Journal of Data Science and Modeling, Vol. 1, No. 1, 11-19, December 2022

Research Manuscript

Predicting the Brexit Outcome Using Singular Spectrum Analysis

Rahim Mahmoudvand *[1] , Paulo Canas Rodrigues[2],

1. Assistant professor of statistics, Bu-Ali Sina University, Hamedan, Iran.

2. Assistant professor of statistics, Federal University of Bahia, Salvador, BA, Brazil.

2. CAST, Faculty of Natural Sciences, University of Tampere, Tampere, Finland.

Received: 3/4/2017 Accepted: 15/7/2017

Abstract:

In a referendum conducted in the United Kingdom (UK) on June 23, 2016, 51.6% of the participants voted to leave the European Union (EU). The outcome of this referendum had a major policy and financial impact for both UK and EU, and was seen as a surprise because the predictions consistently indicate that the "Remain" would get a majority. In this paper, we investigate whether the outcome of the Brexit referendum could have been predictable by polls data. The data consists of 233 polls that have been conducted between January 2014 and June 2016 by YouGov, Populus, ComRes, Opinion, and others. The sample size ranges from 500 to 20058. We used Singular Spectrum Analysis (SSA), which is an increasingly popular and widely adopted filtering technique for both short and long time series. We found that the real outcome of the referendum is very close to our point estimate and within our prediction interval, which reinforces the usefulness of the SSA to predict polls data.

Keywords: Singular spectrum analysis, Recurrent SSA forecasting algorithm, Polls data.

Mathematics Subject Classification (2010): 62G05, 62M10, 62P25.

*Corresponding author: r.mahmoudvand@basu.ac.ir

1. Introduction

Singular Spectrum Analysis (SSA) is a nonparametric time series analysis and forecasting technique, which has become an increasingly popular method for noise reduction and forecasting. Research studies can be found with application to economic and financial data (Hassani et al. (2009, 2013b); Mahmoudvand et al. (2018)), industrial data (Hassani et al. (2013a, 2015); Mahmoudvand and Rodrigues (2016)), medical data (Aydin et al. (2011)), and mortality data (Mahmoudvand et al. (2015, 2017)), among others, which can be been as evidence of the increasing popularity of SSA. In brief, the SSA technique seeks to decompose a time series to identify the trend, signal, harmonic and noise components, and after that reconstructs a new filtered time series which can be used for out-of-sample forecasting. In comparison to classical time series models, the SSA technique has the advantage of not being bound by the assumptions of stationarity or normality, which are highly unlikely to hold in real-world applications.

This paper aims to introduce a new application for SSA for pools data. We consider the polls data in the UK between 2014 and the day before the Brexit referendum, which occurred on June 23, 2016. We assess whether the data from the pools support the final decision of the referendum, i.e., for the UK to leave the European Union.

The remainder of this paper is organized as follows. Section 2 describes the methodology underlying SSA forecasting algorithms. Section 3 gives a description of the data and shows the empirical results of the application. Section 4 reports some concluding remarks.

2. Methodology

Let $Y_N = [y_1, \ldots, y_N]$ denote a sample of a univariate time series with length N. Let us assume that Y_N can be written in terms of a signal plus noise model. Singular spectrum analysis (SSA) starts by denoising the original time series Y_N.

2.1 Denoising Time Series

Considering the window length L, a full augmented trajectory matrix is constructed by a L-dimensional embedding (e.g. Golyandina and Zhigljavsky, 2013) of the time series with lag 1, resulting in a Hankel trajectory matrix $\mathbf{X}$ of dimension $L \times K$, $K = N - L + 1$, as below:

$$\mathbf{X} = \begin{pmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{pmatrix}. \tag{2.1}$$

Denote by $U_j = [u_{1,j}, \ldots, u_{L,j}]'$ for $j = 1, \ldots, d$, the j^{th} left eigenvector of $\mathbf{X}$. A denoised version of original time series can be reconstructed as follows

$$\widetilde{y}_t = \frac{1}{w_t} \sum_{p=s_1}^{s_2} \sum_{i=1}^{L} \sum_{j=1}^{r} u_{i,j} u_{p,j} y_{t+i-p} \ , \ t = 1, \ldots, N. \tag{2.2}$$

where, $w_t = \min\{t, L, N-t+1\}$, $s_1 = \max\{1, t-N+L\}$ and $s_2 = \min\{L, t\}$ and $r < d$ is a cutting point that must be selected by the user. The forecast engine of SSA, which is a linear function of the last L observations of the denoised time series, will be constructed by a projection method. Two different approaches have been introduced to obtain this function: the recurrent forecasting algorithm and the vector forecasting algorithm. Here, we use recurrent method which is detailed in the next subsection.

2.2 Recurrent forecasting algorithm

Let $v^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the eigenvector U_i ($i = 1, \ldots, r$). Moreover, suppose that for any vector $U \in \mathbf{R}^L$, $U^\nabla \in \mathbf{R}^{L-1}$ represents the vector with the first $L-1$ components of the vector U. Let $\widehat{y}_{N+1}, \ldots, \widehat{y}_{N+h}$ be the h steps-ahead forecasts obtained by the SSA recurrent forecast algorithm, which can be obtained by using the following equation.

$$\widehat{y}_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \alpha_j \widehat{y}_{i-j} & \text{for } i = N+1, \ldots, N+h \end{cases}, \tag{2.3}$$

where $\widetilde{y}_i$ ($i = 1, \ldots, N$) represents the reconstructed time series (noise reduced series) and the vector $\mathbf{a} = (\alpha_{L-1}, \ldots, \alpha_1)$ can be computed by:

$$\mathbf{a} = \frac{1}{1-v^2} \sum_{i=1}^{r} \pi_i U_i^\nabla. \tag{2.4}$$

2.3 SSA Choices

As mentioned before, the window length, L, and the number of eigentriples used for reconstruction, r, must be specified by the user to conduct the model fit and model forecasting with SSA. The set of all choices for these quantities are as follows:

$$(L, r) \in \{(i, j) : i = 2, \dots, N-1 \ \text{ and } \ j < i\}. \tag{2.5}$$

Although many attempts have been made to find the optimal values of L and r, it is still a challenge to obtain these values for real data. For model fit, many studies suggested to use L close to half of the series length, $N/2$ (e.g., Golyandina et al. (2001); Hassani et al. (2011)). However, this choice might not be the most appropriated for model forecasting (e.g., Mahmoudvand et al. (2013)). Recently, Mahmoudvand and Rodrigues (2018) provided a new method for using two different values for the window length, one for model fit and another for model forecasting. Graphical tools such as the scree plot of singular values and the plot of the consecutive pairs of eigenvectors can be used to determine the number of eigentriples used for reconstruction r. As a general rule, we have to assess several values of (L, r) to verify which pair produce more consistent results.

2.4 Bootstrap based prediction interval for SSA forecasting

To obtain the bootstrap prediction interval for the h-steps-ahead forecasts, the first step is to obtain the SSA decomposition $Y_N = \widetilde{S}_N + \widetilde{E}_N$, where $\widetilde{S}_N$ is the reconstructed series that approximates the signal S_N of the time series and $\widetilde{E}_N$ is the residual series. Assuming that we have a specific model for the residuals $\widetilde{E}_N$, we can simulate p independent copies $\widetilde{E}_{N,i}$, $i = 1, \dots, p$, of the residual series E_N. Adding each of these residual series to the signal $\widetilde{S}_N$, we can get p series $Y_{N,i} = \widetilde{S}_N + \widetilde{E}_{N,i}$, $i = 1, ..., p$. Applying the recurrent (or vector) SSA forecasting algorithm, while keeping unchanged the window length L and the number r of eigenvalues/eigenvectors used for reconstruction, to the series $Y_{N,i}$, $i = 1, \dots, p$, we can obtain p values for the h-steps-ahead forecasts $\widehat{y}_{N+h,i}$. The empirical $\alpha/2$ and $1-\alpha/2$ quantiles of the h-steps-ahead forecasts $\widehat{y}_{N+h,1}, \dots \widehat{y}_{N+h,p}$, correspond to the bounds of the bootstrap prediction interval with confidence level $1 - \alpha$.

3. Empirical Results

3.1 Description of the Data

The data consist of 233 polls that have been conducted between January 2014 and June 2016 by YouGov, Populus, ComRes, Opinion, and others. The sample sizes range from 500 to 20058. The data was obtained from the Financial Times Brexit poll tracker. Since there were more than one poll for some dates, we pooled these data and computed the pooled proportions. The plot on the left-hand side of Figure 1 shows the time series of proportions for the UK to "Exit" and to "Remain" in the EU. Note that this time series is not defined on equally spaced time points, as more data was observed when the referendum date approaches. Overall it seems that the "Remain" series is above the "Exit" series. However, two changes can be observed in 2015: the "Remain" series increases, and the "Exit" series decreases in the first part of 2015, whereas this trend is reversed at the end of 2015. Also, the ratio between the proportion of "Exit" and the proportion of "Remain" is depicted on the right-hand side plot in Figure 1. This plot has a structure similar to the "Exit" series in the left-hand side plot of this figure.

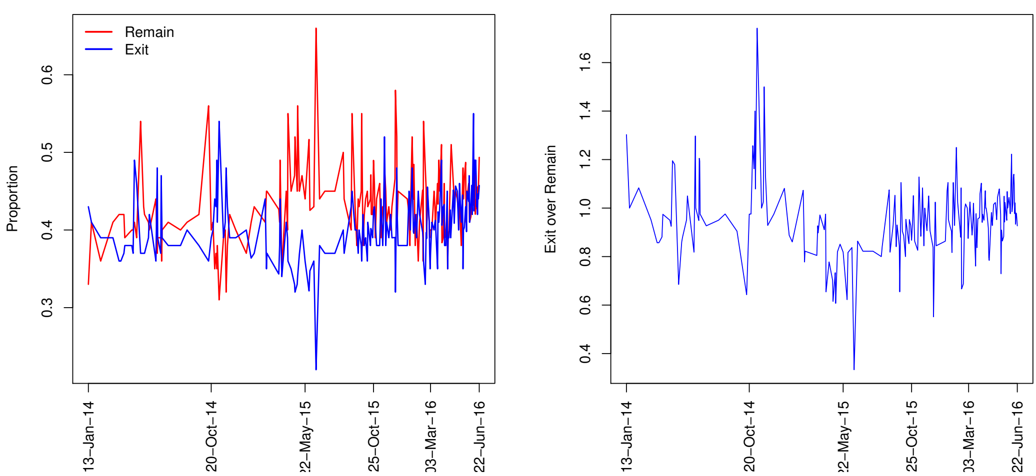

Figure 1: Time series of the Brexit polls data (January 2014 – June 2016).

3.2 Results

First of all, let us assess the trend in this time series, if there is any. To do this, we consider the series, "Exit" divided by the series "Remain". Let the window length be $L = 36$; then we can find the singular values and eigenvectors of the trajectory matrix. The left-hand side plot in Figure 2 shows that, when compared with the first five singular values, the singular values from 6 to 36 are small and without major changes in magnitude between them. In addition, Figure 3 shows that eigenvectors 1 to 3 could not produce harmonic components, whereas eigenvectors 4 and 5 do produce harmonic components. Based on these plots, we decided to consider $r = 5$ engentriples for reconstruction. The plot on the right-hand side of Figure 2 shows the fitted values by SSA considering $L = 36$ and $r = 5$. As it can be seen in this plot, the fitted values start to increase after May 2015, being the last 10 values, respectively: 1.07, 1.06, 1.05, 1.04, 1.03, 1.02, 1.01, 1.01, 1.01, and 1.00. The analysis of the trend of the series based on SSA indicates that there is a real possibility that the outcome of the referendum to be an "exit" from the EU. In addition, using the recurrent forecasting algorithm, we calculated the one-step-ahead forecast, which results in 1.03. Then, the proportion of an exit is estimated by $1.03/2.03 = 0.5074$, meaning that 50.74% of UK citizens support the "Exit" from the EU.

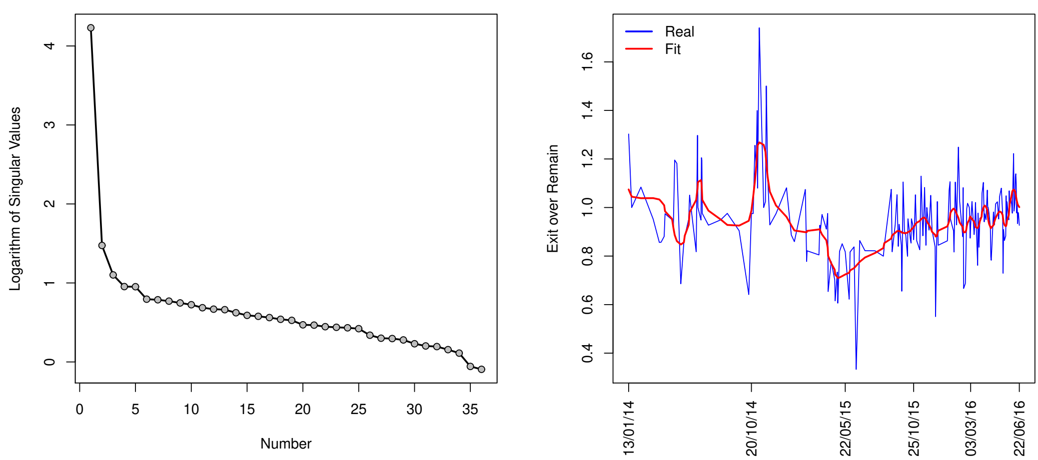

Figure 2: Left: Singular values of the trajectory matrix, Right: Fit by SSA using $L = 36$ and $r = 5$.

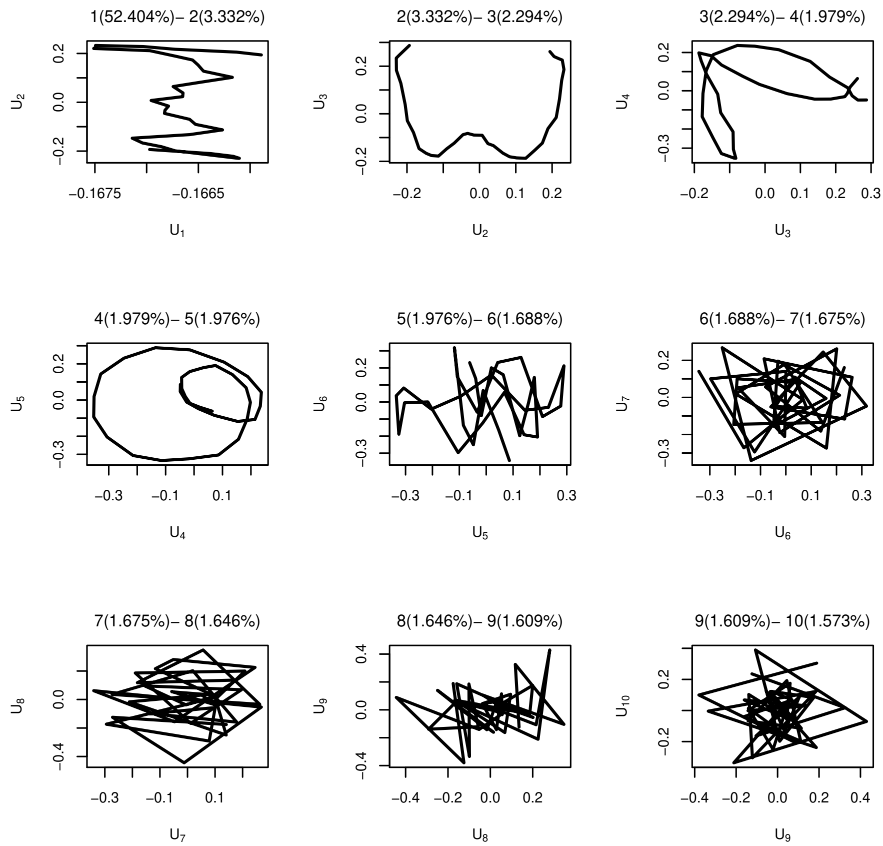

Figure 3: Pair of eigenvectors.

Being 50.74% just is a point estimate, we also decided to compute prediction intervals that can be very useful in assessing the quality of the forecasts. Using bootstrap with 1000 replications, we obtain a prediction interval of (48.58, 52.91), which reinforces that the result from the Brexit poll is uncertain. Moreover, this prediction interval also includes the real outcome of the referendum: 51.9% voted in favour of leaving the European Union and 48.1% voted in favour of remaining a member of the European Union.

4. Conclusions

In this paper, we analyzed the time-series data from the polls about the Brexit referendum between January 2014 and June 2016. The Singular Spectrum Analysis was used for denoising the time series and the recurrent SSA forecasting algorithm used to obtain point and interval predictions. Our conclusions show that, with a point estimate for one-step-ahead forecast of 50.74% and prediction interval of $(48.58, 52.91)$ for the UK to "Exit" the EU, support the idea that, the result of this referendum would not be certain as many advocate. Moreover, the real outcome of the referendum (51.9% voted in favour of leaving the EU) is very close to our point estimate and within our prediction interval, which reinforces the usefulness of SSA to predict polls data.

References

Aydın, S., Saraoglu, H. M., and Kara, S. (2011). Singular Spectrum Analysis of Sleep EEG in Insomnia. *Journal of Medical Systems,* **35**, 457-461.

Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). Analysis of time series structure: SSA and related techniques. Chapman & Hall/CRC.

Hassani, H., Heravi, H., and Zhigljavsky, A. (2009). Forecasting European industrial production with Singular Spectrum Analysis. *International Journal of Forecasting,* **25**, 103-118.

Hassani, H.; Mahmoudvand, R.; and Zokaei, M. (2011). Separability and Window Length in Singular Spectrum Analysis. *Comptes Rendus Mathematique*, **349(17)**, 987-991.

Hassani, H., Heravi, H., and Zhigljavsky, A. (2013a). Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting,* **32**, 395-408.

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

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.

Mahmoudvand, R.; Alehosseini, F.; and Rodrigues, P. C. (2015). Mortality Forecasting with Singular Spectrum Analysis. *RevStat - Statistical Journal*, **13**, 193-206.

Rodrigues, P.C., and Mahmoudvand, R. (2016). Correlation analysis in contaminated data by singular spectrum analysis, *Quality and Reliability Engineering International*, **32**, 2127-2137.

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**, 717-732.

Mahmoudvand, R.; Yarmohammadi, M.; and Rodrigues, P.C. (2018). Forecasting Daily Exchange Rates: A Comparison between SSA and MSSA. *RevStat-Statistical Journal*, **17**(4), 599-616.

Mahmoudvand, R.; Rodrigues, P.C (2018). A New Parsimonious Recurrent Forecasting Model in Singular Spectrum Analysis. *Journal of Forecasting*, **37**, 191-200.

Mahmoudvand, R.; Najari, N.; and Zokaei, M. (2013). On the Parameters for Reconstruction and Forecasting in the Singular Spectrum Analysis. *Communication in Statistics: Simulations and Computations*, **42**, 860-870.