SVD-Based & ARIMA Models for Time Series Evaluation

A Comparative Analysis

APMTH 205 Final Project

By Omar Abdel Haq

Table of Contents

I. Motivation & Research Question

Time series analysis is a potent tool crucial for unraveling temporal dynamics across diverse domains like geology, finance, and astronomy. While the dominant strategy to study and examine time series is the family of AutoRegressive Integrated Moving Average (ARIMA) models, new strategies, such as Recurrent Neural Networks (RNNs) have emerged as especially useful in language processing tasks. This project not only introduces but also critically compares an often-overlooked method in time series analysis: Singular Value Decomposition (SVD)-based techniques. Through a series of inquiries, the project explores various techniques, ranging from examining singular values in matrix-reshaped time series to employing methods like Singular Spectrum Analysis (SSA), positioning them as potential contenders to customary approaches like autocorrelation plots and ARIMA modeling.

As such, the primary research question guiding this project delves into whether the application of SVD matrix decompositions, specifically through singular value inspection and SSA, can indeed compete with traditional time series analysis tools in terms of accuracy, time complexity, and robustness. Rooted in Chapter 3 of Heath's *Scientific Computing*, this research aims to contribute to the broader understanding of whether SVD approaches could emerge as robust tools in the repertoire of a time series analyst.

II. Existing Literature

In the realm of matrix analysis, SVD assumes a pivotal role, providing not only an exact matrix representation but also a mechanism for efficient dimensionality reduction by discarding less significant data. This inherent versatility extends its applications to various fields, enabling functionalities such as image compression and recommendation systems. Moreover, SVD's nature proves instrumental in constructing subspaces defined by matrix columns, playing a crucial role in tasks like digit recognition and image reconstruction. Beyond its role in domain-specific applications, SVD offers valuable assistance in scenarios where determining the "best-fitting" curve for a set of given points is crucial, as exemplified in the task of linear least squares. This capability becomes particularly useful when seeking the "closest" solution in contexts such as understanding relationships between variables or predicting information based on historical data [1].

Despite the extensive literature on SVD applications across various domains, its utilization in time series analysis has received limited attention compared to well-explored approaches such as ARIMA models and RNNs. While ARIMA models are recognized for their simplicity and effectiveness in capturing linear dependencies, and RNNs excel in modeling complex temporal dependencies, especially in natural language, the growing interest in SVD and its derivative, SSA, suggests a growing curiosity about alternative methodologies.

Developed sometime in the 1970s or 80s, SSA has gained traction for its capacity to decompose time series into interpretable components, including trends, periodicities, and noise. Initially applied to the field of climatology, SSA has offered a novel perspective in time series analysis, showcasing its potential expansion into other domains within the time series realm [2]. As the exploration of alternative methodologies continues to evolve, the unique attributes of SVD and SSA present promising avenues for enhancing our understanding and analysis of time-dependent data.

III. Data

The project attempts its analysis using three datasets of varying complexity, all visualized in Figure A. The simple time series is generated by applying the sign function to a sine wave with a frequency of 0.01 cycles per data point, representing a square wave. Random noise with an amplitude of 0.1 is added to this square wave, resulting in the following expression:

$$f(t) = \text{sign}\left(\sin\left(2\pi \cdot 0.01t\right)\right) + \mathcal{N}(0, 0.1)$$

This results in a boxed wave structure with a period of 100. The second time series is generated by combining multiple components waves of varying complexity, as well as dominating additive random noise. The formula for the second dataset is expressed as follows:

$$\begin{aligned} g(t) = &\sin(2\pi \cdot t) \cdot \sin(\pi t) + 0.5\cos(2\pi \cdot 2t) \\ &+0.2\sin^2(2\pi \cdot 8t) + 0.3\sin^3(2\pi \cdot 5t) + \mathcal{N}(0, 1.21) \end{aligned}$$

The third and last time series is a Kaggle dataset showcasing the price of a Microsoft stock at the opening of each day [3]. Notably, the data is much more complex; it is much clearer that there is no singular period that determines the price trend.

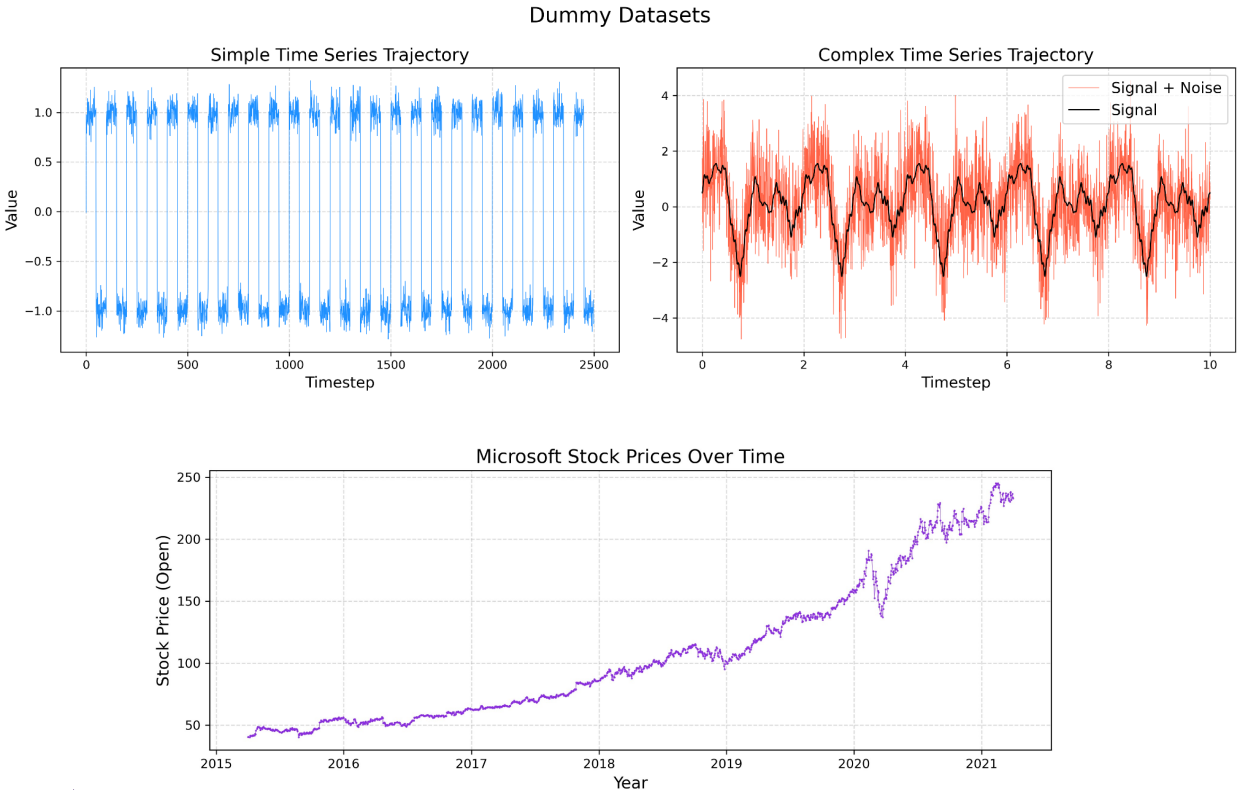

Figure A. *Visualizations of the three datasets used in this project.*

IV. Series Periodicity Estimation

SVD serves as a vital technique in matrix factorization, extensively applied in the realms of linear algebra and numerical analysis. When applied to a matrix A, SVD dissects it into three

fundamental matrices denoted as $U\Sigma V^T$. In this expression, U and V represent orthogonal matrices, and Σ is a diagonal matrix encompassing the singular values of A. The singular values, embedded within Σ, essentially quantify the magnitudes of transformations undergone by the matrix along its principal axes. This information proves invaluable for understanding the inherent structure of the matrix. Furthermore, each set of the ith left and right singular vectors, paired with the corresponding singular value, collectively known as an eigentriple, encapsulates approximating "information" about the original data [4].

In applications such as image compression or Principal Component Analysis, we leverage the k most significant eigentriples — those associated with the k largest singular values — to reconstruct the initial matrix. This process enables us to obtain a proficient approximation of the original data. Additionally, by examining the magnitudes of successive singular values, we gain insights into the significance or utility of each principal component in data reconstruction.

However, the application of this idea to time series still remains unexplained. While time series are conventionally represented as one-dimensional arrays, transforming them into matrices and employing SVD can yield valuable insights. One notable metric in this context is the Singular Value Ratio (SVR), calculated as the ratio of the first and second singular values (as opposed to the condition number, which is the quotient of the largest and smallest singular values). SVR serves as a useful measure for evaluating the periodicity of a time series. The core intuition behind this is as follows: if we select the rows of the matrix such that each represents a single period, then the first principal component of that matrix could do wonders approximating the original series by learning the underlying periodicity, with the others singular values being relatively smaller as they map onto noise. This implies that, if we select the number of columns such that it is equal to the period of the data, we should observe a high SVR [5].

Nevertheless, before attributing high SVR values solely to periodicity, it's crucial to examine the behavior of singular values in random matrices as a baseline. Consider matrices where entries are independent and identically distributed (IID). In this context, the singular values of large IID matrices exhibit a distinctive statistical behavior described by the Marchenko-Pastur distribution. This distribution characterizes the eigenvalue (and singular value) distribution in the limit of large random matrices. The distribution is semi-circular, featuring higher density towards smaller singular values and lower density for larger ones. In the context of SVD, singular values are conventionally ordered by magnitude, revealing a visually interesting pattern. When plotted from largest to smallest, the graph typically forms a downsloping line, a consequence of the semi-circular distribution. The Marchenko-Pastur distribution thus provides a probabilistic lens for understanding the arrangement of singular values in random matrices, offering valuable insights into their statistical properties [6]. Figure B illustrates the mean singular value for a random matrix with IID entries. While the underlying distributions differ, they are chosen to ensure a mean of zero and a variance of one, providing a glimpse into the statistical behavior of singular values in such matrices.

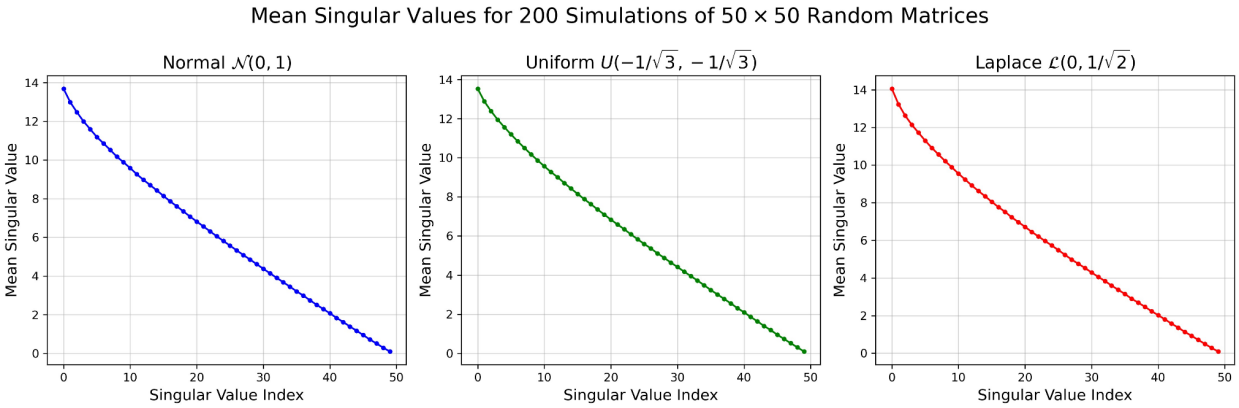

Figure B. *The plots show mean singular values for 200 simulations of 50x50 random matrices drawn from normal, uniform, and Laplace distributions, each with mean 0 and unit variance.*

Given the established mean and variance conditions under which singular values adhere to a known distribution, irrespective of the entry distribution, the crucial question is whether a high SVR unambiguously signals the presence of periodicity or if other scenarios could lead to elevated SVR values. This inquiry prompts a re-examination of two key assumptions: how variations in both mean and variance impact the singular values. Figure C illustrates precisely these two scenarios. Notably, the figure indicates that the first singular component absorbs the mean's multiplicative factor, while the variance primarily extends the range of singular values. Despite the differing data variance, the general trend aligns with the case of mean one and unit variance.

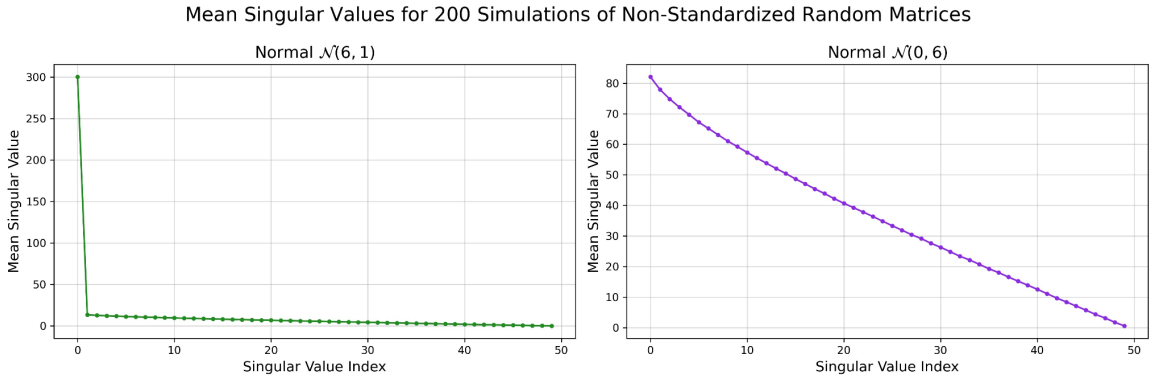

Figure C. *The plots show mean singular values for 200 simulations of 50x50 random matrices drawn from normal distributions with non-zero mean and non-unit variance, respectively.*

Shifting the mean of random variables within a matrix introduces a deterministic component that influences the sum of squares in each column. This shift tends to amplify the magnitudes of the singular values, particularly affecting the first singular value σ_1. As the representative of the most substantial variation in the data, σ_1 is notably sensitive to this mean shift. The additional constant term significantly contributes to the information captured by the first singular value and its associated vectors, causing an inflation in its magnitude compared to the other singular values. Consequently, this inflation affects the SVR, assuming that σ_1 experiences inflation while σ_2 remains relatively unaffected.

Given that changes in variance should have a limited impact on the SVR and that mean changes affecting the SVR can be addressed through data standardization, the primary reason for a high SVR becomes the capturing of significantly more information on the general trend of the data by the first principal component (product of the first singular value and its associated singular vectors). This implies that a high SVR, especially when the number of columns is chosen to match the period, indicates the detection of periodicity [5]. Figure D below illustrates the singular values resulting from different matrix reshaping of the simple dummy dataset used in this project. Notably, the last case, a 25×100 matrix, aligns with the correct underlying period $p = 100$, resulting in an observable increase in the SVR.

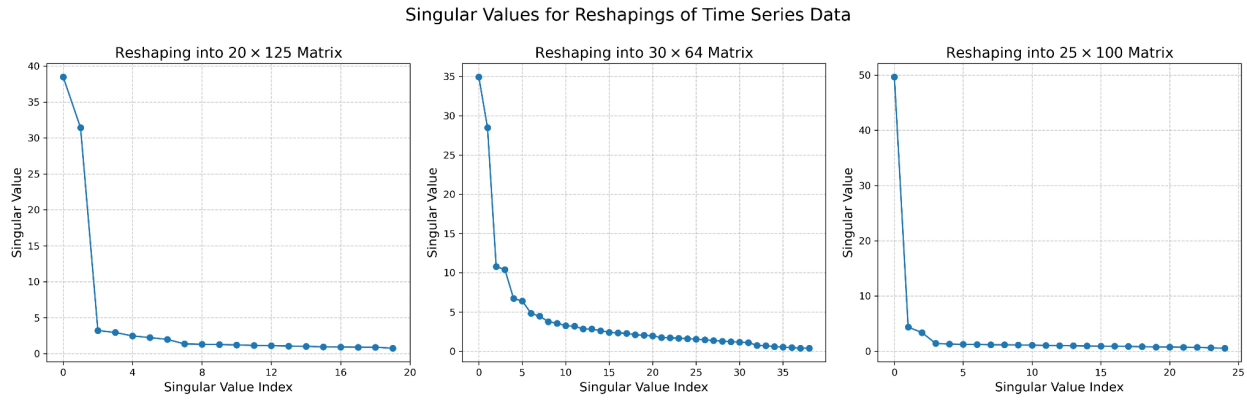

***Figure D**. The plots show the singular values of three different reshapings of the simple time series, with the quotient of the first and second values is largest for the 25x100 case.*

As a result, we can plot the SVR for different potential periods p, and use the value at the highest SVR as the predicted period. The natural counterpart to this approach is to use the autocorrelation function (ACF). By calculating the correlation between the original time series and its lagged values at different intervals, the ACF produces a plot that can be analyzed for significant peaks [7]. These peaks, occurring at regular intervals, suggest potential repeating patterns in the time series. The identified period p corresponds to the lag at which a substantial peak is observed. Figure E presents the findings of both approaches — *very* similar in nature.

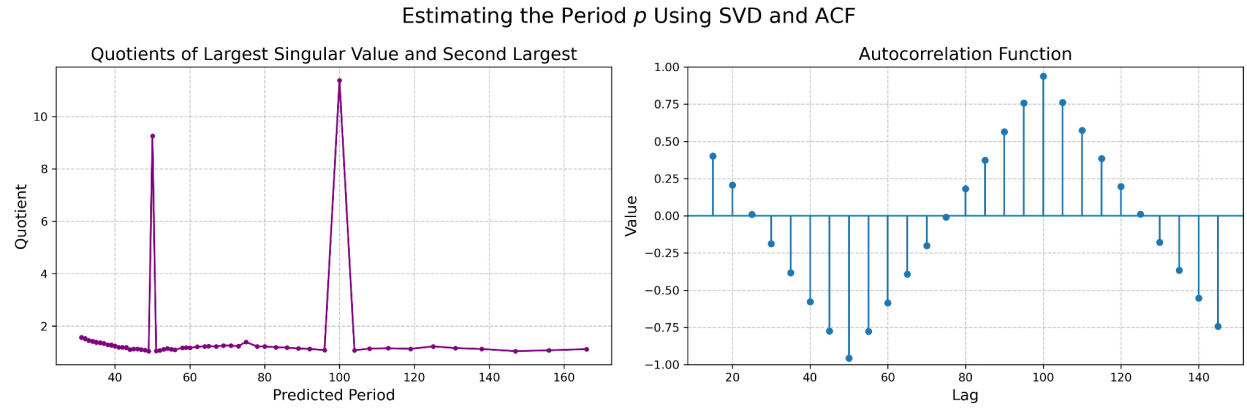

***Figure E**. Estimating the period p using the SVR and ACF approaches.*

Analogously, beyond finding the period p, we can use this approach to create a prediction line for the series at hand. Using our decomposition, we can simply use the first principle

component as our line of best fit, attributing the remaining components to variation due to noise. The dominant strategy used on the other hand is the AutoRegressive Model. The autoregressive (AR) model, denoted as AR(p), represents a time series as a linear combination of its own past values up to order p. The model is defined as:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

Here, c is a constant, $\phi_1, \phi_2, \ldots, \phi_p$ are autoregressive coefficients, $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ are lagged values, and ε_t is a noise term [7]. Figure F displays the results of the AR model as compared to those of using the first principal component; the SVD-based approach overfits much less to the inherent noise of the model as compared to the AR model.

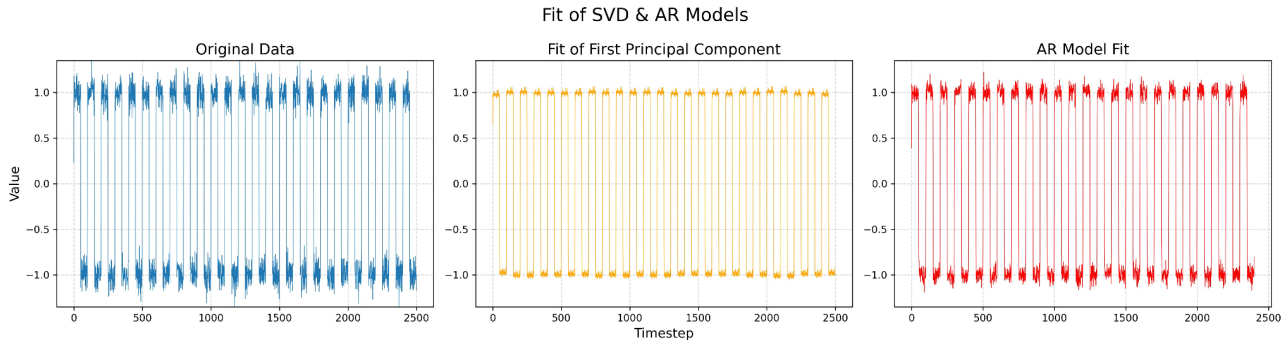

***Figure F**. The fit of the SVD-based and AR model prediction lines.*

The reduced overfitting to noise can be verified by computing the Root Mean Squared Error (RMSE) between the line of best fit from each approach and the noiseless version of the dataset. With an RMSE of 0.106 and 0.246 for the SVD and AR approaches, respectively, the SVD approaches overwhelmingly outperforms the AR model — seemingly significantly more robust to noise.

V. Complex Modeling: Robustness

Questions remain for how the models compare when dealing with time series data that is significantly more complex than a single period and added noise. To do this comparison, the SVD-based arsenal can be expanded to Singular Spectrum Analysis. SSA is a technique for time series analysis that, at its core, breaks down a trend into its component trajectories, similarly to how PCA breaks up a feature space into its component dimensions. SSA starts out with an initial embedding phase, where the time series data of length N, denoted as $X = [x_1, x_2, \ldots, x_N]$, is converted into a Hankel matrix. To do so, we first choose a window length L, which serves as the column length in our matrix, and then lay out values in subsequent rows lagged by one value. Pictorially, the Hankel matrix of window length L is defined as:

$$\mathbf{H} = \begin{bmatrix} x_1 & x_2 & \ldots & x_{N-L+1} \\ x_2 & x_3 & \ldots & x_{N-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \ldots & x_N \end{bmatrix}$$

Intuitively, this matrix produces columns which are different lagged versions of the time series of reference. The next phase, SVD, decomposes the Hankel matrix into component trajectories. The singular values capture the importance of underlying modes in the time series. The singular values, then, can be studied to extract trend from noise, given a few underlying assumptions. Since approximation using patterns/periodicities is easier than approximating noise, this posits that large singular values point to components significant to the underlying trend, while smaller ones point to noise.

Finally, the reconstructed components are derived by averaging along specific diagonals of the component matrices obtained in the decomposition step. These diagonals, known as anti-diagonals, run from the top-right to the bottom-left of the matrices. By averaging along these anti-diagonals, SSA effectively combines information across different time lags, allowing it to capture and highlight distinct patterns and trends present in the original time series data [2]. Noise, even if large, is still filtered out, since intuitively, capturing periodicity makes for a better approximation in the first few "principal trajectories" as opposed to fitting patternless noise. Figure G depicts the use of SSA on the noisy dummy dataset outlined previously, with the first five principal trajectories (chosen arbitrarily) plotted alongside both the signal and noisy data. Remarkably, the SSA approach produces a line of best fit nearly devoid of noise.

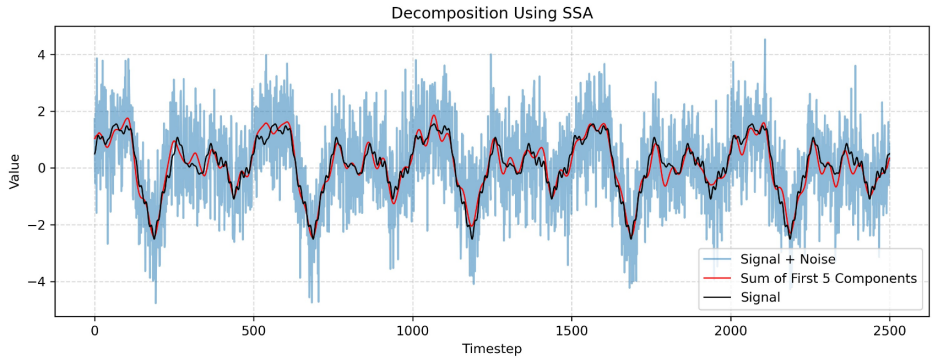

***Figure G**. The SSA line of best fit closely matches the signal, and largely filters out all noise.*

VI. Complex Modeling: Real Data

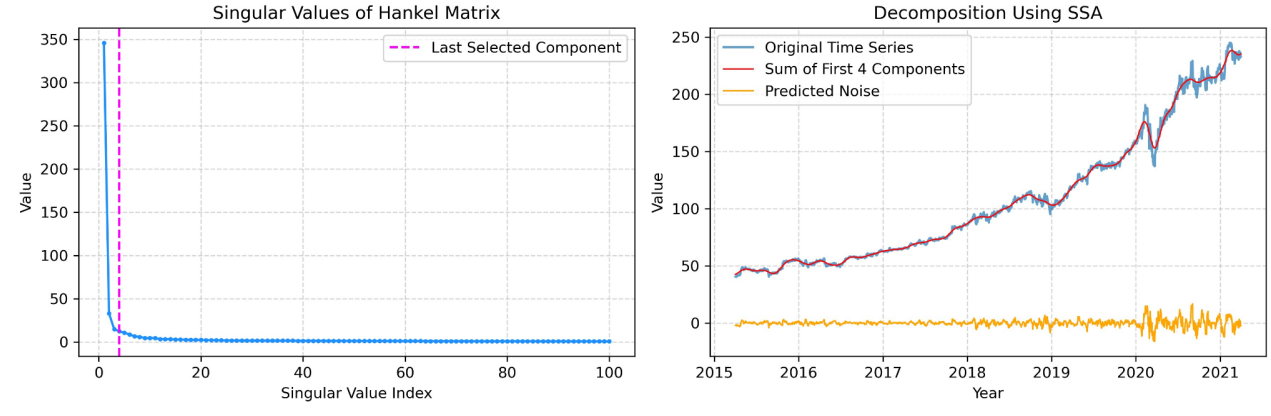

***Figure H**. Applied to the Microsoft stock data, the singular values of the Hankel matrix alongside the line of best fit from the SSA model and the associated "noise".*

In the case of the Microsoft stock data at hand, we can choose the number of principal trajectories based on a plot of the singular values, rather than arbitrarily. Doing so indicates

drastically diminishing returns to using more than the first four components. We can apply the same SSA strategy to this data now too; using the sum of the first four components, and the residuals of the model as the approximate noise, we observe the general trend of the stock data, without overfitting to what seems to be random noise. Figure H above presents these findings.

The counterpart to the SSA procedure in classical time series methods is ARIMA modeling. ARIMA is a popular time series forecasting method used to model and predict future values based on past observations. The model consists of three main components: Autoregressive (AR), Integrated (I), and Moving Average (MA). The parameters p, d, and q in ARIMA represent the key components of the model. The p refers to the autoregressive order, which represents the number of lagged observations included in the model; it captures the relationship between the current observation and its past values. The d represents the differencing order, indicating the number of times the raw observations are differenced to make the time series stationary. A stationary time series is one whose statistical properties, such as mean, variance, and autocorrelation, do not change over time. In other words, the behavior of the time series is consistent and does not exhibit trends or seasonality. Stationarity is crucial for ARIMA, as it assumes that the time series data has a constant mean and variance over time; this assumption is not necessary for SSA, on the other hand. Finally, the q denotes the moving average order, which signifies the number of lagged forecast errors included in the model. It accounts for the influence of past forecast errors on the current value [7].

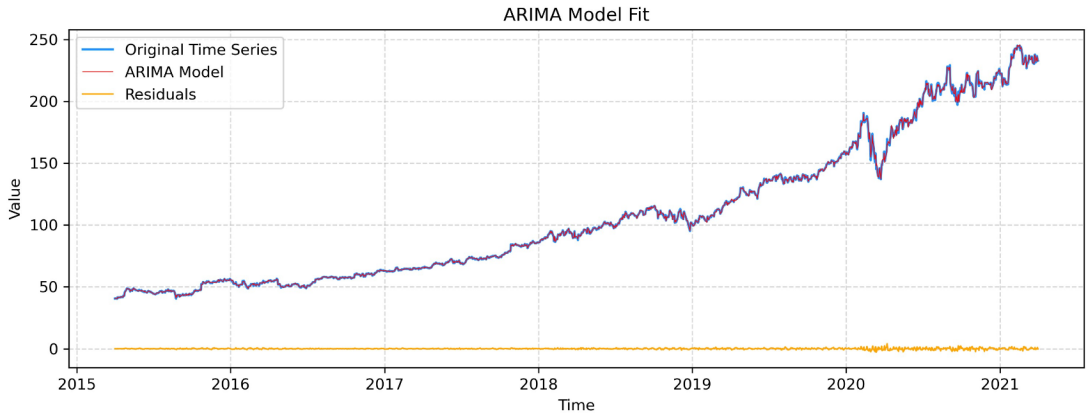

Figure I. *Fit of ARIMA model alongside its residuals.*

In this analysis, a randomized grid search approach was used to find the optimal set of parameters p, d, q. The evaluation metric for selection is the Akaike Information Criterion (AIC). A measure of the relative quality of a statistical model for a given set of data, the AIC takes into account the goodness of fit of the model and penalizes models for having more parameters, thus discouraging overfitting. Figure I presents the original Microsoft stock data, the best fitted ARIMA model and its residuals.

ARIMA, despite exhibiting tendencies toward overfitting, as can be seen above, with both original data and line of best fit nearly equivalent. Despite this disadvantageous characteristic, the model has the particular strength of being able to predict values beyond the initial dataset. In contrast, SSA, a non-parametric method, encounters the notable pitfall of inability to intuitively extend predictions into the future. SSA excels in decomposing time series into meaningful components but lacks the inherent forecasting capabilities that make ARIMA a powerful tool for predictive analytics.

VII. Summary

To summarize, the project introduces and critically compares SVD-based techniques, often overlooked in favor of established methods like ARIMA models and RNNs. The project first delves into the application of SVD in time series period estimation, introducing the Singular Value Ratio as a metric for evaluating periodicity. The robustness of this approach is evaluated and its performance as a data-fitting tool is discussed; the approach seems to work hand in hand with regular period estimation strategies like Autocorrelation Functions, but does a better job at fitting trends and ignoring noise.

Secondly, the investigation extends to more complex models, introducing SSA as a tool to break down time series trends into component trajectories. The steps of SSA, and the rationale behind its robustness to modeling noise are explained at depth. Finally, the comparison between SSA and ARIMA modeling on real data reveals the nuanced strengths of each approach. While ARIMA may exhibit tendencies toward overfitting, as noted from its fit to Microsoft stock data, its strength comes from the fact that its parameters can be used to predict values beyond the initial dataset. On the other hand, SSA, a non-parametric method, stands out in decomposing time series into meaningful components (without the need for stationarity) but lacks the inherent forecasting capabilities of ARIMA. The table below outlines the strengths and weaknesses of each method.

Approach	**SVD / SSA**	**AR / ARIMA**
Period Estimation Approach	Uses the SVR on standardized data.	Uses the Autocorrelation Function on stationary data.
Period Estimation Time Complexity	$O(km^2n)$, where *k* equals the number of reshapings, *m* the number of rows, and *n* the number of columns. [4]	$O(p^2t)$, where *p* is the AR parameter and *t* the length of the time series. [8]
Tends to Overfit in Simple Cases	No	Yes
Robust to Noise	Yes	No
Stationarity Requirement	No	Yes
Predictive Capability	No	Yes

Table A*. A comparison of algorithm characteristics.*

In conclusion, a final thought comes to mind: is it possible to both create a predictive model, but do so without overfitting to noise? Although beyond the scope of this project, it would be interesting to test the following idea: if we first “clean” the signal from the noise using SSA, and then use ARIMA modeling, will the resulting model benefit from the better generalization and forecasting capabilities? This approach, if effective, would make SSA a regularizing method employed before modeling, similar to how principal component analysis, to its own credit, does the same in the case of least squares regression.

References

[1]. Kuang, Zecheng. "Singular-Value Decomposition and Its Applications." University of California San Diego.

[2]. Golyandina, Nina. Particularities and Commonalities of Singular Spectrum Analysis as a Method of Time Series Analysis and Signal Processing, arxiv.org/pdf/1907.02579.pdf.

[3]. Venkitesh, Vijay. "Microsoft Stock Time Series Dataset." Kaggle, 15 June 2021, www.kaggle.com/datasets/vijayvvenkitesh/microsoft-stock-time-series-analysis/.

[4]. "Chapter 3: Linear Least Squares." Scientific Computing: An Introductory Survey, by Michael T. Heath, Society for Industrial and Applied Mathematics, 2018.

[5]. Khoshrou, Abdolrahman. Data-Driven Pattern Identification and Outlier Detection in Time Series, arxiv.org/pdf/1807.03386.pdf.

[6]. Marchenko–Pastur Distribution, www.wolfram.com/language/11/random-matrices/marchenko-pastur-distribution.html.

[7]. Nau, Robert. "ARIMA Models for Time Series Forecasting." Statistical Forecasting: Notes on Regression and Time Series Analysis, people.duke.edu/~rnau/411arim3.htm.

[8]. Wang, Xiaoqian, et al. Distributed ARIMA Models for Ultra-Long Time Series, 2022, arxiv.org/pdf/2007.09577.pdf.

Acronyms

ACF	—	Autocorrelation Function
AIC	—	Akaike Information Criterion
AR	—	AutoRegressive Model
ARIMA	—	AutoRegressive Integrated Moving Average
IID	—	Independent and Identically Distributed
PCA	—	Principal Component Analysis
RMSE	—	Root Mean Squared Error
RNN	—	Recurrent Neural Network
SSA	—	Singular Spectrum Analysis
SVR	—	Singular Value Ratio
SVD	—	Singular Value Decomposition