

## TIME SERIES FORECASTING WITH A PRIOR WAVELET-BASED DENOISING STEP<sup>1</sup>

Milan Bašta\*

### Abstract

We provide an extensive study assessing whether a prior wavelet-based denoising step enhances the forecast accuracy of standard forecasting models. Many combinations of attribute values of the thresholding (denoising) algorithm are explored together with several traditional forecasting models used in economic time series forecasting. The results are evaluated using M3 competition yearly time series. We conclude that the performance of a forecasting model combined with the prior denoising step is generally not recommended, which implies that a straightforward generalisation of some of the results available in the literature (which found the denoising step to be beneficial) is not possible. Even if cross-validation is used to select the value of the threshold, a superior performance of the forecasting model with the prior denoising step does not generally follow.

**Keywords:** wavelets, noise, evaluating forecasts, automatic forecasting

**JEL Classification:** C15, C22, C49, C53

### Introduction

Many business, economic, and financial time series can be considered to be buried in noise. Such a view is in agreement with several traditional approaches to time series modelling which explicitly assume that a time series can be decomposed into a signal (e.g. a deterministic trend and/or deterministic seasonal component) plus noise (an error term). Since the noise is generally considered as unwanted, removing it from a noisy time series (a procedure referred to in this paper as denoising) prior to the estimation of a time series model and forecast construction could potentially enhance the accuracy of the forecast.

The potential benefit of such a denoising step for forecasting purposes can be reported by noting that simple exponential smoothing (which is known to perform well in forecasting economic and business time series) can be thought of as an application of the random walk method to a time series in which noise has been removed by applying a causal exponentially weighted moving average.

<sup>1</sup> 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.

\* University of Economics, Prague, Faculty of Informatics and Statistics (milan.basta@vse.cz mail).

In general, various wavelet-based techniques are considered excellent for time series denoising. The question thus arises whether the prior wavelet-based denoising step can be useful – in terms of improving the forecast accuracy – as a prior step before a time series model is estimated and a forecast constructed. Unfortunately, papers available in the literature (see Section 1 for details) report contradictory results in this aspect. Moreover, these papers are often based only on a few time series and forecasting models, which prevent wider generalisation of the obtained results. Consequently, we conduct an analysis, which should provide generalisable conclusions about the usefulness of the prior wavelet-based denoising step for forecasting. Specifically, (a) we explore several combinations of the attributes of the wavelet-based denoising algorithm, as well as (b) many forecasting models/methods and (c), assess the performance using many economic, demographic, industrial, and financial time series. To our knowledge, no such extensive study in the three aspects (a), (b) and (c) has been conducted previously, Olmeda and Fernández (2000) come the closest to our design although they use different types of forecasting models and assess the results using a different type of data.

More specifically, we employ a subset of M3 competition time series in our analysis. The M3 competition time series is a set of 3003 yearly, quarterly, and monthly time series or those of other frequencies. The time series can be categorised into six fields: micro, industrial, macro, financial, demographic, and other time series. The M3 competition time series have been used in the literature as benchmark time series to assess the forecast accuracy of various forecasting models and methods (Makridakis and Hibon, 2000). Consequently, we have decided to also use these time series in our analysis. The time series are available in the Mcomp R package (Hyndman, 2013b). Two sets of data are available – the historical and the future data – for each time series, which enables us to easily assess the forecast accuracy.

Concerning the forecasting models, the following forecasting models/methods<sup>2</sup> will be used in our paper:<sup>3</sup> Random walk / naïve method (RW), drift method (Drift), simple exponential smoothing (SES, `ses()`), Holt's linear trend method (Holt, `holt()`), AR(1) model with constant (AR1, `ar.ols()` from the stats package), AR(1) model with constant for first differences (AR1diff, `ar.ols()` from the stats package), theta method (Theta, `thetaf()`), automatically selected innovation state space model (`s_ETS`, `ets()`), neural network autoregression with a single hidden layer, one hidden neuron and one lagged input (NnetAR, `nnetar()`) and the automatically selected ARIMA model (`s_ARIMA`, `auto.arima()`).

Furthermore, for those time series and forecasting models for which wavelet-based denoising is detrimental (in terms of the forecast accuracy), none or only a small amount of noise should be removed prior to forecasting, whereas a larger amount of noise should be removed in other instances. Consequently, besides the traditional wavelet-based denoising techniques, we also explore whether wavelet-based denoising combined with time series cross-validation (which is directly related to the assessment of the forecast accuracy) leads

2 A detailed description of the models/methods can be found in Hyndman and Athanasopoulos (2013).

3 The abbreviation of the model's name is given in the parentheses followed by an R function used to estimate the model (if such a function was employed) in our analysis. The functions are from the R forecast package (Hyndman and Khandakar, 2008), unless stated otherwise. The default settings of the functions were used, unless stated otherwise.

to an optimal decision on the amount of noise to be removed. Such an approach is similar to the approach of Lotrič (2004) where noise is removed adaptively.

Finally, since a prior wavelet-based denoising step combined with the random walk method is similar to simple exponential smoothing, the difference being in the way noise is removed (see also Ferbar et al., 2009), we examine the performance of the prior wavelet-based denoising step (namely the one that uses time series cross-validation to decide on the amount of noise to be removed) combined with the random walk method in more detail in our paper.

The paper is organised as follows. Section 1 reviews the literature on the use of wavelet-based techniques for denoising time series prior to forecasting the time series. In Section 2 an introduction to the maximal overlap discrete wavelet transform (MODWT) is given. Denoising techniques based on the MODWT are introduced in Section 3. The evaluation of the performance of the forecasting models combined with the prior denoising step is outlined in Section 4. The results are presented in Section 5.

## 1. Literature Review

There are several papers in the literature which examine the benefit of a prior wavelet-based denoising step for forecasting purposes. For example, Lotrič (2004) uses a wavelet-based unit, which is integrated into a multilayered perceptron. Denoising is not used as a data preprocessing step but as an integral part of the perceptron. The performance of the proposed model is tested using various time series (the Feigenbaum sequence, a second-order process, a quality control time series) and the model is claimed to outperform the traditional multilayered perceptron and its combination with a statistical denoising criterion. Boto-Giralda et al. (2010) use wavelet-based denoising as a part of their algorithm for traffic volume forecasting. The algorithm provides accurate forecasts compared to other benchmark forecasting models.

Bruzda (2014) assumes that the underlying signal is stochastic, and uses scaling of wavelet coefficients for denoising. The performance of the denoising step followed by a forecasting algorithm is studied using simulated time series, namely stochastic stationary and non-stationary processes buried in additive Gaussian white noise, as well as 16 M3 competition time series. The results suggest that wavelet-based denoising followed by an AR(I)MA forecasting procedure can reduce the forecast error in the case of short and medium-length time series. Ferbar et al. (2009) argue that wavelet-based denoising can be used as an alternative to exponential smoothing and may bring benefits to forecasting.

Alrumaih and Al-Fawzan (2002) explore whether wavelet-based denoising used as a preprocessing step can be beneficial for forecasting the Saudi Stock Index. They use the data from the period between January 26, 1994, and November 4, 1999. The authors suggest that including rather than omitting the thresholding step can reduce forecast error. Olmeda and Fernández (2000) consider several settings of the parameters of the wavelet-based denoising algorithm and explore whether the prior denoising step can improve the forecast of daily closing prices of some stock indices. Various autoregressive models are employed to accomplish the forecast. The results indicate that wavelet-based denoising does not lead to more accurate forecasts and, in addition, can often be harmful. Schlüter and Deuschle (2010) assess the contribution of wavelet-based denoising (as one

of the wavelet-based approaches to forecasting) to forecasting West Texas Intermediate oil prices, Deutsche Bank stock prices, the Euro/Dollar exchange rate or the UK day-ahead power prices. Three traditional forecasting models are used to forecast the denoised time series (ARMA, ARIMA and structural time series modelling). The results suggest that denoising can be useful when combined with ARMA and ARIMA modelling.

The above-mentioned literature review suggests that the results on the usefulness of the prior wavelet-based denoising step for forecasting are rather inconclusive and the studies are often conducted on a small number of time series. Consequently, our paper should contribute to the debate about the usefulness of the prior denoising step and should provide results that are generalisable, at least within a certain class of economic, demographic and financial time series.

## 2. Maximal Overlap Discrete Wavelet Transform

We provide an introduction to the maximal overlap discrete wavelet transform (MODWT) filters (Section 2.1) so that we can introduce the MODWT (Section 2.2) and the denoising techniques based on the MODWT (Section 3). Since the theory behind the MODWT is rather complex, interested readers are referred to e.g. Percival and Walden (2006) for further details. The crucial point in our context is that MODWT transforms the input process into a more advantageous form that makes further preprocessing (denoising) possible.

### 2.1 MODWT Filters

The  $j$ th level MODWT wavelet and scaling filters (for  $j = 1, 2, 3, \dots$ ), denoted as  $\{h_{j,l}: l = 0, \dots, L_j - 1\}$  and  $\{g_{j,l}: l = 0, \dots, L_j - 1\}$ , are *causal* linear filters with lengths given as  $L_j = (2^j - 1)(L_1 - 1) + 1$ , where  $L_1$  is the length of the first level filters.<sup>4</sup>  $\{h_{j,l}\}$  (for  $j = 1, 2, 3, \dots$ ) is approximately an ideal band-pass filter with a nominal pass-band  $[2^{-(j+1)}, 2^{-j}]$ ,  $\{g_{j,l}\}$  (for  $j = 1, 2, 3, \dots$ ) being approximately an ideal low-pass filter with a nominal pass-band  $[0, 2^{-(j+1)}]$ . For example, the first level MODWT wavelet filter  $\{h_{1,l}: l = 0, \dots, L_1 - 1\}$  is approximately an ideal high-pass filter which passes frequencies in the range  $[1/4, 1/2]$ . The first level MODWT scaling filter  $\{g_{1,l}: l = 0, \dots, L_1 - 1\}$  is approximately an ideal low-pass filter which passes frequencies in the range  $[0, 1/4]$ .

Interested readers can be referred to e.g. Percival and Walden (2006) for the actual construction of the filters. It is also important to stress that there is not only one set of filters  $\{h_{1,l}\}$ ,  $\{h_{2,l}\}$ , ...,  $\{g_{1,l}\}$ ,  $\{g_{2,l}\}$ , ... , but that there are several such sets called families. The well-known families are e.g. the Haar, D(4) or the LA(8) family.

4 By stating e.g. “filter  $\{h_{j,l}: l = 0, \dots, L_j - 1\}$ ”, we in fact mean “filter with coefficients”. Since it is custom in the given context to state only “filter” (see e.g. Percival and Walden, 2006), we will adhere to this terminology.

## 2.2 MODWT Coefficients and MODWT of Level

Having introduced the MODWT filters, we can introduce the MODWT coefficients. Specifically, let us assume a stochastic process  $\{X_t: t = \dots, -1, 0, 1, \dots\}$ . The  $j$ th level MODWT wavelet and scaling coefficients for  $\{X_t\}$  (for  $j = 1, 2, 3, \dots$ ), denoted as  $\{W_{j,t}: t = \dots, -1, 0, 1, \dots\}$  and  $\{V_{j,t}: t = \dots, -1, 0, 1, \dots\}$ , are obtained by linear filtering  $\{X_t\}$  with  $\{h_{j,l}\}$  and  $\{g_{j,l}\}$ . Namely,

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}, \quad t = \dots, -1, 0, 1, \dots, \quad (1)$$

$$V_{j,t} = \sum_{l=0}^{L_j-1} g_{j,l} X_{t-l}, \quad t = \dots, -1, 0, 1, \dots \quad (2)$$

Since  $\{h_{j,l}\}$  is an approximate band-pass filter with a nominal pass-band  $[2^{-(j+1)}, 2^{-j}]$ ,  $\{W_{j,t}\}$  is associated with the dynamics of  $\{X_t\}$  in the range of frequencies  $[2^{-(j+1)}, 2^{-j}]$ . Analogously, since  $\{g_{j,l}\}$  is an approximate low-pass filter passing frequencies in the range  $[0, 2^{-(j+1)}]$ ,  $\{V_{j,t}\}$  is associated with the dynamics of  $\{X_t\}$  in the range of frequencies  $[0, 2^{-(j+1)}]$ . For example,  $\{W_{1,t}\}$  is associated with high frequencies, namely the frequency range  $[1/4, 1/2]$ , whereas  $\{V_{1,t}\}$  is associated with low frequencies, namely the frequency range  $[0, 1/4]$ .

It is also important to note that  $\{W_{1,t}\}, \{W_{2,t}\}, \dots, \{W_{J,t}\}, \{V_{J,t}\}$  (where  $J \geq 1$ ) are associated with frequency ranges that partition the frequency interval  $[0, 1/2]$  into disjoint segments, namely into segments  $[1/4, 1/2], [1/8, 1/4], \dots, [2^{-(J+1)}, 2^{-J}], [0, 2^{-(J+1)}]$ . The set of coefficients  $\{W_{1,t}\}, \{W_{2,t}\}, \dots, \{W_{J,t}\}, \{V_{J,t}\}$  will be called the MODWT of level  $J$ .

## 2.3 Reconstruction of $\{X_t\}$

In the previous section we showed how to obtain the MODWT coefficients from an input  $\{X_t\}$ . Now, we show that it is also possible to reconstruct  $\{X_t\}$  from the MODWT of level  $J$  (where  $J \geq 1$ ), i.e. from the set of coefficients  $\{W_{1,t}\}, \dots, \{W_{J,t}\}, \{V_{J,t}\}$ . This will be a crucial point in the context of the wavelet-based denoising techniques introduced in Section 3.

First, for the Haar family, the reconstruction of  $\{X_t\}$  is very simple and amounts to summing up the coefficients from the MODWT of level  $J$ , i.e. (see, e.g. Bařta, 2014):

$$X_t = V_{J,t} + \sum_{j=1}^J W_{j,t}, \quad \text{for } t = \dots, -1, 0, 1, \dots \quad (3)$$

Second, for any family, the reconstruction is possible as follows (see, e.g. Percival and Walden, 2006; Bařta, 2014)

$$X_t = S_{J,t} + \sum_{j=1}^J D_{j,t}, \quad \text{for } t = \dots, -1, 0, 1, \dots, \quad (4)$$

where  $\{D_{j,t}: t = \dots, -1, 0, 1, \dots\}$  and  $\{S_{j,t}: t = \dots, -1, 0, 1, \dots\}$  (for  $j = 1, 2, \dots$ ) are the  $j$ th level MODWT detail and smooth, which can be obtained from the  $j$ th level MODWT coefficients by non-causal linear filtering with the non-causal filters being obtained by reflecting the filters  $\{h_{j,l}\}$  and  $\{g_{j,l}\}$  around zero. Specifically (see, e.g. Percival and Walden, 2006),

$$D_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} W_{j,t+l} = \sum_{l=-(L_j-1)}^{L_j-1} d_{j,l} X_{t-l}, \quad \text{for } t = \dots, -1, 0, 1, \dots; \quad (5)$$

$$S_{j,t} = \sum_{l=0}^{L_j-1} g_{j,l} V_{j,t+l} = \sum_{l=-(L_j-1)}^{L_j-1} s_{j,l} X_{t-l}, \quad \text{for } t = \dots, -1, 0, 1, \dots, \quad (6)$$

where the *non-causal* and *symmetric* linear filters  $\{d_{j,l} : l = -(L_j-1), \dots, -1, 0, 1, \dots, L_j-1\}$  and  $\{s_{j,l} : l = -(L_j-1), \dots, -1, 0, 1, \dots, L_j-1\}$ , both of the lengths  $2L_j-1$ , are defined as follows

$$d_{j,l} = \sum_{u=0}^{L_j-1} h_{j,u} h_{j,u+l}, \quad \text{for } l = -(L_j-1), \dots, -1, 0, 1, \dots, L_j-1; \quad (7)$$

$$s_{j,l} = \sum_{u=0}^{L_j-1} g_{j,u} g_{j,u+l}, \quad \text{for } l = -(L_j-1), \dots, -1, 0, 1, \dots, L_j-1. \quad (8)$$

We can see that the reconstruction of  $\{X_t\}$  given in Equation 3 is simpler than the reconstruction of  $\{X_t\}$  given in Equation 4 in the sense that in the former case, the coefficients are directly added when reconstructing  $\{X_t\}$ , whereas in the latter case, a linear filtering operation is applied to the coefficients before they are added.

### 3. Thresholding of Wavelet Coefficients

Denoising based on the thresholding of wavelet coefficients is usually introduced using a version of wavelet transform called the discrete wavelet transform (DWT). The DWT-based approach to thresholding is fundamental to truly understand the very concept of thresholding. However, a detailed exposition of the DWT and DWT-based approaches to thresholding would not fit the scope and extent of this paper. The rationale behind the preference for the MODWT-based rather than DWT-based approaches in this paper follows the reasoning in Coifman and Donoho (1995), who report that the latter approaches to thresholding often lead to unappealing visual artefacts, and suggest a cycle-spinning method be used in order to alleviate those artefacts. As shown in Percival and Walden (2006), cycle-spinning can be implemented using the MODWT. Thus, in this paper, the MODWT-based approach to thresholding represents the implementation of the cycle-spinning method using the MODWT.

To provide an intuitive understanding of the MODWT-based approach to thresholding, let us assume that  $\{X_t\}$  consists of a deterministic signal  $\{Z_t\}$  buried in an additive Gaussian white noise,  $\{\epsilon_t\}$ , i.e.

$$X_t = Z_t + \epsilon_t, \quad t = \dots, -1, 0, 1, \dots \quad (9)$$

The task is to recover (estimate)  $\{Z_t\}$  from  $\{X_t\}$ , the estimate being denoted as  $\{\hat{Z}_t\}$ .  $\{Z_t\}$  is assumed to be of an infinite length. The case of a finite length will be discussed in Section 3.2. The intuition behind the MODWT-based approach to thresholding of wavelet coefficients is the following: The MODWT of level  $J$  ( $J \geq 1$ ) for  $\{X_t\}$  is calculated, i.e. the sequences  $\{W_{1,t}\}$ , ...,  $\{W_{J,t}\}$ , and  $\{V_{J,t}\}$  are obtained. Subsequently, for each sequence  $\{W_{j,t}\}$  (for  $j = 1, \dots, J$ ), it is often the case that coefficients that are large in magnitude are attributable to  $\{Z_t\}$ , whereas the small ones are attributable to  $\{\epsilon_t\}$  (see Percival and Walden, 2006). As a result, noise can be approximately removed from

$\{W_{j,t}\}$  (for  $j = 1, \dots, J$ ) by either setting those coefficients that are small in magnitude to zero and leaving the larger ones untouched (the hard thresholding rule) or by setting those coefficients that are small in magnitude to zero and pushing the larger ones toward zero (the soft thresholding rule). This procedure – called the thresholding of wavelet coefficients – is performed for each sequence  $\{W_{j,t}\}$  separately (for  $j = 1, \dots, J$ ). After the thresholding of wavelet coefficients has finished, the altered coefficient sequences are used either in Equation 3 or Equation 4 to approximately reconstruct  $\{Z_t\}$ , i.e. to obtain  $\{\hat{Z}_t\}$ .

The MODWT-based approach to thresholding is rigorously introduced in Section 3.1. Section 3.2 discusses the situation of thresholding a time series of a *finite* length and Section 3.3 deals with the selection of a threshold – a crucial parameter which facilitates the assessment of whether a given coefficient is large in magnitude or not. Section 3.4 discusses misspecification issues related to the use of the thresholding approach for denoising a time series prior to its forecasting.

### 3.1 MODWT-based Approach to Thresholding

In this section, the MODWT-based approach to the thresholding of wavelet coefficients will be introduced rigorously. It runs as follows:

- Calculate the MODWT of level  $J$  for  $\{X_t: t = \dots, -1, 0, 1, \dots\}$ , resulting in  $J + 1$  sequences, namely  $\{W_{1,t}: t = \dots, -1, 0, 1, \dots\}$ ,  $\dots$ ,  $\{W_{J,t}: t = \dots, -1, 0, 1, \dots\}$ ,  $\{V_{J,t}: t = \dots, -1, 0, 1, \dots\}$ .
- Threshold  $\delta$  has to be specified (see Section 3.3). Due to theoretical issues (Percival and Walden, 2006), level-specific thresholds  $\delta_j$  ( $j = 1, \dots, J$ ) have to be defined as  $\delta_j = \delta/2^{j/2}$ . These level-specific thresholds will be used to assess whether or not a given coefficient  $W_{j,t}$  is large in magnitude.
- A thresholding rule is applied to each  $W_{j,t}$  (for  $j = 1, \dots, J$  and  $t = \dots, -1, 0, 1, \dots$ ). The goal of the rule is to remove that part of  $W_{j,t}$  which is attributable to noise by transforming  $W_{j,t}$  to  $W_{j,t}^{(thr)}$ , called the thresholded coefficient.  $W_{j,t}^{(thr)}$  is based on the comparison of  $W_{j,t}$  with  $\delta_j$ . Various thresholding rules are introduced in Section 3.1.1.
- Level  $J$  scaling coefficients, i.e.  $\{V_{J,t}\}$ , are left untouched since no noise is assumed to be involved in these coefficients since they correspond to the low-frequency dynamics of  $\{X_t\}$ , which is assumed to be free of noise.
- Reconstruction: The reconstruction equations, i.e. either Equation 3 (in the case of the Haar wavelet family) or Equation 4, are applied to the new set of coefficients  $(\{W_{1,t}^{(thr)}: t = \dots, -1, 0, 1, \dots\}, \dots, \{W_{J,t}^{(thr)}: t = \dots, -1, 0, 1, \dots\}, \{V_{J,t}: t = \dots, -1, 0, 1, \dots\})$ , the result of the reconstruction being  $\{\hat{Z}_t: t = \dots, -1, 0, 1, \dots\}$ .

#### 3.1.1 Thresholding Rules

Various thresholding rules can be used to transform  $W_{j,t}$  to  $W_{j,t}^{(thr)}$  by comparing  $W_{j,t}$  with  $\delta_j$ . The simplest thresholding rule is the *hard* thresholding rule which sets those coefficients to zero that are in magnitude less than or equal to  $\delta_j$  and leaves the other coefficients untouched, i.e. (Percival and Walden, 2006):



$$W_{j,t}^{(thr)} = \begin{cases} 0, & |W_{j,t}| \leq \delta_j, \\ W_{j,t}, & \text{otherwise.} \end{cases} \quad (10)$$

Other thresholding rules can be employed besides the hard thresholding rule. Specifically, the *soft* thresholding rule sets those coefficients to zero which are in magnitude less than or equal to  $\delta_j$  and pushes coefficients greater in magnitude than  $\delta_j$  toward zero by  $\delta_j$ , i.e. (Percival and Walden, 2006):

$$W_{j,t}^{(thr)} = \text{sign} \{W_{j,t}\} (|W_{j,t}| - \delta_j)_+, \quad (11)$$

where

$$\text{sign} \{W_{j,t}\} = \begin{cases} +1, & W_{j,t} > 0, \\ 0, & W_{j,t} = 0, \\ -1, & W_{j,t} < 0, \end{cases} \quad (x)_+ = \begin{cases} x, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (12)$$

The *mid* thresholding rule is a compromise between the hard and the soft thresholding rule and is defined (Percival and Walden, 2006) as

$$W_{j,t}^{(thr)} = \text{sign} \{W_{j,t}\} (|W_{j,t}| - \delta_j)_{++}, \quad (13)$$

where

$$(|W_{j,t}| - \delta_j)_{++} = \begin{cases} 2(|W_{j,t}| - \delta_j)_+, & |W_{j,t}| < 2\delta_j, \\ |W_{j,t}|, & \text{otherwise.} \end{cases} \quad (14)$$

### 3.2 Boundaries and Boundary Conditions

The thresholding algorithm presented above can be extended to the case of  $\{X_t; t = 0, \dots, N-1\}$  of a finite length  $N$  given as

$$X_t = Z_t + \epsilon_t, \quad t = 0, \dots, N-1. \quad (15)$$

More specifically, to obtain  $\{\hat{Z}_t; t = 0, \dots, N-1\}$  from  $\{X_t; t = 0, \dots, N-1\}$ , i.e.  $\hat{Z}_t$  for all  $t = 0, \dots, N-1$ , surrogates for the (unavailable) values of  $X_t$  for  $t = -L_j + 1, \dots, -1$  (historical values) must be supplied since these values of  $X_t$  are needed if Equation 3 is used in the reconstruction part of the thresholding procedure. This follows directly from equations 1, 2 and 3. No future values of  $X_t$  (i.e. values of  $X_t$  for  $t = N, N+1, \dots$ ) are needed in this scenario.

On the other hand, if Equation 4 is used in the reconstruction part of the thresholding procedure, surrogates for the (unavailable) values of  $X_t$  for  $t = -L_j + 1, \dots, -1$  (historical values) and  $t = N, \dots, N+L_j-2$  (future values) are needed. This follows directly from equations 4, 5 and 6.

Thus, the former scenario (employing Equation 3 for reconstruction) does *not* require that (surrogates for) the future values of  $X_t$  are supplied during the prior denoising step



(since these future values are not needed). This is not the case in the latter scenario, which requires (surrogates for) the future values. On the other hand, the former scenario is based on *causal* filters (see equations 1, 2 and 3) that implicitly induce a time delay in the denoised time series, which can be considered as unwanted. This is not the case of the latter scenario, which is based on *non-causal* and *symmetric* filters, and, therefore, no time delay is induced in the denoised time series (see equations 4, 5 and 6). Consequently, it is hard to assess which of the two scenarios would be more advantageous since both possess one undesirable property (either a time delay in the denoised time series or a need to forecast the future values of  $X_t$  in the first place). As a result, both scenarios are explored in our paper.

Surrogates for the historical and future values of  $X_t$  which may be required during the prior denoising step, as was just discussed, will be referred to as boundaries. When boundaries are defined, they can be used in combination with either of the two reconstruction equations (Equation 3 or 4). This leads to the definition of the so-called boundary conditions. The next section discusses various types of boundaries and boundary conditions.

### 3.2.1 Periodic, Reflection, and Forecast Boundaries and Boundary Conditions

In this section, three types of boundaries (i.e. three ways of how surrogates for the unavailable historical and future values of  $X_t$  values can be obtained) are introduced, namely the periodic, reflection, and forecast boundaries. Subsequently, various boundary conditions (i.e. combinations of boundaries and reconstruction equations) are introduced.

In the case of *periodic boundaries*,  $X_t$  for  $t < 0$  and  $t > N - 1$  is defined (Percival and Walden, 2006; Bašta, 2014) as follows

$$X_t = X_{t \bmod N}, \quad t < 0 \text{ or } t > N - 1, \quad (16)$$

where *mod* denotes the modulo operation.

*Reflection boundaries* are defined in two steps. In the first step, we define (Percival and Walden, 2006; Bašta, 2014)

$$X_t = X_{2N-1-t}, \quad t = N, \dots, 2N - 1. \quad (17)$$

This leads to the definition of a new time series of length  $2N$ ,  $\{X_t; t = 0, \dots, N - 1, N, \dots, 2N - 1\}$ , consisting of  $X_0, X_1, \dots, X_{N-1}, X_N = X_{N-1}, X_{N+1} = X_{N-2}, \dots, X_{2N-2} = X_1$  and  $X_{2N-1} = X_0$ . Subsequently, periodic boundaries are imposed upon this new time series, which results in the definition of  $X_t$  for all  $t$ .

In the case of *forecast boundaries*,  $X_t$  for  $t > N - 1$  and  $t < 0$  are defined by forecasting and backcasting  $\{X_t; t = 0, \dots, N - 1\}$  with the use of the same forecasting approach that is employed for forecasting the denoised time series.

It is important to note that forecast boundaries need not generally be better than reflection boundaries since the forecast boundaries, while being more flexible, are expected to have less bias but a higher variance than reflection boundaries. Periodic

boundaries are generally expected to provide the poorest results since they will generally be extremely biased.

If periodic, reflection, and forecast boundaries are used in combination with Equation 4, we will refer to this combination as *boundary conditions* Ia, IIa, and IIIa. Analogously, periodic, reflection, and forecast boundaries used in combination with Equation 3 (being applicable only to the Haar wavelets) will be referred to as boundary conditions Ib, IIb, and IIIb.

### 3.3 Threshold Selection

The notion of threshold  $\delta$  was introduced in Section 3.1. The question arises how to determine  $\delta$ . Several approaches to this determination will be described, namely the universal threshold, the infinite threshold, and the cross-validation-based threshold. More thresholds are given in the literature although a larger number of thresholds used in our analysis would lead to a difficult to manage amount of parameter combinations. Consequently, only the three mentioned thresholds will be used.

The *universal threshold* is defined (Donoho and Johnstone, 1994) as

$$\delta^{(u)} = \sigma_\epsilon \sqrt{2 \log(N)}, \quad (18)$$

where  $\log$  stands for the natural logarithm and  $\sigma_\epsilon$  is the standard deviation of the white noise  $\{\epsilon_t\}$  and is often estimated (Percival and Walden, 2006) as

$$\hat{\sigma}_\epsilon = \frac{2^{1/2} \text{median}(|W_{1,0}|, |W_{1,2}|, \dots, |W_{1,N-1}|)}{0.6745}. \quad (19)$$

The first-level MODWT wavelet coefficients of Equation 19 are implicitly assumed to have a zero mean. Consequently, we use D(6) wavelets for the calculation of the first-level MODWT wavelet coefficients in Equation 19 despite the fact that the Haar or D(4) wavelets will be employed in the thresholding algorithm itself. The reason for this is that the D(6) first-level MODWT wavelet coefficients are centred around zero if there is a trend in the time series  $\{X_t\}$ , which is a locally polynomial of degree *less* than three and if the time series is integrated of an order *less* than three (Percival and Walden, 2006), which is not the case with the Haar or D(4) wavelet coefficients.

The *infinite threshold* is defined as  $\delta^{(\infty)} = \infty$ . In the case of the infinite threshold, the thresholding algorithm is equivalent to linear filtering  $\{X_t; t = 0, \dots, N-1\}$  with the low-pass filter  $\{g_{J,l}\}$  or  $\{s_{J,l}\}$  if Equations 3 or 4 are respectively used for reconstruction.

Another approach to determining the value of  $\delta$  may rest on the joint optimisation of  $\delta$  and the other parameters of the forecasting method. This will be done in the following way. The performance of a forecasting model combined with the prior denoising step using a given  $\delta$  value is assessed by time series cross-validation, resulting in a cross-validation error. The cross-validation error is calculated as an average of one-step-ahead squared forecast errors in the rolling-origin-recalibration approach to cross-validation (see Tashman, 2000; Bergmeir and Benítez, 2012). Cross-validation errors are evaluated for several values of  $\delta$ , the optimal value – denoted as  $\delta^{(cv)}$  and called the *cross-validation-based threshold* – being the one which minimises the cross-validation error.

### 3.4 Model Misspecification

In real-life situations,  $\{X_t\}$  is often not generated as dictated by Equation 9. It is common to have correlated noise instead of an i.i.d. one and multiplicative noise instead of an additive one; the noise can be non-Gaussian or the assumption of an underlying deterministic signal need not be correct. It would potentially be possible to use such wavelet-based denoising techniques that could accommodate these different assumptions. However, this would be beyond the scope and extent of the current paper, where we implicitly deal with the simplest setting, namely the assumption of an underlying deterministic signal and an additive Gaussian white noise. It should be stressed, however, that even if  $\{X_t\}$  is not generated as dictated by Equation 9, denoising based on the setting given by this equation can potentially be beneficial. We provide several justifications for such a statement.

First, forecasts from a complex model, which provides a correct description of the data generating mechanism, can potentially be outperformed – because of the bias-variance tradeoff issue – by simpler and misspecified models (Shmueli, 2010; Hastie et al., 2011). The results of the so-called M-Competitions (Makridakis and Hibon, 2000; Makridakis et al., 1982; Makridakis et al., 1993) suggest that this is often the case of economic and business time series forecasting where simple forecasting approaches such as exponential smoothing, various naïve and extrapolation techniques, etc. often outrun more complex forecasting models. The simple forecasting approaches are often formulated as algorithms (methods) used to construct the forecasts rather than as statistical/econometric models that would provide a correct description of the nature of the underlying signal, nature of the noise etc.

Second, simple exponential smoothing – which performs well in forecasting economic and business time series and is formulated as a forecasting algorithm rather than as a statistical/econometric model can be considered to be an application of the random walk method to a time series in which noise has been removed by applying a causal exponentially weighted moving average. A prior wavelet-based denoising step, as introduced in this paper, combined with *the random walk method*, is closely related to simple exponential smoothing – the difference being only in the way noise is removed (see also Ferbar et al., 2009).

Third, for those time series and forecasting models for which denoising (as implemented in our paper) is detrimental, none or only a small amount of noise should be removed, whereas a larger amount of noise should be removed in other instances. Consequently, we can explore whether the time series cross-validation-based approach to threshold selection described in Section 3.3 can be employed to adaptively decide on the optimal amount of noise to be removed. Such an approach is similar to the approach of Lotrič (2004).

It should also be noted that not all forecasting models are expected to work well after a prior wavelet-based denoising step has been applied to the time series. The reason is that some forecasting models (such as AR1, s\_ARIMA etc.) may not generally be suitable to be applied to denoised (smoothed) time series. On the other hand, naïve forecasting models (such as RW, SES etc.) may potentially benefit more from the prior wavelet-based denoising step.

## 4. Evaluation of Forecast Accuracy for Models with the Prior Denoising Step

We use the M3 competition yearly time series to evaluate the forecast accuracy of the forecasting models combined with the prior denoising step. R software (R Core Team, 2015) and the forecast R package (Hyndman and Khandakar, 2008; Hyndman, 2013a) have been used in the analysis.

### 4.1 M3 Competition Time Series

The M3 competition dataset was introduced in the Introduction. In our analysis, only yearly time series are used for the reason that our thresholding algorithm implicitly assumes that the first-level wavelet coefficients for  $\{X_t\}$  are mostly attributable to the noise. Such a condition is better met by yearly rather than quarterly or monthly time series. There are 645 yearly time series in the data set. The mean, median, lower, and upper quartile of the lengths of the 645 time series are 22.4, 19, 15 and 30; the minimum and maximum are 14 and 41.

### 4.2 Evaluation of Forecast Accuracy

Let us assume a forecasting model (either with or without the prior denoising step), the historical data  $\{X_t; t = 0, \dots, N-1\}$  and the forecast horizon  $h$ . Let the symmetric absolute percentage error ( $sAPE_h$ ) and the squared error relative to that of the random walk model ( $RWB_h$ ) be defined (see also Hyndman and Koehler, 2006) as

$$sAPE_h = 200 \frac{|X_{N-1+h} - \hat{X}_{N-1+h|N-1}^{(*)}|}{X_{N-1+h} + \hat{X}_{N-1+h|N-1}^{(*)}}, \quad h = 1, 2, \dots, \quad (20)$$

$$RWB_h = \frac{(X_{N-1+h} - \hat{X}_{N-1+h|N-1}^{(*)})^2}{(X_{N-1+h} - \hat{X}_{N-1+h|N-1}^{(RW)})^2}, \quad h = 1, 2, \dots, \quad (21)$$

where  $\hat{X}_{N-1+h|N-1}^{(*)}$  is an  $h$ -step-ahead point forecast of  $\{X_t; t = 0, \dots, N-1\}$  obtained from the forecasting model (either with or without the prior denoising step) and  $\hat{X}_{N-1+h|N-1}^{(RW)}$  that obtained from the random walk (RW) model *without* the prior denoising step.  $X_{N-1+h}$  is the true future value (which can be found in the testing set of the M3 dataset).  $RWB_h$  can be equal to infinity or undefined in some situations. The 20% trimmed means of  $sAPE_h$  and  $RWB_h$ , denoted as *trimsAPE* and *trimRWB*, respectively, can be calculated (with undefined values omitted) across different time series.

### 4.3 Attributes of the Thresholding Algorithm

The wavelet family (Haar, D(4)),  $J$ , thresholding rule,  $\delta$ ,  $\sigma_\epsilon$  estimator, and boundary conditions have to be chosen as the characteristics of the thresholding algorithm – hereafter called attributes of the thresholding algorithm. Since different thresholds require different

attributes to be selected, we present the attributes and their attainable values for  $\delta^{(u)}$ ,  $\delta^{(x)}$  and  $\delta^{(cv)}$  separately in the columns of Table 1. The total number of combinations studied for  $\delta^{(cv)}$  is rather low since the calculation of  $\delta^{(cv)}$  is computationally very expensive. The table also reveals that unlike I Ib boundary conditions, Ib and IIIb ones are not explored since the choice of the boundaries (periodic, reflection, forecast) is generally not of much importance (as revealed by the Monte Carlo simulations) if Equation 3 is used for reconstruction.

From Table 1 it follows that the number of combinations associated with the  $\delta^{(u)}$ ,  $\delta^{(x)}$  and  $\delta^{(cv)}$  threshold is:

- 2 (wavelet families)  $\times$  3 (values of  $J$ )  $\times$  3 (thresholding rules) = 18 (for  $\delta^{(u)}$ ),
- 2 (wavelet families)  $\times$  3 (values of  $J$ ) = 6 (for  $\delta^{(x)}$ ),
- 1 (wavelet family)  $\times$  1 (value of  $J$ )  $\times$  1 (thresholding rule) = 1 (for  $\delta^{(cv)}$ )

for *each* of the boundary conditions Ia, IIa and IIIa. In total, we thus have 18 + 6 + 1 = 25 combinations for each of the Ia, IIa and IIIa boundary conditions. For the I Ib boundary conditions the number of combinations is less as only Haar wavelets can be used with the I Ib boundary conditions. Namely, the number of combinations associated with the  $\delta^{(u)}$ ,  $\delta^{(x)}$  and  $\delta^{(cv)}$  threshold is:

- 1 (wavelet family)  $\times$  3 (values of  $J$ )  $\times$  3 (thresholding rules) = 9 (for  $\delta^{(u)}$ ),
- 1 (wavelet family)  $\times$  3 (values of  $J$ ) = 3 (for  $\delta^{(x)}$ ),
- 1 (wavelet family)  $\times$  1 (value of  $J$ )  $\times$  1 (thresholding rule) = 1 (for  $\delta^{(cv)}$ )

for the I Ib boundary conditions, which amounts to a total of 9 + 3 + 1 = 13 possible combinations for I Ib boundary conditions. In total (regardless of the boundary conditions), we have 3  $\times$  25 + 13 = 88 possible combinations.

**Table 1 | Attribute values for various thresholds**

attributes/threshold	$\delta^{(u)}$	$\delta^{(x)}$	$\delta^{(cv)}$
wavelet family	Haar, D(4) (not used with I Ib)	Haar, D(4) (not used with I Ib)	Haar
$J$	1, 2, 3	1, 2, 3	1
thresholding rule	hard, soft, mid	–	mid
$\sigma_\epsilon$ estimator	“D6” (see Eq. 19)	–	–
bound. conditions	Ia, IIa, IIIa, I Ib	Ia, IIa, IIIa, I Ib	Ia, IIa, IIIa, I Ib

Source: Author’s own calculations

### 5. Results

Section 5.1 provides descriptive results and Section 5.2 assesses their statistical significance. Further, Section 5.3 provides a comparison of simple exponential smoothing with its wavelet-based analogue.

## 5.1 Descriptive Results

The performance is evaluated conditional to the forecast horizon which is set to either  $h = 1$  or  $h = 5$ . The results for *trimRWB* are given in Figure 1 ( $h = 1$ ) and Figure 2 ( $h = 5$ ), whereas those for *trimsAPE* are given in Figure 3 ( $h = 1$ ) and Figure 4 ( $h = 5$ ). In each of the figures, the results are plotted separately for boundary conditions Ia, IIa, IIIa and IIb.

The results indicate that the attributes of the thresholding algorithm interact in a complex way with the forecasting model and the forecast horizon in determining the forecasting performance. There is no obvious tendency for any of the forecasting models to benefit more from the prior denoising step than the other models. The prior denoising step is generally harmful (measured either by *trimRWB* or *trimsAPE*) for  $h = 1$ , whereas it could be – if the appropriate attribute values of the denoising algorithm are chosen – slightly beneficial for  $h = 5$  for some forecasting models. The significance of these results is assessed in Section 5.2 and is reported in the figures as described in the captions.

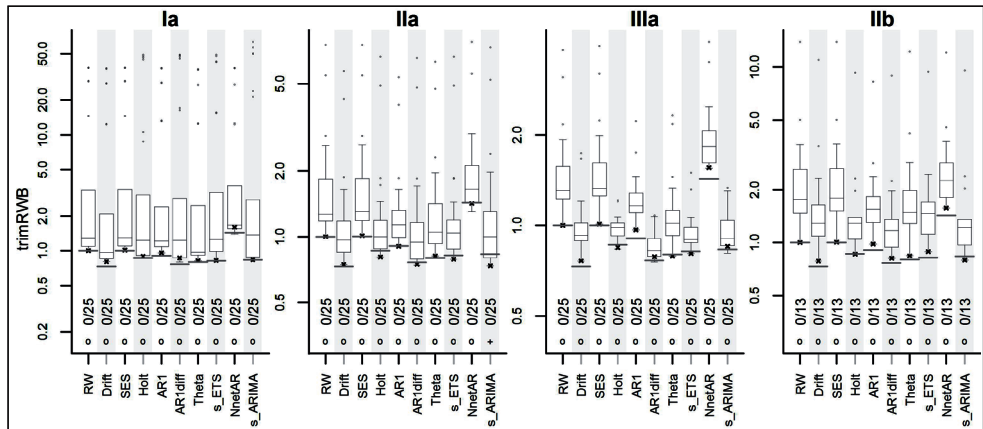
Noting the different ranges of the vertical axes in the subplots associated with different boundary conditions, we can conclude that IIIa boundary conditions seem to be typically the best – IIa boundary conditions being typically comparable to IIIa, IIb boundary conditions being typically slightly worse than IIIa, and Ia boundary conditions performing *typically* the worst (regardless of the forecast horizon and the error measure), which is not surprising and is in agreement with the arguments given in Section 3.2.1. However, an exception to this rule is the Drift method and  $h = 5$ , where the use of Ia boundary conditions produces several significant outcomes (if *trimRWB* error measure is used) and results that outperform not only the Drift method without the denoising step but also other traditional forecasting models/methods without the prior denoising step. This can be explained by observing that for many of the 645 time series, the drift in the time series (i.e. the average change observed in the historical data) becomes weaker in the testing data set. Employing periodic boundaries (which are used in Ia boundary conditions and would normally be considered biased) often induces this weaker drift and thus leads to a better forecast.

Concerning the cross-validation-based approach to threshold selection, which we intended to explore as it provides an adaptive approach to denoising, its performance is close to that of the forecasting method without the prior denoising step in most (but not all) of the cases (see the cross symbols in the figures and compare the position of these symbols with the position of the black segments). The assessment of whether or not the cross-validation-based approach performs significantly better than the model without the prior denoising step is given in Section 5.2 and the results of the assessment are depicted in the figures as described in the captions.

Further, as noted in the introduction, we want to examine the performance of the prior wavelet-based denoising step (namely the one which uses cross-validation to threshold selection) combined with the random walk method and compare its performance with the performance of simple exponential smoothing without the prior denoising step. We can note that if *trimRWB* is used as the error measure, the performance of the RW method combined with the prior cross-validation-based denoising step is very similar to the performance of simple exponential smoothing (without any prior denoising step) for both the forecast horizons ( $h = 1$  and  $h = 5$ ). On the other hand, if *trimsAPE* is used as the error measure, the RW method combined with the prior cross-validation-based denoising step (and boundary conditions IIa or IIIa) seems to be slightly superior to simple exponential smoothing for  $h = 1$  but not for  $h = 5$ . The significance of these results

is assessed in Section 5.2 and is depicted in the figures as described in the captions. A further and more detailed comparison is provided in Section 5.3.

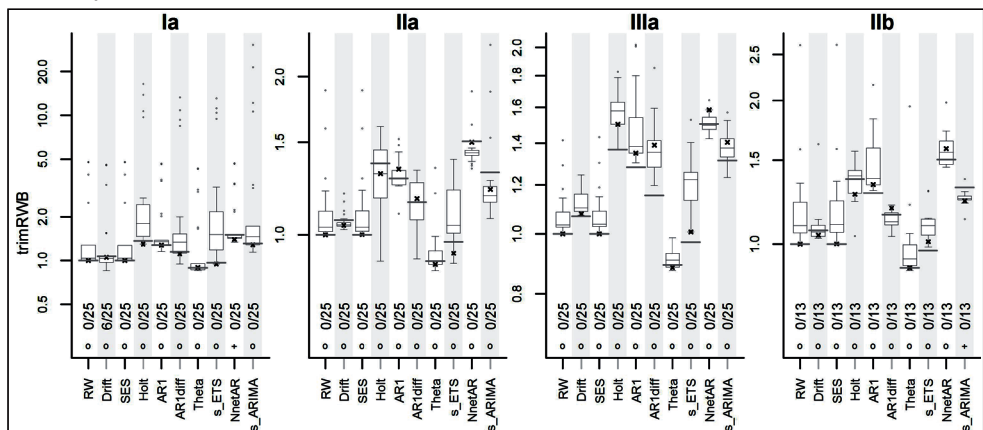
**Figure 1 | Results for *trimRWB* and  $h = 1$**



Note: Subplots correspond to boundary conditions Ia, IIa, IIIa, and IIb. For each subplot, various forecasting models are explored (x-axis). For a given forecasting model, there are 25 *trimRWB* values associated with each of the boundary conditions Ia, IIa, and IIIa, and 13 *trimRWB* values associated with boundary conditions IIb (see Section 4.3). Each of the *trimRWB* values relates to one specific combination of the attribute values from Table 1 and is calculated from the corresponding  $RWB_{h=1}$  errors across all the 645 time series. The boxplot of the 25 (or 13) *trimRWB* values is plotted (for each forecasting model). Further, the position of the *trimRWB* value for the single combination of the attribute values which employs cross-validation is denoted by the cross symbol. Further, a black horizontal segment is added for each forecasting model depicting *trimRWB* of the model without the prior denoising step calculated for  $h = 1$  across all the 645 time series. The significance of the superior performance of the models employing the prior denoising step is assessed as described in Section 5.2, the results being presented using fractions, and the “o” and “+” signs positioned close to x-axis labels.

Source: Author's own calculations

**Figure 2 | Results for *trimRWB* and  $h = 5$**

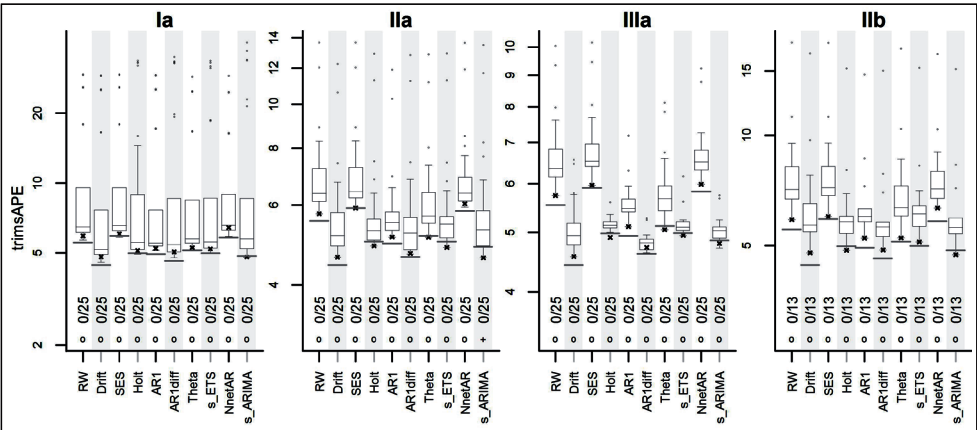


Note: The meaning of the plots is analogous to that in Figure 1 with the exception that  $RWB_{h=5}$  errors are used in the evaluation.

Source: Author's own calculations



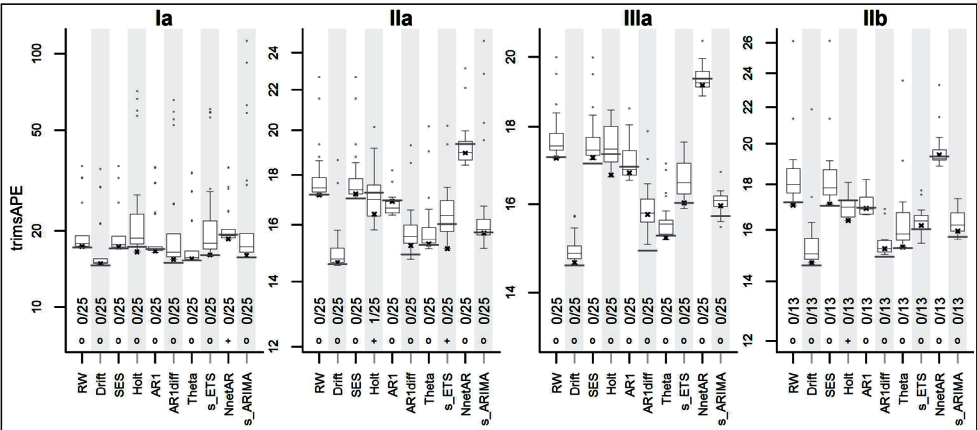
**Figure 3 | Results for *trimsAPE* and  $h = 1$**



Note: The meaning of the plots is analogous to Figure 1 with the exception that *trimsAPE* is assumed as the error measure.

Source: Author's own calculations

**Figure 4 | Results for *trimsAPE* and  $h = 5$**



Note: The meaning of the figure is analogous to Figure 2 with the exception that the *trimsAPE* error measure is assumed.

Source: Author's own calculations

## 5.2 Significance of the Results

In each figure of Section 5.1, the significance of the superior forecasting performance of the model with the prior denoising step (compared to the model without the step) is assessed for a given forecasting model (RW, Drift, etc.) and the given boundary conditions (Ia, IIa, IIIa or IIb) as follows:<sup>5</sup> For *each single* combination of the attribute values

<sup>5</sup> We assume that the 645 time series we have used in the analysis are a random sample from a population of (economic, financial, demographic etc.) time series.

of the thresholding algorithm, we conduct an *individual* test (either *trimRWB* or *trimsAPE* error measure is assumed)

$$H_0: \text{TRIMRWB}^{(d)} = \text{TRIMRWB}, \quad (H_0: \text{TRIMSAP}^{(d)} = \text{TRIMSAP}), \quad (22)$$

$$H_1: \text{TRIMRWB}^{(d)} < \text{TRIMRWB}, \quad (H_1: \text{TRIMSAP}^{(d)} < \text{TRIMSAP}), \quad (23)$$

where  $\text{TRIMRWB}^{(d)}$  ( $\text{TRIMSAP}^{(d)}$ ) and  $\text{TRIMRWB}$  ( $\text{TRIMSAP}$ ) are the true (population) 20% trimmed means of *RWB* (*sAPE*) for the forecasting model with and without the prior denoising step. The test is conducted using the Yuen method (for paired data) presented in Wilcox (2012). There are 25 such tests (for a given model) for each of the boundary conditions Ia, IIa, and IIIa, and 13 tests for IIb boundary conditions and each of the tests corresponds to a different combination of attribute values (see also Section 4.3). The false discovery rate approach to multiple hypothesis testing (assuming dependent tests, see Miller et al., 2001) is used to decide about the rejections of the null hypotheses in the 25 (or 13) tests. The settings are such that

$$E \left( \frac{R_f}{R_f + R_c} \right) \leq 0.05, \quad (24)$$

where  $R_f$  and  $R_c$  are the numbers of false and correct rejections. A fraction – close to the relevant x-axis label – is reported in the figures, with the fraction's numerator being the number of rejected null hypotheses and the denominator being the total number tested. This approach to multiple hypothesis testing is appealing since its results indicate whether it is easy to find a combination of attribute values which would, if used in the prior denoising step, enhance the forecast accuracy of a given forecasting model. Since the number of rejected hypotheses is zero in most of the instances (with the exception of the Drift method combined with Ia boundary conditions for  $h = 5$  and the *trimRWB* error measure, and the Holt method combined with IIa boundary conditions for  $h = 5$  and the *trimsAPE* error measure), it is obvious that an *ad hoc* use of the prior denoising step is generally not recommended.

Since the assessment of the performance of the cross-validation-based approach to threshold selection (the cross symbol in the figures) has been selected as one of our major goals (see Introduction), the results of the individual hypothesis test associated with this approach (there is only one such test for a given forecasting model and given boundary conditions) are reported explicitly. Specifically, should the p-value of the test be less than or equal to 0.05, a + symbol is reported close to the relevant x-axis label in the figures, otherwise a o symbol is reported. For *trimRWB* errors, there is one significant outcome for  $h = 1$  (s\_ARIMA combined with IIa boundary conditions) and two for  $h = 5$  (NnetAR combined with Ia boundary conditions, and s\_ARIMA combined with IIb boundary conditions), which does not support the view that the prior denoising step employing cross-validation for threshold selection is helpful. Analogous results can be reported for *trimsAPE* errors where one significant outcome is present for  $h = 1$  and four for  $h = 5$ .

### 5.3 Comparison of SES and RW Combined with Cross-validation

Random walk (RW) combined with the prior cross-validation-based denoising step can be considered a wavelet-based analogue to simple exponential smoothing (SES) without the prior denoising step (Ferbar et al., 2009). To compare the performance (measured either by *trimRWB* or *trimsAPE*) of these two forecasting approaches, point estimates of the difference between the performances and the 95% upper confidence bounds for the difference are reported in Table 2. If *trimRWB* is used as an error measure, RW combined with the prior cross-validation-based denoising step performs significantly better (in the statistical sense) than SES without the prior denoising step both for  $h = 1$  and  $h = 5$ , the difference is, however, negligible from the practical point of view (since the upper confidence bound, though being negative, is practically very close to zero). If *trimsAPE* is used to measure the performance, RW combined with the prior cross-validation-based denoising step performs significantly better than SES without the prior denoising step for  $h = 1$  if boundary conditions IIa and IIIa are utilised; no similar significant results are reported for  $h = 5$ .

**Table 2 | The difference between the error calculated for RW combined with the prior cross-validation-based denoising step and the error for SES without the prior denoising step**

	Ia	IIa	IIIa	IIb
<i>trimRWB</i> , $h = 1$	$-2.0 \times 10^{-4}$ ( $-1.5 \times 10^{-4}$ )	$-2.0 \times 10^{-4}$ ( $-1.5 \times 10^{-4}$ )	$-2.0 \times 10^{-4}$ ( $-1.5 \times 10^{-4}$ )	$-2.0 \times 10^{-4}$ ( $-1.5 \times 10^{-4}$ )
<i>trimRWB</i> , $h = 5$	$-2.2 \times 10^{-5}$ ( $-1.0 \times 10^{-5}$ )	$-2.2 \times 10^{-5}$ ( $-1.0 \times 10^{-5}$ )	$-2.2 \times 10^{-5}$ ( $-1.0 \times 10^{-5}$ )	$-2.2 \times 10^{-5}$ ( $-1.0 \times 10^{-5}$ )
<i>trimsAPE</i> , $h = 1$	0.00 (0.16)	-0.17 (-0.02)	-0.17 (-0.02)	-0.02 (0.13)
<i>trimsAPE</i> , $h = 5$	0.32 (0.62)	0.13 (0.42)	0.13 (0.42)	0.03 (0.34)

Note: Negative values imply a lower error of the former method. Various error measures and forecast horizons (rows) and various boundary conditions (columns) are explored. The first number in each cell represents the point estimate of the difference (i.e. the estimated difference for the 645 time series). The number in the parentheses stands for the 95% upper confidence bound for the difference.

Source: Author's own calculations

### Conclusions

We have used M3 competition yearly time series to explore the usefulness of the prior wavelet-based denoising step. The outcomes of the analysis suggest that the step is generally not recommended, which implies that a straightforward generalisation of some of the results available in the literature (which found the denoising step to be beneficial) is

not possible. Even if cross-validation is used to select the value of the threshold, a superior performance of the forecasting model with the prior wavelet-based denoising step does not generally follow. The performance of a random walk model combined with the prior denoising step employing cross-validation for threshold selection is similar to simple exponential smoothing.

Even though an ad hoc use of a prior wavelet-based denoising step is generally not recommended, it may be possible that the use of the denoising step could be beneficial (if appropriate attributes of the denoising algorithm were used) for some forecasting models if it was used only with those time series that possess some specific features. We have already conducted a preliminary analysis (not presented in this paper), which suggests that this could be the case. More research is needed in this direction.

## Acknowledgment

The author of the paper thanks Karel Helman, PhD, (Private University College of Economic Studies) for kind assistance in proofreading the text.

## References

- Alrumaih, R. M. and Al-Fawzan, M. A. (2002). Time Series Forecasting Using Wavelet Denoising: An Application to Saudi Stock Index. *Journal of King Saud University. Engineering Sciences*, 2(14), pp. 221-234.
- Bašta, M. (2014). Additive Decomposition and Boundary Conditions in Wavelet-based Forecasting Approaches. *Acta Oeconomica Pragensia*, 22(2), pp. 48-70, <https://doi.org/10.18267/j.aop.431>
- Bergmeir, C. and Benítez, J. M. (2012). On the Use of Cross-validation for Time Series Predictor Evaluation. *Information Sciences*, 191, pp. 192-213, <https://doi.org/10.1016/j.ins.2011.12.028>
- Boto-Giralda, D., Díaz-Pernas, F. J., González-Ortega, D., Díez-Higuera, J. F., Antón-Rodríguez, M., Martínez-Zarzuela, M. and Torre-Díez, I. (2010). Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, 25(7), pp. 530-545, <https://doi.org/10.1111/j.1467-8667.2010.00668.x>
- Bruzda, J. (2014). Forecasting via wavelet denoising: the random signal case. In: M. Gallegatti and W. Semmler, eds., *Wavelet applications in economics and finance*, 1st ed. New York: Springer, pp. 187-225.
- Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant de-noising. In: A. Antoniadis and G. Oppenheim G., eds., *Wavelets and Statistics. Lecture Notes in Statistics*, vol. 103. New York: Springer, pp. 125-150, [https://doi.org/10.1007/978-1-4612-2544-7\\_9](https://doi.org/10.1007/978-1-4612-2544-7_9)
- Donoho, D. L. and Johnstone, J. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, 81(3), pp. 425-455, <https://doi.org/10.1093/biomet/81.3.425>
- Ferbar, L., Čreslovnik, D., Mojškerc, B. and Rajgelj, M. (2009). Demand Forecasting Methods in a Supply Chain: Smoothing and Denoising. *International Journal of Production Economics*, 118(1), pp. 49-54, <https://doi.org/10.1016/j.ijpe.2008.08.042>
- Hastie, T., Tibshirani, R. and Friedman, J. (2011). *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York: Springer.
- Hyndman, R. J. and Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Melbourne: OTexts.

- Hyndman, R. J. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 27(3), <https://doi.org/10.18637/jss.v000.i00>
- Hyndman, R. J. and Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, 22(4), pp. 679-688, <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Hyndman, R. J. (2013a). forecast: Forecasting functions for time series and linear models. R package version 4.7. Available at: <https://cran.r-project.org/web/packages/forecast/index.html> [Accessed 15 Apr. 2015].
- Hyndman, R. J. (2013b). Mcomp: Data from the M-competitions. R package version 2.05. Available at: <https://cran.r-project.org/web/packages/Mcomp/index.html> [Accessed 25 Nov. 2015].
- Lotrič, U. (2004). Wavelet Based Denoising Integrated into Multilayered Perceptron. *Neurocomputing*, 62, pp. 179-196, <https://doi.org/10.1016/j.neucom.2004.02.003>
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, 1(2), pp. 111-153, <https://doi.org/10.1002/for.3980010202>
- Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. and Simmons, L. F. (1993). The M2-competition: A Real-Time Judgmentally Based Forecasting Study. *International Journal of Forecasting*, 9(1), pp. 5-22, [https://doi.org/10.1016/0169-2070\(93\)90044-N](https://doi.org/10.1016/0169-2070(93)90044-N)
- Makridakis, S. and Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting*, 16(4), pp. 451-476, [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)
- Miller, C. J., Genovese, C., Nichol, R. C., Wasserman, L., Connolly, A., Reichart, D., Hopkins, A., Schneider, J. and Moore, A. (2001). Controlling the False-discovery Rate in Astrophysical Data Analysis. *The Astronomical Journal*, 122, pp. 3492-3505, <https://doi.org/10.1086/324109>
- Olmeda, I. and Fernández, E. (2000). Filtering With Wavelets May Be Worse Than You Think. *Computing in Economics and Finance 2000*, 150, Society for Computational Economics.
- Percival, D. B. and Walden, A. T. (2006). *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press.
- R Core Team (2015). R: A Language and Environment for Statistical Computing, [online]. *R Foundation for Statistical Computing*, Vienna, Austria. Available at: <https://www.R-project.org> [Accessed 13 Jan. 2016].
- Schlüter, S. and Deuschle, C. (2010). Using Wavelets for Time Series Forecasting: Does It Pay Off? *FAU Discussion Papers in Economics* 04/2010, Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics.
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), pp. 289-310, <https://doi.org/10.1214/10-ST5330>
- Tashman, L. J. (2000). Out-of-sample Tests of Forecasting Accuracy: An Analysis and Review. *International Journal of Forecasting*, 16(4), pp. 437-450, [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0)
- Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. 3rd ed. Oxford: Academic Press.