Forecasting with Auxiliary Information in Forecasts using Multivariate Singular Spectrum Analysis

Emmanuel Sirimal Silva[1], Hossein Hassani[2], Mansi Ghodsi[2], and Zara Ghodsi[3]

[1] *Fashion Business School, London College of Fashion, University of the Arts London*

[2] *Research Institute of Energy Management and Planning, University of Tehran, Tehran 1417466191, Iran*

[3] *Statistical Research Centre, Faculty of Management, Bournemouth University, UK*

November 29, 2018

Abstract

The internet gives us free access to a variety of published forecasts. Motivated by this increasing availability of data, we seek to determine whether there is a possibility of exploiting auxiliary information contained within a given forecast to generate a new and more accurate forecast. The proposed theoretical concept requires a multivariate model which can consider data with different series lengths as forecasts are predictions into the future. Following applications which consider published forecasts generated via unknown time series models and forecasts from univariate models, we achieve promising results whereby the proposed multivariate approach succeeds in extracting the auxiliary information in a given forecast for generating a new and more accurate forecast, along with statistically significant accuracy gains in certain cases. In addition, the impact of filtering and the use of Google Trends within the proposed methodology is also considered. Overall, we find conclusive evidence which suggests a sound opportunity to exploit the forecastability of auxiliary information contained within existing forecasts.

Keywords: Forecasting; auxiliary information; published forecasts; Google Trends; Multivariate Singular Spectrum Analysis.

1 Introduction

The highly volatile economic, social, political and environmental conditions across the globe have increased the prolific importance of accurate forecasts for planning and decision making. In response, a wide range of forecasts are published regularly via online platforms (see for example, Office for National Statistics in UK, U.S. Energy Information Administration and Central Banks across the globe), whilst the emergence of Big Data continues to provide new insights and opportunities for improving the accuracy of forecasts within a multivariate framework [19]. There is also evidence of researchers actively engaging in exploiting such freely available, publicly accessible data. For example, in [2], the authors exploit such auxiliary information to analyse the statistical profile of jobs whilst in [53] the authors note that advances in mobile devices and sensors have increased interest in multivariate time series modelling. Moreover, there is also a growing interest in developing new forecasting methods and procedures, see for example [9–11]. Motivated by such advances, we seek to show how auxiliary information in the form of forecasts can also be exploited further.

Prior to diving deeper into the gist of this research, it is important to note the difference between univariate and multivariate forecasting. Historically, univariate forecasting, the use of historical data for a given variable to obtain a forecast for that same variable, has been the most popular norm with a wide range of successful applications (see for example, [21, 47, 54]). However, in the recent past, emphasis has been placed on multivariate forecasting, which is the use of multiple variables to generate a forecast for a particular variable of interest (see for example, [38,44,49]). This distinction is important as the published forecasts which are used as examples in this paper are a result of complex multivariate approaches and are widely regarded as being highly accurate.

Given the increasing availability of forecasts, our interest lies in investigating the possibility of exploiting the auxiliary information in forecasts. That is, once any forecast has been generated, is there a possibility of extracting the auxiliary information contained within a given forecast in a multivariate framework, to create a new and improved forecast which outperforms the accuracy of the initial forecast? In other words, the proposed theoretical development suggests the use of historical data for a given variable with a forecast for that same variable in a multivariate framework (there is room for additional explanatory variables including the forecast) to produce a new forecast. In light of the information age and increasing availability of forecasts from different sources for the same object, methods for combining this information and delivering a superior forecast can be of obvious value.

It is hypothesized that, given a forecast is accurate to some extent, then a comparatively more accurate forecast can be generated via a multivariate modelling process which can exploit the auxiliary information already contained within the original forecast. Historically, many authors have successfully considered building forecasting models which consider time lags into the past (see, [8,27,33,46,50] for some recent examples). However, modelling a forecast invariably means that one must be able to model data which captures information into the future. Herein lies the complication as there exists no published research on combining and modelling historical data with forecasts into the future, and to the best of our knowledge there is no published research which seeks to exploit the forecastability of auxiliary information in forecasts.

This research considers forecasts from various models as benchmarks. The choice of benchmarks are justified as the tool used for exploiting the forecastability of auxiliary information in forecasts is of little or no use unless it can outperform the accuracy of existing forecasts based on some accuracy criterion. It is important to note that; the usual multivariate modelling problem involves using two different time series with equal series length and extracting any useful information for improving the accuracy of forecasts for both variables or one of the two variables. In contrast, in this research we consider historical data for the variable of interest, along with a forecast for that same variable, and seek to generate a new forecast which can outperform the accuracy of the original forecast. Therefore, not all multivariate forecasting techniques can exploit this new idea as not all of them are able to model time series of different lengths.

Accordingly, the theoretical concept proposed herewith requires a model which can meet two conditions:

1. It should be able to consider data with different series lengths.

2. It should be able to model forecasts, which are essentially data into the future.

As such, any model which meets these two criteria can be used to exploit the forecastability of auxiliary information in forecasts.

To the best of our knowledge, the Multivariate Singular Spectrum Analysis (MSSA) [17] technique is the only contender for this approach at present. This is because of its capability at modeling multiple time series with different series lengths, which indicates that it could model data into the future. In brief, the MSSA process filters the data and extracts signals which can be used to generate a new time series that is less noisy, and then uses this less noisy, reconstructed series for generating a forecast [40]. In comparison to its univariate counterpart, Singular Spectrum Analysis (SSA) [4, 5, 14], which has a variety of applications [1, 13, 21, 22, 26, 30–32, 34, 37, 41, 43, 51], MSSA is yet to be exploited on a similar scale. As a result there have been comparatively few applications of MSSA, see for example [6, 20, 25, 35, 39, 42, 44].

It is noteworthy that the MSSA technique also has the advantage of being nonparametric. As such, MSSA is not bound by the parametric assumptions of normality and stationarity of residuals, or linearity of the model [40]. This feature enables one to model the data without any transformations, which in turn helps ensure a comparatively more accurate approximation to the real situation with no loss of information [25].

At this stage, it is also pertinent to distinguish between the proposed theory for exploiting the forecastability of auxiliary information in forecasts and the long existing field of forecast combination in time series literature. For example, in [7] the authors consider combining competing forecasts to develop a new and improved forecast via variance-covariance based methods or regression based methods. In another example, the authors create a multivariate combined forecast by merging principal component regression method, the partial least squares regression method and the modified partial least squares regression method [52] whilst in [48] the authors considered a semi-heterogeneous approach to combining crude oil price forecasts. However, the proposed approach is completely different and novel as it combines historical data with a forecast and seeks to extract the auxiliary information contained within the forecast using a single multivariate model and produces a new and improved forecast.

It is not the intention of this paper to claim any methodological advances at this stage, and we subscribe to the view that it is possible to build more sophisticated forecasting models than those used here. However, we strongly believe that the approach described in this paper can serve as a foundation to aid forecasters in exploiting their own modelling and analytical skills to develop methodological advances refined for specific applications. In brief, the findings of our research provides substantial evidence supporting the applicability and feasibility of the proposed theoretical development along with statistically significant results in certain cases. The applications we present not only considers exploiting the forecastability of auxiliary information in published forecasts, but also forecasts from other time series analysis and forecasting models. Overall, the findings indicate that there exists a sound possibility for exploiting the forecastability of auxiliary information in forecasts.

The remainder of this paper is organized as follows. In Section 2 we introduce the theory underlying exploiting the forecastability of auxiliary information in forecasts. Within the theoretical framework we also introduce the MSSA forecasting algorithms used in this research. Section 3 is dedicated to applications and a concise discussion is presented in Section 4. The paper concludes in Section 5.

2 Theoretical Development

Assume that we have a time series $Y_N^{(1)}$ of length N, and further auxiliary information through a h-step ahead forecast for $Y_N^{(1)}$ contained in Ω. Thus, Ω is a $h \times 1$ vector of future information about $Y_N^{(1)}$. Note that $Y_N^{(1)}$ and Ω are time series with different series lengths as shown below. The data in Ω can represent any forecast for $Y_N^{(1)}$, attained using any approach (from experts, forecasting models, or news). Our hypothesis is that, provided the information contained in Ω is accurate to a certain degree, then we can utilize this information alongside historical information in $Y_N^{(1)}$ within a multivariate framework to develop an all new forecast for $Y_N^{(1)}$ which will outperform the accuracy of the forecast in Ω.

For explanation purposes, let us assume $Y_N^{(1)}$ is the actual data for a variable and Ω is the h-step ahead forecast for that same variable, such that:

$$Y_N^{(1)} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \; and \; \Omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_h \end{pmatrix}. \tag{1}$$

Then, a new $N + h \times 1$ vector can be constructed by incorporating the historical values with the forecasted values such that, $Y_{N+h}^{(2)} = (Y_N^{(1)}, \Omega)$:

$$Y_h^{(2)} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \\ \omega_1 \\ \vdots \\ \omega_h \end{pmatrix}. \tag{2}$$

From this point onwards, we suggest exploiting the MSSA technique [17]. MSSA can model time series with different series lengths to extract any auxiliary information contained within Ω and then use this information in combination with the historical data in $Y_N^{(1)}$, to produce a new forecast which can outperform the accuracy of the forecast in Ω.

The entire MSSA process is dependent on the choices of the Window Length L (an integer such that $2 \leq L \leq N/2$) and the number of eigenvalues r. Those interested in a discussion on the selection of these two choices are referred to [40]. The first stage in MSSA is called Decomposition, and the first step here is embedding which maps a one dimensional time series into a multidimensional time series [40]. We can define the trajectory matrices $\mathbf{X}^{(i)}$ $(i = 1, 2)$ of the one-dimensional time series $Y_{N_i}^{(i)} (i = 1, 2)$ with different series length. Thus, applying the above procedure to $Y_N^{(1)}$ and $Y_{N+h}^{(2)}$ separately provides 2 different $L_i \times K_i$ trajectory matrices $\mathbf{X}^{(i)}$ $(i = 1, 2)$, such that

$$\mathbf{X}^{(1)} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{pmatrix}, \tag{3}$$

where $K = N - L + 1$. However, our interest is in modelling $Y_N^{(1)}$ and $Y_{N+h}^{(2)}$. As such, we need to perform embedding over $Y_N^{(2)}$. Thus, $\mathbf{X}^{(2)} = [X_1, ..., X_K, \ldots, X_{K+h}]$,

$$\mathbf{X}^{(2)} = (x_{ij})_{i,j=1}^{L,K+h} = \begin{pmatrix} y_1 & y_2 & \cdots & y_K & y_{K+1} & \cdots & \omega_{K+h} \\ y_2 & y_3 & \cdots & y_{K+1} & y_{K+2} & \cdots & \omega_{K+h+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N & \omega_{N+1} & \cdots & \omega_{N+h} \end{pmatrix}. \tag{4}$$

Following the embedding process, we organise a new block Hankel matrix. Here, we use the MSSA approach in vertical form (VMSSA). However, there are some restrictions in the selection of the values of K and it is required that $K_1 = K_2 = K$ [17]. Accordingly, the VMSSA approach enables us to have various window length L_i and different series length N_i, but as we mentioned above, the same K_i for all series. Note also that if one wishes to use the horizontal form, similar L and different K should be used. Those interested in a detailed discussion are referred to Sanei and Hassani [40]. However, we find it pertinent to follow Hassani and Mahmoudvand [17] and present the reader with a summary of the similarities and differences between the VMSSA and HMSSA forecasting algorithms via Table 1.

Table 1: Similarities and dissimilarities between the VMSSA and HMSSA forecasting algorithms.

Method	Series Length	L_i	K_i	Number of λ_i	LRF
VMSSA-R	Different	Different	Equal	$\sum L_i$	Different
HMSSA-R	Different	Equal	Different	L	Equal

Then, the block Hankel trajectory matrix can then be defined as

$$\mathbf{X}_V = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}, \tag{5}$$

where, $\mathbf{X}_V$ indicates that the output of the embedding step is in *vertical* form. It is important to note that the dimensions in $\mathbf{X}_V$ are compatible because the VMSSA process allows one to have various L_i and different N_i whilst restricting the selection of values of K_i, which must be the same for all series. Next, we obtain the SVD of $\mathbf{X}_V$ which is the second step in MSSA. Denote $\lambda_{V_1}, \ldots, \lambda_{V_{L_{sum}}}$ as the eigenvalues of $\mathbf{X}_V \mathbf{X}_V^T$, arranged in decreasing order $\left(\lambda_{V_1} \geq \ldots \lambda_{V_{L_{sum}}} \geq 0\right)$ and $U_{V_1}, \ldots, U_{V_{L_{sum}}}$, the corresponding eigenvectors, where $L_{sum} = L_1 + L_2$. Note also that the structure of the matrix $\mathbf{X}_V \mathbf{X}_V^T$ is as follows:

$$\mathbf{X}_V \mathbf{X}_V^T = \begin{bmatrix} \mathbf{X}^{(1)}\mathbf{X}^{(1)T} & \mathbf{X}^{(1)}\mathbf{X}^{(2)T} \\ \mathbf{X}^{(2)}\mathbf{X}^{(1)T} & \mathbf{X}^{(2)}\mathbf{X}^{(2)T} \end{bmatrix}. \tag{6}$$

The structure of the matrix $\mathbf{X}_V \mathbf{X}_V^T$ is similar to the variance-covariance matrix in the classical multivariate statistical analysis literature. The matrix $\mathbf{X}^{(i)}\mathbf{X}^{(i)T}$, which is used in SSA, for the series $Y_{N_i}^{(i)}$, appears along the main diagonal and the products of two Hankel matrices $\mathbf{X}^{(i)}\mathbf{X}^{(j)T} (i \neq j)$, which are related to the series $Y_N^{(1)}$ and $Y_{N+h}^{(2)}$, appears in the off-diagonal. The SVD of $\mathbf{X}_V$ can be written as $\mathbf{X}_V = \mathbf{X}_{V_1} + \cdots + \mathbf{X}_{V_{L_{sum}}}$, where $\mathbf{X}_{V_i} = \sqrt{\lambda}_i U_{V_i} {V_{V_i}}^T$ and $V_{V_i} = \mathbf{X}_V^T U_{V_i} / \sqrt{\lambda}_{V_i}$ ($\mathbf{X}_{V_i} = 0$ if $\lambda_{V_i} = 0$). Also noteworthy is that, if we used SSA, whereby we consider information from one variable, then we have singular values in the form of $\lambda_i = \lambda_1, \lambda_2, \ldots, \lambda_L$ $(i = 1, \ldots, L)$. In contrast, with MSSA, the singular values capture additional information contained within the additional variable(s) included in the model, such that we obtain $\lambda_i^* = \lambda\omega_1, \lambda\omega_2, \ldots, \lambda\omega_L$, where $\lambda_i^* \neq \lambda_i + \lambda\omega_i$.

Having extracted the singular values, we move on to the second stage of MSSA entitled Reconstruction, which incorporates the steps of Grouping and Diagonal Averaging [17]. Grouping corresponds to splitting the matrices $\mathbf{X}_{V_1}, \ldots, \mathbf{X}_{V_{L_{sum}}}$ into several disjoint groups and summing the matrices within each group, such that the split of the set of indices $\{1, \ldots, L_{sum}\}$ into disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation $\mathbf{X}_v = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. The procedure of choosing the sets $I_1, \ldots, I_m$ is called grouping. The purpose of this step is to analyze the singular values and differentiate between signal and noise. On the other hand, Diagonal Averaging transforms the reconstructed matrix $\widehat{\mathbf{X}}_{V_i}$ into the form of a Hankel matrix, which can be subsequently converted to a time series for forecasting future data points. Those interested in a detailed description of these two steps are referred to [17]. In what follows, we briefly outline the VMSSA forecasting algorithms and in doing so we mainly follow [17].

2.1 VMSSA Recurrent Forecasting Algorithm (VMSSA-R)

Let us have two series with different length $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})$ and corresponding window length L_i, $1 < L_i < N_i, i = 1, 2$. The VMSSA-R forecasting algorithm for the h-step ahead forecast is as follows.

1. For a fixed value of K, construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L_i,K}$ for each single series $Y_{N_1}^{(1)}$, and $Y_{N_2}^{(2)}$ separately.

2. Construct the block trajectory matrix $\mathbf{X}_V$ as follows:

$$\mathbf{X}_V = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}. \tag{7}$$

3. Denote $\lambda_{V_1} \geq \ldots \geq \lambda_{V_{L_{sum}}} \geq 0$ are the eigenvalues of the $\mathbf{X}_V \mathbf{X}_V^T$, where $L_{sum} = L_1 + L_2$.

4. Let $\mathbf{U}_{V_j} = (U_j^{(1)}, U_j^{(2)})^T$ be the j^{th} eigenvector of the $\mathbf{X}_V \mathbf{X}_V^T$, where $U_j^{(i)}$ with length L_i corresponds to the series $Y_{N_i}^{(i)}$ $(i = 1, 2)$.

5. Consider $\widehat{\mathbf{X}}_V = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} U_{V_i} U_{V_i}^T \mathbf{X}_V$ as the reconstructed matrix achieved from r eigentriples:

$$\widehat{\mathbf{X}}_V = \begin{bmatrix} \widehat{\mathbf{X}}^{(1)} \\ \widehat{\mathbf{X}}^{(2)} \end{bmatrix}. \tag{8}$$

6. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathcal{H}\widehat{\mathbf{X}}^{(i)}$ $(i = 1, 2)$ as the result of the Hankelization procedure of the matrix $\widehat{\mathbf{X}}^{(i)}$ obtained from the previous step, where $\mathcal{H}$ is a Hankel operator.

7. Assume $U_j^{(i)\nabla}$ denotes the vector of the first $L_i - 1$ components of the vector $U_j^{(i)}$ and $\pi_j^{(i)}$ is the last component of the vector $U_j^{(i)}$ $(i = 1, 2)$.

8. Select the number of r eigentriples for the reconstruction stage that can also be used for forecasting purpose.

9. Define matrix $\mathbf{U}^{\nabla(1,2)} = \left(U_1^{\nabla(1,2)}, \ldots, U_r^{\nabla(1,2)}\right)$, where $U_j^{\nabla(1,2)}$ is as follows:

$$U_j^{\nabla(1,2)} = \begin{bmatrix} U_j^{(1)\nabla} \\ U_j^{(2)\nabla} \end{bmatrix}. \tag{9}$$

10. Define matrix $\mathbf{W}$ as follows:

$$\mathbf{W}=\begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \cdots & \pi_r^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \cdots & \pi_r^{(2)} \end{bmatrix}. \tag{10}$$

11. If the matrix $\left(\mathbf{I}_{2\times 2}-\mathbf{W}\mathbf{W}^T\right)^{-1}$ exists and $r \le L_{sum}-2$, then the h-step ahead VMSSA forecasts exist and is achieved by the following formula:

$$\left[\hat{y}_{j_1}^{(1)}, \hat{y}_{j_2}^{(2)}\right]^T = \begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \tilde{y}_{j_2}^{(2)}\right], & j_i=1,\ldots,N_i \\ \\ \left(\mathbf{I}_{2\times 2}-\mathbf{W}\mathbf{W}^T\right)^{-1}\mathbf{W}\mathbf{U}^{\nabla 2 T}\mathbf{Z}_h, & j_i=N_i+1,\ldots,N_i+h, \end{cases} \tag{11}$$

where, $\mathbf{Z}_h=\left[Z_h^{(1)}, Z_h^{(2)}\right]^T$ and $Z_h^{(i)}=\left[\hat{y}_{N_i-L_i+h+1}^{(i)},\ldots,\hat{y}_{N_i+h-1}^{(i)}\right]$ $(i=1,2)$. It should be noted that equation (11) indicates that the h-step ahead forecasts of the refined series $\hat{Y}_{N_i}^{(i)}$ are obtained by a multi dimensional linear recurrent formula (LRF). For the univariate case, there is only a one dimensional LRF.

2.2 VMSSA Vector Forecasting Algorithm (VMSSA-V)

Let us have items 1-10 of VMSSA-R. Consider the matrix:

$$\mathbf{\Pi}=\mathbf{U}^{\nabla}\mathbf{U}^{\nabla T}+\mathcal{R}\left(\mathbf{I}_{2\times 2}-\mathbf{W}\mathbf{W}^T\right)\mathcal{R}^T, \tag{12}$$

where, $\mathcal{R}=\mathbf{U}^{\nabla}\mathbf{W}^T\left(\mathbf{I}_{2\times 2}-\mathbf{W}\mathbf{W}^T\right)^{-1}$. Let $\mathbf{\Pi}=\left(\mathbf{\Pi}^{(1)},\mathbf{\Pi}^{(2)}\right)^T$ and $\mathcal{R}=\left(\mathcal{R}^{(1)},\mathcal{R}^{(2)}\right)^T$, where $\mathbf{\Pi}^{(i)}$ with dimension $(L_i-1)\times(L_{sum}-2)$ and $\mathcal{R}^{(i)}$ $(i=1,2)$ with length $L_{sum}-2$ correspond to the series $Y_{N_i}^{(i)}$.

Then, Theorem 1 in [17] indicates that the linear projection $\mathcal{P}^{(\nu)}:\mathfrak{L}_r\mapsto\mathbb{R}^{L_{sum}-2}$ by the following formula provides the continuation vectors for the multivariate V-forecasting.

$$\mathcal{P}^{(\nu)}Y=\begin{bmatrix} \mathbf{\Pi}^{(1)}Y_{\triangle} \\ \mathcal{R}^{(1)T}Y_{\triangle} \\ \mathbf{\Pi}^{(2)}Y_{\triangle} \\ \mathcal{R}^{(2)T}Y_{\triangle} \end{bmatrix}, Y\in\mathfrak{L}_r, \tag{13}$$

where, $Y_{\triangle}^T=\left(Y_{\triangle}^{(1)},Y_{\triangle}^{(2)}\right)$ such that $Y_{\triangle}^{(i)}(i=1,2)$ denotes the last L_i-1 entities of Y_i with length L_i. Using above notations, the following algorithm is proposed for calculating the VMSSA-V forecasts.

1. Define vectors Z_i as follows:

$$Z_i=\begin{cases} \widetilde{X}_i & \text{for } i=1,\ldots,k \\ \mathcal{P}^{(\nu)}Z_{i-1} & \text{for } i=k+1,\ldots,k+h+L_{\max}-1, \end{cases} \tag{14}$$

   where, $L_{\max}=\max\{L_1,L_2\}$.

2. Constructing the matrix $\mathbf{Z}=[Z_1:\ldots:Z_{K+h+L_{\max}-1}]$ and making its hankelization. Using this calculation we obtain $\hat{y}_1^{(i)},\ldots,\hat{y}_{N+h+L_{\max}}^{(i)}$ $(i=1,2)$.

3. The numbers $\hat{y}_{N_i+1}^{(i)},\ldots,\hat{y}_{N_i+h}^{(i)}$ $(i=1,2)$ form the h step ahead VMSSA-V forecasts.

3 Applications

In this section we consider applications of the proposed theoretical development to show its validity and feasibility in practice. In the real world, forecasters (for example, consider official forecasts) are usually interested in providing predictions for the coming year (i.e. 12 steps-ahead for monthly data and 4 steps-ahead for quarterly data). As such, we consider applications which provide out-of-sample forecasts for the next year. For example, if we are dealing with monthly data, the last 12 observations for which 12 forecasted data are available are set aside as the out-of-sample data and the remainder is used for training and testing the forecasting models. Where 12 observations are forecasted, this means the first forecasted data point is the $h = 1$ step-ahead forecast, the second forecasted data point is the $h = 2$ steps-ahead forecast and so on up until the final forecasted data point which represents the $h = 12$ steps-ahead forecast or the 12 months ahead value for a given variable.

3.1 Metrics

We use the Root Mean Squared Error (RMSE), which is one of the most frequently cited loss functions in forecasting literature [12, 23, 24, 45], and the Ratio of the RMSE (RRMSE) to evaluate and distinguish between the forecasts obtained via different models.

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right)^{\frac{1}{2}}, \tag{15}$$

where, Y_i is the actual value, $\hat{Y}_i$ refers to a forecast from a given model, and n is the number of the forecasts. On the other hand, the RRMSE enables one to quantify the percentage gain or loss made by one model in comparison to another model in terms of its forecast accuracy. Accordingly, if $\frac{MSSA}{SSA}$ is less than 1, then this shows that MSSA outperforms SSA by 1-$\frac{MSSA}{SSA}$ percent and vice versa.

3.2 Benchmarks

In addition to considering various published forecasts as benchmarks, this paper also considers model based forecasts from SSA [40], ARIMA [29], Exponential Smoothing [28], and Holt-Winters [36] as benchmarks for comparison against forecasts derived through the newly proposed VMSSA approach. It is noteworthy that the chosen model based benchmarks are frequently used in time series literature, and represent both parametric and nonparametric time series analysis and forecasting models.

The applications which follow considers different scenarios. Initially, we begin by considering the possibility of exploiting the forecastability of auxiliary information in published forecasts. This is followed by applications which consider exploiting the forecastability of auxiliary information in forecasts generated via other univariate time series analysis and forecasting models. All outcomes are evaluated for statistical significance using not only the modified Diebold-Mariano (DM) test in [15], but also the Hassani-Silva (HS) test [18] which is based on the Kolmogorov-Smirnov test and principles of stochastic dominance. It should be noted that as a result of a small number of observations for evaluating forecast accuracy, it is not entirely surprising that this research results in minimal statistically significant outcomes at present. Also, all references to VMSSA-R and VMSSA-V forecasting results show the RMSE or RRMSE obtained via the newly proposed theory for exploiting the forecastability of auxiliary information in forecasts.

Scenario 1: Published Forecasts as Additional Information

Considered as the first example is the possibility of exploiting the forecastability of auxiliary information in published forecasts. Here, we consider forecasts obtained via the U.S. Energy Information Administration (EIA)[1] for a variety of energy variables, and a forecast for inflation by a group of non-financial service providers which includes manufacturers, universities, forecasting firms, investment advisors, pure research firms and consulting firms. The EIA time series are shown in Figure 1. The last year is considered as out-of-sample data and the results from the forecasting exercise are reported in Tables 2 and 3.

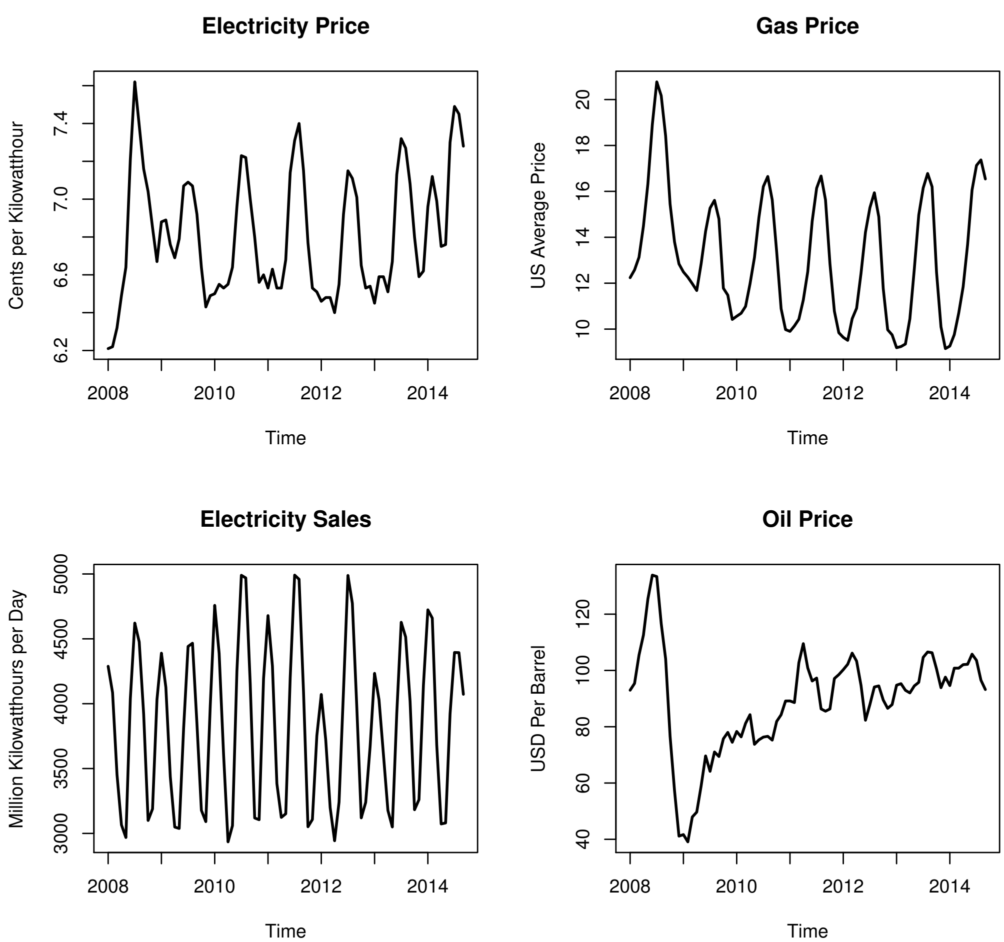

Figure 1: The EIA time series on energy used as examples.

The first application considers modelling the U.S. industrial sector average regional electricity prices. The RMSE values in Table 2 show that both VMSSA models are outperforming the EIA and SSA forecasts. We go a step further and calculate the RRMSE values and then compare the forecast errors from each model for statistically significant differences. As seen in Table 3, the RRMSE criterion indicates that both VMSSA forecasts are 4% better than the EIA forecast

[1]http://www.eia.gov/forecasts/steo/outlook.cfm

and 26% better than SSA-R and SSA-V forecasts. In this case, there is evidence of the newly proposed VMSSA approach outperforming SSA forecasts with statistically significant results and thus provides sound evidence for the validity of the proposed theory in practice.

The next application considers average residential natural gas prices in the United States. Again, based on the RMSE values, the VMSSA forecasts outperform the EIA and SSA forecasts. The VMSSA-V model reports the lowest RMSE (Table 2). The RRMSE values show that the VMSSA-V forecast is 6% better than the EIA forecast, and 9% better than the SSA-V forecast, whereas the VMSSA-R forecast is 5% better than the EIA forecast and 8% better than the SSA-R forecast. However, we find no evidence of statistically significant differences between any of the competing forecasts in this application.

Table 2: RMSE when using official forecasts for forecasting last year of each data set.

Series	OF	SSA-V	SSA-R	VMSSA-V	VMSSA-R
EIA					
Electricity Price	0.24	0.31	0.31	0.23	0.23
Gas Price	0.87	0.90	0.90	0.82	0.83
Electricity Sales	253.46	392.69	306.44	253.95	248.53
Oil Price	4.34	5.42	10.01	4.25	4.32
PF					
CPI	0.53	1.17	2.40	0.31	0.40

Note: OF: Official forecast. PF: Professional forecast. Electricity Price - Industrial Sector Average Regional Electricity Prices. Gas Price - Average Residential Natural Gas Price. Electricity Sales - Residential Sector Total Electricity Sales. Oil Price - West Texas Intermediate Spot Average Crude Oil Price. PF: Professional forecast from group of non-financial service providers. CPI - Consumer Price Index.

The third application looks at data on total electricity sales in the U.S. residential sector. The RMSE results in Table 2 show that the VMSSA-R model can provide the forecast with the lowest error, whilst the VMSSA-V forecast is on par with the EIA forecast. The RRMSE values in Table 3 indicates that the VMSSA-V forecast is 35% better than the SSA-V forecast whilst the VMSSA-R forecast is 2% better than the EIA forecast and 19% better than the SSA-R forecast. Once again, we find no evidence of statistically significant differences between the competing forecasts.

The fourth application looks at the oil price (West Texas Intermediary) series. The RMSE results in Table 2 show that the VMSSA forecast outperforms the EIA and SSA forecasts, along with the VMSSA-V model reporting the lowest RMSE. In terms of the RRMSE criterion, as reported in Table 3, the VMSSA-V forecast is 2% better than the official forecast, and 22% better than the SSA-V forecast. Likewise, the VMSSA-R forecasts are 1% better than the official forecast, and 57% better than the SSA-R forecast. However, when tested for statistically significant differences between the forecasts, evidence was only found for significant differences between the VMSSA-R and SSA-V forecasts. Figure 2 shows the out-of-sample forecasts related to the EIA applications.

The final application considers forecasting the last four quarters of the quarterly consumer price index growth rate series. Here, a forecast by professionals is used as more information and the resulting RMSE is reported in Table 2. Given that there are only four out-of-sample observations, it is not realistic to expect statistically significant differences between the forecasts in this case. However, the RRMSE results in Table 3 can provide a reasonable indication of the comparative performance. The RMSE shows that both VMSSA models outperform not only SSA, but also the professional forecast, whilst VMSSA-V reports the lowest RMSE. The

Table 3: RRMSE when using official forecasts for forecasting last year of each data set.

Series	$\frac{VMSSA-V}{OF}$	$\frac{VMSSA-R}{OF}$	$\frac{VMSSA-V}{SSA-V}$	$\frac{VMSSA-R}{SSA-R}$
EIA				
Electricity Price	0.96	0.96	0.74^{*}	$0.74^{*,\dagger}$
Gas Price	0.94	0.95	0.91	0.92
Electricity Sales	1.00	0.98	0.65	0.81
Oil Price	0.98	0.99	0.78	$0.43^{*,\dagger}$
PF				
CPI	0.58	0.75	0.26	0.17

Note: * indicates a statistically significant difference between the two forecasts based on the modified Diebold-Mariano test at $p = 0.10$. † indicates a statistically significant difference between the two forecasts based on the Hassani-Silva test at $p = 0.10$.

RRMSE criterion shows that the VMSSA-V forecast is 42% better than the professional forecast and 74% better than the SSA-V forecast. Likewise, the VMSSA-R forecast is 25% better than the professional forecast and 83% better than the SSA-R forecast. Figure 3 provides a graphical representation of the out-of-sample forecasts. It is evident that the VMSSA-V forecast is the only one which remains comparatively aligned with the actual inflation values.

Scenario 2: Forecasts from Other Forecasting Models

Here, we look at improving the accuracy of forecasts from a variety of other time series analysis models such as ARIMA, Exponential Smoothing and Holt-Winters. This is important as especially in government organizations, methods such as ARIMA and Holt-Winters are widely accepted and continue to be used owing to traditions and familiarity with such models. Figure 4 plots the time series used here as examples. Each time series has been obtained via the Datamarket[2].

These monthly time series include the popular U.S. accidental deaths time series, milk production, number of city births in New York over time and residential electricity usage in Iowa, U.S. It is clear via Figure 4 that the chosen series captures the effects of stationarity, non-stationarity, increasing trends, seasonality and structural breaks. In reality, we are likely to be faced with such varying time series and it is therefore important to consider such phenomenons as examples. In each case, the last 12 monthly observations are left aside as out-of-sample and the models are trained over the remainder of the observations. The results from the applications are presented in Table 4.

[2]https://datamarket.com/

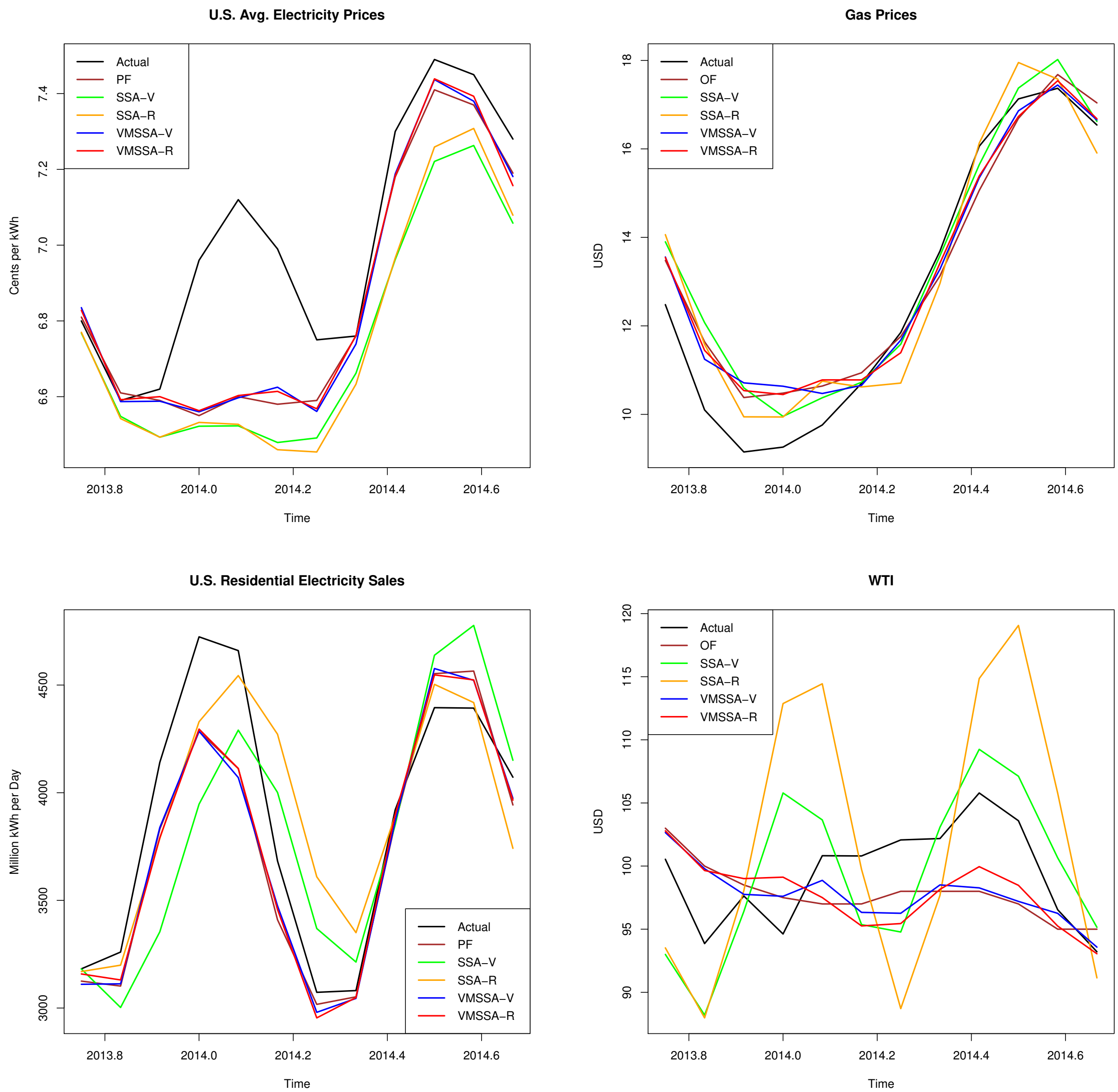

Figure 2: Out-of-sample forecasts for the last 12 months.

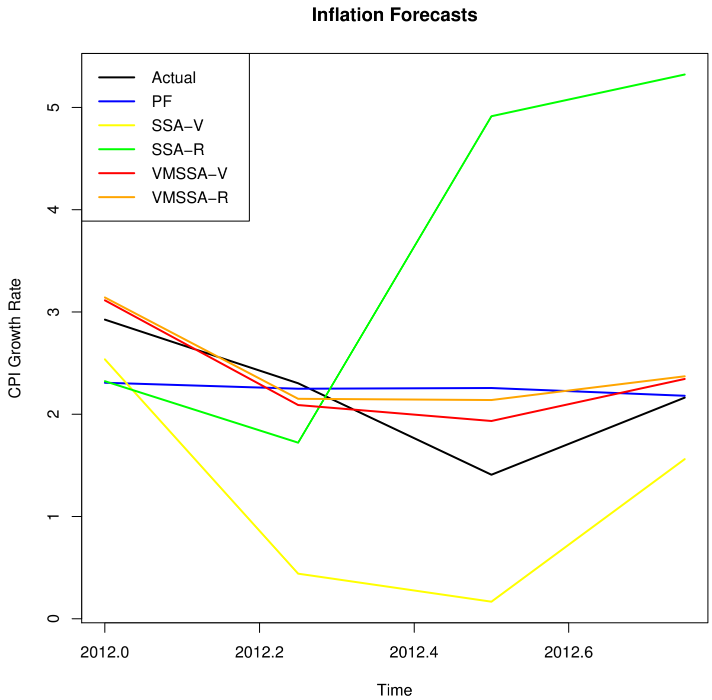

Figure 3: Out-of-sample forecasts for the last four quarters of the CPI.

The out-of-sample forecasting RMSE is obtained via ARIMA, HW, ETS, SSA-V, SSA-R, VMSSA-V and VMSSA-R for each data set. Note that when modelling with VMSSA, forecasts from the univariate model which reports the lowest in-sample forecasting RMSE for the training data is selected as more information in the VMSSA model to generate the multivariate out-of-sample forecasts.

The U.S. accidental deaths series has been widely adopted in time series literature, see for example [16] and [3]. For the death series, ARIMA provided the lowest in-sample forecasting RMSE and therefore the out-of-sample forecasts from ARIMA were selected as the additional information for the VMSSA model. It is important to evaluate whether the newly proposed VMSSA approach can result in forecasts which not only outperform the accuracy of the best forecast, but also forecasts from its univariate counterpart SSA. Based on the RMSE criterion and results in Table 4, it is evident that VMSSA can provide forecasts with a lower RMSE in comparison to all other models for this series, and VMSSA-V in particular reports the lowest RMSE. However, these forecasting differences can be attributed to chance occurrences. In order to evaluate if the VMSSA forecast is significantly better, all outcomes are tested for statistically significant differences with the results reported in Table 5 along with the RRMSE.

Table 4: RMSE for forecasting last year of each data set.

Series	ARIMA	HW	ETS	SSA-V	SSA-R	VMSSA-V	VMSSA-R
Death	332	432	338	736	624	312	327
Milk Prod.	14.1	14.8	8.63	19.5	13.7	7.28	7.69
NY Births	0.91	1.06	1.13	1.38	1.46	0.85	0.88
Elec. Use	51.6	78.1	39.9	57.3	53.7	38.73	36.40

Note: Forecasts from the univariate model providing the lowest in-sample forecasting RMSE are selected as the additional information for the MSSA model.

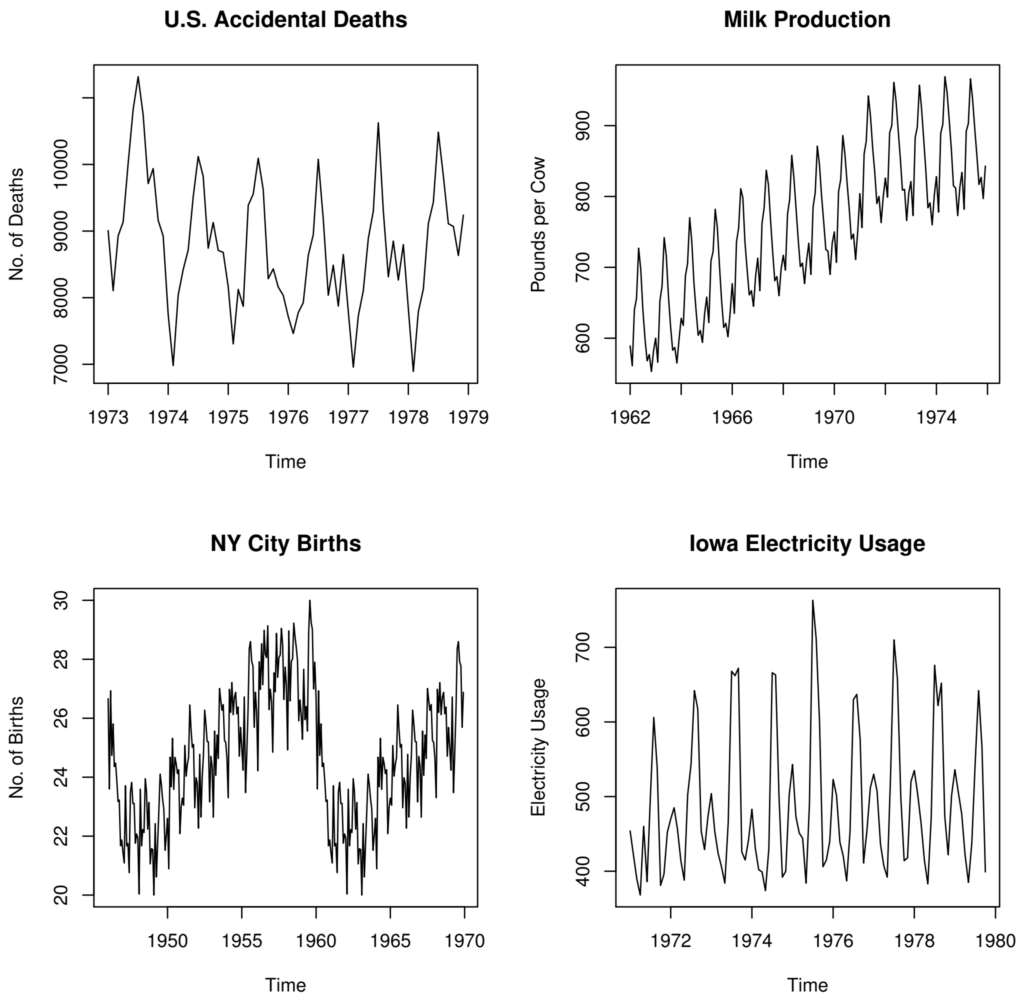

Figure 4: Four time series used as examples.

Based on the RRMSE, the VMSSA-V forecasts are 6% better than ARIMA, 28% better than HW, 8% better than ETS, and 58% better than SSA-V forecasts. Likewise, VMSSA-R forecasts are 2% better than ARIMA, 24% better than HW, 3% better than ETS, and 48% better than SSA-R forecasts. In this case, based on both DM and HS tests, we find evidence of statistically significant differences between the forecasts of VMSSA-V and SSA-V, and VMSSA-R and SSA-R at a 10% significance level. However, we do not find similar evidence in relation to the other models. Yet, the fact that VMSSA forecasts are significantly better than the SSA forecasts indicate that the proposed approach is viable.

Next, we consider the monthly milk production series. In this case, ETS forecasts were found to be best in-sample and were selected as the additional information for the VMSSA model. Once again, based on the RMSE in Table 4, it is clear that the VMSSA forecasts can outperform the rest of the models considered here. Likewise, based on the RRMSE criterion, we can conclude that VMSSA-V forecasts are 48% better than ARIMA, 51% better than HW, 16% better than ETS, and 63% better than SSA-V forecasts (Table 5). Moreover, VMSSA-R forecasts are 45% better than ARIMA, 48% better than HW, 11% better than ETS, and 44% better than SSA-R forecasts (Table 5). Interestingly, in relation to the previous application,

Table 5: RRMSE for forecasting last year of each data set.

Series	$\frac{VMSSA-V}{ARIMA}$	$\frac{VMSSA-R}{ARIMA}$	$\frac{VMSSA-V}{HW}$	$\frac{VMSSA-R}{HW}$	$\frac{VMSSA-V}{ETS}$	$\frac{VMSSA-R}{ETS}$	$\frac{VMSSA-V}{SSA-V}$	$\frac{VMSSA-R}{SSA-R}$
Death	0.94	0.98	0.72	0.76	0.92	0.97	$0.42^{*,\dagger}$	$0.52^{*,\dagger}$
Milk Prod.	$0.52^{*,\dagger}$	0.55^{*}	$0.49^{*,\dagger}$	$0.52^{*,\dagger}$	0.84	0.89	$0.37^{*,\dagger}$	0.56^{*}
NY Births	0.93	0.97	0.80	0.83	0.75	0.76	0.62	0.60
Elec. Use	0.75	0.71	$0.50^{*,\dagger}$	$0.47^{*,\dagger}$	0.97	$0.91^{\dagger}$	$0.68^{*,\dagger}$	$0.68^{*,\dagger}$

Note: * indicates a statistically significant difference between the two forecasts based on the modified Diebold-Mariano test at $p = 0.10$. † indicates a statistically significant difference between the two forecasts based on the Hassani-Silva test at $p = 0.10$.

there are a higher number of statistically significant outcomes in this case. In fact, VMSSA forecasts via the proposed approach are significantly better than ARIMA, ETS, SSA-V and SSA-R forecasts.

The third application considers monthly city births in New York. In this instance, ARIMA provided the best in-sample forecast and was therefore selected as the model which will provide more information for the VMSSA process. Table 4 shows that VMSSA forecasts once again outperforms all models based on the RMSE, and that the VMSSA-V forecast records the lowest RMSE. The RRMSE values in Table 5 indicates that VMSSA-V forecasts are 7%, 20%, 25% and 38% better than ARIMA, HW, ETS and SSA-V forecasts respectively, whilst VMSSA-R forecasts are 3%, 7%, 24% and 40% better than ARIMA, HW, ETS and SSA-V forecasts, respectively. Regardless of the gains suggested via the RRMSE criterion, there is no sufficient evidence of statistically significant differences between VMSSA and competing forecasts in this case. Given the comparatively large gains reported here, the inability of the statistical tests at picking up significant differences could be a result of small sample sizes.

The final application relating to the use of forecasts from other models as more information, considers monthly average residential electricity usage in Iowa. ETS provided the best in-sample forecast for this series and therefore its out-of-sample forecast was considered as more information within the VMSSA framework. As reported in Table 4, once again the VMSSA models outperform the rest based on the RMSE criterion. The RRMSE indicates that the VMSSA-V forecast reports gains of 25%, 50%, 3% and 32% in relation to forecasts from ARIMA, HW, ETS and SSA-V, respectively. At the same time, the VMSSA-R forecast reports gains of 29%, 53%, 9%, 32% in relation to the forecasts from ARIMA, HW, ETS and SSA-R, respectively. The tests for statistical significance indicates there exists significant differences between VMSSA and HW forecasts, and VMSSA and SSA forecasts. In addition, there is a statistically significant difference between the VMSSA-R and ETS forecasts. Figure 5 provides a graphical representation of the out-of-sample forecasts from all four applications.

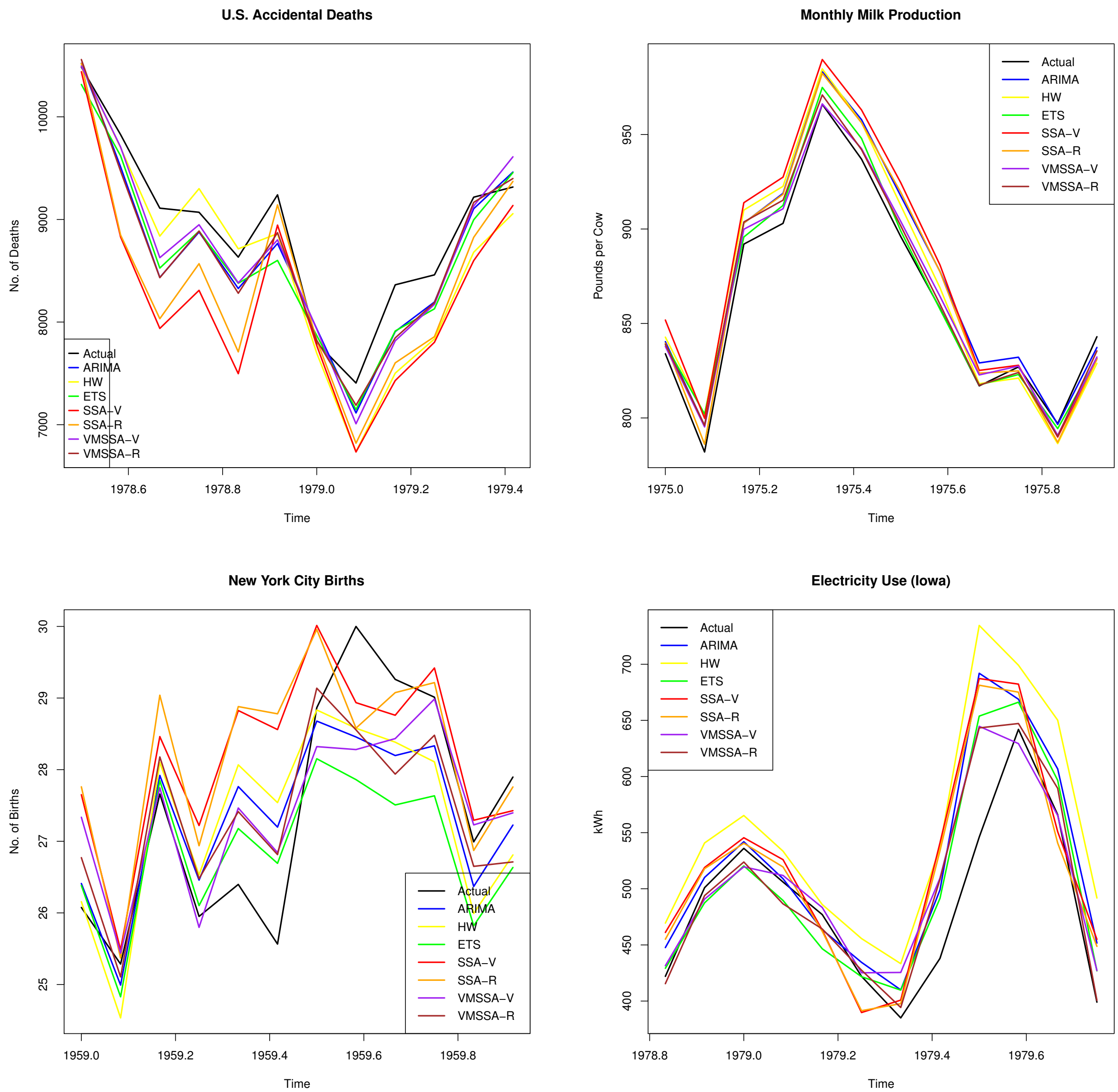

Figure 5: Out-of-sample forecasts for the last 12 months.

4 Discussion

4.1 Noise Reduction and Forecastability of Auxiliary Information in Forecasts

If we consider the results generated via the exploitation of auxiliary information contained within published forecasts (Table 3) and purely model based forecasts (Table 5), these show very different outcomes. Note how the RRMSE's in Table 5 show comparatively larger gains when purely model based forecasts are considered, in relation to the RRMSE's in Table 3 where the EIA's forecasts are used as auxiliary information.

This discrepancy suggests two possibilities. Firstly, it shows that published forecasts are likely to be considerably more accurate than univariate model based forecasts, which can be one explanation as to why the proposed methodology for exploiting the forecastability of auxiliary information in forecasts cannot provide significantly large gains when published forecasts are taken into consideration. In other words, this implies that there is comparatively less auxiliary information contained within published forecasts in relation to purely model based forecasts, i.e. the errors of published forecasts are likely to be much smaller than the errors from a model based forecast. Secondly, if there is comparatively less auxiliary information in a published forecast, then the gains made by MSSA in exploiting the forecastability of auxiliary information in forecasts can be a result of filtering which leads to noise reduction.

In order to evaluate whether noise reduction is likely to be an explanation, we consider the weighted correlation (w-correlation) statistic which is a measure used to determine separability between signal and noise [40]. According to [16], the w-correlation statistic can be computed as follows for two series $Y_N^{(1)}$ and $Y_N^{(2)}$, where $Y_N^{(1)}$ represents the signal and $Y_N^{(2)}$ represents the noise:

$$\rho_{12}^{(w)} = \frac{\left(Y_N^{(1)}, Y_N^{(2)}\right)_w}{\parallel Y_N^{(1)} \parallel_w \parallel Y_N^{(2)} \parallel_w},$$

where $Y_N^{(1)}$ and $Y_N^{(2)}$ are two time series, $\parallel Y_N^{(i)} \parallel_w = \sqrt{\left(Y_N^{(i)}, Y_N^{(i)}\right)_w}$, $\left(Y_N^{(i)}, Y_N^{(j)}\right)_w = \sum_{k=1}^{N} w_k y_k^{(i)} y_k^{(j)} \quad (i, j = 1, 2)$, w_k=min$\{k, L, N-k\}$ (here, assume $L \leq N/2$).

If the w-correlation between two reconstructed components are close to zero, this implies that the corresponding series are w-orthogonal, and suggests that the two components are well separable. Table 6 below reports the w-correlations following MSSA filtering in the case of published forecasts. As all w-correlations are considerably small and close to zero, it indicates that the MSSA decompositions have led to a sound separation between signal and noise. These results show that even when a forecast is of a very high quality (such that there is likely to be a very small error when compared with actual data), there is still room for exploiting the forecastability of auxiliary information in forecasts, if the tool being used has filtering capabilities.

Table 6: The w-correlation between signal and residuals for the EIA data.

Series	VMSSA-V	VMSSA-R
Electricity Price	0.001	0.001
Gas Price	0.004	0.004
Electricity Sales	0.006	0.006
Oil Price	0.010	0.009

4.2 A Hybrid Model for the Forecastability of Auxiliary Information in Forecasts

Up to this point, we have considered published forecasts and purely model based forecasts as auxiliary information for exploiting the forecastability of auxiliary information in forecasts. However, it is noteworthy that the proposed theoretical development enables the generation of hybrid forecasts too. This is because, in addition to considering published forecasts as auxiliary information alongside historical data for a given variable, there is also scope to consider trends based on news, expert opinions and judgements, and even other indicators as additional auxiliary information. Therefore, we find it pertinent to discuss whether the inclusion of additional indicators and trends, alongside the proposed theory can help improve the accuracy of forecasts further, beyond the current levels.

4.2.1 Indicators as Auxiliary Information

As the first example, we consider re-modelling the EIA Electricity Sales series. However, in this instance, we include the historical data for Gas Prices from the EIA as an indicator, alongside the initial framework, such that a total of three variables are considered in the multivariate model. Given that the model which follows has additional independent variables, we would expect a natural improvement in the forecast performance. The resulting output from the forecasting exercise is reported via Table 7.

Table 7: RRMSE for exploiting the forecastability of EIA's Electricity Sales forecast alongside historical Gas Price data.

Series	$\frac{VMSSA-V*}{OF}$	$\frac{VMSSA-R*}{OF}$	$\frac{VMSSA-V*}{VMSSA-V}$	$\frac{VMSSA-R*}{VMSSA-R}$
EIA				
Electricity Sales	0.79	0.73	0.79	0.74

Note: VMSSA-V* and VMSSA-R* indicates the multivariate model with Gas Price as an indicator variable.

Recall the results for this same variable in Table 3. There, the initial VMSSA-V model reported no gain or loss in relation to the EIA's official forecast for the Electricity Sales, whilst the VMSSA-R model was only able to report a small 2% gain over the EIA forecast. In contrast, by incorporating an indicator variable in the form of historical data for Gas Price into the multivariate model, and modelling this data alongside the historical data and the EIA forecast for Electricity Sales has resulted in considerable gains as per the RRMSE. In fact, the newly built VMSSA-V* model reports forecasts which are 21% better than the official forecast whilst the VMSSA-R* model reports forecasts which are 27% better than the official forecast. In addition, the newly built VMSSA-V* and VMSSA-R* models record forecasts which are 21% and 26% better than the previous VMSSA-V and VMSSA-R forecasts, respectively. However, this exercise did not yield evidence for statistically significant differences between the forecasts based on both DM and HS tests. Figure 6 shows the out-of-sample forecasts generated via the various models. Notice clearly how the VMSSA-V* and VMSSA-R* models provide much better forecasts for this variable over majority of the 12 month period in relation to both the EIA forecast and the previous multivariate forecasts generated through this study.

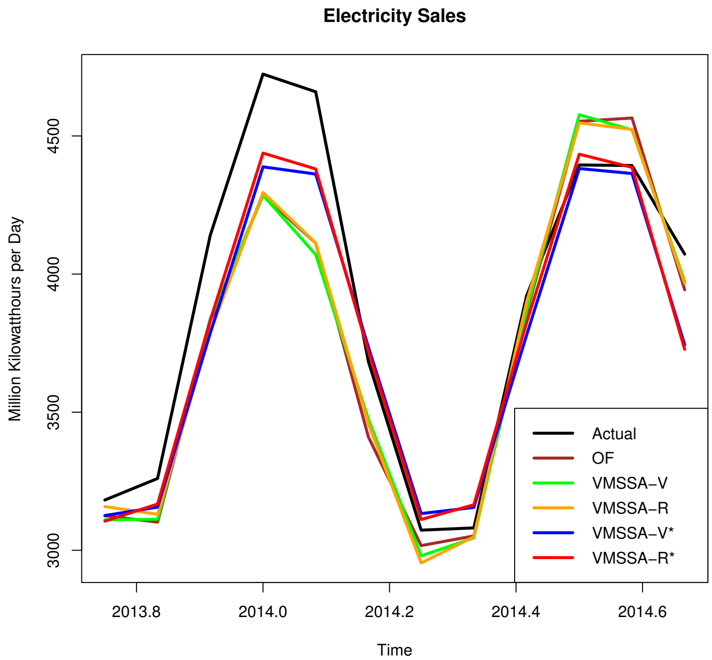

Figure 6: Out-of-sample forecasts for the last 12 months of the Electricity Sales series.

4.2.2 Google Trends as Auxiliary Information

Next, we consider Google Trends as auxiliary information for building a hybrid forecast for the Average Residential Natural Gas Price. Google Trends (`https://www.google.com/trends`) are freely accessible historical information pertaining to Google searches. Here, we consider the Google Trend for the search term 'Natural Gas' as auxiliary information alongside the EIA forecast for Gas Price to generate a hybrid forecast.

Figure 7 shows the actual data for Gas Price, the Google Trends for Natural Gas and the SSA trend for both data sets. Based on this figure, it is evident that both the actual data and Google Trends show seasonality and that the addition of the SSA trend to these graphs are able to show a declining trend over time for both data sets. Recall the results from the earlier forecasting exercise for the variable Gas Price as reported in Table 3. There, the VMSSA-V forecast was 6% better than the official forecast whilst the VMSSA-R forecast was 5% better. The output from the hybrid forecasting exercise is reported via Table 8. In this case, we notice that the inclusion of Google Trends within the proposed theoretical framework has improved the forecast accuracy such that the VMSSA-V* forecast is 24% better than the official EIA forecast whilst the VMSSA-R* forecast is 22% better than the EIA forecast. In comparison to the original MSSA models, the VMSSA-V* forecast has improved by 19% whilst the VMSSA-R forecast has improved by 18%. Once again, we do not find evidence of statistically significant differences between the forecasts.

Finally, Figure 8 provides a graphical representation of the out-of-sample forecasts generated via the various models for Gas Price. It is clear that the inclusion of Google Trends and generation of a hybrid forecast has improved the accuracy of the forecast attainable via the proposed model for exploiting the forecastability of auxiliary information in forecasts.

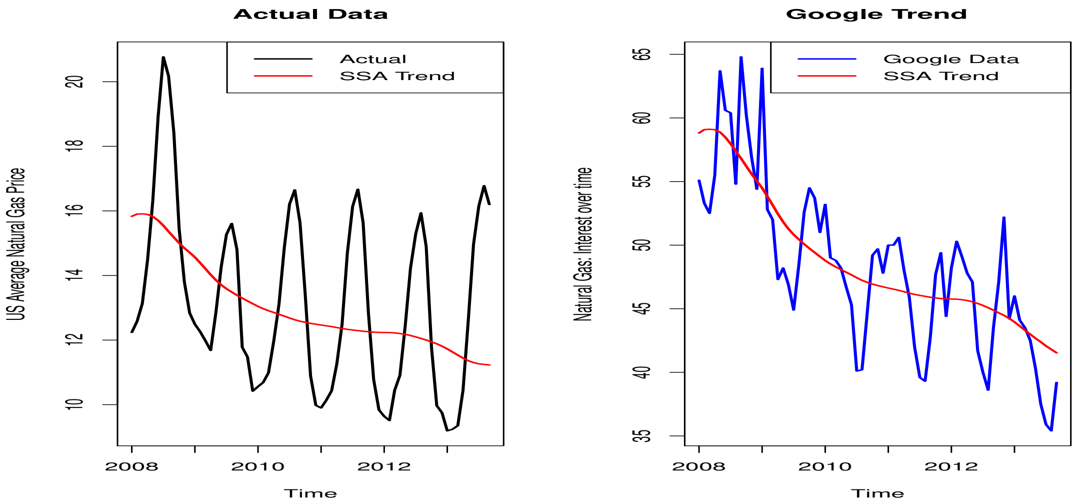

Figure 7: Historical data, Google Trend and SSA Trend for Natural Gas.

Table 8: RRMSE for exploiting the forecastability of EIA's NGRCUUS forecast alongside Google Trends.

Series	$\frac{VMSSA-V*}{OF}$	$\frac{VMSSA-R*}{OF}$	$\frac{VMSSA-V*}{VMSSA-V}$	$\frac{VMSSA-R*}{VMSSA-R}$
EIA				
Gas Price	0.76	0.78	0.81	0.82

Note: VMSSA-V* and VMSSA-R* indicates the multivariate model with Google Trends as auxiliary information.

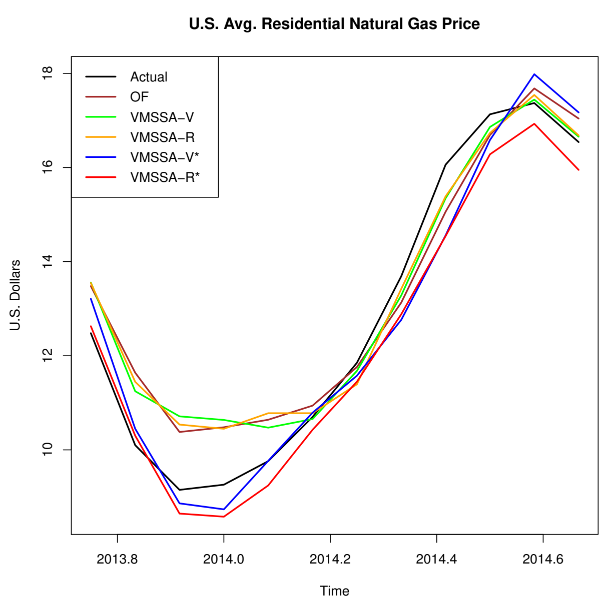

Figure 8: Out-of-sample forecasts for the Gas Price.

5 Conclusion

Historically, the Granger idea used time lags into the past for causality detection. In contrast, this paper begins with the aim of introducing a theoretical framework for exploiting the forecastability of auxiliary information in forecasts. That is, it seeks to answer the question as to once a forecast is generated, is there any possibility of exploiting the information contained within a given forecast for generating a new and more accurate forecast? Answering this question requires a tool which can consider modelling multiple series with different series lengths, as a forecast represents data into the future. Whilst any model which meets the above criteria can be used, here we use the MSSA technique as a tool for achieving the objective of this research.

The proposed methodology seeks to exploit forecasts and couples this information with historical data pertaining to the same variable in order to generate a new and improved forecast. The only condition is that the original forecast must have some level of good accuracy as otherwise there would not be any useful auxiliary information that can be extracted from the forecast. Following the theoretical development it is applied to several real world applications. It is noteworthy that the proposed theory can be used with any time series analysis and forecasting technique that can handle multiple time series with different series lengths. The MSSA model we have chosen for this purpose is one such example of a time series analysis and forecasting technique which can handle data with different series lengths as explained to the reader in Section 2. The results indicate that VMSSA forecasts which exploit the proposed theory are able to outperform its univariate counterpart, SSA in all instances (with statistically significant results in some cases). Moreover, there has always been at least one VMSSA model which can outperform published or purely model based forecasts in all cases based on the RMSE criterion. The low number of out-of-sample forecasts available for comparison purposes makes it an arduous task for the statistical tests to pick up significant differences. However, the RRMSE criterion is able to show that in certain cases the VMSSA models report gains of well over 20% in relation to a competing forecast.

The discussion which has been presented enlightens the reader on the positive impact of noise reduction on exploiting the forecastability of auxiliary information in forecasts. In addition, the discussion also covers the inclusion of additional indicators and information for generating hybrid forecasts combined with the theory for exploiting the forecastability of auxiliary information in forecasts. The two example hybrid forecasts which consider an indicator and Google Trends have illustrated how the accuracy of published forecasts can be greatly improved via the inclusion of such variables in the modelling process.

In conclusion, through this paper we have presented a novel theoretical development and shown that there is indeed an opportunity to exploit the forecastability of auxiliary information in forecasts for improving the accuracy of existing forecasts. The introductory nature of this theoretical concept opens up new research avenues with specific interests for the future. For example, given the findings of this study, further research and development should focus on developing linear models with new adaptations to enable other models to exploit the approach proposed in this paper. Moreover, future research should also consider using this development alongside SSA change point detection and with automated algorithms which will promote the effective use of this new theoretical development. For example, firstly, an automated algorithm should be developed for extracting the VMSSA parameters for a given data set. Thereafter, extensive simulation studies which takes into account different noise levels, stationarity and non-stationarity amongst other time series features should be carried out to provide more justification for this theory. It is expected that this new theoretical development will be of utmost importance whilst motivating researchers at Central Banks, Governments and Professional Forecasters across the globe.

References

[1] Abdollahzade, M., Miranian, A., Hassani, H., and Iranmanesh, H. (2015). A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting. *Information Sciences*, **295**, 107–125.

[2] Alam, M., Shakil, K. A., and Sethi, S. (2016). Analysis and clustering of workload in google cluster trace based on resource usage. *In:* Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) and 15th International Symposium on Distributed Computing and Applications for Business Engineering (DCABES), 2016 IEEE Intl Conference on (pp. 740-747). IEEE.

[3] Brockwell, P. J., and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. 2nd Edition. Springer.

[4] Broomhead, D. S., and King, G. P. (1986). On the Qualitative Analysis of Experimental Dynamical Systems. *Nonlinear Phenomena and Chaos*, ed. S. Sarkar, Adam Hilger, Bristol.

[5] Broomhead, D. S., and King, G. P. (1986). Extracting Qualitative Dynamics from Experimental Data. *Physica D: Nonlinear Phenomena*, **20**(2-3), 217–236.

[6] de Carvalho, M., and Rua, A. (2017). Real-time nowcasting the US output gap: Singular spectrum analysis at work. *International Journal of Forecasting*, **33**(1), 185–198.

[7] Diebold, F. X., and Lopez, J. A. (1996). 8 Forecast Evaluation and Combination. In: *Handbook of Statistics*, eds. Maddala, Statistical Methods in Finance, **14**, 241–268.

[8] Donate, J. P., and Cortez, P. (2014). Evolutionary optimization of sparsely connected and time-lagged neural networks for time series forecasting. *Applied Soft Computing*, **23**, 432–443.

[9] Efendi, R., Arbaiy, N., and Deris, M. M. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. *Information Sciences*, **441**, 113–132.

[10] Furong, Y., Liming, Z., Defu, Z., Hamido, F., and Zhiguo, G. (2016). A novel forecasting method based on multi-order fuzzy time series and technical analysis. *Information Sciences*, **367-368**, 41–57.

[11] Galicia, A., Torres, J. F., Martnez-Àlvarez, F., and Troncos, A. (2018). A novel spark-based multi-step forecasting algorithm for big data time series. *Information Sciences*, **476**, 800–818.

[12] Ghodsi, Z., Hassani, H., Kalantari, M., and Silva, E. S. (2018). Estimation of protein diffusion parameters. *Stat*, **7**(1), e192.

[13] Ghodsi, M., Hassani, H., Rahmani, D., and Silva, E. S. (2018). Vector and recurrent singular spectrum analysis: which is better at forecasting? *Journal of Applied Statistics*, **45**(10), 1872–1899.

[14] Golyandina, N., Nekrutkin, V. and Zhiglovsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman and Hall/CRC.

[15] Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). Newbold, P. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, **13**(2), 281–291.

[16] Hassani, H. (2007). Singular Spectrum Analysis: Methodology and comparison. *Journal of Data Science*, **5**(2), 239–257.

[17] Hassani, H. and Mahmoudvand, R. (2013). Multivariate Singular Spectrum Analysis: A general view and new vector forecasting approach. *International Journal of Energy and Statistics*, **1**(1), 55–83.

[18] Hassani, H., and Silva, E. S. (2015). A Kolmogorov-Smirnov Based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts. *Econometrics*, **3**(3), 590–609.

[19] Hassani, H., and Silva, E. S. (2015). Forecasting with Big Data: A Review. *Annals of Data Science*, **2**(1), 5–19.

[20] Hassani, H., and Silva, E. S. (2018). Forecasting UK consumer price inflation using inflation forecasts. *Research in Economics*, **72**, 367–378.

[21] Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., and Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, **63**, 112–127.

[22] Hassani, H., Silva, E. S., and Ghodsi, Z. (2017). Optimizing bicoid signal extraction. *Mathematical Biosciences*, **294**, 46–56.

[23] Hassani, H., Silva, E. S., Gupta, R., and Das, S. (2018). Predicting global temperature anomaly: A definitive investigation using an ensemble of twelve competing forecasting models. *Physica A*, **509**, 121–139.

[24] Hassani, H., Silva, E. S., Gupta, R., and Segnon, M. K. (2015). Forecasting the price of gold. *Applied Economics*, **47**(39), 4141–4152.

[25] Hassani, H., Soofi, A. S., and Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society A*, **176**(3), 743–760.

[26] Hassani, H., Webster, A., Silva, E. S., and Heravi, S. (2015). Forecasting U.S. Tourist Arrivals using Optimal Singular Spectrum Analysis. *Tourism Management*, **46**, 322–335.

[27] Herrera, A. M., Hu, L., and Pastor, D. (2018). Forecasting crude oil price volatility. *International Journal of Forecasting*, **34**(4), 622–635.

[28] Hyndman, R. J., and Athanasopoulos, G. (2013). *Forecasting: Principles and Practice.* OTexts, Australia. Available via: `https://otexts.org/fpp2/` [Accessed: 12.11.2018].

[29] Hyndman, R. J., and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, **27**(3), 1–22.

[30] Khan, M. A. R., and Poskitt, D. S. (2017). Forecasting stochastic processes using singular spectrum analysis: Aspects of the theory and application. *International Journal of Forecasting*, **33**(1), 199–213.

[31] Lahmiri, S. (2017). Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. *Applied Mathematics and Computation*, **320**, 444–451.

[32] Lai, L., and Guo, K. (2017). The performance of one belt and one road exchange rate: Based on improved singular spectrum analysis. *Physica A: Statistical Mechanics and its Applications*, **483**, 299–308.

[33] Li, L-L., Sun, J., Wang, C-H., Zhou, Y-T., and Lin, K-P. (2018). Enhanced Gaussian process mixture model for short-term electric load forecasting. *Information Sciences*, **477**, 386–398.

[34] Ma, X., Jin, Y., and Dong, Q. (2017). A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Applied Soft Computing*, **54**, 296–312.

[35] Mahmoudvand, R., Konstantinides, D., and Rodrigues, P. C. (2017). Forecasting mortality rate by multivariate singular spectrum analysis. *Applied Stochastic Models in Business and Industry*, **33**(6), 717–732.

[36] Meyer, D. (no date). Holt-Winters Filtering. Available via: `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/HoltWinters.html` [Accessed: 12.11.2018].

[37] Plaza, E. G., and Lopez, P. J. N. (2017). Surface roughness monitoring by singular spectrum analysis of vibration signals. *Mechanical Systems and Signal Processing*, **84**(Part A), 516–530.

[38] Pravilovic, S., Bilancia, M., Appice, A., and Malerba, D. (2017). Using multiple time series analysis for geosensor data forecastin. *Information Sciences*, **380**, 31–52.

[39] Rodrigues, P. C., and Mahmoudvand, R. (2018). The benefits of multivariate singular spectrum analysis over the univariate version. *Journal of the Franklin Institute*, **355**(1), 544–564.

[40] Sanei, S., and Hassani, H. (2015). *Singular Spectrum Analysis of Biomedical Signals*, CRC Press: United States.

[41] Shen, Y., Guo, J., Liu, X., Kong, Q., Guo, L., and Li, W. (2017). Long-term prediction of polar motion using a combined SSA and ARMA model. *Journal of Geodesy*, **92**(3), 333–343.

[42] Silva, E. S., Ghodsi, Z., Ghodsi, M., Heravi, S., and Hassani, H. (2017). Cross country relations in European tourist arrivals. *Annals of Tourism Research*, **63**, 151–168.

[43] Silva, E. S. and Hassani, H. (2015). On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession. *International Economics*, **141**, 34–49.

[44] Silva, E. S., Hassani, H., and Heravi, S. (2018). Modeling European industrial production with multivariate singular spectrum analysis: A cross-industry analysis. *Journal of Forecasting*, **37**(3), 371–384.

[45] Silva, E. S., Hassani, H., Otero, J., and Beneki, C. (2018). Forecasting Inflation Under Varying Frequencies. *Electronic Journal of Applied Statistical Analysis*, **11**(1), 307–339.

[46] Smeekes, S., and Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting*, **34**(3), 408–430.

[47] Wichern, D. W. (1973). Modeling and forecasting discrete univariate time series with applications. *Information Sciences*, **6**, 247-264.

[48] Wang, J., Li., X., Hong, T., and Wang, S. (2018). A semi-heterogeneous approach to combining crude oil price forecasts. *Information Sciences*, **460-461**, 279–292.

[49] Xu, M., Han, M., and Lin, H. (2019). Wavelet-denoising multiple echo state networks for multivariate time series prediction. *Information Science*, **465**, 439–458.

[50] Xuefan, D., Ying, L., and Yijun, L. (2018). Small and multi-peak nonlinear time series forecasting using a hybrid back propagation neural network. *Information Sciences*, **424**, 39–54.

[51] Yi, C., Lv, Y., Dang, Z., Xiao, H., and Yu, X. (2017). Quaternion singular spectrum analysis using convex optimization and its application to fault diagnosis of rolling bearing. *Measurement*, **103**(2), 321–332.

[52] Yin, S., Liu, L., and Hou, J. (2016). A multivariate statistical combination forecasting method for product quality evaluation. *Information Sciences*, **355-356**, 229-236.

[53] Yu, C., Luo, L., Chan, L. L-H., Rakthanmanon, T., and Nutanong, S. (2019). Wavelet-denoising multiple echo state networks for multivariate time series prediction. *Information Sciences*, **476**, 337–356.

[54] Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. *Information Sciences*, **177**(23), 5329–5346.