5-2016

Takens Theorem with Singular Spectrum Analysis Applied to Noisy Time Series

Thomas K. Torku
East Tennessee State University

Follow this and additional works at: https://dc.etsu.edu/etd

Part of the Algebra Commons, Applied Mathematics Commons, Longitudinal Data Analysis and Time Series Commons, Other Statistics and Probability Commons, and the Statistical Models Commons

Recommended Citation

Takens Theorem with Singular Spectrum Analysis Applied to Noisy Time Series

—————————

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

—————————

by

Thomas Torku Kofi

May 2016

—————————

Jeff Knisley, Ph.D. Chair

Anant Godbole, Ph.D.

Michele Joyner, Ph.D.

Nicole Lewis, Ph.D.

Keywords: Singular Value Decomposition, Singular Spectrum Analysis, Takens

Vectors, Entropy, Mutual Information, Financial time series.

ABSTRACT

Takens Theorem with Singular Spectrum Analysis Applied to Noisy Time Series

by

Thomas Torku Kofi

The evolution of big data has led to financial time series becoming increasingly complex, noisy, non-stationary and nonlinear. Takens theorem can be used to analyze and forecast nonlinear time series, but even small amounts of noise can hopelessly corrupt a Takens approach. In contrast, Singular Spectrum Analysis is an excellent tool for both forecasting and noise reduction. Fortunately, it is possible to combine the Takens approach with Singular Spectrum analysis (SSA), and in fact, estimation of key parameters in Takens theorem is performed with Singular Spectrum Analysis. In this thesis, we combine the denoising abilities of SSA with the Takens theorem approach to make the manifold reconstruction outcomes of Takens theorem less sensitive to noise. In particular, in the course of performing the SSA on a noisy time series, we branch off into a Takens theorem approach. We apply this approach to a variety of noisy time series.

DEDICATION

To my parents

Mr. and Mrs. Torku

and

good friend

Akutcha Samuel

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The financial world is evolving with many of its institutions having to deal with the analysis of big financial time series data. Stock markets in particular are using ideas from fields like data science, data compression, machine learning, predictive analysis, information theory and among others to analyse these data. Thus the need for a good modelling and forecasting technique cannot be overemphasized.

Previous studies [37] show that classical modeling and forecasting techniques do not have good performance due to factors like non-stationarity, volatility of the stock market, policy changes, technological changes, the presence of noise, etc. The main assumptions underlying these techniques are linearity and normality which are restrictive and limited in scope. As a result, in many cases, poor approximations of the actual data occur when these assumptions are applied.

The presence of noise significantly reduces the efficiency level of the technique used in analyzing the data. Let $x_t = s_t + \varepsilon_t$ $(t = 1, ..., n)$ be a time series which behaves as a stochastic dynamic system. Then the deterministic part is s_t [36]. The term ε_t refers to the variability or the noise in the time series x_t and the portion which makes it stochastic instead of entirely deterministic. There are two major methods for analyzing noisy time series. According to the first method, the noise is ignored and a forecasting model is fitted directly from the noisy data with the hope of extracting the underlying deterministic dynamics. The second method entails filtering the noisy data to reduce the noise levels before forecasting. Examples of both linear and nonlinear noise reduction methods include simple nonlinear filtering, singular value decomposition (SVD), local projective, and autoregressive moving averages (ARMA)

model. According to [51] SVD-based approaches are more effective in noise reduction than the other ones.

Research has revealed that most financial time series are nonlinear. This calls for the application of nonlinear methods [16, 39, 51]. The Takens approach is used to analyze and forecast nonlinear financial data but only if it is not noisy. In contrast, the Singular Spectrum Analysis (SSA) technique is a powerful tool that can be used for noise reduction and forecasting of both linear and nonlinear, stationary and non-stationary time series [37]. It is interesting to note that SSA incorporates the elements of classical time series analysis, multivariate statistics, dynamic systems, multivariate geometry and signal processing [32]. The SSA also incorporates filtering of the time series and SVD [36].

It is possible to combine the Takens approach with SSA to analyze financial time series. It is evident that we can estimate the key parameters in Takens theorem using SSA. In this work, we combine the denoising abilities of the SSA together with Takens approach to make the reconstruction of the Takens manifold less sensitive to noise. This work is made up of five chapters including the introduction. In Chapter 2, we provide the background to the study. In Chapters 3, we extensively present our methodology. In Chapter 4, we make applications to a given data set and present the results. The conclusions and discussion are given in Chapter 5.

2 BACKGROUND

The novel Singular Spectrum Analysis came to the limelight when Broomhead and King [14, 13, 32] and Broomhead *et al.* [11, 32] made their first publication on it. Since their publication, SSA has become popular in the statistical world and beyond. SSA has myriads of applications including finance, data science, statistics, geosciences, climate change and many others. For example, SSA has now become a standard tool for the analysis of climatic, meterorological and geophysical time series, according to, Vautard and Ghil [54, 32], Ghil and Vautard [31, 32], and Yiou *et al.* [56, 32].

The methods and procedures involved in SSA have currently engendered many authors to write several papers and publications. This tells us how powerful this statistical tool is. For example, authors like Vautard *et al.* [55, 32], Ghil and Taricco [30, 32], Allen and Smith [7, 32], Danilov and Zhigljavsky [20, 32], Yiou and *et al.* [57, 32] made massive references from the first publication on SSA [32]. The first introductory book on this subject was published by Elsner and Tsonis [25, 32].

The main goal of the basic SSA is to decompose the original series into a sum of series so that each component in this sum can be identified as either a trend, periodic or quasi-periodic component or noise [35, 32]. The capabilities of basic SSA include ascertaining the solution to the following problems: change-point detection; determining structure in short time series; smoothing; seasonal component extraction; periodicities with different amplitudes extraction; extracting complex trends and others [35].

The basic SSA primarily involves two stages – decomposition and reconstruction.

The decomposition stage is divided into two steps: embedding and singular value decomposition (SVD). The reconstruction stage has two parts: grouping and diagonal averaging. These two stages make up the basic SSA algorithm [32]. The summary of this algorithm is as follows. First, a one-dimensional time series is transformed into a higher dimension representation whose dimension is called the window length. This higher dimensional time series is called the trajectory matrix. The second step involves computing the SVD of the trajectory matrix. The next step is the grouping step which involves splitting the matrices from the SVD into several groups. By taking the average along the diagonals of each group and combining them into one tine series we obtain the approximated time series of the original or initial series [32].

2.1 Time Series

A time series is the collection of quantitative observations that are measured successively and evenly spaced in time intervals [8]. Data that can be classified as time series include annual rainfall, daily or weekly closing price of stock, number of death cases in the year and recording of temperature. A time series that can be measured as a single variable is termed as univariate. If two or more variables are measured then we call it multivariate. When we measure time series at discrete or finite steps or points, then we have a discrete time series. The data set used in this work is a discrete time series. The mathematical expression for a discrete time series is $x_t, t = 0, 1, 2, \ldots$. In effect, x_t is considered to be a random variable. However, observations which are measured over a specified interval is known as a continuous time series [5].

Generally speaking, a time series has four major components, namely: *seasonal, cyclical, trend and irregular.* Seasonal variation in time series occurs as a result of changes in the weather and climate conditions. For example, the increase in the sales of winter clothes is caused by seasonal variation. Repeated patterns or cycles due to medium term changes as seen in the financial markets lead to cyclical variation. Eventualities which are not repetitive in nature like earthquakes, flood, war and other natural disasters create a scenario in time series referred to as irregular or random fluctuations. A trend in time series occurs when there is a pattern of continuous increase, decrease or stagnation over time. For example, there is an upward trend in the rent of city apartments and a downward trend in birth rates [32, 5].

2.1.1 Time Series and Stochastic Process

A stochastic process is a non-deterministic process [5]. It involves some randomness and probability. Time series is also stochastic in nature in which case it is assumed to follow some probability model [18] that explains the joint distribution of a random variable. This means that there is uncertainty associated with the predictability of future values of the series. Each time series variable x_t is considered by assumption to be independent and identically distributed (i.i.d) and follows a normal distribution [5]. This may not always be the case. Statistical properties such as mean μ and variance σ^2, which are not contigent upon time, are very useful in describing the stationarity of a stochastic process. It must be observed that the square root of the variance gives us the standard deviation σ. There are two types of stationary process: weak and strong, which is also known as strict [38].

13

Definition 2.1 *A strong stationary process occurs if the joint probability distribution function of $\{x_{t-\tau}, x_{t-\tau+1}, \ldots x_t, \ldots x_{t+\tau+1}, x_{t+\tau}\}$ does not depend on the concept of time t for all τ [38].*

Definition 2.2 *A process $\{x_t, t = 0, 1, 2, 3, \ldots\}$ is said to be weak stationary if the statistical properties or moments of the process depend only on the time differences [18, 38].*

Definition 2.3 *Non-stationarity in time series occurs when the mean and the variance of the stochastic process change over time [38].*

2.1.2   Linear and Non-linear Time Series

A time series is either linear or nonlinear. A major part of time series analysis is the white noise or the error term.

Definition 2.4 *White noise is the error term emanating from a stationary random process. So a stationary time series denoted as ρ_t is said to be white noise if the $Corr(\rho_t, \rho_{t+\tau})$=0 given that $t \neq t+\tau$. The main assumption underlying ρ_t is that the expected value or the mean of all the errors is zero. The mathematical expression is*

$$E(\rho_t) = E(\rho_t|\rho_{t+1}, \rho_{t+2}, \ldots.) = 0 \qquad (1)$$

Definition 2.5 *A linear time series is any time series that can be expressed in the form*

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \phi_j \rho_{t+j} \qquad (2)$$

where μ is the mean of x_t and $\{\rho_t\}$ are the (i.i.d) random variables [4].

Definition 2.6 *A time series is nonlinear if it does not satisfy Equation (2).*

2.1.3   Financial Time Series

Financial time series are the financial data obtained through the financial markets such as the stock exchange market. They relate to financial asserts such as stock, bonds, equities and others. Price, both historical and current, and returns are key factors that influence the stock exchange market. Due to the volatility nature of stock market, prices of financial assets fluctuate. This makes financial time series mostly nonlinear, complex and non-stationary. There is also the issue of uncertainty in the financial market such that good prediction models must be constructed to predict future prices.

2.2   Singular Value Decomposition

The Singular Value Decomposition (SVD) is a matrix factorization technique in which a given matrix is factorized into product of matrices [9]. The SVD is widely used in linear algebra, data compression, and linear least-squares problems [6].

Definition 2.7 *A matrix formed by changing the rows of the original matrix into columns is called the transpose matrix. The transpose of matrix X is written as X^T.*

Definition 2.8 *A matrix X is orthogonal if $XX^T = X^TX = I$, where I is the identity matrix.*

Definition 2.9 *If X is a matrix, then its singular values are the roots of its covariance matrix, X^TX.*

Definition 2.10 *Let X be an $m \times n$ matrix. A number $s \geq 0$ is a **singular value** of X, if there exist $u \neq 0$ such that*

$$X^TXu = s^2u \tag{3}$$

or equivalently, there exist $v \neq 0$ such that

$$XX^Tv = s^2v. \tag{4}$$

*Let $v_1, ..., v_m$ be an orthonormal set of left **singular vectors** and let $u_1, ..., u_n$ be an orthonormal set of right **singular vectors** corresponding to **singular values** $s_1 \geq s_2 \geq ... \geq s_r > 0$. Then $V = [v_1|...|v_m]$ and $U = [u_1|...|u_n]$ are both orthogonal. We use s_r because r is the rank of X.*

Definition 2.11 *If X is an $m \times n$ real matrix then its SVD is*

$$X = U\tilde{\Sigma}V^T \tag{5}$$

where U and V^T are both orthogonal matrices (of dimensions $m \times m$ and $n \times n$ respectively) and

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \tag{6}$$

is an $m \times n$ matrix, where

$$\Sigma = \begin{bmatrix} s_1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & s_r \end{bmatrix} \tag{7}$$

is a diagonal matrix of nonzero singular values $s_1 \geq s_2 \geq ... \geq s_r > 0$ of X [9].

2.3 Entropy

Shannon [49] introduced the idea of information entropy. Entropy can be seen as the average information from a process. It is usually measured in bits and related to probability. A bit (binary unit) is a parameter **b** whose values are either 0 or 1. For example, the entropy of tossing a fair coin is 1 bit, because all the outcomes are 1 bit length. Probability and information are closely tied together. The former is based on counting occurrences in n trials as n approaches infinity (∞). The latter is based on the average occurrences of the patterns in n trials for very large n [2].

Definition 2.12 *Let X be a discrete random variable. Then the entropy of X, denoted as $H(X)$ is defined as the average information in all the possible outcomes of X produced over n trials as n becomes arbitrarily large.*

Definition 2.13 *If $X = U\Sigma V^T$ and if we let*

$$p_j = \frac{s_j}{\sum s_j} \tag{8}$$

then the singular entropy of X is

$$H(X) = -\sum_{i=1}^{n} p_j \log_2(p_j) \tag{9}$$

in units of bits, where p_j is the individual probability for each x.

2.4   Mutual Information

Given two variables whose entropy [19] can be defined, the amount that one variables information is decreased by knowing the others is known as the mutual information (MI) between the two variables [19, 50]. There is more than one way to measure statistical dependence. While some measures (like the correlation coefficient) are limited and restricted to linear models, others (like mutual information and global correlation coefficient) capture both linear and nonlinear relationship in any given data set without any restrictions on the model. Many authors [33, 22] say that mutual information satisfies some desirable properties of a good measure of dependence. Mutual information can be estimated for both discrete and continuous variables.

Suppose X and Y are both discrete variables. Then the mutual information for these two variables is defined as

$$I(X,Y) = \sum\sum p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{10}$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution of X and Y respectively [44].

18

Mutual information can also be expressed in terms of entropy as

$$I(X,Y) = H(X) - H(X|Y) \tag{11a}$$

$$= H(X) - H(Y|X) \tag{11b}$$

$$= H(X) + H(Y) - H(X,Y) \tag{11c}$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies and $H(X,Y)$ is the joint entropy of X and Y [49]. Statistical dependence occurs between X and Y if and only if $I(X,Y) > 0$, where $H(X) \geq H(X|Y)$. If $I(X,Y) = 0$, then we have statistical independence [44].

2.4.1 Limitations of the Mutual Information

There are inherent difficulties in the estimation and usage of mutual information. The unknown relevant probability density function (p.d.f) is one difficulty. Although approximating densities using histogram sounds like a good way to address this difficulty, doing this poses another problem of underestimating and overestimating the empirical mutual information. Thus authors like Darbellay and Wuertz [21] proposed a method called marginal equiquantization to tackle this problem.

Another difficulty is that the Equation in (10) takes values between 0 and infinity thereby making it difficult to compare different samples. Other articles like [34, 33, 22, 23] proposed a standardized measure of mutual information called global correlation coefficient.

2.5 Takens Theorem and Phase Space Reconstruction

The complexity and nonlinearity inherent in financial time series make it difficult to describe all the features of such a series. Thus an ideal situation is to decompose the original series into state components such that by describing a state component, we in turn describe the original series [40]. The technique used to accomplish this task is the manifold-learning based phase space reconstruction. Manifold learning is the embedding of a one-dimensional time series onto a higher dimensional manifold (Ω) implied by the series [24, 40]. Manifold learning seeks to achieve three goals: extraction of internal structures from the reconstructed phase space, reduction of noise in the data, improvement in the quality of the reconstructed phase space [24]. After the manifold learning is used to embed or map a one-dimensional time series onto a higher dimension, phase space reconstruction (PRS) is used to describe and recover all the features of the original series. The two key elements in PRS are the embedding dimension K and time delay τ [40].

Theorem 2.14 *Takens Theorem*

Suppose $x(t) = v_j(t)$ for some $j = 1, .., n$, where $v(t) = (v_1(t), .., v_n(t))$ is a curve on a manifold Ω. Suppose $v(t)$ visits each part of Ω which means that $v(t)$ is dense in Ω under its topology. Then there exist $\tau > 0$, $K \in Z$ such that the corresponding vectors $(x(t), x(t + \tau), ..., x(t + K\tau))$ are on a manifold topologically equivalent to Ω.

Takens Theorem and PRS can be applied to chaotic systems. Examples of such systems are the Lorenz, Rössler and Hénon systems. We demonstrate this concept

by first considering the Lorenz system

$$
\begin{cases}
\frac{dx}{dt} = \sigma(y - x), \\[2mm]
\frac{dy}{dt} = x(\rho - z) - y, \\[2mm]
\frac{dz}{dt} = xy - \beta z
\end{cases}
\tag{12}
$$

where x, y and z are initial conditions and σ, ρ and β are parameters of the system. A trajectory is a set of state values that are ordered over a given period of time. By varying the values of the initial conditions we obtain different trajectories.

Well known initial conditions are $x = -8.0$, $y = 8.0$ and $z = 27.0$ with standard parametric values of $\sigma = 10$, $\rho = 8/3$ and $\beta = 28$ [3].

To illustrate, we used Takens theorem to reconstruct the Lorenz attractor from $x(t)$ alone. From Figure 1, we can infer that the reconstructed system looks somewhat different from the original but Takens theorem ensures that the reconstructed series retains most of the features of the original system or series.

Original and Reconstructed System of the Lorenz Attractor
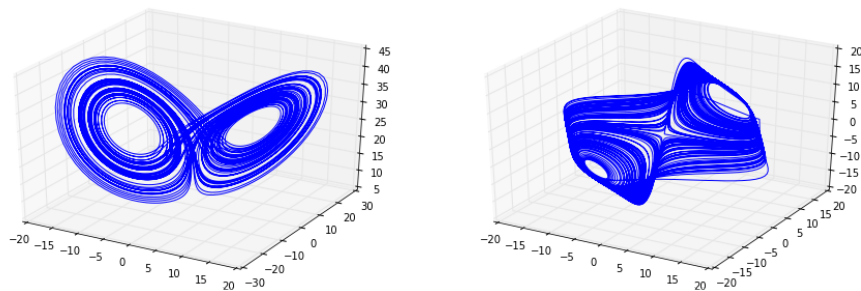
Figure 1: Lorenz Attractor: Original and Reconstructed System

The next system to consider is the Hénon system. This system is made up two parameters: a used to control the amount of stretching and b used to control the thickness of folding. Thus it aims to capture the folding and stretching of dynamic chaotic systems [4]. The Hénon system is given by the following equations

$$
\begin{cases}
x_{n+1} = 1 - a(x_n)^2 + y_n \\
\\
y_{n+1} = bx_n
\end{cases}
\tag{13}
$$

Figure 2 shows the Hénon attractor and a Takens theorem reconstruction from the one dimensional x_n data.

Original and Reconstructed System of the Henon Map
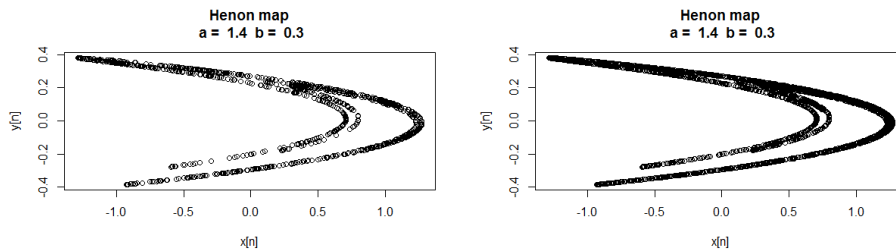
Figure 2: Henon Map: Original and Reconstructed System

A natural question is how do we obtain the most optimal time delay and embedding dimension such that the reconstructed system retains most, if not all, the features of the original system [4].

2.5.1 Ascertaining Time Delay and the Embedding Dimension

Time Delay

The PSR begins by estimating time delay τ. This influences the choice of the embedding dimension. It is a well known fact that if τ is very small then there is redundancy

22

[17]. That is, there would be very high correlation among coordinates causing the embedding to cluster around the diagonal of the embedding space [41]. Similarly, a very high τ leads to irrelevance [53, 41]. In this case, the structure of the reconstructed space will not be revealed [3].

Literature is replete with so many methods in finding a good τ. Notable among them are the autocorrelation function and mutual information methods. In using the autocorrelation function, we can use methods such as First Zero and First.e.decay to ascertain τ [4].

First Minimum of Average Mutual Information (FMoAMI)

This method is the generalization of the autocorrelation function which seeks to bring to bear the needed statistical independence between successive lagged values [4]. This method involves plotting the time-delayed average mutual information versus the delay and identifying the first minimum in the resulting curve [27, 10].

The mathematical equation for finding τ is given as follows:

$$I(\tau) = \sum_{t, t+\tau} P(x_t, x_{t+\tau}) \log_2 \left[\frac{P(x_t, x_{t+\tau})}{P(x_t) P(x_{t+\tau})} \right] \tag{14}$$

where $P(x_t)$ is the individual probability, $P(x_t, x_{t+\tau})$ is the joint probability density function, x_t is any given time series and $x_{t+\tau}$ is the τ delayed version of x_τ. $I(\tau)$ returns a good choice for τ [4]. According to [47, 42, 43], the first minimum of $I(\tau)$ is the time lag which yields the least redundancy or where $x_{t+\tau}$ adds maximal information to the information from x_t. It follows that $I(\tau)$ is both a probabilistic measure and a measure of information about $x_{t+\tau}$ contained in x_t [4].

Embedding Dimension

There are a number of methods in the literature regarding how to find a good em-

bedding dimension. The method of false nearest neighbors is commonly used. This method looks at the extent of close neighbor points in lower dimension and how they do not change over time in the next higher dimension [43]. The inherent problem in this method is the arbitrary selection of some tolerance level. Cao [15] addressed this problem by proposing a modified method called the average false neighbors. Given a good choice of τ, this method produces a good embedding dimension that reduces the prediction error [52].

However, in this work we use the singular value decomposition in finding the embedding dimension K. According to this approach, it was stated in [12, 46] that a sufficient embedding dimension K is the same as the number of linearly independent vectors derived from the columns of trajectory matrix Y. That is, K is the rank of Y. First the trajectory matrix is found using an arbitrarily large K. Then by taking the SVD of this matrix we obtain the singular values in the form $s_1 \geq s_2 \geq ... \geq s_r \geq 0$. Ideally, we choose K to be r, the rank of Y. However, noise often causes r to have a very high value with many singular values close 0. In this thesis, we explore methods for choosing a value of K that is less than r.

3 METHODOLOGY

3.1 The Singular Spectrum Analysis

In this work, we use Taken's Theorem with the basic SSA process to analyse our data. We decompose a one dimensional time series using embedding step and SVD. The SVD of the SSA is used to find both the time delay τ and embedding dimension K. During this process, noise is removed. Then by Takens theorem and PSR, we reconstruct the series. We construct confidence intervals for both the original and the reconstructed series for forecasting purposes.

3.1.1 The Embedding

The embedding step is the first step in the SSA algorithm. As is standard, we begin with an arbitrary large value for the window length K. Then we transform a one-dimensional time series of length N into a K-dimension series of length L. Thus, we obtain a trajectory matrix of size $L \times K$ [32]. The trajectory matrix is of the form:

$$
Y = \begin{bmatrix}
x_0 & x_1 & x_2 & \dots & x_{K-1} \\
x_1 & x_2 & x_3 & \dots & x_K \\
x_2 & x_3 & x_4 & \dots & x_{K+1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
x_{L-1} & x_L & x_{L+1} & \dots & x_{N-1}
\end{bmatrix}
\tag{15}
$$

In (15) we observe that, the matrix has equal elements along the South-West to North-East diagonals ($i + j = const$). This type of matrix is called a *Hankel matrix* [32, 44]. In this work, we introduced a parameter called the time delay τ (or sampling lag) and take the window length K to be the same as the the embedding dimension

d_E [15]. Thus the trajectory matrix will be of the form:

$$Y = \begin{bmatrix} x_0 & x_\tau & \dots & x_{(K-1)\tau} \\ x_1 & x_{1+\tau} & \dots & x_{1+(K-1)\tau} \\ x_2 & x_{2+\tau} & \dots & x_{2+(K-1)\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_N & x_{N+\tau} & \dots & x_{N-(K-1)\tau} \end{bmatrix} \tag{16}$$

When $K = 1$, then we have the form of trajectory matrix expressed in Equation (15). The general case for $K > 1$ is expressed in Equation (16). The first minimum of average mutual information and first left singular vector $u_1(t)$ are both used to ascertain the time delay τ and embedding dimension K is ascertained by SVD.

How to find τ using FMoAMI

Given $x(t)$, how much information on average can be predicted about $x(t + \tau)$? By calculating the mutual information $I(\tau)$ in Equation (14), we answer this question. The interpretation of the graph of $I(\tau)$ is as follows: A graph of $I(\tau)$ begins off very high (given a measurement $x(t)$); we know as many bits as possible about $x(t + 0) = x(t)$), where $\tau = 0$. As τ is increases, $I(\tau)$ decreases, then usually rises again. Fraser and Swinney [27] suggest using the first minimum in $I(\tau)$ to select τ.

How to find τ using the first left singular vector

We use the $u_1(t)$ which is the first left singular vector from the SVD because it is less sensitive to noise. From the previous approach, we know that the $u_1(t)$ provides as much information as possible about the $u_1(t + \tau)$. Therefore the graph of $I(\tau)$ starts at its maximum, as τ increases, $I(\tau)$ decreases and rises again. The first minimum is the value of τ that avoids noise.

3.1.2 The Singular Value Decomposition

We decompose Equation (16) into the Singular Value Decomposition given by Equation (5). More about this step is in [44, 32]. The main goal in this step is to use the singular values from SVD to ascertain the actual embedding dimension K.

3.1.3 Separating the signal from the noise

By applying statistical independence, we separate the signal from the noise. We use a 95% confidence criteria to select the data that would be used for analytical purposes and set 5% as noise. To accomplish this goal, we explore two criteria [44]:

- The ratio-contribution criterion $\left(\frac{s_K}{\sum s_i}\right)$ where s_i is the singular values obtained from $\tilde{\Sigma}$.

- The co-variance criterion $\left(\frac{s_K^2}{\sum s_i^2}\right)$ where s_i is the singular values obtained from $\tilde{\Sigma}$.

3.1.4 Decomposing de-noised data into an approximated trajectory matrix

The de-noised data is now transformed into an approximated trajectory matrix of the form:

$$Y_{app} = \begin{bmatrix} \hat{x}_0 & \hat{x}_\tau & \hat{x}_{2\tau} & ... & \hat{x}_{K\tau} \\ \hat{x}_1 & \hat{x}_{\tau+1} & \hat{x}_{2\tau+1} & ... & \hat{x}_{K\tau+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{\tau-1} & \hat{x}_{2\tau-1} & \hat{x}_{3\tau-1} & ... & \hat{x}_{K\tau-1} \\ \hat{x}_\tau & \hat{x}_{2\tau} & \hat{x}_{3\tau} & ... & \hat{x}_{K\tau} \\ \hat{x}_{\tau+1} & \hat{x}_{2\tau+1} & \hat{x}_{3\tau+1} & ... & \hat{x}_{K\tau+1} \end{bmatrix}. \tag{17}$$

3.1.5 Reconstruction of the original Series

Applying Taken's Theorem and phase space reconstruction, we can reconstruct the manifold from the denoised data. This step involves the decomposition of (17) into several matrices according to the length or size of τ. These matrices are of the following form:

$$
\begin{bmatrix} \hat{x}_0 & \hat{x}_\tau & \hat{x}_{2\tau} & ... \\ \hat{x}_\tau & \hat{x}_{2\tau} & \hat{x}_{3\tau} & ... \\ \hat{x}_{2\tau} & \hat{x}_{3\tau} & \hat{x}_{4\tau} & ... \end{bmatrix}, \begin{bmatrix} \hat{x}_1 & \hat{x}_{\tau+1} & \hat{x}_{2\tau+1} & ... \\ \hat{x}_{\tau+1} & \hat{x}_{2\tau+1} & \hat{x}_{3\tau+1} & ... \\ \hat{x}_{2\tau+1} & \hat{x}_{3\tau+1} & \hat{x}_{4\tau+1} & ... \end{bmatrix} ... \begin{bmatrix} \hat{x}_{\tau-1} & \hat{x}_{2\tau-1} & \hat{x}_{3\tau-1} & ... \\ \hat{x}_{2\tau-1} & \hat{x}_{3\tau-1} & \hat{x}_{4\tau-1} & ... \\ \hat{x}_{3\tau-1} & \hat{x}_{4\tau-1} & \hat{x}_{5\tau-1} & ... \end{bmatrix}.
$$
$$(18)$$

Since these matrices are Hankel matrices, we diagonalize over the averages along the South-West to North-East diagonals [44] to obtain both the averages (which will be the approximated time series) and standard deviations. Another name for this step is the diagonal averaging [32, 44]. The resultant series is an array of approximated time series corresponding to each τ in the form

$$
[\hat{x}_0, \hat{x}_\tau, \hat{x}_{2\tau}, ...], [\hat{x}_1, \hat{x}_{\tau+1}, \hat{x}_{2\tau+1}, ...]...[\hat{x}_{\tau-1}, \hat{x}_{2\tau-1}, \hat{x}_{3\tau-1}, ...].
$$
$$(19)$$

We repeat the same procedures for obtaining standard deviations of the approximated series.

$$
[0, \sigma(\hat{x}_\tau), \sigma(\hat{x}_{2\tau}), ...], [\sigma(\hat{x}_1), \sigma(\hat{x}_{\tau+1}), ...]...[\sigma(\hat{x}_{\tau-1}), \sigma(\hat{x}_{2\tau-1}), ..., 0].
$$
$$(20)$$

The next thing we do is to group all the approximated series in (19) into one-dimensional time series of the same length as the original series. It looks like the following:

$$
[\hat{x}_0, \hat{x}_1, ..., \hat{x}_{\tau-1}, \hat{x}_\tau, \hat{x}_{\tau+1}, ..., \hat{x}_{N-(k-1)\tau}].
$$
$$(21)$$

The approximated standard deviations for the time series will be of the form

$$[0, \sigma(\hat{x}_1), ..., \sigma(\hat{x}_{\tau-1}), \sigma(\hat{x}_\tau), \sigma(\hat{x}_{\tau+1}), ..., 0]. \tag{22}$$

3.1.6 Confidence Interval

A confidence interval supplies bounds within which we can predict future values for the original time series. From (21) and (22), we construct confidence interval bounds for the original time series as well as the approximated series. Based on these bounds we can also forecast the future values of the series.

3.2 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a network model that mimics how the human brain functions. It is widely used in science, mathematics, economics, biology, finance, and elsewhere. Its computational capabilities combine big data analysis, machine learning algorithm, and data mining techniques [29].

A basic ANN is made of three layers: the input layer, the hidden layer and the output layer [4]. Since it is a network, it is made up of nodes called neurons and the connections between the nodes. Signals are transmitted from the input layer to hidden layer. Some weights are assigned to the transmitted signals to increase or decrease their strength depending on the value of the weight. The weighted signals are then transferred to an activation function which eventually gives out the desired output of the signal [1, 4, 29]. Figure 3 explains how the ANN reflects the human brain processes a received signal.
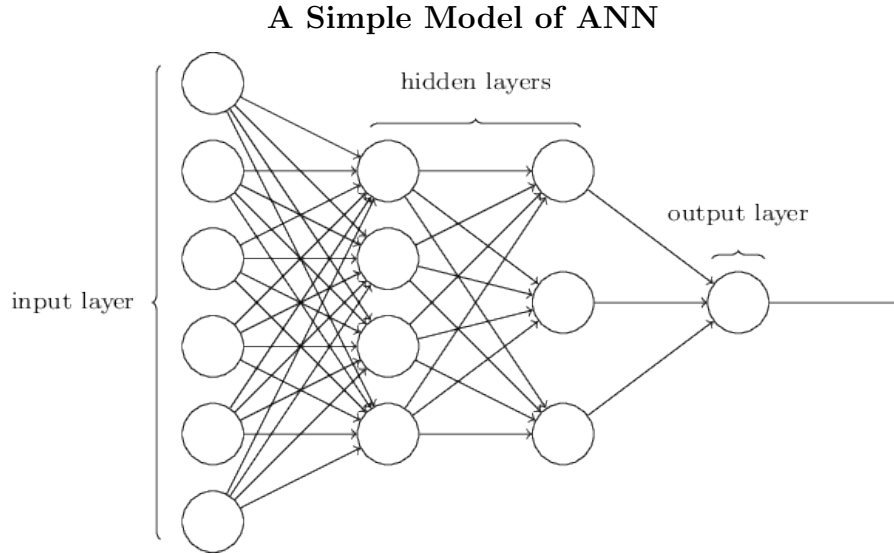
A Simple Model of ANN

Figure 3: A simple model of ANN

3.2.1 The Mathematical Model for ANN

Lets consider the graph in Figure 4

The parameters in the mathematical model are the arbitrary weight assignments denoted as w_{ij}, the transfer function $\sum$ which incorporates some bias term b_k, the activation function φ and the output term O_j. The bias term is sometimes called the threshold term denoted as θ_j. The network model could be single-layer or multiple-layer. The activation function could also includes: linear function, tangent hyperbolic function and sigmoid function (logistics) [45, 1].
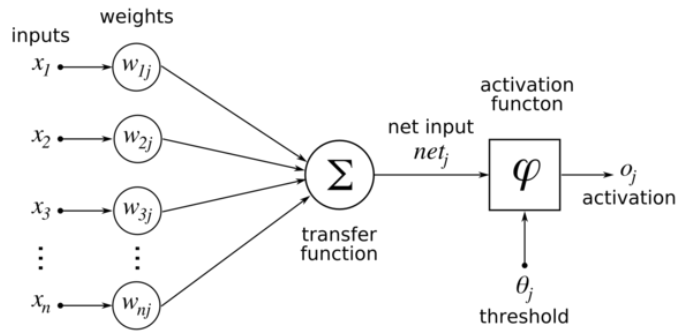
The Mathematical Model of ANN

Figure 4: The mathematical model of ANN

The most used activation function is the sigmoid function defined as

$$\varphi(x) = \frac{1}{1 + e^{-x}}.$$

(23)

We can learn from the ANN algorithm. The most commonly used algorithm for learning from ANN such that weights and bias are adjusted until error is minimal is the backpropagation algorithm. It is also called the feed-forward layered algorithm [29]. It was first introduced in the 1970s but was fully recognized when David Rumelhart, Geoffrey Hinton, and Ronald Williams made their famous publication on it [48].

This algorithm entails the transmission of received signal from inputs down to the hidden layer where weights are assigned. Then the activation function releases the output through the output layer. There is always a predicted output (by the training data) and a desired output. An error occurs when the desired output is different from the predicted output. The gradient descent algorithm is used to update all the weights and biases in the system in order to minimize the error associated with each node in the both the hidden and output layers . The adjusted weights are then fed back into the network system iteratively until the error is zero or minimal [29, 4, 1].

31

It must be noted that since the concept of gradients and derivatives are so useful in optimization, we can find the rate of change of each node in either the hidden or the output layer with respect to corresponding weight assigned. By differentiating the sigmoid function in Equation (23) and simplifying, we have

$$\varphi'(x) = \varphi(x)\left[1 - \varphi(x)\right]. \tag{24}$$

Definition 3.1 *output layer is defined as*

$$\delta_{ok} = \varphi'(v_k)(d_k - y_k) \tag{25}$$

where d_k is the kth desired output, y_k is the kth predicted output and $\varphi'(v_k)$ is the derivative in the sigmoid function corresponding to each node [1].

Definition 3.2 *hidden layer is also defined as*

$$\delta_{hk} = \varphi'(v_k)(\delta_{ok}w_{ij}) \tag{26}$$

where the w_{ij} is the connective weights from the ith node in the input layer to the jth node in the hidden layer [1].

Adjusting both the weights and bias functions includes [1]:

Definition 3.3 *weight function*

$$w(n+1) = w(n) + \alpha w(n-1) + \eta\delta(n)y \tag{27}$$

where $w(n+1)$ is the new weight of the nth iteration, $w(n)$ is the current weight, α is some mobility factor and η is the learning parameter [1].

Definition 3.4 *bias function*

$$b(n+1) = b(n) + \alpha b(n-1) + \eta \delta(n)1 \tag{28}$$

where $b(n+1)$ is the new bias of the nth iteration, $b(n)$ is the current bias, α is some mobility factor and η is the learning parameter [1].

3.2.2 Machine Learning with an ANN

A given data set in a machine learning algorithm is partitioned into training, testing and validation sets. We can also use just the training and the testing sets depending on the classifier. In the ANN algorithm, the training data is used for the training phase of the algorithm. That is, the desired classes are known or determined for each input or observation and the output nodes can be assigned also the desired outputs d_k. This makes it possible to evaluate the difference between the desired outputs and the predicted outputs y_k. Thus, we can find the errors associated with trained data point as $e_k = d_k - y_k$. All the error terms are then fed back into the network to be adjusted by adjusting the corresponding weights. This can be done iteratively until the minimal error is achieved or the desired output is equal to the predicted output [4, 1, 29] .

3.2.3 Stock Market Predicting Using ANN

Most financial time series exhibit nonlinear patterns. Hence the ANN remains the most powerful tool for predicting the stock market. Again, the main merit of using ANN in predicting stock is its ability to handle noisy data and accurately classify patterns in untrained data set [4].

3.2.4 Model Appraisal

It is a common and intuitive practice to assess and appraise a built model. The goal of doing this is to ensure that the application of the model produces good results [4]. The most widely used method for this assessment is the Receiver Operating Characteristics (ROC) curve. This curve can also be used to organize classifiers and visualize the performance. It can be used in decision making [26].
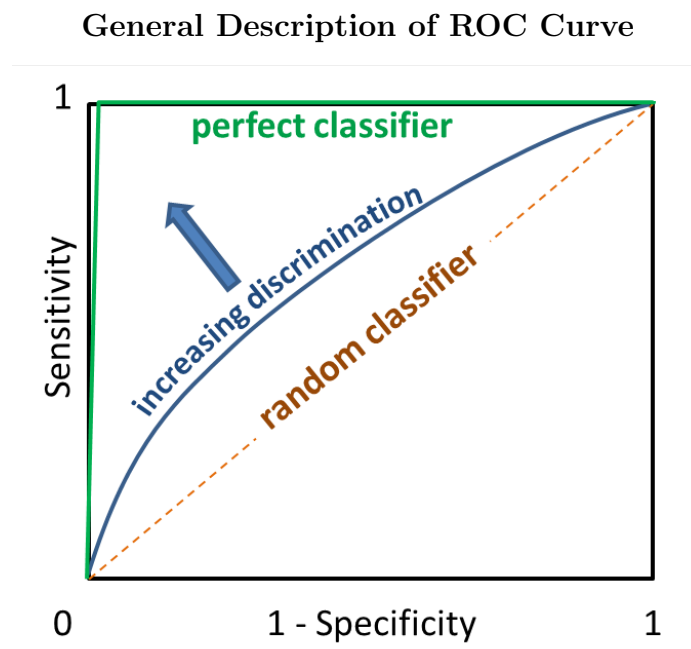
General Description of ROC Curve

Figure 5: ROC Curve for Model Appraisal

An ANN is a discrete classifier which outputs only one class label. This single-point output is obtained from the pair of false positive rate (sensitivity) and true positive rate (1-specificity) as shown in Figure 5. The curve is always bounded by 1 [4]. The broken dashed red lines in Figure 5 is a 45 degree diagonal indicating a random classifier. A perfect or best classifier is obtained when the ROC gives an area under

curve (AUC) of 1. The AUC is the probability of accurately classifying a randomly

chosen observation given that the training set is relatively balanced [28].

4 APPLICATION: DATA AND RESULTS

This section entails the application of our approach to a given data set. We selected a section of Apple Stock from Yahoo! finance. In the first place, we apply the Takens theorem in conjunction with Singular Spectrum Analysis (SSA) to denoise the data. Both the principle of mutual information and first left singular vector $u_1(t)$ from SVD are used to find a suitable sampling lag or time delay τ and the SVD in SSA is used to find the embedding dimension K for the reconstruction of the denoised data.

The data set consists of the daily *Close, Open, High, Low, Volume, Adjusted Close* prices. A little over 5-year consecutive previous values of the data set is used (June 8, 2009 to August 6, 2014) in this work.

The histograms and descriptive statistics of the data set are given in Figure 6-8 and Table 1 respectively.
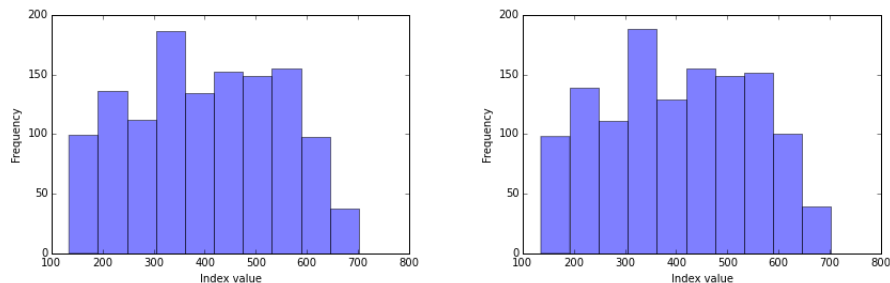
Histograms for Close Price and Open Price

Figure 6: Histograms: Close and Open

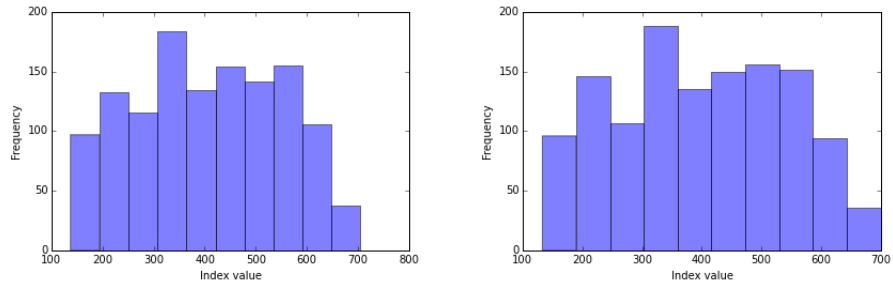**Histograms for High Price and Low Price**

Figure 7: Histograms: Low and High

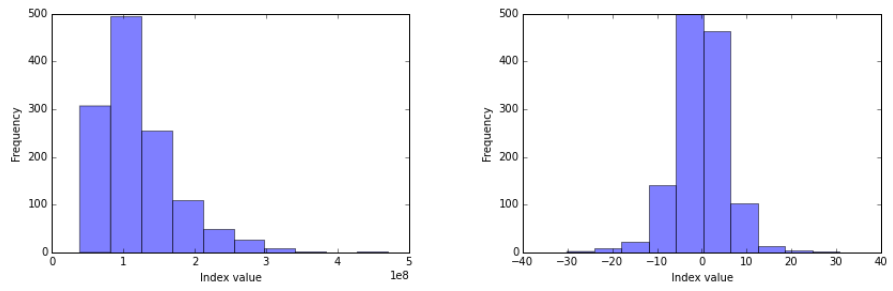**Histograms for Volume and Close Minus Open**

Figure 8: Histograms: Volume and Close Minus Open

Descriptive Statistics

Statistics	N	Mean	St. Dev.	Min	Max
Open	1260	401.75	363.59	135.42	702.41
High	1260	405.41	145.39	136.93	705.07
Low	1260	397.53	144.01	132.88	699.57
Close	1260	401.55	145.34	134.01	702.10
Volume	1260	122092900	356728330	39373600	470249500
Adjusted Close	1260	55.15	20.25	18.21	95.80

4.0.5 Specific Application to Closing Price

We first apply the model to the Closing Price. We begin by applying the Takens Theorem with SSA to find a suitable time delay τ and embedding dimension K. The two graphs in Figure 9 show how mutual information and first left singular vector $u_1(t)$ are both used to find τ.

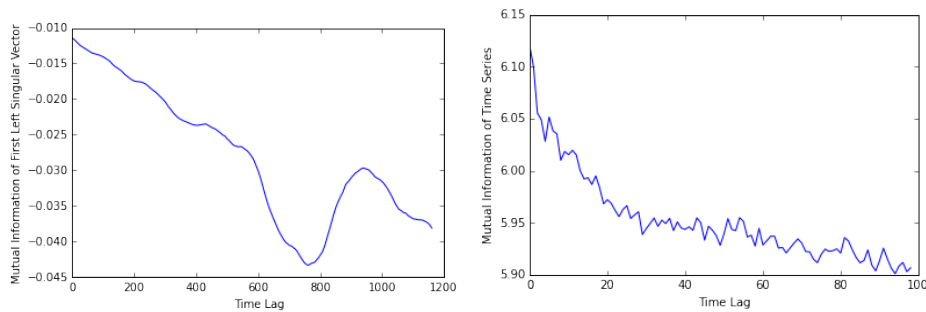**Graphs showing how to find the time lag**

Figure 9: Finding τ using mutual information and first left singular vector

From Figure 9, we find the suitable sampling lag by looking at the first minimum of the two curves. Interesting, they both have their first minimum at 5. Hence our sampling lag is 5. That is, the two curves start off from their maximum located at the mutual information axis. As τ (time lag axis) begins to increase from 0, the curves

of the $I(\tau)$ begin to decrease. The first minimum value of τ as the $I(\tau)$ decrease is 5 (This value corresponds to one of the values on the time lag axis). Now, we use the SVD to find the corresponding embedding dimension K. We choose a large arbitrary value for K, say $K = 100$. The actual K is bounded by the arbitrary value for K. The singular values are arranged in non-decreasing order. Then we use the criteria discussed in section 3.1.3 to estimate K. Thus $K = 4$. In the process of finding the actual K, we end up denoising the data. We proceed to reconstruct the manifold using the parameters that we have found. Figure 10 reveals how the original closing price resembles the reconstructed closing price.
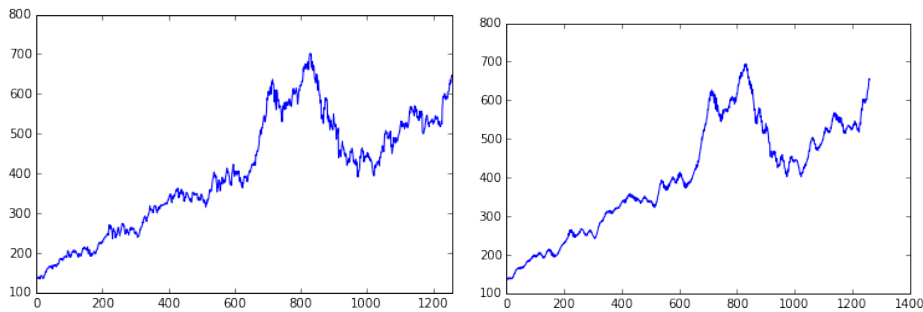
Original and Reconstructed Closing Price

Figure 10: Plot for Original and Reconstructed Closing Price

Finally, we construct the confidence interval to observe patterns within the closing price and prediction purposes.

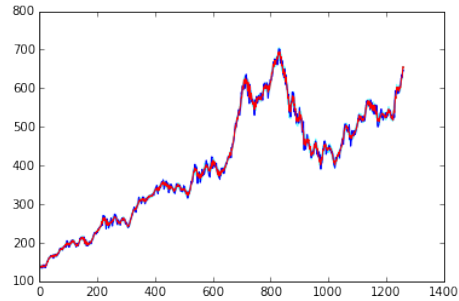**Confidence Interval for Predicting Closing Price**

Figure 11: Confidence Interval for Predicting Closing Price

We apply our model to the Open price and Close Minus Open sections of the data. The sampling lag τ and embedding dimension K for Open price are 7 and 4 respectively. For Close Minus Open, we obtain $\tau = 4$ and $K = 10$. The graphs for Open price and Close Minus Open are under Appendix 1 and 2 respectively.

4.0.6 Results

We evaluate how well our model can denoise the noisy time series, reconstruct the series and give accurate predictions. So, in this section, we only test the performance of the model by using the reconstructed Close Minus Open data. This process of evaluation and assessment involves feeding the reconstructed data into the ANN [4] for training and prediction. The heuristic approach for partitioning the data set is $75 : 15 : 15$. This means that 75% of the data is treated as the training set, 15% of data is used for validating the model and the remaining 15% is used for testing the model [4].

The goal is to predict the next day Close price by looking at the relationship between the Close Price and Open Price. In this work, we define two classes :

- Predict Class 1 if Close > Open.

- Predict Class 0 if Close ≤ Open.

After defining our classes, we select some section of the reconstructed data from Close Minus Open data for our machine learning algorithm. The selection spans from June 8, 2008 to February 28, 2011. This constitutes 600 days of the entire data set. The model for the 600 days produces an area under the curve (AUC) of 0.80. This implies that we have a good model for both denoising and predicting.
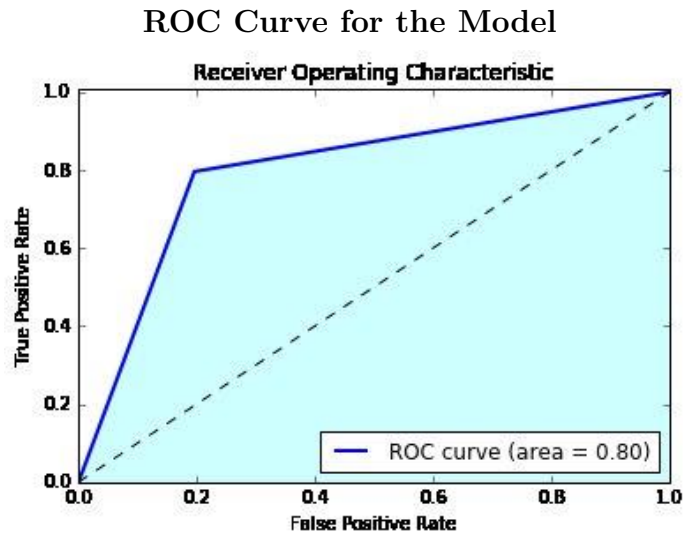
ROC Curve for the Model

Figure 12: ROC curve for the Model

5   CONCLUSION AND REMARKS

The financial market offers a lot of prospects for all stakeholders including brokers, investors, agents and others. The recent economic meltdown in 2008 caused more harm than good to this sector of the economy. Many financial institutions are becoming more global and thus financial data increasingly is becoming complex and noisy. The need for a good and accurate model to predict future values or price of financial assets is becoming a great concern to all stakeholders and players.

The traditional linear models for noisy data are limited in both scope and application. Thus we need a model that captures both linear and non-linear patterns in data without necessarily assuming any structure. Thanks to the novel basic SSA, we can effectively handle noisy data. In this work, we apply Taken's theorem with SSA to denoise a noisy data. While removing the noise, we find the sampling lag τ via mutual information and first left singular vector and the embedding dimension K by SVD.

We reconstruct the series based on the assumption that the error term (noise) is independent of the series. Thus the mutual information between the reconstructed data and the error term is zero. The Taken's theorem guarantees that we are able to reconstruct the denoised series to have at least the same properties as the original series. We construct confidence interval based on the reconstructed data and the original series. A 95% confidence interval indicates the bounds within which we can predict the next day value.

After subjecting our model through an evaluation process, we observe that our model is a good one (an AUC of 0.80). This implies that using Taken's theorem with

SSA we can effectively denoise any given noisy data and make good predictions about the future prices.

The following include some remarks, observations and future work.

1. We only evaluated the model based on 600 data points instead of the entire 1260 data points. The reason for this choice is to create cases for future study. We can compare models based on just few data points and the entire data points.

2. We only used the covariance-approach in our criteria for denoising the data. The other criteria include ratio-contribution approach and mutual information.

3. Future work could be to compare the best model from using all three criteria of denoising noisy data:

    - ratio-contribution criterion $\left(\frac{s_K}{\sum s_i}\right)$ where s_i is the singular values obtained from $\tilde{\Sigma}$.

    - co-variance criterion $\left(\frac{s_K^2}{\sum s_i^2}\right)$ where s_i is the singular values obtained from $\tilde{\Sigma}$.

    - mutual information used in finding K.

BIBLIOGRAPHY

[1] *Backpropation algorithm.* https://www.youtube.com/watch?v=I2I5ztVfUSE. Accessed: 03-14-2016.

[2] *Maximum Entropy Principle and Information.* https://www.youtube.com/watch?v=a06NGTXcrsw. Accessed: 02-23-2016.

[3] *Tutorials on Time Embedding Dimension.* http://www.node99.org/tutorials/ar/. Accessed: 02-23-2016.

[4] A.-S. B. ABDULAI, *Predicting intraday financial market dynamics using takens' vectors; incorporating causality testing and machine learning techniques,* (2015).

[5] R. ADHIKARI AND R. AGRAWAL, *An introductory study on time series modeling and forecasting,* arXiv preprint arXiv:1302.6613, (2013).

[6] A. G. AKRITAS AND G. I. MALASCHONOK, *Applications of singular-value decomposition (svd),* Mathematics and computers in simulation, 67 (2004), pp. 15–31.

[7] M. R. ALLEN AND L. A. SMITH, *Monte carlo ssa: Detecting irregular oscillations in the presence of colored noise,* Journal of Climate, 9 (1996), pp. 3373–3404.

[8] Z. BAI, W.-K. WONG, AND B. ZHANG, *Multivariate linear and nonlinear causality tests,* Mathematics and Computers in Simulation, 81 (2010), pp. 5–17.

44

[9] K. Baker, *Singular value decomposition tutorial*, The Ohio State University, 24 (2005).

[10] A. Basharat and M. Shah, *Time series prediction by chaotic modeling of nonlinear dynamical systems*, in Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 1941–1948.

[11] D. Broomhead, R. Jones, G. King, and E. Pike, *Singular system analysis with application to dynamical systems*, Chaos, noise and fractals, (1987), pp. 15–27.

[12] D. Broomhead, R. Jones, and G. P. King, *Comment on singular-value decomposition and embedding dimension*, Physical review A, 37 (1988), p. 5004.

[13] D. Broomhead and G. P. King, *On the qualitative analysis of experimental dynamical systems*, Nonlinear phenomena and chaos, 113 (1986), p. 114.

[14] D. S. Broomhead and G. P. King, *Extracting qualitative dynamics from experimental data*, Physica D: Nonlinear Phenomena, 20 (1986), pp. 217–236.

[15] L. Cao, *Practical method for determining the minimum embedding dimension of a scalar time series*, Physica D: Nonlinear Phenomena, 110 (1997), pp. 43–50.

[16] L. Cao and A. S. Soofi, *Nonlinear deterministic forecasting of daily dollar exchange rates*, International Journal of Forecasting, 15 (1999), pp. 421–430.

[17] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, *State space reconstruction in the presence of noise*, Physica D: Nonlinear Phenomena, 51 (1991), pp. 52–98.

[18] J. H. Cochrane, *Time series for macroeconomics and finance*, Manuscript, University of Chicago, (2005).

[19] T. M. Cover and J. A. Thomas, *Information theory and statistics*, Elements of Information Theory, (1991), pp. 279–335.

[20] D. Danilov and A. Zhigljavsky, *Principal components of time series: the caterpillarmethod*, St. Petersburg: University of St. Petersburg, (1997), pp. 1–307.

[21] G. A. Darbellay and D. Wuertz, *The entropy as a tool for analysing statistical dependences in financial time series*, Physica A: Statistical Mechanics and its Applications, 287 (2000), pp. 429–439.

[22] A. Dionisio, R. Menezes, and D. A. Mendes, *Mutual information: a measure of dependency for nonlinear time series*, Physica A: Statistical Mechanics and its Applications, 344 (2004), pp. 326–329.

[23] A. Dionísio, R. Menezes, and D. A. Mendes, *Entropy-based independence test*, Nonlinear Dynamics, 44 (2006), pp. 351–357.

[24] J.-P. Eckmann and D. Ruelle, *Fundamental limitations for estimating dimensions and lyapunov exponents in dynamical systems*, Physica D: Nonlinear Phenomena, 56 (1992), pp. 185–187.

[25] J. B. Elsner and A. A. Tsonis, *Singular spectrum analysis: a new tool in time series analysis*, Springer Science & Business Media, 2013.

[26] T. Fawcett, *Roc graphs: Notes and practical considerations for researchers*, Machine learning, 31 (2004), pp. 1–38.

[27] A. M. Fraser and H. L. Swinney, *Independent coordinates for strange attractors from mutual information*, Physical review A, 33 (1986), p. 1134.

[28] J. Garland and E. Bradley, *Predicting computer performance dynamics*, in Advances in Intelligent Data Analysis X, Springer, 2011, pp. 173–184.

[29] C. Gershenson, *Artificial neural networks for beginners*, arXiv preprint cs/0308031, (2003).

[30] M. Ghil and C. Taricco, *Advanced spectral analysis methods*, Past and present variability of the solar-terrestrial system: measurement, data analysis and theoretical models, (1997), pp. 137–159.

[31] M. Ghil and R. Vautard, *Interdecadal oscillations and the warming trend in global temperature time series*, Nature, 350 (1991), pp. 324–327.

[32] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of structure of time series: Ssa and related techniques*, 2001.

[33] C. Granger and J.-L. Lin, *Using the mutual information coefficient to identify lags in nonlinear models*, Journal of time series analysis, 15 (1994), pp. 371–384.

[34] C. Granger, E. Maasoumi, and J. Racine, *A dependence metric for possibly nonlinear processes*, Journal of Time Series Analysis, 25 (2004), pp. 649–669.

[35] H. HASSANI, *Singular spectrum analysis: methodology and comparison*, Journal of Data Science, 5 (2007), pp. 239–257.

[36] H. HASSANI, A. DIONISIO, AND M. GHODSI, *The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets*, Nonlinear Analysis: Real World Applications, 11 (2010), pp. 492–502.

[37] H. HASSANI AND D. THOMAKOS, *A review on singular spectrum analysis for economic and financial time series*, Statistics and its Interface, 3 (2010), pp. 377–397.

[38] K. W. HIPEL AND A. I. MCLEOD, *Time series modelling of water resources and environmental systems*, vol. 45, Elsevier, 1994.

[39] D. A. HSIEH, *Chaos and nonlinear dynamics: application to financial markets*, The journal of finance, 46 (1991), pp. 1839–1877.

[40] Y. HUANG AND G. KOU, *Manifold learning-based phase space reconstruction for financial time series.*, in PACIS, 2014, p. 364.

[41] R. J. HYNDMAN, *Cran task view: Time series analysis*, (2015).

[42] H. KANTZ AND T. SCHREIBER, *Nonlinear time series analysis*, vol. 7, Cambridge university press, 2004.

[43] R. KŘÍŽ, *Finding chaos in finnish gdp*, International Journal of Automation and Computing, 11 (2014), pp. 231–240.

[44] R. Menezes, A. Dionísio, and H. Hassani, *On the globalization of stock markets: An application of vector error correction model, mutual information and singular spectrum analysis to the g7 countries*, The Quarterly Review of Economics and Finance, 52 (2012), pp. 369–384.

[45] C. Özkan and F. S. Erbek, *The comparison of activation functions for multi-spectral landsat tm image classification*, Photogrammetric Engineering & Remote Sensing, 69 (2003), pp. 1225–1234.

[46] M. Paluš and I. Dvořák, *Singular-value decomposition in attractor reconstruction: pitfalls and precautions*, Physica D: Nonlinear Phenomena, 55 (1992), pp. 221–234.

[47] M. Qing-Fang, P. Yu-Hua, and X. Pei-Jun, *A new method of determining the optimal embedding dimension based on nonlinear prediction*, Chinese Physics, 16 (2007), p. 1252.

[48] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Cognitive modeling, 5 (1988), p. 1.

[49] C. Shannon, *Ba mathematical theory of communication,[bell system tech*, 1948.

[50] C. E. Shannon, *The mathematical theory of communication. 1963.*, MD computing: computers in medical practice, 14 (1996), pp. 306–317.

[51] A. S. Soofi and L. Cao, *Nonlinear forecasting of noisy financial data*, in Modelling and Forecasting Financial Data, Springer, 2002, pp. 455–465.

[52] A. S. Soofi, A. Galka, Z. Li, Y. Zhang, and X. Hui, *Applications of methods and algorithms of nonlinear dynamics in economics and finance*, in Complexity in Economics: Cutting Edge Research, Springer, 2014, pp. 1–30.

[53] F. Strozzi, J.-M. Zaldivar, and J. P. Zbilut, *Application of nonlinear time series analysis techniques to high-frequency currency exchange data*, Physica A: Statistical Mechanics and its Applications, 312 (2002), pp. 520–538.

[54] R. Vautard and M. Ghil, *Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series*, Physica D: Nonlinear Phenomena, 35 (1989), pp. 395–424.

[55] R. Vautard, P. Yiou, and M. Ghil, *Singular-spectrum analysis: A toolkit for short, noisy chaotic signals*, Physica D: Nonlinear Phenomena, 58 (1992), pp. 95–126.

[56] P. Yiou, E. Baert, and M.-F. Loutre, *Spectral analysis of climate data*, Surveys in Geophysics, 17 (1996), pp. 619–663.

[57] P. Yiou, D. Sornette, and M. Ghil, *Data-adaptive wavelets and multi-scale singular-spectrum analysis*, Physica D: Nonlinear Phenomena, 142 (2000), pp. 254–290.

APPENDIX

0.1   Appendix A: Specific Application to Open Price
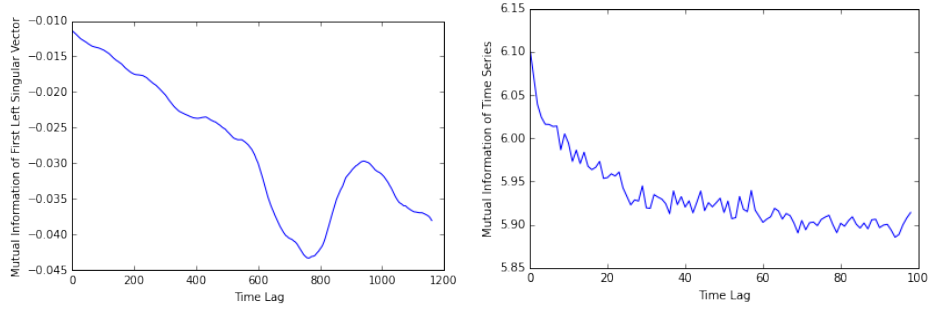
Graphs showing how to find time lag

Figure 13: Finding τ using mutual information and first left singular vector

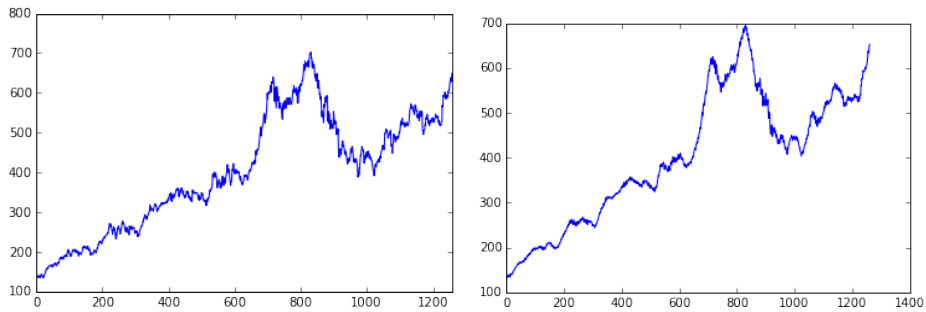**Original and Reconstructed Open Price**

Figure 14: Plot for Original and Reconstructed Open Price
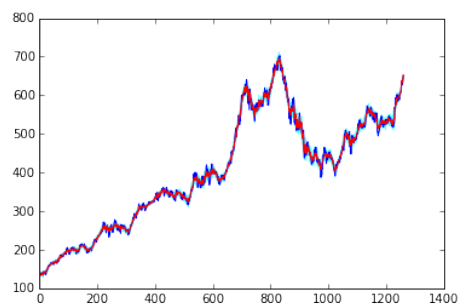
Confidence Interval for Predicting Open Price

Figure 15: Confidence Interval for Predicting Open Price

0.2 Appendix B: Specific Application to Close Minus Open

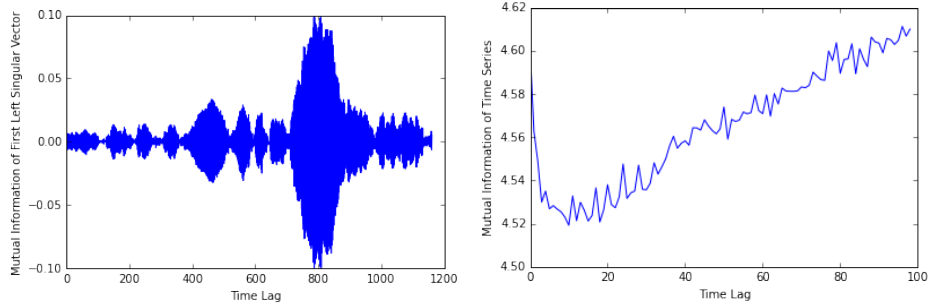### Graphs showing how to find time lag

Figure 16: Finding τ using mutual information and first left singular vector

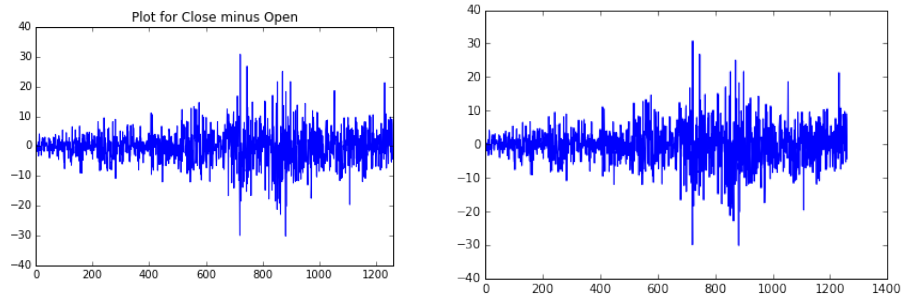### Original and Reconstructed Close Minus Open

Figure 17: Plot for Original and Reconstructed Close Minus Open

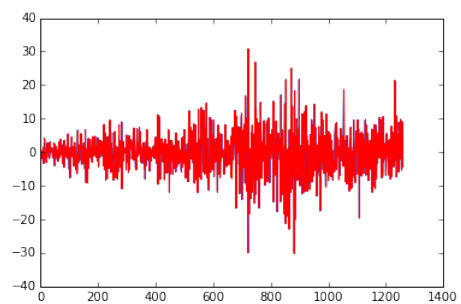**Confidence Interval for Predicting Close Minus Open**

Figure 18: Confidence Interval for Predicting Close Minus Open

VITA

THOMAS KOFI TORKU

Education:	B.S. Mathematics with Business,
	University of Cape Coast,
	Cape Coast, Ghana 2012
	M.S. Mathematics (Applied Mathematics),
	East Tennessee State University,
	Johnson City, Tennessee 2016
Professional Experience:	Graduate Teaching Assistant and Associate,
	East Tennessee State University,
	Johnson City, Tennessee, 2014–present