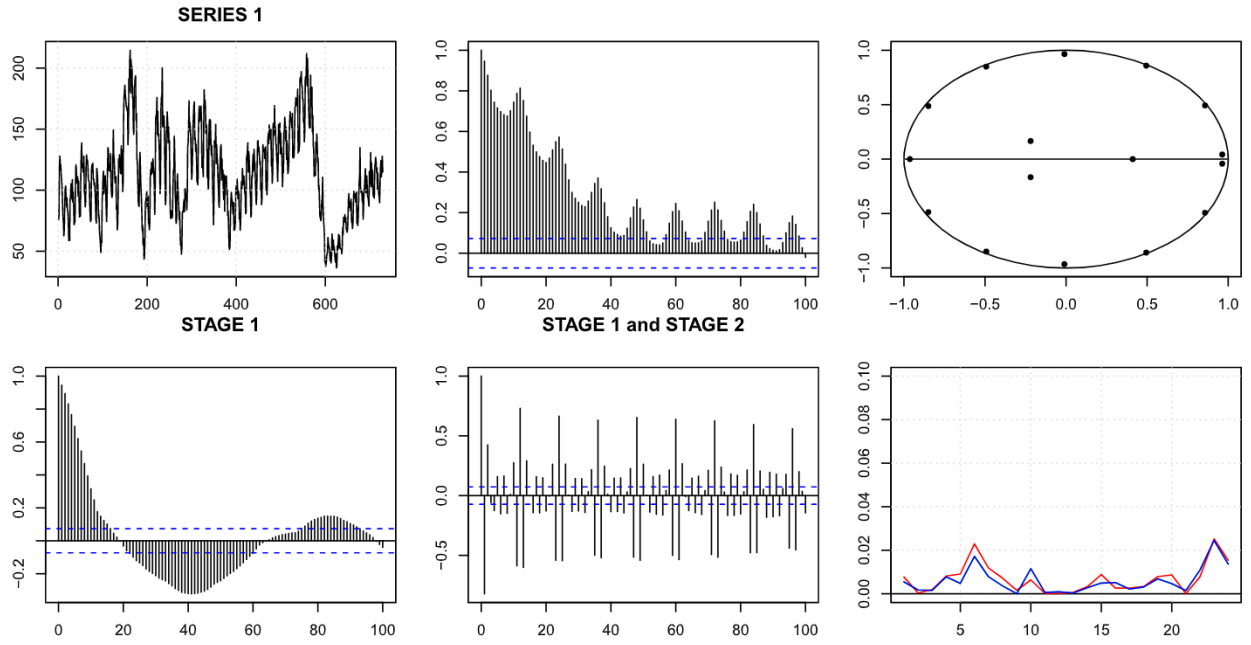

Fig. 5. Top: time series plot of Series 1, sample ACF of the linearized Series 1, estimated inverse roots of the prospective differencing polynomial $\delta^{(0)}(z)$ at the frequencies of interest in the complex plane. Bottom: sample ACF of the Series 1 based on the MG procedure's differencing, sample ACF of the Series 1 based on the ZT procedure's differencing, time series plot of linearized data, and squared out-of-sample forecast errors for MG (green), ROOT (blue), and ZT (red). (For interpretation of the colors in the figures, the reader is referred to the web version of this article.)

appending previous forecasts to the end of the extended series, and applying the 1-step ahead filter – see Proposition 10.11.3 of McElroy and Politis (2020).

In this manner forecasts are generated for the 24 withheld values (where the AR model is fitted to this reduced span), and the forecast errors can be directly computed. The squared forecast errors, by forecast horizon (between 1 and 24), are plotted in the bottom right panels of Figs. 5-6; red corresponds to the ZT procedure, with blue for ROOT and green for MG. Note that if insufficient differencing were supplied to the series, to some extent the fitted AR model will rectify this problem, because OLS is used as a fitting method (OLS allows for the estimation of unit and explosive autoregressive roots). For Series 1 the ROOT procedure (which yields a specification of $\delta(L)$ that is identical to that of the MG procedure) has superior performance at lower leads, but the ZT specification is competitive overall, and competitive at some of the higher horizons. Note that the difference in the two specifications occurs at the seasonal frequencies $2\pi \cdot 3/12$, $2\pi \cdot 4/12$, $2\pi \cdot 5/12$, and $2\pi \cdot 6/12$. Examining the sample ACF based on the ZT procedure's differencing (central bottom panel of Fig. 5), we see a high degree of persistency at lags 11, 12, and 13 (and multiples thereof). The sample ACF based on the MG procedure's differencing, in contrast, does not have such persistencies – this indicates a possible decision error in the ZT procedure, whereby zeroes in the spectral density were fallaciously identified. In contrast, Series 2 shows an improvement of the ZT method over the MG and ROOT procedures' specifications, since the ZT has lower forecast error at most forecast leads. Comparing the sample ACF plots (bottom panels of Fig. 6), it seems that the ZT procedure benefits from not identifying a unit root at $2\pi \cdot 5/12$.

5. Conclusion

Unit root identification is an important aspect of time series model selection. In this article we approach this problem by potentially over-specifying the differencing operator needed to reduce a process to stationarity, followed by testing for the presence of zeroes in the spectral density. This is akin to the moving average unit root testing procedures advocated in the time series literature (Tanaka, 1996); in order to avoid the impact of mis-specification of the stationary part of the model – which generates size and power distortions – we estimate the spectral density nonparametrically, utilizing a tapered sequence of sample autocovariances. The quantiles of the limiting distribution are estimated using the subsampling methodology.

Classical approaches to moving average unit root detection, including the methodology of Lacroix (1999), yield a degenerate limit that necessitates further circumlocutions, ultimately requiring the user to choose further tuning parameters. Our method, in contrast, circumvents this degeneracy by employing fixed-b bandwidth asymptotics. The subsampling methodology provides asymptotically correct critical values no matter the user's choice of taper and bandwidth fraction. Whereas subsampling involves the further selection of a subsampling window size B, the procedure of Götze and Račkauskas (2001) and Bickel and Sakov (2008) provides a data-driven rule for choosing the subsampling size B. A further contribution of

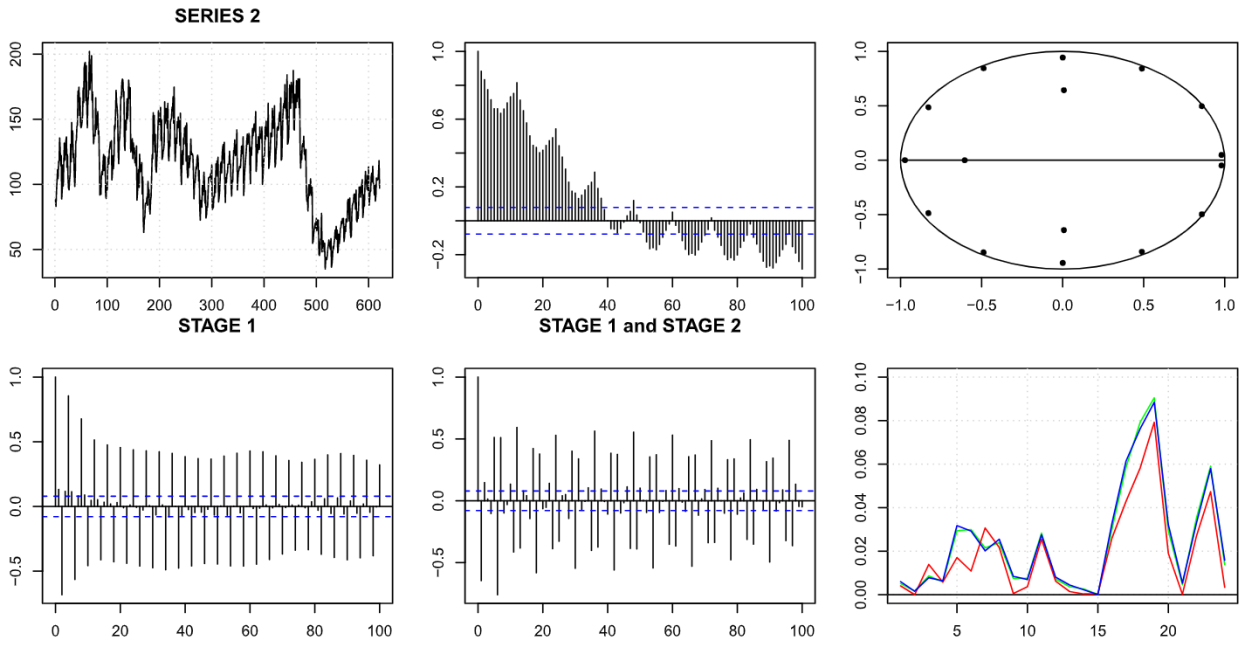

Fig. 6. Top: time series plot of Series 2, sample ACF of the linearized Series 2, estimated inverse roots of the prospective differencing polynomial $\delta^{(0)}(z)$ at the frequencies of interest in the complex plane. Bottom: sample ACF of the Series 2 based on the MG procedure's differencing, sample ACF of the Series 2 based on the ZT procedure's differencing, time series plot of linearized data, and squared out-of-sample forecast errors for MG (green), ROOT (blue), and ZT (red).

our paper is the construction of a correct differencing operator – although this is a direct application of prior zero-finding methodologies, we explicitly determine the differencing polynomial through the careful articulation of a backwards deletion algorithm.

A first step of this methodology is to identify a candidate differencing operator, and this could be done with exploratory analysis, or through simply specifying seasonal differencing; alternatively, a modification of the procedure of Gómez (2013) could be used. Following this latter approach, as a second step we apply a backwards deletion algorithm to this MG (Modified Gómez) differencing operator, calling this the ZT (Zero Testing) procedure. The article provides distribution theory and simulations for the ZT procedure (conditional on the first step being given), focusing on a range of processes with both trend and seasonal effects present. It was found that performance in simulation is good when the sample size is at least $T = 240$ and no trend unit roots are present, i.e., the purely seasonal case works better as compared to inclusion of trend unit roots. Some deterioration in performance occurs when trend unit roots are present. Furthermore, the simulations provide some guidance about the choice of taper and bandwidth size.

We illustrate this methodology on two series where the unit root identification is not self-evident, and make comparisons to the MG procedure and the AR root (ROOT) procedure of McElroy (2021). For the first time series the ZT procedure identifies less unit roots than both MG and ROOT, and the resulting forecast performance is similar. For the second series only one of the unit roots identified by MG is discarded by ZT, and the forecast performance is improved. It is possible to change the ZT procedure's decisions by altering the Type I error rate α; for instance, increasing α may yield a rejection of the null hypothesis, resulting in the corresponding factor being included in $\delta(L)$.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2022.107580.

References

Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T.S., 2012. A review of some modern approaches to the problem of trend extraction. Econom. Rev. 31 (6), 593–624.
Bell, W.R., Hillmer, S.C., 1984. Issues involved with the seasonal adjustment of economic time series. J. Bus. Econ. Stat. 2 (4), 291–320.
Bickel, P.J., Sakov, A., 2008. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. Stat. Sin. 18, 967–985.
Box, G., Jenkins, G., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
Chen, M.C., Davis, R.A., Song, L., 2011. Inference for regression models with errors from a non-invertible MA (1) process. J. Forecast. 30 (1), 6–30.
Chiu, S.T., 1988. Weighted least squares estimators on the frequency domain for the parameters of a time series. Ann. Stat. 16 (3), 1315–1326.
Dagum, E.B., Bianconcini, S., 2008. The Henderson smoother in reproducing kernel Hilbert space. J. Bus. Econ. Stat. 26 (4), 536–545.
Dagum, E.B., Bianconcini, S., 2016. Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation. Springer International Publishing.
Davis, R.A., Dunsmuir, W.T., 1996. Maximum likelihood estimation for MA (1) processes with a root on or near the unit circle. Econom. Theory 12 (1), 1–29.

Davis, R.A., Song, L., 2011. Unit roots in moving averages beyond first order. Ann. Stat. 39 (6), 3062–3091.
De Jong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H., 1992. Integration versus trend stationarity in time series. Econometrica 60, 423–433.
Deo, R.S., Chen, W.W., 2000. On the integral of the squared periodogram. Stoch. Process. Appl. 85 (1), 159–176.
Dickey, D.A., Bell, W.R., Miller, R.B., 1986. Unit roots in time series models: tests and implications. Am. Stat. 40, 12–26.
Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13, 253–263.
Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.
Findley, D., 1990. Making difficult model comparisons. SRD Research Report no. RR90/11. U.S. Census Bureau.
Findley, D.F., 1991. Convergence of finite multistep predictors from incorrect models and its role in model selection. Note Mat. XI, 145–155.
Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C., 1998. New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. J. Bus. Econ. Stat. 16 (2), 127–152.
Gómez, V., 2013. A strongly consistent criterion to decide between $I(1)$ and $I(0)$ processes based on different convergence rates. Commun. Stat. 42, 1848–1864.
Götze, F., Račkauskas, A., 2001. Adaptive choice of bootstrap sample sizes. In: Klaassen, Chris, van der Vaart, Aad, de Gunst, Mathisca (Eds.), State of the Art in Probability and Statistics. In: IMS Lecture Notes Monogr. Ser., vol. 36. Cambridge University Press, Cambridge, pp. 286–309.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.
Hannan, E.J., Rissanen, J., 1982. Recursive estimation of mixed autoregressive-moving average models. Biometrika 66, 265–270.
Lacroix, R., 1999. Testing for zeros in the spectrum of an univariate stationary process: part I. Banque de France Working Paper No. 70, available at SSRN: https://ssrn.com/abstract=1734311 or https://doi.org/10.2139/ssrn.1734311, 1999.
Larsson, R., 2014. A likelihood ratio type test for invertibility in moving average processes. Comput. Stat. Data Anal. 76, 489–501.
Lin, W., Huang, J.Z., McElroy, T., 2020. Time series seasonal adjustment using regularized singular value decomposition. J. Bus. Econ. Stat. 38 (3), 487–501.
Ling, S., Li, W.K., 1998. Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. Ann. Stat. 26 (1), 84–125.
McElroy, T., 2008. Matrix formulas for nonstationary ARIMA signal extraction. Econom. Theory 24 (4), 988–1009.
McElroy, T., 2021. A diagnostic for seasonality based upon polynomial roots of ARMA models. J. Off. Stat. 37 (2), 1–28.
McElroy, T., Jach, A., 2019. Testing collinearity of vector time series. Econom. J. 22 (2), 97–116.
McElroy, T., Politis, D., 2014. Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics. J. Econom. 182, 211–225.
McElroy, T., Politis, D., 2020. Time Series: a First Course with Bootstrap Sampler. Chapman and Hall, New York.
McElroy, T., Roy, A., 2021. A Review of Seasonal Adjustment Diagnostics. Int. Stat. Rev. (published online).
Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.
Pantula, S.G., 1991. Asymptotic distributions of unit-root tests when the process is nearly stationary. J. Bus. Econ. Stat. 9, 63–71.
Perron, P., Ng, S., 1996. Useful modifications to unit root tests with dependent errors and their local asymptotic properties. Rev. Econ. Stud. 63, 435–465.
Politis, D., Romano, J., Wolf, M., 1999. Subsampling. Spinger, New York, NY.
Rivers, D., Vuong, Q., 2002. Model selection tests for nonlinear dynamic models. Econom. J. 5, 1–39.
Sax, C., Eddelbuettel, D., 2018. Seasonal adjustment by X-13ARIMA-SEATS in R. J. Stat. Softw. 87 (11), 1–17.
Schwert, G.W., 1989. Tests for unit roots: a Monte Carlo investigation. J. Bus. Econ. Stat. 7, 147–159.
Tam, W.K., Reinsel, G.C., 1997. Tests for seasonal moving average unit root in ARIMA models. J. Am. Stat. Assoc. 92 (438), 725–738.
Tanaka, K., 1996. Time Series Analysis: Nonstationary and Noninvertible Distribution Theory. John Wiley and Sons, New York, NY.
Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer-Verlag, New York City, NY.
U.S. Census Bureau, 2017. X-13ARIMA-SEATS Reference Manual. U.S. Census Bureau, U.S. Department of Commerce, Washington, DC. Available at https://www.census.gov/ts/x13as/docX13ASHTML.pdf.