Wavelet-Based Test for Time Series Non-Stationarity[1]

Milan Bašta[2] | *University of Economics, Prague, Czech Republic*

Abstract

In the present paper, we propose a wavelet-based hypothesis test for second-order stationarity in a Gaussian time series without any deterministic components or seasonality. The null hypothesis is that of a second-order stationary process, the alternative hypothesis being that of a non-stationary process with a time-varying autocovariance function (excluding processes with unit roots). The test is based on the smoothing of the series of squared maximal overlap discrete wavelet transform coefficients employing modern techniques, such as robust filtering and cross-validation. We propose several test statistics and use bootstrap to obtain their distributions under the null hypothesis. We examine the test in settings that may mimic the properties of economic time series, showing that it enjoys reasonable size and power characteristics. The test is also applied to a data set of the U.S. gross domestic product to demonstrate its practical usefulness in an economic time series analysis.

Keywords

Wavelets, time series, non-stationarity, bootstrap, hypothesis test, gross domestic product

JEL code

C 12, C 15, C 22, C 49, E 23

INTRODUCTION

When referring to stationarity in this paper, we will mean second-order stationarity (see Section 2). Processes that are not stationary will be called non-stationary.

In the analysis of economic, financial and demographic time series, non-stationary time series are often assumed to have either unit roots and/or deterministic trends. The approach to unit root non-stationarity testing was pioneered by Dickey, Fuller (1979). Similar or extended approaches can be found in Said, Dickey (1984), Said, Dickey (1985) or Phillips, Perron (1988). Non-stationarity is, however, a much broader term. In fact, any time series, whose mean function or autocovariance function (to be defined in Section 2) are time-varying, is necessarily non-stationary.

We propose a bootstrap wavelet-based hypothesis test where the null hypothesis is that of stationarity and the alternative hypothesis that of a non-stationary process with a time-varying autocovariance function (generally excluding processes with unit roots).[3] The size and power of the test are estimated by Monte Carlo simulations. The results are compared with those of a test available in the literature. A data set of the U.S. gross domestic product is used to illustrate the implementation of the test.

[1] This paper was written with the support of the Czech Science Foundation (GA CR), grant no. P402/12/P507, Modelling of Financial and Economic Time Series – Application and Comparison of Wavelet and Traditional Methods.

[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. E-mail: milan.basta@vse.cz.

[3] The reason for this exclusion will be made clear in Section 3 and Section 4.

We run the Monte Carlo simulations and implement the test utilizing R software (R Core Team, 2014). The following contributed R packages were widely used during the preparation of the paper: *wmtsa* (Constantine, Percival, 2013), *locits* (Nason, 2013a) and *forecast* (Hyndman, 2015).

The paper is organized as follows. A literature review is given in Section 1. Section 2 defines the notion of second-order stationarity. Section 3 provides a short introduction to the maximal overlap discrete wavelet transform (MODWT) and the properties of MODWT coefficients. The hypothesis test is introduced in Section 4. The properties of the test (size and power) are studied in Section 5. The test is applied to the first difference of the logarithm of the U.S. gross domestic product in Section 6.

1 LITERATURE REVIEW

Regarding the tests for stationarity other than unit root ones, we can mention Grinsted et al. (2004), who followed the idea of Torrence, Compo (1998), having applied the continuous wavelet transform to an input time series, calculating the so-called wavelet power and comparing the power values to critical values obtained under the assumption that the input time series was generated by a stationary AR(1) process. This type of test detects the instances of non-stationarity localized in time and scale (in an "otherwise stationary" process).[4]

Another wavelet-based test for stationarity has been proposed by von Sachs, Neumann (2000), who used localized versions of periodogram and Haar wavelet coefficients of the periodogram to decide about the stationarity of the underlying stochastic process. A test similar to that of von Sachs, Neumann (2000) has been introduced by Nason (2013b), who studied whether a specific linear transformation of the evolutionary wavelet spectrum (Nason et al., 2000) is time-varying or not. This was accomplished by exploring Haar wavelet coefficients of the empirical wavelet periodogram. If the null hypothesis of stationarity is rejected in the test by Nason (2013b), a locally stationary wavelet model with a time-varying autocovariance function is suggested as an alternative.

Variability in the series of smoothed or averaged squared wavelet coefficients is perceived – in an exploratory and descriptive sense – as qualitative evidence against stationarity also in other papers (see, e.g., Jensen, Whitcher, 2014, Whitcher et al., 2000, Nason et al., 2000, Fryzlewicz, 2005). None of these studies, however, include a hypothesis test that would provide the significance of this evidence.

Similarly to the papers mentioned above, we also smooth the series of squared wavelet coefficients using, however, a different smoothing approach. More specifically, we note that the median function of the logarithm of squared MODWT wavelet coefficients is constant over time for stationary Gaussian processes, being, however, generally time-varying for non-stationary Gaussian processes with a time-varying autocovariance function. We propose to use practices common in the field of statistical learning (see, e.g., Hastie et al., 2011) and non-parametric regression, such as cross-validation and robust filtering, to estimate the median function. Moreover, we do not downsample the series of the logarithm of squared MODWT wavelet coefficients in order not to lose any valuable information. Further, we propose several measures of non-constancy of the estimated median function to be used as the test statistic. Because of the complexity of the estimation procedure and analytical intractability of the distribution of the test statistic under the null hypothesis, approximate p-values are found by bootstrap. This leads to a computationally expensive test which may – as demonstrated by the results of the Monte Carlo simulations – enjoy better size and power properties than the test proposed by Nason (2013b) for time series lengths common in economics. Moreover, the time series length is not required to be a power of two, which is the requirement for the practical implementation of the test by Nason (2013b) in R *locits* package

4 As will be discussed later, in hypothesis testing in general, the alternative hypothesis need not be necessarily well-specified. This is also the case of the test of Grinsted et al. (2004), where no explicit statistical model associated with the alternative hypothesis is given.

(Nason, 2013a). Moreover, our test provides a single p-value (for a given test statistic), differing from the test by Grinsted et al. (2004) where each "time-scale cell" is tested for stationarity individually and where no "global" decision about stationarity is provided.

2 SECOND-ORDER STATIONARITY

Following Brockwell, Davis (2002), let us assume a stochastic process $\{X_t: t = \ldots, -1, 0, 1, \ldots\}$ with $E(X_t^2) < \infty$, $t = \ldots, -1, 0, 1, \ldots$. Further, let $\mu_t \equiv E(X_t)$, $t = \ldots, -1, 0, 1, \ldots$, be the mean function and

$$\gamma(t+h,t) \equiv E\left[(X_{t+h} - \mu_{t+h})(X_t - \mu_t)\right], \quad t,h = \ldots,-1,0,1,\ldots, \tag{1}$$

the autocovariance function of the process, the variance function being defined as the autocovariance function for $h = 0$. The process is defined to be second-order stationary if both the mean and autocovariance functions (the latter for each h) are independent of time t.

When referring to stationarity in further parts of this paper, we will mean second-order stationarity. Processes that are not stationary will be called non-stationary.

3 MAXIMAL OVERLAP DISCRETE WAVELET TRANSFORM

In this section, the notion of MODWT coefficients is introduced together with the properties of these coefficients for stationary, integrated and locally stationary wavelet processes. These properties provide the basis for the hypothesis test.

3.1 MODWT wavelet filters and coefficients

The jth level ($j = 1, 2, 3, \ldots$) MODWT wavelet filter, denoted as $\{h_{j,l}: l = 0, \ldots, L_j - 1\}$, where L_j is the filter length, is[5] an approximately ideal linear filter for the frequency range $[1/2^{j+1}, 1/2^j]$. The filter is constructed in a very special way, fulfilling the following properties:

$$\sum_{l=0}^{L_j-1} h_{j,l} = 0, \quad \sum_{l=0}^{L_j-1} h_{j,l}^2 = \frac{1}{2^j}, \quad \sum_{l=0}^{L_j-1} h_{j,l} h_{j,l+2^j n} = 0 \text{ (for } n \neq 0). \tag{2}$$

There are various "families" of filters, such as the Haar, D(4), LA(8), etc.

Let $\{X_t: t = \ldots, -1, 0, 1, \ldots\}$ be a stochastic process which need not be stationary. The jth level ($j = 1, 2, 3, \ldots$) MODWT wavelet coefficients for $\{X_t\}$ are denoted as $\{W_{j,t}: t = \ldots, -1, 0, 1, \ldots\}$ and obtained by linear filtering $\{X_t\}$ with $\{h_{j,l}\}$, i.e.

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}, \quad t = \ldots,-1,0,1,\ldots . \tag{3}$$

Percival, Walden (2002) show that the jth level MODWT wavelet coefficients are closely related to *changes* between two adjacent weighted averages of $\{X_t\}$ values, the weighted averages being calculated on an effective scale 2^{j-1}.

From Equation 3 and the first identity given in Equation 2, it follows that the constant mean function of $\{X_t\}$ is a sufficient condition for the zero mean function of $\{W_{j,t}\}$ (for $j = 1, 2, 3, \ldots$). Equation 3 also directly implies that $\{W_{j,t}\}$ (for j = 1, 2, 3, …) is a Gaussian process provided so is $\{X_t\}$.

[5] $\{h_{j,l}\}$ characteristics are based on Percival, Walden (2002, Ch. 5).

In the next sections, we will discuss the variance function of $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$) for stationary, integrated and locally stationary wavelet processes. The characteristics of the variance function for these types of processes will provide the basis for our hypothesis test.

3.2 Variance function of $\{W_{j,t}\}$ for stationary $\{X_t\}$

Let us assume a stationary stochastic process $\{X_t\}$. Since $\{W_{j,t}\}$ is the output from linear filtering of $\{X_t\}$ with $\{h_{j,l}\}$, it is stationary too and has a zero mean function. The variance of $W_{j,t}$ is called the wavelet variance and is denoted by v_j^2, i.e.

$$\nu_j^2 = \mathrm{var}(W_{j,t}) = E(W_{j,t}^2), \quad j = 1,2,\dots;\ t = \dots,-1,0,1,\dots . \tag{4}$$

Percival, Walden (2002, pp. 296) reveal that

$$\mathrm{var}(X_t) = \sum_{j=1}^{\infty} \nu_j^2, \quad t = \dots,-1,0,1,\dots . \tag{5}$$

v_j^2 generally differs across various stationary stochastic processes.

3.3 Variance function of $\{W_{j,t}\}$ for integrated processes

Since a stationary $\{X_t\}$ has a stationary $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$), it follows that a non-stationary $\{W_{j,t}\}$ necessarily implies a non-stationary $\{X_t\}$. However, not all non-stationary processes have non-stationary $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$). Percival, Walden (2002, Ch. 8.2) show that $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$) for processes integrated of order d is stationary provided "Daubechies" filters[6] with $L_1 \geq 2d$ are used in the analysis.[7] Moreover, this implies that the variance function of $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$) is constant over time in such a situation.

3.4 Variance function of $\{W_{j,t}\}$ for locally stationary wavelet processes

In this section, we discuss a certain class of non-stationary processes, namely locally stationary wavelet processes as well as the variance function of $\{W_{j,t}\}$ for such processes. Since a thorough discussion might be too theoretical exceeding the scope and extent of this paper, the interested reader is referred to Nason et al. (2000), where details can be found.[8]

Nason et al. (2000, Def. 1) construct locally stationary wavelet processes making use of wavelet filters with random amplitudes as building blocks. Locally stationary wavelet processes are defined in a way implying that their mean function is zero (see Nason et al., 2000, pp. 274) and so is therefore the mean function of $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$). Locally stationary wavelet processes are characterized by the so-called (evolutionary) wavelet spectrum (Nason et al., 2000, Def. 2), which is generally time-varying. Nason et al. (2000, pp. 278) reveal that the generally time-varying autocovariance function of locally stationary wavelet processes tends (in a sense rigorously described in Nason et al., 2000) to the so-called local autocovariance defined in Nason et al. (2000, Def. 4) which depends on the wavelet spectrum and is time-varying in general.

Since the mean function of $\{W_{j,t}\}$ (for j = 1, 2, 3, …) is zero for locally stationary wavelet processes,[9] the variance function of $\{W_{j,t}\}$ is equal to the mean function of $\{W_{j,t}^2\}$. Nason et al. (2000, Prop. 4) show

6 Daubechies filters include, among others, the Haar, D(4) as well as LA(8) family of filters.

7 $L_1 = 2$ for Haar, $L_1 = 4$ for D(4) and $L_1 = 8$ for LA(8) filters.

8 Nason et al. (2000) utilize a different notation than that used in our paper.

9 It follows from Section 3.1 that the mean function of $\{W_{j,t}\}$ (for $j = 1, 2, 3, \dots$) would be zero even if a non-zero constant was added to the locally stationary process.

that the mean function of $\{W_{j,t}^2\}$ is (up to a remainder term) equal to a special linear transformation[10] of the wavelet spectrum. Because the wavelet spectrum is generally time-varying for locally stationary wavelet processes, so is the linear transformation. Let us denote this transformation by $\varphi_{j,t}$.

Nason et al. (2000, Prop. 3a) show that all stationary processes with an absolutely summable autocovariance function are locally stationary wavelet processes. Nason et al. (2000, Prop. 3b) also prove that any locally stationary wavelet process that has a wavelet spectrum independent of time – and fulfills an additional restriction stated in Nason et al. (2000, Prop. 3b) – is stationary with an absolutely summable autocovariance function. The constancy of the wavelet spectrum over time also implies the constancy of the local autocovariance as well as that of $\varphi_{j,t}$ over time which turns into v_j^2, as introduced in Section 3.2.

3.5 Smoothing of the series of squared wavelet coefficients

Let us assume that the mean function of $\{W_{j,t}\}$ (for $j = 1, 2, 3, \ldots$) is zero (see Section 3.1 for the sufficient condition for the zero mean function). From Section 3.2, it further follows that for stationary $\{X_t\}$, the mean function of $\{W_{j,t}^2\}$ (for $j = 1, 2, 3, \ldots$) is constant over time. From Section 3.3, it follows that for integrated processes, the mean function of $\{W_{j,t}^2\}$ is also constant over time provided a long enough wavelet filter is employed in the calculation of wavelet coefficients. From Section 3.4, it follows that the mean function of $\{W_{j,t}^2\}$ tends to $\varphi_{j,t}$ and is generally time-varying for locally stationary wavelet processes.

As a result, the characteristics of the mean function of $\{W_{j,t}^2\}$ distinguish between stationary and non-stationary processes such as locally stationary wavelet ones, but generally not between stationary and integrated processes. Exploring the behavior of squared wavelet coefficients can thus be used as a means of non-stationarity detection. Such reasoning has also been applied in Torrence, Compo (1998) or Grinsted et al. (2004) even though a different type of wavelet transform than MODWT has been used.

Since the task is to estimate the mean function of $\{W_{j,t}^2\}$ from the squared wavelet coefficients, various smoothing and averaging techniques can be utilized for this purpose. Nason et al. (2000) used denoising based on wavelet transform. We can also mention various approaches to smoothing and averaging of squared wavelet coefficients utilized in other papers as an exploratory and descriptive tool for the detection of non-stationarity. See, for example, the smoothing of squared wavelet coefficients applied in Jensen, Whitcher (2014), the averaging of squared wavelet coefficients separately within different calendar seasons (resulting in the so-called seasonal wavelet variances) utilized in the study of Whitcher et al. (2000), the application of cross-validation to the choice of the smoothing parameter while smoothing the downsampled squared wavelet coefficients in Fryzlewicz (2005). In all of these studies, the non-constancy of the smoothed series of squared wavelet coefficients or the differences between seasonal variances have been interpreted as evidence against stationarity. However, a hypothesis test is not included in these studies that would provide the significance of this evidence. Nason (2013b) does not explicitly smooth the series of squared wavelet coefficients, but calculates its Haar wavelet coefficients and provides a formal hypothesis test for stationarity. However, as demonstrated in Section 5, the test of Nason (2013b) does not seem to be suitable – due to a relatively low power – for time series lengths typical in economics.

3.6 Synchronization of wavelet coefficients and boundary effects

$\{h_{j,l}\}$ is a causal linear filter. Consequently, $\{W_{j,t}\}$ (for $j = 1, 2, 3, \ldots$) is not synchronized with $\{X_t\}$, lagging behind it. Advancing $\{W_{j,t}\}$ by $\delta_j \geq 0$ time units is a way to approximately synchronize $\{W_{j,t}\}$ with $\{X_t\}$. The value of δ_j (given in Wickerhauser, 1994, p. 341; or in Percival, Walden, 2002, p. 118) depends on both the shape of the first-level wavelet filter and j. As a result, the process $\{w_{j,t}: t = \ldots, -1, 0, 1, \ldots\}$ (for $j = 1, 2, 3, \ldots$) defined as

[10] The weights of the linear transformation are time-independent.

$$w_{j,t} \equiv W_{j,t+\delta_j}, \quad t = \ldots,-1,0,1,\ldots \tag{6}$$

is approximately synchronized with $\{X_t\}$.

It is noteworthy that the characteristics of $\{W_{j,t}\}$, which can be used to distinguish a stationary process from a locally stationary wavelet process with a time-varying autocovariance function (such as a constant vs. time-varying mean function of $\{W_{j,t}^2\}$), are shared by $\{w_{j,t}\}$.

In real life applications, the observed time series of a *finite* length N can be considered the realization of a *portion* of a stochastic process; let this portion be denoted as $\{X_t: t = 0, \ldots, N-1\}$. As a consequence, the coefficient $W_{j,t}$ of Equation 3 can be defined (if no further ad hoc assumptions on X_t for $t < 0$ are introduced) only for times $t = L_j - 1, \ldots, N - 1$. Further, the coefficient $w_{j,t}$ can be defined only for times $t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j$. Moreover, since L_j increases with j, an upper limit on the j value, denoted as J, is obtained by requiring N to be larger than L_j. As a result, j = 1, 2, …, J in real life applications.

4 HYPOTHESIS TEST

In Section 4.1, we discuss the null and alternative hypothesis of our test. In Section 4.2, we propose an approach to smoothing the series of squared wavelet coefficients. The smoothed series will establish the basis for the calculation of the test statistic (Section 4.3). Bootstrap will be used to obtain the approximation of the test statistic distribution under the null hypothesis (Section 4.4).

4.1 Null and alternative hypothesis

The null hypothesis of stationarity (Section 2) will be tested. In hypothesis testing in general, the alternative hypothesis need not be well-specified (see, e.g., Efron, Tibshirani, 1994, Ch. 16 and p. 233), the hypothesis test providing evidence whether or not the data are in agreement with the null hypothesis not necessarily pointing to any specific alternative model.

However, we can make the alternative hypothesis more specific in our test. We will *assume* that the mean function of $\{X_t\}$ is constant over time, that $\{X_t\}$ is Gaussian and contains no seasonality.[11] As argued in Section 4.2, under these assumptions, the rejection of the null hypothesis will point to a process with a time-varying autocovariance function – except for integrated processes. A locally stationary wavelet process with a time-varying autocovariance function can thus provide an excellent example of the model under the alternative hypothesis.

4.2 Smoothing in the logarithmic scale

Let the jth level (for $j = 1, 2, \ldots, J$) wavelet coefficients for $\{X_t\}$ be calculated and synchronized with $\{X_t\}$. Sequences $\{w_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ (for $j = 1, 2, \ldots, J$) are obtained as a result.

We assume that the mean function of $\{X_t\}$ is constant over time, $\{X_t\}$ being Gaussian. Consequently, $\{w_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ (for $j = 1, 2, \ldots, J$) is a Gaussian sequence with a zero mean function (see Section 3.1). We can thus write (for $j = 1, 2, \ldots, J$; $t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j$)

$$w_{j,t}^2 = \mathrm{var}(w_{j,t})U_{j,t}^2 = E(w_{j,t}^2)U_{j,t}^2 = E(w_{j,t}^2) + E(w_{j,t}^2)\left(U_{j,t}^2 - 1\right), \tag{7}$$

where $U_{j,t}$ is a zero-mean unit-variance Gaussian variable. From Equation 7 it follows that $\{w_{j,t}^2\}$ is heteroskedastic, the variance function of $\{w_{j,t}^2\}$ being proportional to the square of the mean function

[11] The assumption of no seasonality is in agreement with the fact that the mean function of $\{X_t\}$ is supposed to be constant over time, which would not be the case if deterministic seasonality was present. Moreover, for seasonally integrated processes, $\{W_{j,t}\}$ is non-stationary with a variance function varying over time. However, the proposed test is neither intended nor designed to test for seasonal unit roots in $\{X_t\}$. The assumption of no seasonality in $\{X_t\}$ thus avoids the need to deal with the seasonal pattern.

of $\{w_{j,t}^2\}$. In such a situation, logarithmic transformation stabilizes the variance (for a general discussion on variance-stabilizing transformations, see, e.g., Rawlings et al., 2001, Ch. 12.3), i.e.

$$z_{j,t} \equiv \log(w_{j,t}^2) = \log\left(E(w_{j,t}^2)\right) + \log\left(U_{j,t}^2\right), \quad (8)$$

where log stands for the natural logarithm.

Because of the properties of the mean function of $\{w_{j,t}^2\}$, which follow from Section 3, we can note that the distributions of $z_{j,t}$ (for $t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j$) generally differ only in their locations for locally stationary wavelet processes with a time-varying autocovariance function and are identical for stationary processes (and potentially also for integrated ones).

$\{w_{j,t}\}$ is correlated. The correlation, however, is very close to zero for lags greater or equal to 2^j, which is often utilized in the sense that if downsampled by 2^j, $\{w_{j,t}\}$ can be approximately treated as a sequence of uncorrelated random variables (see, e.g., Fryzlewicz, 2005, pp. 213–214, or Percival, Walden, 2002, Ch. 9). Consequently, since $\{w_{j,t}\}$ is Gaussian, $w_{j,t}$ (for a given t) can be assumed to be approximately independent of the set $\{w_{j,t+\tau}: |\tau| \geq 2^j\}$. Thus, $z_{j,t}$ (for a given t) can be assumed to be approximately independent of the set $\{z_{j,t+\tau}: |\tau| \geq 2^j\}$. Similar arguments are applied by Fryzlewicz (2005, p. 214).

4.2.1 Robust smoothing

There are alternative ways how smoothing and averaging of wavelet coefficients can be performed, as demonstrated by the examples given in Section 3.5 and noted by Fryzlewicz (2005).

We propose to work in the logarithmic scale since the logarithmic transformation leads to variance stabilization which will be useful while implementing cross-validation in Section 4.2.2. In fact, Nason et al. (2000, p. 282) also noted that logarithmic transformation could be a useful transformation applicable prior to the smoothing procedure even though they used a different approach to smoothing in the end.

It follows from the previous section that exploring the constancy of any measure of location of $\{z_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ can serve as a means to distinguish between stationary and locally stationary wavelet processes with a time-varying autocovariance function. We propose to study the constancy of the median function[12] of $\{z_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$. This choice is determined by the fact that the median function can be *robustly* estimated using a locally weighted median smoother (Härdle, Gasser, 1984, see also Fried et al., 2007, Sec. 2.1 and 2.2 for the notion of the weighted median and weighted median filtering). The robust aspect of the estimation is appealing and necessary because of a heavy left tail[13] of $z_{j,t}$. More specifically, a symmetric weighted *median* filter of an odd length D is used in the paper, with weights $\{u_k: k = -(D-1)/2, \ldots, 0, \ldots, (D-1)/2\}$ given as[14]

$$u_k = \frac{1}{\sqrt{1+|k|}}, \quad k = -\frac{D-1}{2}, \ldots, 0, \ldots, \frac{D-1}{2}. \quad (9)$$

Since the median of $z_{j,t}$ differs from $\log(E(w_{j,t}^2))$ only by a constant independent of t, the changes in the median of $z_{j,t}$ over time are directly related to those in $\log(E(w_{j,t}^2))$ over time, and thus also to the (percentage) changes in $E(w_{j,t}^2)$ over time. This enables us to qualitatively explain the temporal

[12] The median function of a sequence of random variables is defined to be a sequence of medians of the random variables.

[13] In real life applications, $w_{j,t}^2$ can also be equal to zero due to rounding issues, which consequently leads to "minus infinite" values of $z_{j,t}$.

[14] At the boundaries of $\{z_{j,t}\}$, the symmetric weighted median filter cannot be fully employed and an asymmetric weighted median filter is used instead, being constructed by assuming only that part of the symmetric filter for which data are available.

variations in the median of $z_{j,t}$ by means of the temporal variations in $E(w_{j,t}^2)$; this approach being applied in some of the following parts of the paper.

4.2.2 Cross-validation

We use cross-validation to select an "optimal" span over which smoothing is to be performed. Such an approach to smoothing is considered a very flexible "statistical learning" technique in general (Hastie et al., 2011), its use being also advocated by Fryzlewicz (2005) for the smoothing of squared wavelet coefficients. We differ from Fryzlewicz (2005) by performing smoothing in the logarithmic scale. Further, in contrast to Fryzlewicz (2005), we do not downsample the series by 2^j, but work with dependent data not to lose any information. We note that such an approach is legitimate if a modified version of cross-validation, such as the "leave-$(2r + 1)$-out" cross-validation, is utilized (see, e.g., Arlot, Celisse, 2010, Sec. 8.1 or Chu, Marron, 1991), where the validation set is constructed so that it can be considered independent of the training set. In accordance with the previous discussion on (approximate) independence in the sequence $\{z_{j,t}\}$, we put r equal to $2^j - 1$. Moreover, we do not use the usual "least squares" cross-validation criterion because of the heavy-tailed nature of $z_{j,t}$. Instead, we utilize a robust cross-validation criterion (for ideas on robust cross-validation see, e.g., Morell et al., 2013).

More specifically, we implement cross-validation as follows. For a given length of the weighted median filter and a given m, for $m = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j$, a total number of $(2r + 1)$ observations, namely $z_{j,m-r}, \ldots, z_{j,m}, \ldots, z_{j,m+r}$, are left out[15] from the sequence $\{z_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$. A prediction from the median filter for time m is constructed using the remaining observations from the sequence, e_m (the prediction error) being defined as $z_{j,m}$ minus the prediction. Subsequently, the sequence $\{e_m: m = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ is sorted in the ascending order, the lowest 12.5% and the highest 12.5% of observations in this sorted set being removed. The mean of the *absolute values* of the remaining e_m values is calculated and serves as a cross-validation criterion. The optimal D is selected as such a length of the weighted median filter which minimizes the criterion. The weighted median filter of the optimal length is used to smooth $\{z_{j,t}\}$.

4.3 Test statistic

Let $\{\mathrm{med}(z_{j,t}): t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ denote the median function of $\{z_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ and let $\{q_{j,t}: t = L_j - 1 - \delta_j, \ldots, N - 1 - \delta_j\}$ be the median function estimate made using the approach described above. As will be explained in this section, we can also be interested in a linear combination of median functions such as

$$\sum_{j=1}^{J} \alpha_j \mathrm{med}(z_{j,t}), \quad t = L_{j^*} - 1 - \delta_{j^*}, \ldots, N - 1 - \delta_{j^*}, \tag{10}$$

where α_j, for $j = 1, \ldots, J$, are real-valued weights of the linear combination and j^* is the maximum value of j for which α_j is non-zero.[16] The linear combination can be estimated by $\{q_t^{\alpha}: t = L_{j^*} - 1 - \delta_{j^*}, \ldots, N - 1 - \delta_{j^*}\}$ defined as

$$q_t^{\alpha} = \sum_{j=1}^{J} \alpha_j q_{j,t}, \quad t = L_{j^*} - 1 - \delta_{j^*}, \ldots, N - 1 - \delta_{j^*}. \tag{11}$$

[15] For $m < L_j - 1 - \delta_j + r$ or $m > N - 1 - \delta_j - r$, the observations have to be left out asymmetrically.

[16] The value of $L_j - 1 - \delta_j$ is increasing with j, whereas that of $N - 1 - \delta_j$ is decreasing with j.

Efron, Tibshirani (1994, Ch. 16) reveal that the two quantities are the components of a bootstrap hypothesis test, namely a test statistic (which need not be an estimate of any parameter) and an estimate of the probability model under the null hypothesis. The choice of the test statistic determines the power of the test. Despite some arbitrariness in its choice, the test statistic has to measure the discrepancy between the null hypothesis (stationarity) and data at hand. Consequently, any reasonable measure of non-constancy of $\{q_t^{\alpha}\}$ can be suggested as a test statistic in our case. Let the statistic be generally denoted as s.

For example, the test statistic can be defined as the standard deviation of $\{q_t^{\alpha}\}$, i.e.

$$s \equiv \sqrt{\frac{1}{N - L_{j*}} \sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} \left(q_t^{\alpha} - \overline{q}_t^{\alpha}\right)^2}, \tag{12}$$

where

$$\overline{q}_t^{\alpha} \equiv \frac{1}{N - L_{j*} + 1} \sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} q_t^{\alpha}. \tag{13}$$

Analogously, the Spearman's rank correlation coefficient between $\{q_t^{\alpha}\}$ and time t can be used as the test statistic, i.e.

$$s \equiv \frac{\sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} \left(R_t - \overline{R}\right)\left(t - \bar{t}\right)}{\sqrt{\sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} (R_t - \overline{R})^2 \sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} \left(t - \bar{t}\right)^2}}, \tag{14}$$

where

$$R_t \equiv \text{rank of } q_t^{\alpha} \text{ in} \{q_t^{\alpha} : t = L_{j*} - 1 - \delta_{j*}, \ldots, N - 1 - \delta_{j*}\}, \tag{15}$$

$$\overline{R} \equiv \frac{1}{N - L_{j*} + 1} \sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} R_t, \tag{16}$$

and

$$\bar{t} \equiv \frac{1}{N - L_{j*} + 1} \sum_{t=L_{j*}-1-\delta_{j*}}^{N-1-\delta_{j*}} t = \frac{L_{j*} - 1 - \delta_{j*} + (N - 1 - \delta_{j*})}{2}. \tag{17}$$

The choice of the weights α_j (for $j = 1, \ldots, J$) in the linear combination of Equation 11 can influence the power of the test and is dictated by the research purpose. The basic version of the test corresponds to the situation where α_j is non-zero for a particular j only and zero for other j values. If we assume that the variance of large-scale changes[17] associated, for example, with $j = 4$, decreases, whereas the variance of short-scale changes, associated with $j = 1$, increases, we can use the following weights:

[17] See Section 3.1 for the relation of wavelet coefficients to changes occurring on various scales.

$\alpha_1 = 1$, $\alpha_4 = -1$ and $\alpha_j = 0$ for j equal neither to 1 nor 4. These weights contrast the dynamics on large and short scales. Further illustrations are provided in Section 5 and Section 6.

The test statistic of Equation 12 goes hand in hand with an exploratory part of the analysis where the variability of the smoothed series (or a linear combination of the smoothed series) of (the logarithm of) squared wavelet coefficients is generally perceived as evidence against non-stationarity (see also Section 3.5). The choice of the Spearman's rank correlation coefficient as the test statistic (see Equation 14) is useful in situations where the occurrence of a non-constant, close-to-monotone (not necessarily highly variable) median function or a non-constant, close-to-monotone linear combination of median functions is expected (see also Section 5 and Section 6 for further illustrations).

4.4 Bootstrap approximation of the p-value

A range of well-founded practices used in the field of statistical learning and non-parametric regression has been employed to smooth the series of the logarithm of squared wavelet coefficients. Due to the complexity of the smoothing approach, its properties are not analytically tractable. And neither is the distribution of the test statistic under the null hypothesis. Analytical intractability is common in complex, albeit standard procedures (see e.g. Faraway, 1992 for an illustration in regression). Efron, Tibshirani (1994, Preface and Ch. 1) and Davison, Hinkley (2009, Ch. 1) suggest, however, that bootstrap can be used to tackle analytically intractable problems. They reveal that bootstrap opens the door for the assessment of complex, even though useful practices by utilizing computer power instead of traditional statistical theory. This aspect of bootstrap can be considered as its strength, not weakness. They also stress that bootstrap avoids often potentially "harmful" and unnecessary oversimplification of the problem that would make it analytically tractable. As a result, some heuristics is commonly seen in the application of bootstrap approaches and bootstrap hypothesis testing in general, such as testing for the co-movement of two time series in time and scale (Grinsted et al., 2004, Ch. 3.4), or in other cases presented, e.g., in Efron, Tibshirani (1994, Ch. 16) or Davison, Hinkley (2009, Ch. 4).

The stochastic process, which represents our test input, is assumed to contain neither deterministic components nor seasonality. Since the test lacks power against unit root non-stationarity and is not intended to be used as a unit root test, differencing can be applied to remove potential unit roots prior to the test. Thus, procedures such as differencing, detrending and removing seasonality are assumed to have already been applied before the analysis if necessary. Further, let $\{X_t: t = 0, \ldots, N-1\}$ be the result of these potential procedures and let a stationary Gaussian ARMA model be assumed to fit $\{X_t\}$ reasonably well. Model-based resampling (see, e.g., Davison, Hinkley, 2009, Ch. 8.2.2) which employs this stationary ARMA model is used to generate B bootstrap time series and B bootstrap replications of the test statistic, s_b^* for $b = 1, \ldots, B$. The bootstrap approximation of the hypothesis test p-value is obtained as (see, e.g., Davison, Hinkley, 2009, Ch. 4.2.3)

$$\frac{1 + \sum_{b=1}^{B} I(s_b^* \geq s)}{B+1}, \tag{18}$$

where $I(s_b^* \geq s)$ is equal to one if $s_b^* \geq s$, and equal to zero otherwise.

5 SIZE AND POWER OF THE TEST

In this section, the size and power of the proposed test will be examined.

5.1 Size of the test

The following four stationary Gaussian processes are assumed, namely

1. an AR(1) process[18] with the autoregressive parameter equal to 0.9,
2. an AR(1) process with the autoregressive parameter equal to –0.9,
3. an MA(1) process[19] with the parameter equal to 0.8,
4. an MA(1) process with the parameter equal to –0.8.

Two patterns of α_j (for $j = 1, \ldots, J$) of Equation 11 are examined (denoted as A and B). Namely:
A: $\alpha_1 = 1$ and $\alpha_j = 0$ for $j = 2, \ldots, J$,
B: $\alpha_1 = -1$, $\alpha_3 = 1$ and $\alpha_j = 0$ for $j = 2, 4, \ldots, J$.

Since economic time series are often rather short, their length usually being of order of tens, the following two choices of N are assumed:[20] $N = 64$ and $N = 32$. Further, the two possible test statistics are assumed, namely the standard deviation (see Equation 12) and the Spearman's rank correlation coefficient (see Equation 14). 1 000 realizations are generated for each combination of the process type, α_j pattern, N and test statistic.

The hypothesis test is performed for each of the realizations and a decision made about the rejection or non-rejection of the null hypothesis – significance levels 0.01, 0.05 and 0.1 being used. Haar filters are employed. A stationary ARMA model which fits the realization is found using the *auto.arima()* function from R *forecast* package (Hyndman, 2015) and employing the following function arguments: max.p = 1,

Table 1 Size of the test estimated for various process types, series lengths (N), test statistics and α_j patterns. Each inner cell provides an estimate of the size of the test for significance levels 0.01, 0.05 and 0.1. The results are rounded to two decimal places

Process type, N = length of the series	Standard deviation, pattern A	Standard deviation, pattern B	Spearman, pattern A	Spearman, pattern B	Nason (2013b), Bonferroni	Nason (2013b), FDR
1, 32	0.01 0.04 0.09	0.01 0.04 0.11	0.01 0.05 0.09	0.01 0.05 0.10	0.00 0.00 0.00	0.00 0.00 0.00
2, 32	0.01 0.06 0.12	0.01 0.05 0.11	0.01 0.04 0.08	0.01 0.05 0.09	0.00 0.00 0.00	0.00 0.00 0.00
3, 32	0.00 0.04 0.09	0.01 0.05 0.09	0.02 0.05 0.11	0.00 0.04 0.10	0.00 0.00 0.00	0.00 0.00 0.00
4, 32	0.01 0.04 0.09	0.00 0.03 0.09	0.01 0.04 0.10	0.01 0.05 0.10	0.00 0.00 0.00	0.00 0.00 0.00
1, 64	0.01 0.04 0.07	0.01 0.04 0.09	0.01 0.06 0.10	0.01 0.05 0.10	0.00 0.00 0.00	0.00 0.00 0.00
2, 64	0.00 0.04 0.10	0.01 0.05 0.12	0.01 0.03 0.08	0.01 0.03 0.08	0.00 0.00 0.02	0.00 0.00 0.02
3, 64	0.01 0.04 0.09	0.00 0.04 0.10	0.01 0.05 0.10	0.01 0.04 0.09	0.00 0.00 0.00	0.00 0.00 0.00
4, 64	0.00 0.03 0.07	0.01 0.05 0.10	0.01 0.04 0.10	0.01 0.04 0.09	0.00 0.00 0.00	0.00 0.00 0.00

Source: Own construction

[18] $X_t = \phi X_{t-1} + a_t$, where $\{a_t\}$ is a unit-variance Gaussian white noise sequence, ϕ being a parameter.
[19] $X_t = a_t + \theta a_{t-1}$, where $\{a_t\}$ is a unit-variance Gaussian white noise sequence, θ being a parameter.
[20] Powers of two are assumed to allow a comparison with the test by Nason (2013b) (see below in this section for further details).

max.q = 1, max.d = 0, seasonal = FALSE, allowdrift = FALSE. The number of bootstrap replications is[21] $B = 99$. The results are presented in Table 1.

A comparison of the results with those obtained by the test proposed in Nason (2013b) is provided.[22] Nason (2013b) test utilizes, among others, multiple hypothesis testing, suggesting two approaches, namely the Bonferroni method and the false discovery rate (FDR) approach of Benjamini, Hochberg (1995).

The size of our test seems to be reasonably close to the nominal level (0.01, 0.05 or 0.1), no large deviation between the estimated size and the nominal level occurring in any of the simulation combinations. On the other hand, the test proposed in Nason (2013b) seems to be extremely conservative for the settings used in our simulation, the estimated size being far below the nominal level for all the simulation combinations.

5.2 Power of the test

The power of the test will be assessed under the conditions when the true data-generating process is either an AR(1) or MA(1) process with a time-varying coefficient. The values of the coefficient are stored in the sequence $\{\phi_t: t = 0, \ldots, N-1\}$ and are given as

$$\phi_t = 0.95 \cos(2\pi F t / N), \quad t = 0, \ldots, N-1, \tag{19}$$

where $F > 0$ is the frequency of the cosine. Two possible values can be attained in the simulation: $F = 0.5$ and $F = 1$. The process is thus defined either as (AR(1))

$$X_0 = \frac{a_0}{\sqrt{1-\phi_0^2}} \quad \text{and} \quad X_t = \phi_t X_{t-1} + a_t,\ t = 1, \ldots, N-1, \tag{20}$$

or as (MA(1))

$$X_0 = a_0 \sqrt{1+\phi_0^2} \quad \text{and} \quad X_t = a_t + \phi_t a_{t-1},\ t = 1, \ldots, N-1, \tag{21}$$

where $\{a_t: t = 0, \ldots, N-1\}$ is unit-variance Gaussian white noise. The transition from high values (+0.95) to low values (–0.95) of the AR(1) (or MA(1)) coefficient is associated with a decrease in the variance of large-scale changes and an increase in the variance of short-scale changes. Both the models of non-stationary time series (AR(1) and MA(1)) are statistical ones, resembling also some of the non-stationary models used in Nason (2013b).

The other simulation settings are the same as in Section 5.1. The estimates of the power of the test are presented in Table 2. Again, a comparison with the test of Nason (2013b) is made.

As expected, higher values of N (ceteris paribus) mostly lead to a higher power of the test. It is also not surprising to often (but not always) find a higher power for the AR process than for the MA one (ceteris paribus). This is due to the fact that the time-varying coefficient is generally accompanied by more pronounced variations in the variance of changes associated with the examined scales in the case of the AR process.

[21] General considerations of Davison, Hinkley (2009, Ch. 4.2.5) on the choice of B in bootstrap hypothesis testing suggest that too small values of B can lead to a loss of the size and power of the test. Davison, Hinkley (2009, Ch. 4.2.5) conclude that 99 bootstrap replications should generally be sufficient provided the significance level is greater or equal to 0.05. A little larger loss of size and power, though not a serious one, can occur if 99 bootstrap replications are used with the 0.01 significance level.

[22] The test proposed in Nason (2013b) is implemented in *hwtos2()* function in R *locits* package (Nason, 2013a). Default settings of the function parameters are used. This function works only with a time series whose length is a power of two.

Table 2 Power of the test estimated for various process types, values of *F*, series lengths (*N*), test statistics and α_j patterns. Each inner cell provides an estimate of the power of the test for significance levels 0.01, 0.05 and 0.1. The results are rounded to two decimal places

Process type, *F*, *N* = length of the series	Standard deviation, pattern A	Standard deviation, pattern B	Spearman, pattern A	Spearman, pattern B	Nason (2013b), Bonferroni	Nason (2013b), FDR
AR, 0.5, 32	0.05 0.22 0.35	0.17 0.44 0.57	0.06 0.18 0.27	0.14 0.35 0.49	0.00 0.00 0.00	0.00 0.00 0.00
MA, 0.5, 32	0.01 0.08 0.17	0.02 0.13 0.24	0.03 0.10 0.17	0.06 0.19 0.30	0.00 0.00 0.00	0.00 0.00 0.00
AR, 1, 32	0.08 0.26 0.41	0.10 0.34 0.47	0.00 0.01 0.02	0.00 0.01 0.02	0.00 0.00 0.00	0.00 0.00 0.00
MA, 1, 32	0.01 0.09 0.18	0.01 0.08 0.16	0.00 0.02 0.06	0.01 0.04 0.07	0.00 0.00 0.00	0.00 0.00 0.00
AR, 0.5, 64	0.13 0.39 0.53	0.35 0.62 0.74	0.12 0.27 0.38	0.25 0.49 0.64	0.00 0.01 0.05	0.00 0.01 0.05
MA, 0.5, 64	0.03 0.12 0.21	0.10 0.29 0.41	0.06 0.14 0.24	0.13 0.30 0.44	0.00 0.00 0.00	0.00 0.00 0.00
AR, 1, 64	0.17 0.45 0.58	0.36 0.61 0.70	0.00 0.00 0.01	0.00 0.00 0.01	0.00 0.04 0.07	0.00 0.04 0.08
MA, 1, 64	0.02 0.09 0.19	0.05 0.18 0.29	0.00 0.02 0.04	0.00 0.01 0.03	0.00 0.00 0.00	0.00 0.00 0.00

Source: Own construction

A change in the pattern of α_j values from A to B, ceteris paribus, has generally resulted in an increase in the power of the test in our simulations. This is due to the fact that if the variance of short-scale changes increases, the variance of large-scale changes decreases in our case, and vice versa. Consequently, a more powerful test can be obtained using pattern B which contrasts the dynamics on short and large scales.

The test employing the Spearman's rank correlation coefficient as the test statistic is much more powerful for $F = 0.5$ than $F = 1$ (ceteris paribus). The reason for this behavior is obvious since $\{q_t^{\alpha}\}$ (or $\{q_{j,t}\}$) is expected to be close to monotone for $F = 0.5$, while no such monotone behavior can be expected for $F = 1$ because, for $F = 1$, the time-varying coefficient starts to revert back to its original value at a time half-way from the start and so does the variance of changes associated with individual scales. It is also interesting to note that the test using Spearman's rank correlation is more powerful than the one using the standard deviation provided $F = 0.5$, the process type being MA. This can presumably be explained by the fact that $\{q_t^{\alpha}\}$ (or $\{q_{j,t}\}$) tends to be quite close to monotone in such a situation ($F = 0.5$, process type = MA) and not too variable (in terms of the standard deviation).

For the settings used in our simulation, the power of the test proposed by Nason (2013b) is often inferior to ours. Nason (2013b) also experimented with various non-stationary models and reports very good power characteristics of his test. This can be explained by the fact that much longer time series were used in his simulation, namely the length of 512 was assumed. To further support this claim, we have run additional minor Monte Carlo simulations. More specifically, having assumed an AR(1) model with a time-varying autoregressive coefficient and $F = 0.5$ (see Equations 19 and 20), we studied the power of the test proposed by Nason (2013b) in dependence upon the length of the series (N), using the 0.05 significance level. The following power estimates have been obtained from 1 000 simulations (the first number corresponding to the Bonferroni method, the second to the FDR approach):

0.00 and 0.00 for $N = 32$, 0.01 and 0.01 ($N = 64$), 0.36 and 0.38 ($N = 128$), 0.94 and 0.97 ($N = 256$), 1.00 and 1.00 ($N = 512$). These additional results suggest that a reasonable power of the test proposed by Nason (2013b) can be obtained provided that the time series is long enough. We have not included such long time series in our major simulation since our test aims – by its design, where cross-validation, robust filtering and bootstrapping is utilized – at rather short time series, not being directly applicable to very long time series due to extensive computations that would be required. For short time series, such as those often occurring in economic settings, our test is, however, expected to enjoy reasonably good size and power characteristics, having reasonable computational demands.

6 APPLICATION OF THE TEST TO THE U.S. GROSS DOMESTIC PRODUCT

We illustrate the hypothesis test using the yearly time series of the U.S. gross domestic product (GDP) retrieved from the U.S. Bureau of Economic Analysis, FRED (2015). The time series is given in current prices in billions of dollars, measuring the value of goods and services in each year's prices. GDP measured in constant prices (i.e. adjusted for inflation) is preferable when the focus is on actual productivity growth, GDP in current prices being of potential interest, for instance, for monetary policy objectives (see, e.g., Feldstein, Stock, 1994; Bernanke, Mishkin, 1997). The decision to use current prices instead of constant ones in this paper is due to the fact that the former facilitate a better demonstration of the various aspects and settings of the hypothesis test. Moreover, the yearly nature of the time series also avoids the need to deal with the seasonal pattern.

The time series will be denoted as $\{Y_t: t = -1, 0, \ldots, N-1\}$, where $N = 85$, time $t = -1$ corresponding to the year 1929 and $t = 84$ to 2014, the total length of the time series being $N + 1 = 86$. The hypothesis test will be performed on $\{X_t: t = 0, \ldots, N-1\}$ which is defined as the first difference of the natural logarithm of $\{Y_t\}$, i.e.

$$X_t \equiv \log(Y_t) - \log(Y_{t-1}), \quad t = 0,\ldots, N-1. \tag{22}$$

Haar filters are employed. A stationary ARMA model which fits $\{X_t\}$ is found using the *auto.arima()* function from R *forecast* package (Hyndman, 2015) – the function is called with the following arguments: max.p = 4, max.q = 4, max.d = 0, seasonal = FALSE, allowdrift = FALSE. The maximum possible orders of the AR and MA part of the model are chosen to be equal to four (i.e. max.p = 4, max.q = 4) so that the model can be flexible enough if needed. The number of bootstrap replications is $B = 499$.

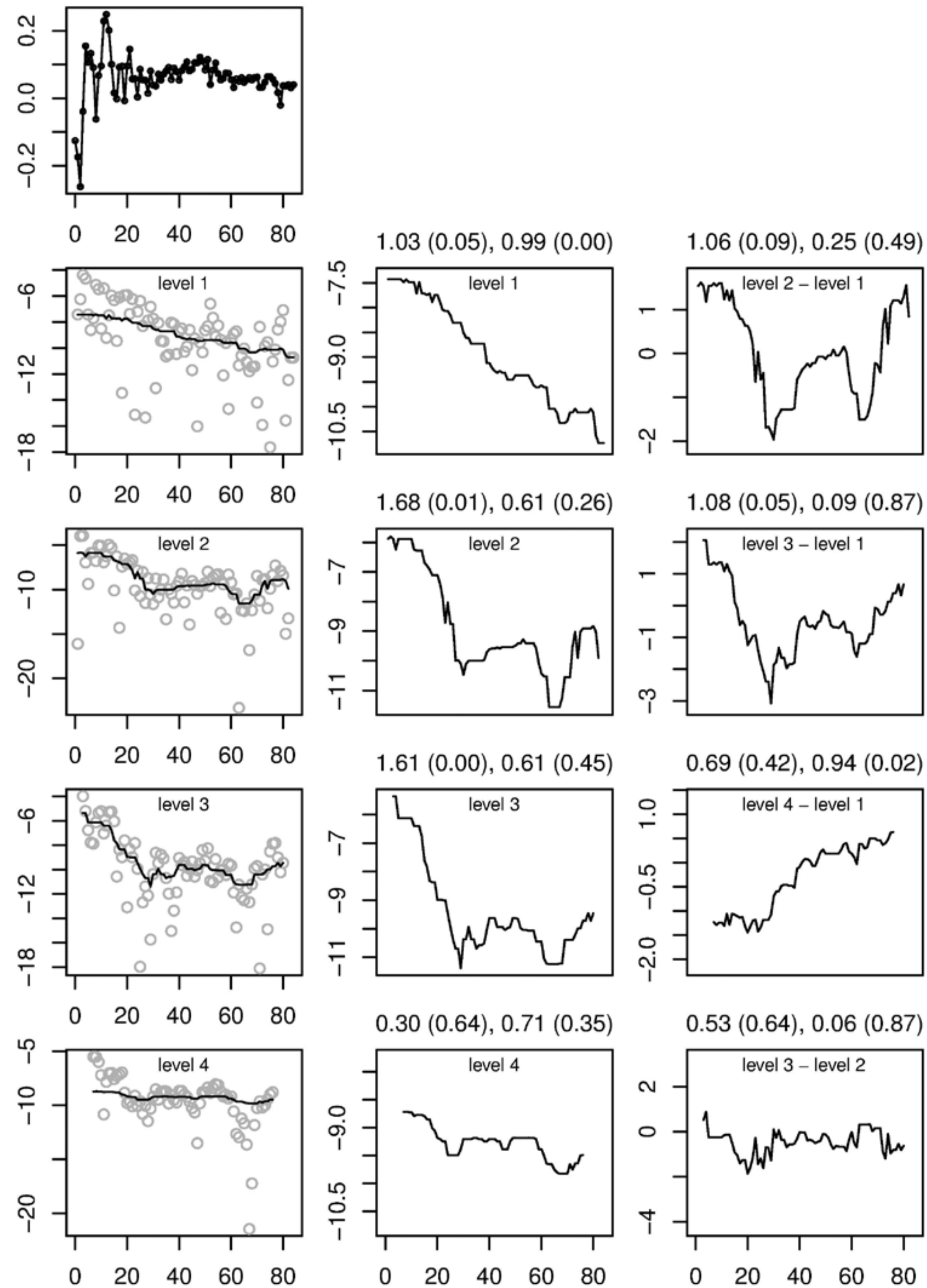
The results are presented in Figure 1. More specifically, the uppermost subfigure in the first column displays the time series $\{X_t\}$. Below this column subfigure, $\{z_{j,t}\}$ ($j = 1, \ldots, 4$) are presented (gray color) in separate subfigures, together with $\{q_{j,t}\}$ ($j = 1, \ldots, 4$) (black color). The asymmetric nature of $z_{j,t}$ with a heavy left tail can be clearly observed.

The subfigures in the second column display $\{q_{j,t}\}$ ($j = 1, \ldots, 4$) again. The y-axis range in these subfigures is set in such a way that the ratio of this range to the range of $\{q_{j,t}\}$ is an increasing function of p-value. This leads to the visual perception of more variable $\{q_{j,t}\}$ in the case of a more significant outcome – a hypothesis test employing the standard deviation of $\{q_{j,t}\}$ as the test statistic is used. The value of the test statistic is presented above each subfigure, the p-value following in parentheses. A test employing the Spearman's rank correlation coefficient between $\{q_{j,t}\}$ and time as the test statistic is also performed. The test statistic is given after the comma above each of the subfigures, the p-value being shown in parentheses.

The subfigures in the third column display $\{q_t^{\alpha}\}$ (see Equation 11) with four patterns of α_j values, namely:

- $\alpha_2 = 1$, $\alpha_1 = -1$ and $\alpha_j = 0$ for j equal neither 2 nor 1,
- $\alpha_3 = 1$, $\alpha_1 = -1$ and $\alpha_j = 0$ for j equal neither 3 nor 1,
- $\alpha_4 = 1$, $\alpha_1 = -1$ and $\alpha_j = 0$ for j equal neither 4 nor 1,
- $\alpha_3 = 1$, $\alpha_2 = -1$ and $\alpha_j = 0$ for j equal neither 3 nor 2.

Figure 1 Illustration of the hypothesis test for U.S. gross domestic product data. See the text for a detailed description



Source: Own construction using data retrieved from U.S. Bureau of Economic Analysis, FRED (2015)

Hypothesis tests employing the standard deviation of $\{q_t^{\alpha}\}$ and the Spearman's rank correlation coefficient between $\{q_t^{\alpha}\}$ and time are performed for each of the four patterns. Similarly to the second column, the test statistics and p-values are presented above the subfigures in the third column.

It follows from all the subfigures of Figure 1 that the rejection or non-rejection of the null hypothesis of stationarity is influenced by the choice of α_j values in $\{q_t^{\alpha}\}$ (or the level j if $\{q_{j,t}\}$ is used) and the chosen test statistic. All these choices have to be made before the analysis. A discussion of the results corresponding to particular choices is presented in the paragraphs to follow.

For example, $\{q_{1,t}\}$ is almost monotonically decreasing over time, which implies that the variance of changes associated with the shortest scale is almost monotonically decreasing over time too. If we decided a priori to use the Spearman's rank correlation coefficient between $\{q_{1,t}\}$ and time as the test statistic, we would "definitely" reject the null hypothesis of stationarity at the 5% significance level (see the subfigure called "level 1" in the second column). On the other hand, if the standard deviation of $\{q_{1,t}\}$ was used as the test statistic, the p-value of the hypothesis test would be close to 0.05 and the decision about stationarity or non-stationarity at the 5% significance level would not be so clear.

There seems to be a decrease in the variance of changes, not only on scales associated with the first level but also on those associated with the second and third level. In contrast to the first level, the decrease of the variance associated with the second and third level is far from being monotone. Consequently, the tests using Spearman's rank correlation between $\{q_{2,t}\}$ and time and $\{q_{3,t}\}$ and time, respectively, are not significant at 5% levels, whereas those using the standard deviation of $\{q_{2,t}\}$ and $\{q_{3,t}\}$, respectively, are significant (see the subfigures called "level 2" and "level 3" in the second column).

Despite the significance of the two lastly mentioned tests, the hypothesis test associated with the subfigure called "level 3 – level 2" in the third column is non-significant. This is due to the fact that the time-varying patterns in $\{q_{2,t}\}$ and $\{q_{3,t}\}$ are rather "synchronous". On the other hand, the test employing Spearman's rank correlation in the subfigure called "level 4 – level 1" is significant. This suggests that non-stationarity manifests itself in different ways on short (j = 1) and large (j = 4) scales.

The tests provide a decision about the rejection or non-rejection of the null hypothesis. Moreover, visual inspection of the plots similar to those of Figure 1 may supply additional information about the character of non-stationarity and the size and importance of effects.

CONCLUSION

We have introduced a new wavelet-based hypothesis test for second-order stationarity which is based on exploring the variability of the smoothed series of squared MODWT wavelet coefficients. Having noted that there are alternative techniques for smoothing available in the wavelet literature, we decided to use robust filtering and modified cross-validation. Even though cross-validation is widely used and generally recognized as a flexible tool in the statistical learning literature, it has not been employed much in the resources on wavelets, not being applied at all in formal wavelet-based tests for stationarity.

Further, we have proposed several test statistics that explicitly answer important questions on whether the variability (or the close-to-monotone behavior) observed in the smoothed series represents a significant effect, or whether the characteristic of non-stationarity is scale-specific.

We have used bootstrap to approximate the distribution of the test statistic under the null hypothesis. In agreement with the literature on bootstrap, we have preferred flexible, well-established smoothing techniques and appealing test statistic to the analytical tractability of the procedure. Although the test is computationally expensive, it enjoys reasonable size and power properties for lengths of time series typical in economics. We consider the properties of the test for these lengths superior to those of the test proposed by Nason (2013b).

Our test was also used to assess the stationarity of the time series of the first difference of the logarithm of the U.S. gross domestic product. The results suggest that the variance of changes associated with various scales alters over time, making the time series non-stationary. In particular, the variance of changes associated with the shortest scales exhibits a significant close-to-monotone variation over time. This pattern is not present on larger scales.

ACKNOWLEDGEMENTS

This paper was written with the support of the Czech Science Foundation (GA CR), grant no. P402/12/P507, Modelling of Financial and Economic Time Series – Application and Comparison of Wavelet and Traditional Methods. We acknowledge the linguistic assistance provided by Karel Helman, Sr. (Private University College of Economic Studies Ltd., Prague). We would also like to thank the anonymous reviewers whose comments helped to improve the paper.

References

ARLOT, S., CELISSE, A. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 2010, 4, pp. 40–79.

BENJAMINI, Y., HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B*, 1995, 57, pp. 289–300.

BERNANKE, B., MISHKIN, F. *Inflation Targeting: A New Framework for Monetary Policy? NBER Working paper No. 5893.* National Bureau of Economic Research, 1997.

BROCKWELL, P., DAVIS, R. *Introduction to Time Series and Forecasting.* 2nd edition, Springer, 2002.

CHU, C.-K., MARRON, J. S. Comparison of Two Bandwidth Selectors with Dependent Errors. *The Annals of Statistics*, 1991, 19, pp. 1906–1918.

CONSTANTINE, W., PERCIVAL, D. *wmtsa: Wavelet Methods for Time Series Analysis. R package version 2.0-0*, 2013.

DAVISON, A. C., HINKLEY, D. V. *Bootstrap Methods and Their Applications.* 11th printing, Cambridge University Press, 2009.

DICKEY, D. A., FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 1979, 74, pp. 427–431.

EFRON, B., TIBSHIRANI, R. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1994.

FARAWAY, J. On the Cost of Data Analysis. *Journal of Computational and Graphical Statistics*, 1992, 1, pp. 213–229.

FELDSTEIN, M., STOCK, J. The use of a monetary aggregate to target nominal GDP. In *Monetary policy*, The University of Chicago Press, 1994, pp. 7–69.

FRIED, R., EINBECK, J., GATHER, U. Weighted Repeated Median Smoothing And Filtering. *Journal of the American Statistical Association*, 2007, 102, pp. 1300–1308.

FRYZLEWICZ, P. Modelling and Forecasting Financial Log-returns as Locally Stationary Wavelet Processes. *Journal of Applied Statistics*, 2005, 32, pp. 503–528.

GRINSTED, A., MOORE, J. C, JEVREJEVA, S. Application of the Cross-Wavelet Transform and Wavelet Coherence to Geophysical Time Series. *Nonlinear Processes in Geophysics*. 2004, 11, pp. 561–566.

HÄRDLE, W., GASSER, T. Robust Non-parametric Function Fitting. *Journal of the Royal Statistical Society. Series B*, 1984, 46, pp. 42–51.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, 2011.

HYNDMAN, R. J. *forecast: Forecasting Functions For Time Series and Linear Models. R package version 5.8.*, 2015.

JENSEN, M., WHITCHER, B. Measuring the Impact Intradaily Events Have On The Persistent Nature of Volatility. In GALLEGATI, M., SEMMLER, W., eds. *Wavelet Applications in Economics and Finance*. Springer, 2014.

MORELL, O., OTTO, D., FRIED, R. On Robust Cross-validation for Non-parametric Smoothing. *Computational Statistics*, 2013, 28, pp. 1617–1637.

NASON, G. *locits: Tests of Stationarity and Localized Autocovariance. R package version 1.4*, 2013a.

NASON, G. A Test For Second-Order Stationarity and Approximate Confidence Intervals for Localized Autocovariances for Locally Stationary Time Series. *Journal of the Royal Statistical Society, Series B*, 2013b, 75, pp. 879–904.

NASON, G., VON SACHS, R., KROISANDT, G. Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum. *Journal of the Royal Statistical Society, Series B*, 2000, 62, pp. 271–292.

PERCIVAL, D. B., WALDEN, A. T. *Wavelet Methods for Time Series Analysis (reprint)*. Cambridge University Press, 2002.

PHILLIPS, P. C. B., PERRON, P. Testing for a unit root in time series regression. *Biometrika*, 1988, 75, pp. 335–346.

R CORE TEAM R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

RAWLINGS, J., PANTULA, S., DICKEY, D. *Applied Regression Analysis: A Research Tool.* 2nd edition, Springer, 2001.

SAID, S. E., DICKEY, D. A. Hypothesis testing in ARIMA(p,1,q) models. *Journal of the American Statistical Association*, 1985, 80, pp. 369–374.

SAID, S. E., DICKEY, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 1984, 71, pp. 599–607.

TORRENCE, C., COMPO, G. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 1998, 79, pp. 61–78.

U.S. Bureau of Economic Analysis, *Gross Domestic Product* (GDPA) [online]. Retrieved from FRED, Federal Reserve Bank of St. Louis. [10.8.2015]. <https://research.stlouisfed.org/fred2/series/GDPA>.

VON SACHS, R., NEUMANN, M. H. A Wavelet-Based Test for Stationarity. *Journal of Time Series Analysis*, 2000, 21, pp. 597–613.

WHITCHER, B., GUTTORP, P., PERCIVAL, D. Wavelet Analysis of Covariance with Application to Atmospheric Time Series. *Journal of Geophysical Research: Atmospheres*, 2000, 105, pp. 14941–14962.

WICKERHAUSER, M. *Adapted Wavelet Analysis from Theory to Software,* Wellesley, MA: A. K. Peters, 1994.