RESEARCH ARTICLE

Stock price prediction using principal components

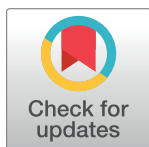**Mahsa Ghorbani**[1]*, **Edwin K. P. Chong**[2]

1 PhD Student, Department of Systems Engineering, Colorado State University, Fort Collins, Colorado, United States of America, **2** Professor, Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, Colorado, United States of America

* Mahsa.Ghorbani@colostate.edu

Abstract

The literature provides strong evidence that stock price values can be predicted from past price data. Principal component analysis (PCA) identifies a small number of principle components that explain most of the variation in a data set. This method is often used for dimensionality reduction and analysis of the data. In this paper, we develop a general method for stock price prediction using time-varying covariance information. To address the time-varying nature of financial time series, we assign exponential weights to the price data so that recent data points are weighted more heavily. Our proposed method involves a dimension-reduction operation constructed based on principle components. Projecting the noisy observation onto a principle subspace results in a well-conditioned problem. We illustrate our results based on historical daily price data for 150 companies from different market-capitalization categories. We compare the performance of our method to two other methods: Gauss-Bayes, which is numerically demanding, and moving average, a simple method often used by technical traders and researchers. We investigate the results based on mean squared error and directional change statistic of prediction, as measures of performance, and volatility of prediction as a measure of risk.

Introduction

Predicting future stock price values is a very challenging task. There is a big body of literature on different methods and different predictors to incorporate into those methods to predict the future values as closely as possible. The literature provides strong evidence that past price/ return data can be used to predict future stock prices. Some studied have found significant auto-correlation for returns over a short period of time. French and Roll find negative correlation for individual securities for daily returns [1]. Some other studies show there is a positive correlation for returns over the period of weeks or months [2]. Studies also demonstrate stock return correlation over the period of multiple months or years. Fama and French report that the auto-correlation is stronger for longer periods, three to five years, compared to daily or weekly periods [3]. Cutler et al. report positive auto-correlation over the horizon of several months and negative auto-correlation over the horizon of three to five years [4]. There are

some other studies that also show correlation in stock returns over a multiple year interval [5, 6] which all confirm that price/return values are predictable from past price/return values.

Bogousslavsky shows that trading by investors with heterogeneous rebalancing horizons can give rise to autocorrelation in the returns at different frequencies [7]. Chowdhury et al. investigate the autocorrelation structure of seven Gulf Cooperation Council (GCC) stock markets. All the markets except for Dubai and Kuwait show significant first-order autocorrelation of returns. They also find that autocorrelation between weekdays is usually larger than that between the first and last trading days of the week [8]. Li et al. study the nonlinear autoregressive dynamics of stock index returns in seven major advanced economies (G7) and China using the quantile autoregression model. For the stock markets in the seven developed economies, the autoregressive parameters generally follow a decreasing pattern across the quantiles with significant portions outside the ordinary least squares estimate intervals [9]. Another study investigates the autocorrelation structure of stock and portfolio returns in the unique market setting of Saudi Arabia [10]. Their results show that there is significantly positive autocorrelation in individual stock and market returns. Another study applies the threshold quantile autoregressive model to study stock return autocorrelations in the Chinese stock market [11]. They report negative autocorrelations in the lower regime and positive autocorrelations in the higher regime.

Other fundamental or macroeconomic factors can also be used in predicting future stock price values. Macroeconomic factors such as interest rates, expected inflation, and dividend can be used in stock return predictions models [3, 12]. Also fundamental variables such as earnings yield, cash flow yield, size and book to market equity [13, 14] have been found to have estimation power in predicting future price/return values.

Silvennoinen and Teräsvirta report correlation between individual U.S. stocks and the aggregate U.S. market [15]. Dennis et al. study the dynamic relation between daily stock returns and daily volatility innovations, and they report negative correlations [16]. Another study investigates the effect of common factors on the relationship among stocks and on the distribution of the investment weights for stocks [17]. They report that market plays a dominant role in both structuring the relationship among stocks and in constructing a well-diversified portfolio. Dimic et al. examine the impact of global financial market uncertainty and domestic macroeconomic factors on stock–bond correlation in emerging markets [18]. In another study, the focus is analyzing the impact of oil price shocks on the interactions of oil-stock prices [19]. The results show that negative changes in oil prices have a significant impact on the stock market.

In this paper, we describe a general method for predicting future stock price values based on historical price data, using time-varying covariance information. When the number of observations is large compared to the number of predictors, the maximum-likelihood covariance estimate [20] or even the empirical covariance is a good estimate of the covariance of the data, but that is not always the case. When the number of observations is smaller than the matrix dimension, the problem is even worse because the matrix is not positive definite [21]. This problem, which happens quite often in finance, gives rise to a new class of estimators such as shrinkage estimators. For example Ledoit and Wolf, shrink the sample covariance towards a scaled identity matrix using a shrinkage coefficient that minimizes the mean squared error of the prediction [22]. Some other studies in this field include [23–25]. In our numerical evaluations in this paper we have sufficient empirical data to reliably track the covariance matrix over time.

Momentum-based forecasting relies on prices following a trend, either upwards or downwards. Based on the assumption that trends like this exist and can be exploited, momentum is used as a heuristic rule for forecasting and is probably the most popular technical indicator

used by traders; in particular, the method of *Direction Movement Index (DMI)*, due to Wilder [26]. This kind of heuristic is a special case of pattern-based forecasting, where, in the case of momentum, the pattern is simply the upward or downward trend. Our method is a systematic method to capture arbitrary patterns, not just upward or downward trends. Indeed, we compute prevalent patterns in the form of eigenvectors (or "eigen-patterns") of the local covariance matrix. As such, we are able to exploit more general patterns that are prevalent (but not necessary known beforehand) in price time series.

The mean squared error (MSE) measures the distance between predicted and real values and is a very common metric to evaluate the performance of predictive methods [27]. Multivariate conditional mean minimizes the mean squared error [28] and is a good estimator for future price values. However, numerical results using this method cannot always be trusted because of associated ill-conditioning issues. In this paper we introduce a method with similar estimation efficiency that does not suffer from this issue.

Principal component analysis (PCA), which is a method for dimensionality reduction of the data, is used in different fields such as statistical variables analysis [29], pattern recognition, feature extraction, data compression, and visualization of high dimensional data [30]. It also has various application in exploring financial time series [31], dynamic trading strategies [32], financial risk computations [32, 33], and statistical arbitrage [34]. In this work, we implement PCA in estimating future stock price values.

Yu et al. introduce a machine-learning method to construct a stock-selection model, which can perform nonlinear classification of stocks. They use PCA to extract the low-dimensional and efficient information [35]. In another study, three mature dimensionality reduction techniques, PCA, fuzzy robust principal component analysis, and kernel-based PCA, are applied to the whole data set to simplify and rearrange the original data structure [36]. Wang et al. present a stochastic function based on PCA developed for financial time-series prediction [37]. In another study, PCA is applied to three subgroups of stocks of the Down Jones Industrial (DJI) index to optimize portfolios [38]. Narayan et al. apply PCA to test for predictability of excess stock returns for 18 emerging markets using a range of macroeconomic and institutional factors [39].

Factor analysis is a technique to describe the variability of observed data through a few factors and is in some sense similar to PCA. There is a long debate in the literature on which method is superior [40, 41]. Factor analysis begins with the assumption that the data comes from a specific model where underlying factors satisfy certain assumptions [42]. If the initial model formulation is not done properly, then the method will not perform well. PCA on the other hand involves no assumption on the form of the covariance matrix. In this paper, we focus on developing an algorithm that can ultimately be used in different fields without prior knowledge of the system, and therefore PCA is the method of choice. In the case study presented in the following section, although only price data is used, it would have been also possible to include multiple predictors to estimate futures values of stock prices.

Our method bears some similarity with subspace filtering methods. Such methods assume a low-rank model for the data [43]. The noisy data is decomposed onto a signal subspace and noise based on a modified singular value decomposition (SVD) of data matrices [44]. The orthogonal decomposition can be done by an SVD of the noisy observation matrix or equivalently by an eigenvalue decomposition of the noisy signal covariance matrix [43].

We compare the performance of our proposed methods in terms of MSE and directional change statistic. Stock-price direction prediction is an important issue in the financial world. Even small improvements in predictive performance can be very profitable [45]. Directional change statistic calculates whether our method can predict the correct *direction* of change in

price values [46]. It is an important evaluation measure of the performance because predicting the direction of price movement is very important in some market strategies.

Another important parameter that we are interested in is standard deviation, one of the key fundamental risk measures in portfolio management [47]. The standard deviation is a statistical measure of volatility, often used by investors to measure the risk of a stock or portfolio.

As mentioned above, in this paper we focus on forecasting stock prices from daily historical price data. In Section, we introduce our technical methodology, and in particular estimation techniques using covariance information. In Section, we describe our method for processing the data and estimating the time-varying covariance matrix from empirical data, including data normalization. We also demonstrate the performance of our method.

Theoretical methodology

Estimation techniques

In this section we introduce a new computationally appealing method for estimating future stock price values using covariance information. The empirical covariance can be used as an estimate of the covariance matrix if enough empirical data is available, or we can use techniques similar to the ones introduced in the previous section, though the time-varying nature of the covariance must be addressed.

Suppose that we are given the stock price values for M days. Our goal is to predict company stock prices for $M + 1$ to N trading days, using the observed values of the past consecutive M days. The reason for introducing N will be clear below.

Gauss-Bayes or conditional estimation of z given y. Suppose that x is a random vector of length N. Let $M \leq N$ and suppose that the first M data points of vector x represent the end-of-day prices of a company stock over the past M consecutive trading days. The multivariate random vector x and can be partitioned in the form

$$x = \begin{bmatrix} y & z \end{bmatrix}. \tag{1}$$

Let random vector y represent the first M data points and z the price of the next $N - M$ days in the future. We wish to estimate z from y.

The covariance matrix for the random vector x can be written as

$$\Sigma_{xx} = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}, \tag{2}$$

where Σ_{yy} is the covariance of y and Σ_{zz} is the covariance of z. Assuming that y and z are jointly normally distributed, knowing the prior distribution of $x = [y, z]$, the Bayesian posterior distribution of z given y is given by

$$\begin{aligned} \hat{z}_{z|y} &= \Sigma_{zy}\Sigma_{yy}^{-1}y \\ \hat{\Sigma}_{z|y} &= \Sigma_{zz} - \Sigma_{zy}\Sigma_{yy}^{-1}\Sigma_{yz}. \end{aligned} \tag{3}$$

The $\hat{\Sigma}_{z|y}$ matrix, representing the conditional covariance of z given y, is also called the Schur complement of Σ_{yy} in Σ_{xx}. Note that the posterior covariance does not depend on the specific realization of y.

The Gauss-Bayes point estimator for the price prediction, the conditional mean $\hat{z}_{z|y}$, minimizes the mean squared error of the estimate in the Gaussian case [28]. Moreover, in the Guassian case, for a specific observation y, the inverse of the conditional covariance is the Fisher

Information matrix associated with estimating z from y, and therefore $\hat{\Sigma}_{z|y}$ is the lower bound on the error covariance matrix for any unbiased estimator of z [28].

The same set of equations arise in Kalman's filtering. Kalman's own view of this process is as a completely deterministic operation [48], and does not rely on assuming normality. Although the point estimator $\hat{z}_{z|y}$ is optimal in term of mean squared error, in practice there are numerical complications involved in this method: The matrix Σ_{yy} is typically not well conditioned, so the numerical calculation of Σ_{yy}^{-1} cannot always be trusted. To overcome this problem, we propose a better conditioned estimator, which has a behavior close to Gauss-Bayes.

Principal components and estimation in lower dimension. Principal component analysis (PCA) is a well-established mathematical procedure for dimensionality reduction of data and has wide applications across various fields. In this work, we consider its application in forecasting stock prices.

Consider the singular value decomposition (SVD) of Σ_{xx}:

$$\Sigma_{xx} = VSV', \tag{4}$$

where S is a diagonal matrix of the same dimension as x with non-negative diagonal elements in decreasing order, and V is a unitary matrix ($VV' = I_N$). The diagonal elements of S are the eigenvalues of Σ_{xx}.

In general, the first few eigenvalues account for the bulk of the sum of all the eigenvalues. The "large" eigenvalues are called the principal eigenvalues. The corresponding eigenvectors are called the principal components.

Let $L < N$ be such that the first L eigenvalues in S account for the bulk part (say 85% or more) of the sum of the eigenvalues. Let V_L be the first L columns of unitary matrix V. Then the random vector x is approximately equal to the linear combination of the first L columns of V:

$$x \approx V_L \alpha, \tag{5}$$

where α is a random vector of length L. Because L is a small number compared to N, Eq (5) suggests that a less "noisy" subspace with a lower dimension than N can represent most of the information. Projecting onto this principle subspace can resolve the ill-conditioned problem of Σ_{yy}. The idea is that instead of including all eigenvalues in representing Σ_{xx}, which vary greatly in magnitude, we use a subset which only includes the "large" ones, and therefore the range of eigenvalues is significantly reduced. The same concept is implemented in speed signal subspace filtering methods, which are based on the orthogonal decomposition of noisy speech observation space onto a signal subspace and a noise subspace [43]. Let $V_{M,L}$ be the first M rows and first L columns of V. We have

$$y = V_{M,L} \alpha + \text{Noise}. \tag{6}$$

Mathematically resolving noisy observation vector y onto the principle subspace can be written as a filtering operation in the form of

$$w = Gy, \tag{7}$$

where G is given by

$$G = (V'_{M,L} V_{M,L})^{-1} V'_{M,L}. \tag{8}$$

The vector w is actually the coordinates of the orthogonal projection of y onto the subspace equal to the range of $V_{M,L}$. We can also think of w as an estimate of α based on least squares.

Substituting y by w in (3) leads to a better conditioned set of equations:

$$\begin{aligned}
\hat{z}_{z|w} &= \Sigma_{zw}\Sigma_{ww}^{-1}w \\
\hat{\Sigma}_{z|w} &= \Sigma_{zz} - \Sigma_{zw}\Sigma_{ww}^{-1}\Sigma_{wz},
\end{aligned}$$

(9)

because the condition number of Σ_{ww} is much lower than that of Σ_{yy}, as we will demonstrate later. In (9) we have

$$\Sigma_{zw} = E[zw'] = \Sigma_{zy}G',$$

(10)

and

$$\Sigma_{ww} = E[ww'] = G\Sigma_{yy}G'.$$

(11)

If the posterior distribution of z estimated based on (9) has a similar behavior to the distribution estimated by (3), it can be considered a good substitute for the Gauss-Bayes method. Our numerical results demonstrate that this is indeed the case, which we will show in Section.

Moving average. Technical traders and investors often use technical trading rules, and one of the most popular methods used by technical traders and researchers are the moving average (MA) rules [49, 50]. Satchell investigates the reason general MA trading rules are widely used by technical analysts [51]. He shows that autocorrelation amplification is one of the reasons such trading rules are popular. Using simulated results, we show that the MA rule may be popular because it can identify the price momentum and is a simple way of assessing and exploiting the price autocorrelation without necessarily knowing its precise structure. Moving average, which is the average of prices over a period of time, is probably the simplest estimator for z:

$$\hat{z}_{\text{MA}} = \frac{1}{K_{MA}}\sum_{i=N-K_{MA}+1}^{N} x_i$$

(12)

where the quantity K_{MA} is the number of data points included to calculate the average, and $\hat{z}_{MA}$ is the average of the most recent K_{MA} price values.

There are different possible values of K_{MA} for calculating the average, from short to medium to long term periods. Here we use periods of 10 and 50 days, which are typical short and mid-term values used in the literature. We will use the moving average estimator for comparison purposes, as we will see in Section below.

Fig 1 shows an example of our stock predictions. Assume that we are given the price values for the past 20 days ($M = 20$), and we want to use those values to predict the future prices over the next 10 business days, from day $M + 1$ to day N ($N = 30$). In our reduced-dimension technique, we can get a relatively smooth plot of the predicted value for a relatively small L, to a plot almost the same as Gauss-Bayes, for larger values of L, as we can can see in Fig 1.

Performance metrics

Mean squared error. To compare the performance of the methods described above, we evaluate the expected value of the squared error between the actual and estimated values. The mean squared error of an estimate $\hat{z}$ is given by:

$$MSE = E\left[\|z - \hat{z}\|^2\right] = E\left[\|z\|^2\right] + E\left[\|\hat{z}\|^2\right] - 2E\left[\|z'\hat{z}\|\right]$$.

The MSE can be expressed in terms of the covariance matrices in (2), by substituting the appropriate form of $\hat{z}$. Alternatively, the mean squared error of an estimator $\hat{z}$ can be written
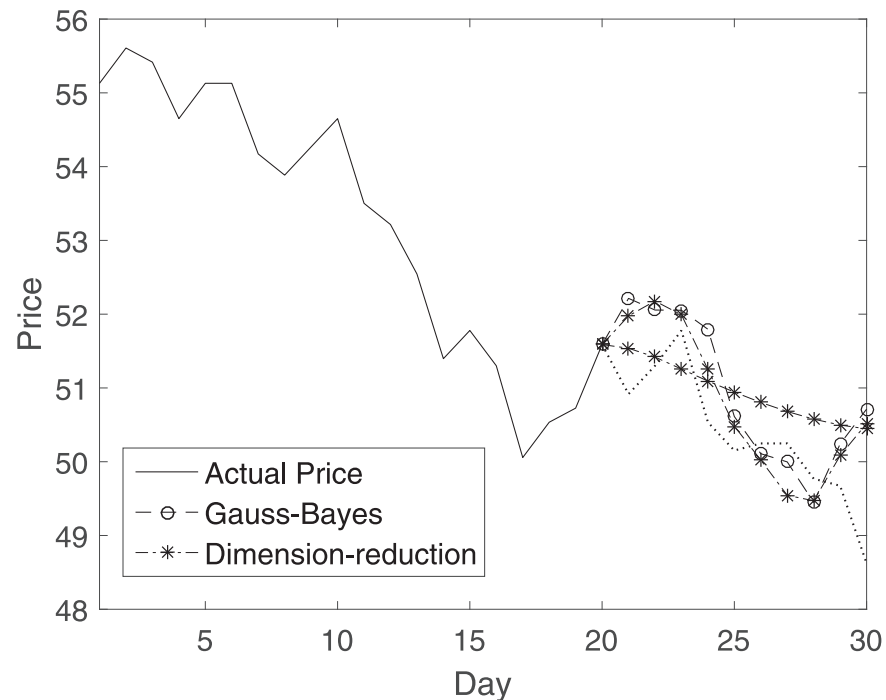
Fig 1. Predicting price for $M + 1$ to N days, actual price: Solid line, GB: $-o-$, RD: $-^*-$ (two lines, one for a small value of L, and one for a relatively large number).

in terms of the variance of the estimator plus its squared bias. The conditional MSE given x is written as

$$MSE_{\hat{z}|z} = E\big[\|z - \hat{z}\|^2 \mid z \big] = trace(\Sigma_{\hat{z}|z}) + \|E\big[\, \hat{z} \mid z \,\big] - z\|^2.$$

The first term is called the variance, and the second term is the squared bias. The expected value of MSE over all observations is the actual MSE, which can be calculated by taking expectations on both sides:

$$MSE = E\Big[E\big[\|z - \hat{z}\|^2 \mid z \big] \Big] = trace(E\|\Sigma_{\hat{z}|z}\|) + E\big[\|E\big[\, \hat{z} \mid z \,\big] - z\|^2 \big]. \tag{13}$$

It turns out that Gauss-Bayes estimator is unbiased, which means that the second term is 0, while the proposed reduced-dimension methods is a biased estimator.

Directional change statistic.

$$b_{ij} = \begin{cases} 1, & \text{if } (z_{ij} - z_0)(\hat{z}_{ij} - z_0) > 0 \\ 0, & \text{otherwise} \end{cases}. \tag{14}$$

Then D_j, the direction statistic for day j, averaged over K samples, is equal to

$$D_j \quad = \frac{1}{K} \sum_{i=1}^{K} b_{ij}, \tag{15}$$

which is a number between 0 and 1 (the higher the better).

Empirical methodology and results

In this section we describe how we estimate the covariance matrix based on a normalized data set, and we evaluate the performance of our method using empirical data.

General setting

Suppose that we have K samples of vector data, each of length N, where $N < K$. Call these row vectors $x_1, x_2, \ldots, x_K$, where each $x_i \in \mathbb{R}^N (i = 1, \ldots, K)$ is a row vector of length N:

$$x_i = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{iN} \end{bmatrix}. \tag{16}$$

We assume that the vectors $x_1, x_2, \ldots, x_K$ are drawn from the same underlying distribution. We can stack these vectors together as rows of a $K \times N$ matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ x_{K1} & x_{K2} & \cdots & x_{KN} \end{bmatrix}.$$

Let $M \le N$ and suppose that we are given a vector $y \in \mathbb{R}^M$ representing the first M data points of a vector we believe is drawn from the same distribution as $x_1, x_2, \ldots, x_K$. Again, these M data points represent the end-of-day prices of a company stock over the past M consecutive trading days. Let z be the price of the next $N - M$ days in the future. We wish to estimate z from y.

Since the vector x_i is a multivariate random vector that can be partitioned in the form

$$x_i = \begin{bmatrix} y_i & z_i \end{bmatrix}, \tag{17}$$

where y_i has length M and z_i has length $N - M$, accordingly the data matrix X can be divided into two sub-matrices Y and Z as follow:

$$X = \begin{bmatrix} Y & Z \end{bmatrix}.$$

We can think of Y as a data matrix consisting of samples of historical data, and Z as a data matrix consisting of the corresponding future values of prices.

Normalizing and centering the data

In the case of stock-price data, the vectors $x_1, x_2, \ldots, x_K$ might come from prices spanning several months or more. If so, the basic assumption that they are drawn from the same distribution may not hold because the value of a US dollar has changed over time, as a result of inflation. To overcome this issue, a scaling approach should be used to meaningfully normalize the prices (we will deal with the time-varying nature of the covariance later). One such approach is presented here. Suppose that $t_i = [t_{i1}, t_{i2}, \ldots, t_{iN}]$ is a vector of "raw" (unprocessed) stock prices over N consecutive trading days. Suppose that $Q \le N$ is also given. Then we apply the following normalization to obtain x_i:

$$x_i = \frac{t_i}{t_i(Q)}. \tag{18}$$

This normalization has the interpretation that the x_i vector contains stock prices as a fraction of the value on the Qth day, and is meaningful if we believe that the pattern of such fractions over the days $1, \ldots, N$ are drawn from the same distribution. Note that $x_i(Q) = 1$.

We believe normalizing the data with this method captures the pattern in the price data better than simply using return data. Although similar to return, the resulting time series still

suffers from being non-stationary over time. We propose to resolve this issue by using a weighting averaging method as explained in the next section.

For the purpose of applying our method based on PCA, we assume that the vectors x_1, x_2, ..., x_K are drawn from the same underlying distribution and that the mean, $\bar{x}$, is equal to zero. However because x_i represents price values, in general the mean is not zero. The mean $\bar{x}$ can be estimated by averaging the vector $x_i \in \mathbb{R}^N (i = 1, \ldots, K)$,

$$\bar{x} = \frac{1}{K} \sum_{i=1}^{K} x_i, \tag{19}$$

and then this average vector is deducted from each x_i to center the data.

Even though this normalization makes the data stationary in the mean, since stock prices are very volatile, there is no guarantee that the covariance of the data would be stationary as well. In order to address this issue, we assign exponential weights $(\gamma^0, \gamma^1, \cdots, \gamma^k)$ to observations, where $0 < \gamma < 1$, to emphasize the most recent periods of data. Using an exponential weighting approach to deal with volatility of financial data has been suggested in multiple studies such as [52]. For each observation x_i, the last K samples prior to that observation are transformed into a Hankel matrix and normalized. Then (decreasing) exponential weights are assigned to the K samples and numerical results are calculated. This process, creating the matrix of data, normalizing, and assigning weights, is repeated for each observation.

To select the value of K we use

$$K = \min \{k : \gamma^k < 10^{-3}\}. \tag{20}$$

Experiments

The daily historical price data for 150 different companies from different market-capitalization categories were downloaded from finance.yahoo.com. Market capitalization is a measure of the company's wealth and refers to the total value of all a company's shares of stock. We randomly select 50 stocks from each of the three market capitalization (cap for short) categories: Big market-cap (125 B\$ to 922 B\$), Mid market-cap (2 B\$ to 10 B\$) and Small market-cap (300 M\$ to 1.2 B\$). The stocks from the Big market-cap category are normally the most stable ones relative to the Small-cap stocks, which have the most volatility. Historical data for four market indexes, S&P500 (GSPC), Dow Jones Industrial Average (DJI), NASDAQ Composite (IXIC), and Russell 2000 (RUT), were also included in this study. The data was transformed into matrices with different sizes as explained in next section. In each case, the daily price value for next 10 days are predicted and the estimation methods are compare based on their out-of-sample performance.

Constructing data matrix

The daily stock price data is transformed into a matrix with K rows, samples of vector data, each of length N. We get that by stacking K rows (K samples), each one time shifted from the previous one, all in one big matrix, called the Hankel matrix.

More precisely, the Hankel matrix for this problem is constructed in the following format:

$$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_K \end{bmatrix} = \begin{bmatrix} P(1) & P(2) & \cdots & P(N) \\ P(2) & P(3) & \cdots & P(N+1) \\ \cdots & \cdots & \cdots & \cdots \\ P(K) & P(K+1) & \cdots & P(K+N-1) \end{bmatrix},$$

where $P(i)$ represents the price for day i. This is our matrix of data, before normalization and centering.

We first normalize each row (observation) by Qth entry, as described earlier, and then subtract the average vector $\bar{x}$ from each row. The prediction is done using the processed data. After doing the prediction, we add back the average vector $\bar{x}_{N-M}$ (last $N - M$ components of $\bar{x}$) from days $M + 1$ through N and also multiply the result by the value of Qth that was used for normalizing to get back to actual stock prices. We tested different values for Q in terms of MSE and estimation variance. For the purpose of this study, we chose $Q = M$ because it shows the best results in this setting. Recall that $x_i(M) = 1$. This column is removed from the data matrix because it does not provide any information. From now on matrix X represents normalized and centered price data.

To account for the nonstationarity of the covariance, we use an exponential averaging method as mentioned before. For this purpose, $\gamma = 0.98$ was selected and the weights smaller than 10^{-3} were considered zero. Then the sample covariance matrix is calculated as

$$\Sigma_{xx} = \left(\frac{1 - \gamma}{1 - \gamma^{k+1}}\right) X' \operatorname{diag}(\gamma^0, \gamma^1, \cdots, \gamma^k) X, \tag{21}$$

where $\operatorname{diag}(\gamma^0, \gamma^1, \cdots, \gamma^k)$ is a diagonal matrix with $(\gamma^0, \gamma^1, \cdots, \gamma^k)$ as the diagonal elements.

We obtained end-of-day stock prices for General Electric and converted this time series into Hankel matrices with different lengths as described above. 2000 samples were used to evaluate the out-of-sample performance of the methods. The values corresponding with the performance metrics presented in this section converge after a few hundred samples. We construct data matrices with 9 different sizes, M from 50 to 530 with a 60 day interval, to investigate the effect of length of observation vector on performance.

Fig 2 shows the histogram of normalized data as a representation of the distribution of normalized data; the curve resembles a bell shape.

MSE performance

Three different estimation methods are implemented for each of the data matrices constructed above. The goal is to predict future price values for the next 10 days (days $M + 1$ to N). when it comes to reduced-dimension method, for each M we try different values of L, the number of principle components. The general goal, as mentioned above, is an estimation technique that has a similar behavior as an ideal Gauss-Bayes estimator but does not have the associated calculation difficulties resulting from ill-conditioning.

We use General Electric price data to calculate the values illustrated in this section. We calculate the squared error (SE) for 2000 samples to evaluate the performance of the methods. We implement our reduced-dimension technique for different Ms, and for different numbers of principal eigenvalues, L.

Fig 3 shows the empirical Cumulative distribution function (CDF) of the SE for 2 different values of M, together with two-standard-deviation confidence interval. Note that to make our comparisons fair and meaningful, we normalized the results from the moving average predictors so that their values are equally normalized with the values from our RD method. When it comes to out-of-sample performance, the numerical complications compromise the estimation accuracy of Gauss-Bayes, causing the SE values for this method to become even worse than the SE plot for the moving average estimators. As we can see, in both plots, our reduced-dimension method is superior to the other two methods. For $M = 110$ some lines are relatively close together. As M gets larger, the plot for the reduced-dimension method improves and the plot for Gauss-Bayes gets worse.
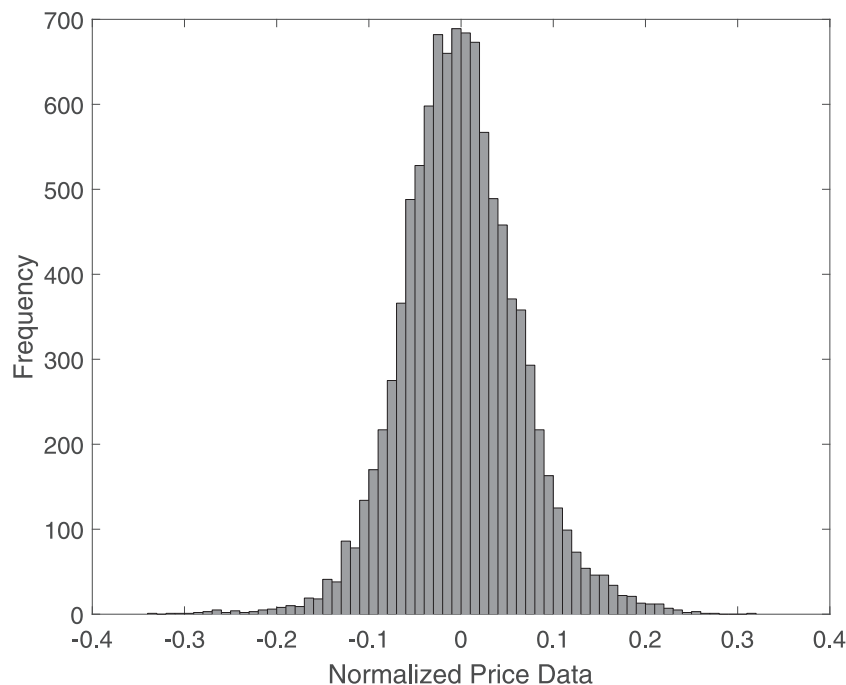
Fig 2. Histogram graph for normalized data.

Another point worth mentioning is that although adding more data improves the performance of our proposed method, that is not the case for the moving average estimator. As the arrow on the plot on the bottom indicates, by adding more data, moving from $z_{MA_{10}}$ to $z_{MA_{50}}$, the performance of the moving average estimator deteriorates. This behavior is expected since the moving average relies on the momentum, in contrast to the reduced-dimension method, which extracts the essence of the information by projecting onto a smaller subspace.

Fig 4 shows the values of MSE over all days of estimation versus the value of L, for 9 different M, lengths of observation vector, from 50 to 530. As we can see, the MSE value is insensitive to the value of L for sufficiently large L. For small values of L, the MSE values fall quickly, but then eventually increase. So if we have a particular constraint on the condition number, we do not lose much in terms of MSE by choosing a reduced-dimension subspace, which leads to a better conditioned problem. After a certain point, adding more data is actually adding noise and the MSE values get worse.

The metric we are looking for is the sum of MSE values over all days of estimation.

For each length of M, the values for MSE are captured based on different constraints of the condition number of Σ_{ww}. The MSE values in the reduced-dimension method are significantly smaller relative to the other two methods.

Fig 5 shows the relative percentage of improvement (RPI) in the reduced-dimension method compared to the other two methods, calculated as

$$RPI_{\text{GB/MA}} = \frac{-100(MSE_{\text{RD}} - MSE_{\text{GB/MA}})}{MSE_{\text{GB/MA}}}. \tag{22}$$

Note that since the denominator in the equation is $MSE_{\text{GB/MA}}$, the improvement percentage does not exceed 100% but the actual MSE values are further apart in absolute terms than illustrated here. For example for $M = 350$, the MSE value for reduced-dimension is between 0.0052
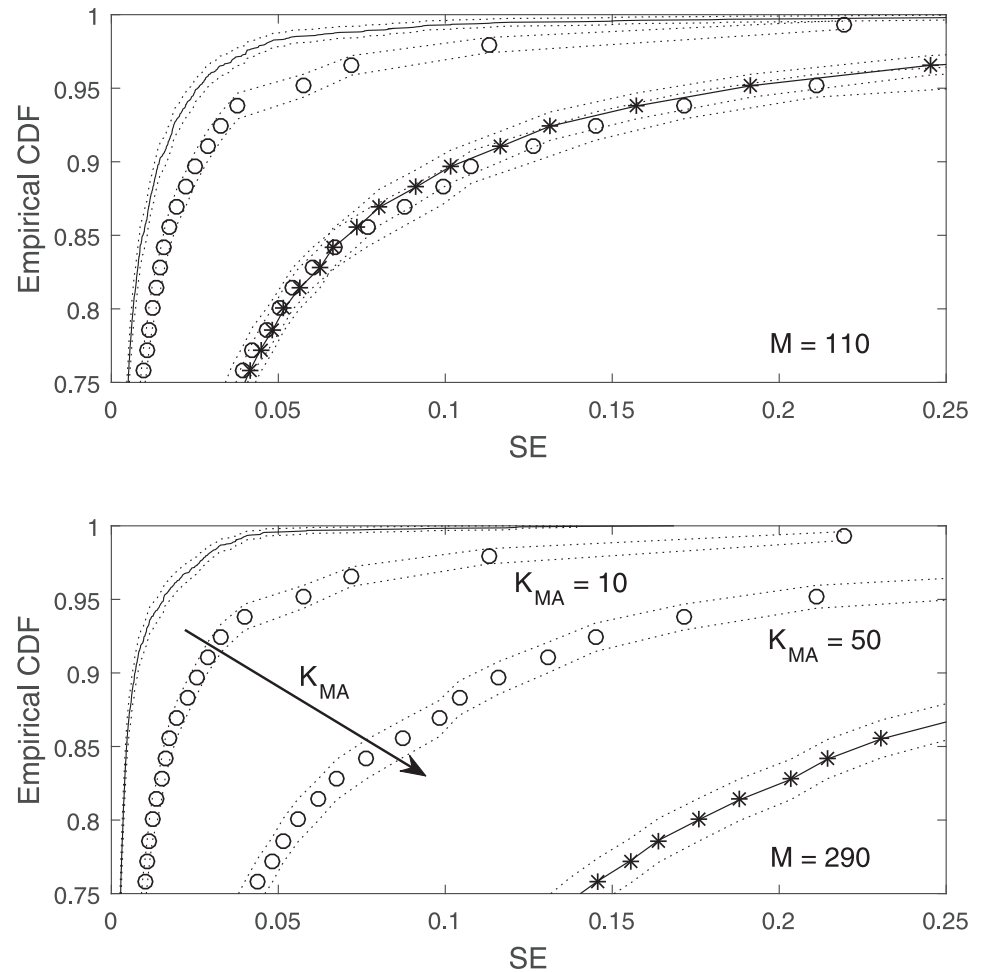
Fig 3. Empirical CDF of SE corresponding to M = 110 and 290. MA_{20} and MA_{50}: $-o-$, GB: $-*-$, RD: Solid lines. Dashed lines illustrate a two standard deviation confidence interval. Plots toward the top and left represent better performance.

https://doi.org/10.1371/journal.pone.0230124.g003

to 0.018, while the MSE in Gauss-Bayes is around 6.33×10^6. The three (overlapping and there-fore appears as only a single plot) lines on top ($-*-$) of Fig 5 compare the reduced-dimension to Gauss-Bayes (RPI_{GB}). The three lines on top ($..o..$) correspond to the comparison of the reduced-dimension and moving average ($RPI_{MA_{50}}$) and the three lines on the bottom ($..o..$) cor-respond to ($RPI_{MA_{10}}$). In each case the three lines are subject to different upper limits on the condition number (10^2, 10^3, and 10^4). It is worth mentioning that the condition number of Σ_{yy} starts from 10^3 for $M = 50$ and goes up to 10^{19} for $M = 530$. The upper limit on the condition number of Σ_{ww} changes from 10^2, associated with the lines on the bottom in each case, to 10^4, the lines on top, for all values of M.

In general, by increasing M, more information is available in each observation, resulting in better performance of the prediction in terms of smallest MSE values. This can be observed easily in the RPI plots in Fig 5 in comparison to the moving average cases since the MSE values in the those cases are almost constant for different values of M. The percent of improvement of MSE values corresponding to the reduced-dimension method increases as M increases. This is as expected since more information is available in each observation, resulting in better
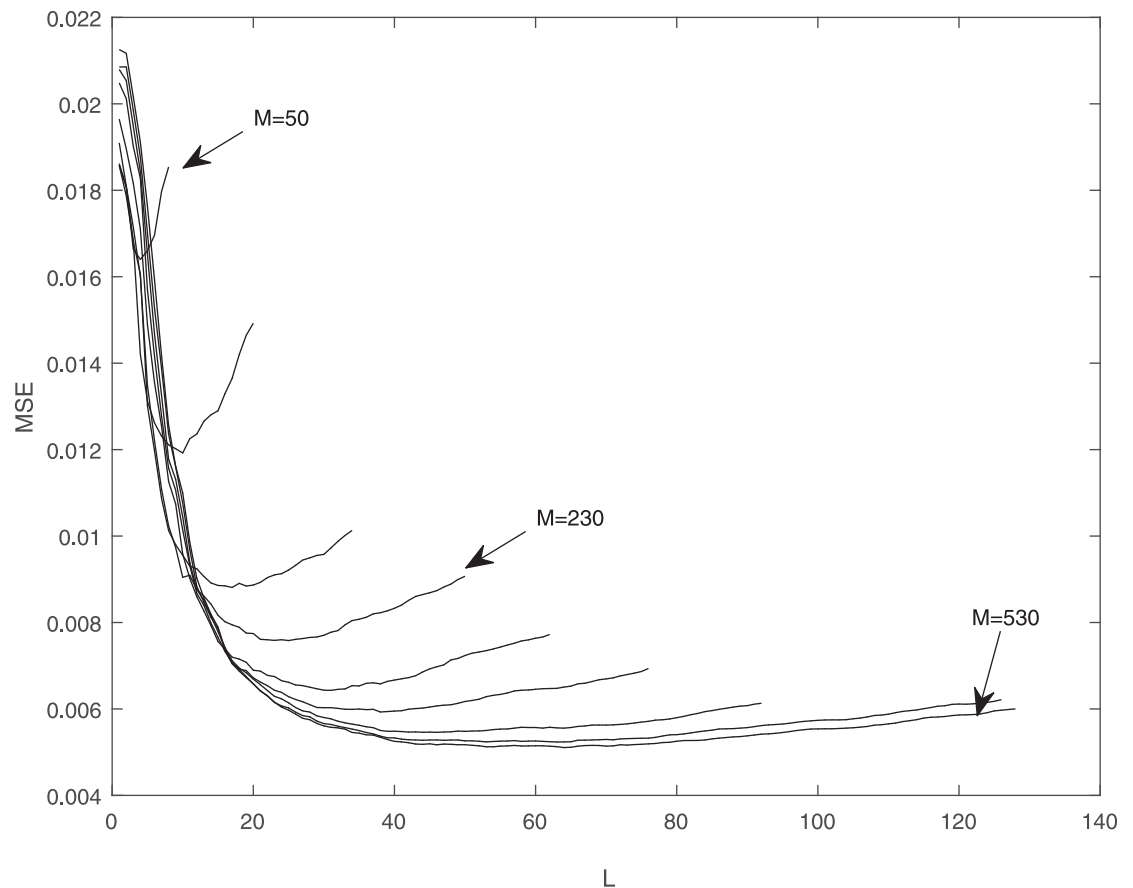
**Fig 4. MSE versus *L* in the normalized domain for different *M*s.**

performance. However after a certain point the RPI flattens out suggesting adding more data at this point is increasing the noise and does not improve the performance.

As we can see, in some cases there is a slight decrease in the improvement rate of the reduced-dimension method compared to the moving average method. A possible explanation for this observation is that when we fix some constraint on condition number, we are actually limiting the value of *L*, and by increasing *M*, after a certain point, we mostly increase the noise, and the MSE value gets worse, which is consistent with Fig 4. Table 1 shows the average RPI values for all stocks in different market-cap categories and average RPI values for market indexes. The reduced-dimension method consistently shows better performance than the other two methods.

Matlab's two-sample t-test function was used to determine the MSE values from our proposed method for 50 stocks in each market-cap category is significantly smaller than the average of the MSE values generated for the same sample using other methods at 5% significance level ($\alpha = 0.05$). When $p < \alpha$ and $h = 1$, the null hypothesis that the two samples have the same mean is rejected, concluding that the difference between the averages of the two sets of samples is statistically significant at α significance level. As shown in Table 2, the results indicate that the average of the MSE values for predictions from our method is significantly smaller than the average of MSE values from other competing methods at 0.05 significance level.

Recall that *L* represents the number of eigenvalues required from the diagonal matrix *S* to represent the bulk part of the information carried in *x*. Fig 6 investigates the dimension of the
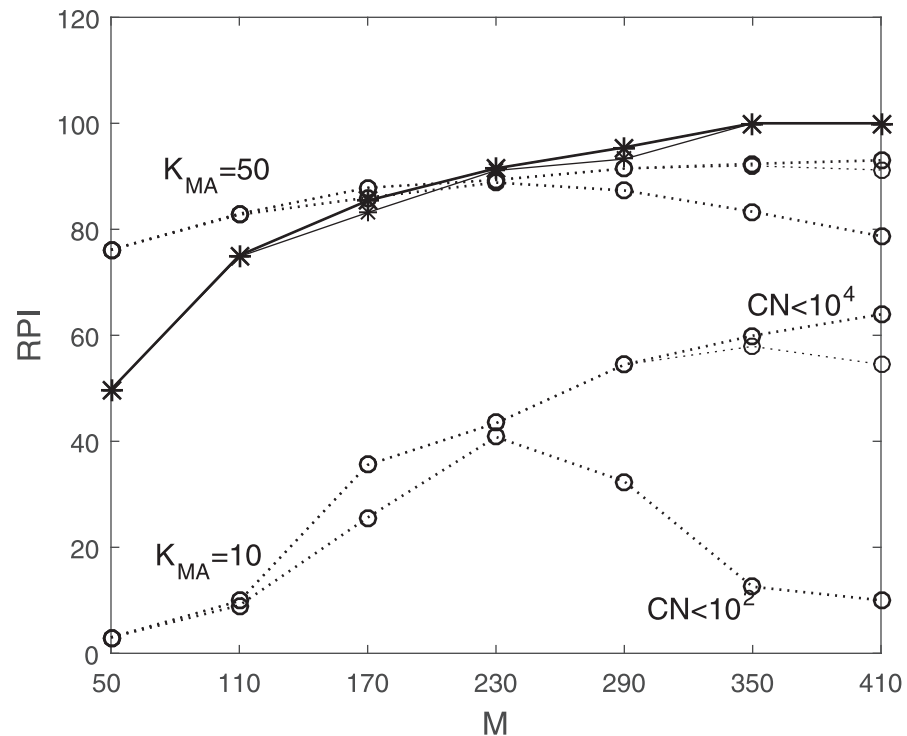
Fig 5. RPI values, subject to different upper limit on condition number of Σ_{ww}, in each case 10^2 associated with the line on the bottom, to 10^4 associated with the line on top, RPI_{MA}: ..o.., RPI_{GB}: −*−. Higher plots represent worse relative performance (relative to RD).

https://doi.org/10.1371/journal.pone.0230124.g005

target subspace by plotting the value of L corresponding to best MSE for different Ms, subject to different limits on condition number (the same case as in Fig 5).

As the upper limit on condition number increases, the value of MSE improves as M increases, and we need a bigger subspace, bigger L, to extract the information. However, as the bottom three plots in Fig 6 show, the value for best L flattens out after a certain point.

Directional change statistic performance

The other evaluation metric that we are interested in is the directional statistic which measures the matching of the actual and predicted values in terms of directional change. Fig 7 shows the average directional statistic over 10 days of estimation using the same $K = 2000$ samples. As the plot indicates, the reduced-dimension method is superior in terms of directional change statistic. It is interesting to note that the directional statistic improves as M increases, and then eventually flattens out, consistent with previous plots.

Table 1. Average RPI values for stocks in different market-cap categories and average RPI values for market indexes ($M = 350$).

MSE	RPI_{GB}	$RPI_{MA_{10}}$	$RPI_{MA_{50}}$
Small-Cap	100%	51%	88%
Mid-Cap	100%	54%	88%
Big-Cap	100%	56%	89%
Market indexes	100%	55%	88%

https://doi.org/10.1371/journal.pone.0230124.t001

Table 2. Statistical analysis for MSE values for stocks in different market cap categories ($M = 350$).

T-test	against	MSE_{GB}	$MSE_{MA_{10}}$	$MSE_{MA_{50}}$
Small-Cap	p-value	0.0024	0.0075	0.00068
	h	1	1	1
Mid-Cap	p-value	0.0283	0.0066	0.000038
	h	1	1	1
Big-Cap	p-value	0.0021	0.00048	0.00001
	h	1	1	1

Table 3 shows the average value for directional statistic for stocks in different market cap categories and indexes for $M = 350$ for Σ_{ww} condition number limited to 10^4. The reduced-dimension method is superior to the other two methods in terms of directional change estimation. It is important to note that the values represented in Table 3 are associated with a specific M for all companies. In practice, it is recommended to tailor the value of M for each company to get the best results.

Matlab's two-sample t-test function was used to determine if the average of the directional statistics from our method for 50 stocks is significantly larger than the average of directional statistics from other methods. Table 4 lists the p-value and h-statistic for each test. The results also indicate that the average of directional statistics from our method is significantly larger than the average of the directional statistics from other competing methods at 5% significance level.

Volatility

Another important parameter that we estimate is the volatility of the prediction, measured in terms of its standard deviation. The square root of the diagonal elements of the estimated
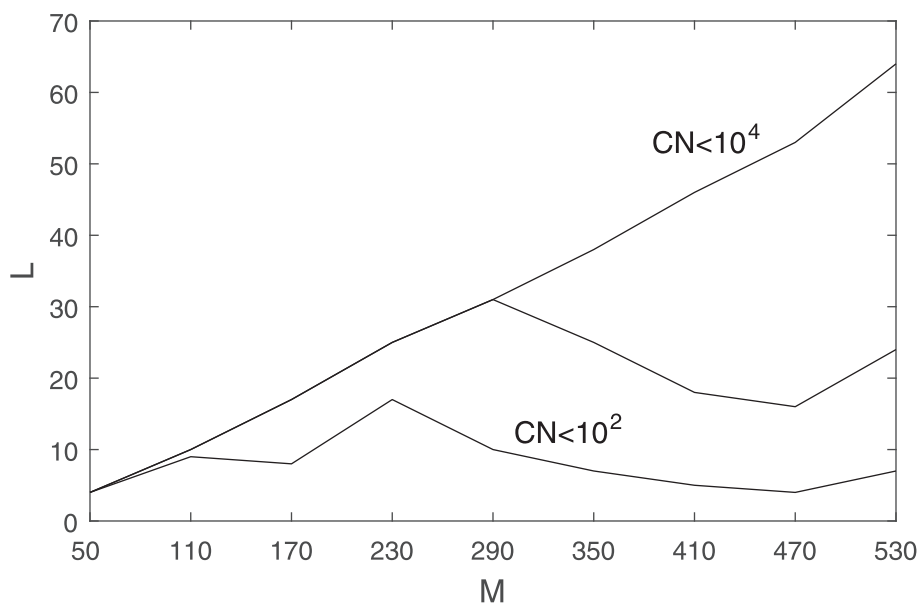

Fig 6. Best L corresponding to best MSE values subject to different limits on condition number, 10^2 associated with the line on the bottom, 10^3 associated with the line in the middle, and 10^4 associated with the line on top.
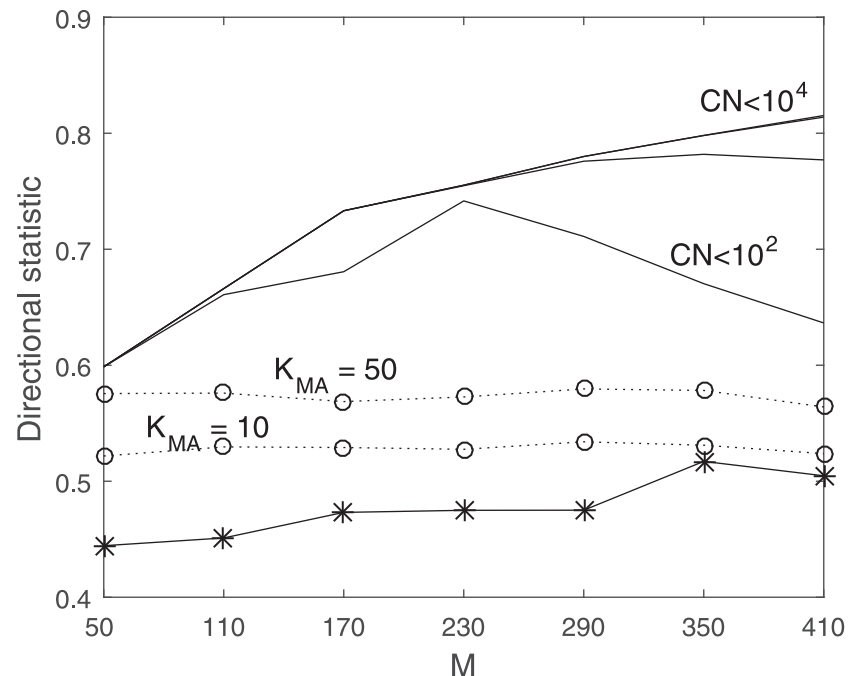
Fig 7. Best directional statistics subject to different upper limit on condition number of Σ_{yy}, 10^2 associated with the line on the bottom, to 10^4 associated with the line on top, GB: $-^*-$, RD: Solid lines, MA: ..o.. (MA$_{10}$ on the bottom and MA$_{50}$ on top). Higher plots represent better performance.

https://doi.org/10.1371/journal.pone.0230124.g007

covariance, $\hat{\Sigma}_{zz}$, are the estimated standard deviations for individual days of estimation. The estimate of the covariance in each method is

$$\begin{aligned}
\hat{\Sigma}_{GB} &= \hat{\Sigma}_{z|y} = \Sigma_{zz} - \Sigma_{zy}\Sigma_{yy}^{-1}\Sigma_{yz}, \\
\hat{\Sigma}_{RD} &= \hat{\Sigma}_{z|w} = \Sigma_{zz} - \Sigma_{zw}\Sigma_{ww}^{-1}\Sigma_{wz},
\end{aligned} \tag{23}$$

However, note that because of the poor conditioning of Σ_{yy}, using the formula above for Σ_{GB} has numerical issues. Hence, we omit their values here. In general the standard deviation values increase moving from day 1 to day 10 of prediction, since less uncertainty is involved in the estimation of stock prices of days closer to the current day. In Fig 8, the standard deviation for individual days of estimation, days 1 to 10, are plotted versus M, the length of observation vector, for the reduced-dimension method. In the reduced-dimension method, the standard deviation values decrease as M increases because more information is provided in each observation. For sufficiently large Ms, the standard deviation values for different days are very close.

Table 3. Average directional statistics for stocks in different market cap categories ($M = 350$).

Directional Statistic	MA$_{10}$	MA$_{50}$	GB	RD
Small-Cap	0.56	0.61	0.51	0.78
Mid-Cap	0.58	0.62	0.51	0.79
Big-Cap	0.60	0.66	0.51	0.80
Market Indexes	0.63	0.70	0.50	0.79

https://doi.org/10.1371/journal.pone.0230124.t003

Table 4. Statistical analysis for directional statistics values for stocks in different market-cap categories ($M = 350$).

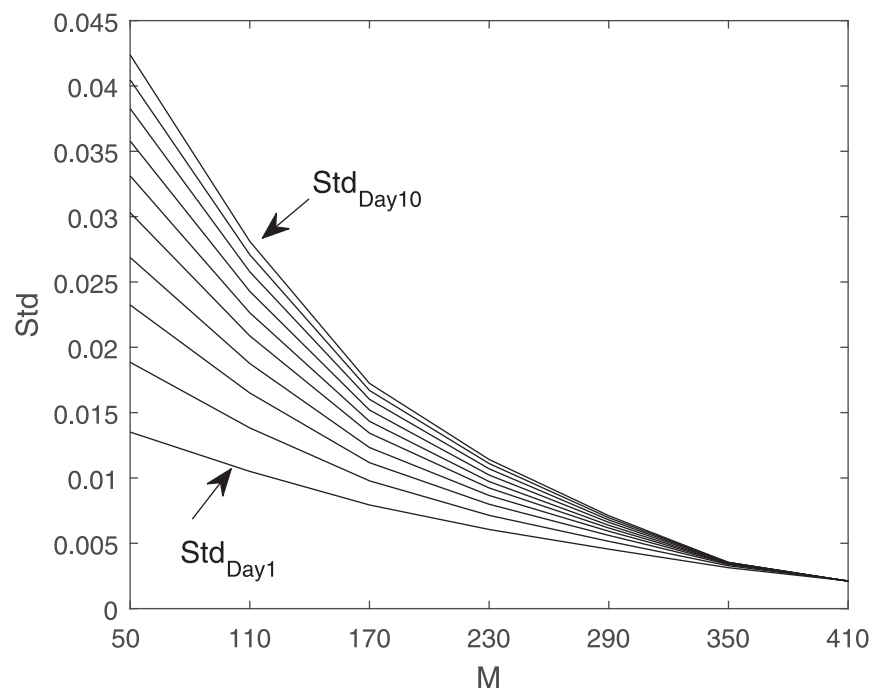| T-test	against	D_{GB}	$D_{MA_{10}}$	$D_{MA_{50}}$
Small-Cap	p-value	$<10^{-10}$	$<10^{-10}$	$<10^{-10}$
	h	1	1	1
Mid-Cap	p-value	$<10^{-10}$	$<10^{-10}$	$<10^{-10}$
	h	1	1	1
Big-Cap	p-value	$<10^{-10}$	$<10^{-10}$	$<10^{-10}$
	h	1	1	1

Fig 8. Standard deviation of individual days of estimation, RD: Solid line.

Conclusion

In this paper we introduced a new method for predicting future stock price values based on covariance information. We develop this method based on a filtering operation using principle components to overcome the numerical complications of conditional mean. We also introduced a procedure for normalizing the data. The matrix of data was constructed in different sizes to investigate the effect of length of observation vector on prediction performance. Our method has showed consistently better out-of-sample performance than Gauss-Bayes (multivariate conditional mean), a numerically challenged estimator, and moving average, an easy to use estimator, for 5 different companies in terms of mean squared error and directional change statistic.

The proposed method can be modified to include multiple predictors. The significance of the proposed approach will be even more apparent when using multiple predictors because where observation vectors are longer it becomes almost impossible to rely on conditional mean due to the severe ill-conditioning of the covariance matrix.

Supporting information

S1 File.
(XLSX)

Author Contributions

Conceptualization: Mahsa Ghorbani, Edwin K. P. Chong.

Data curation: Mahsa Ghorbani.

Formal analysis: Mahsa Ghorbani, Edwin K. P. Chong.

Funding acquisition: Mahsa Ghorbani.

Investigation: Mahsa Ghorbani.

Methodology: Mahsa Ghorbani, Edwin K. P. Chong.

Project administration: Mahsa Ghorbani.

Resources: Mahsa Ghorbani.

Software: Mahsa Ghorbani.

Supervision: Edwin K. P. Chong.

Validation: Mahsa Ghorbani.

Visualization: Mahsa Ghorbani.

Writing – original draft: Mahsa Ghorbani.

Writing – review & editing: Mahsa Ghorbani.

References

1. French KR, Roll R. Stock return variances: The arrival of information and the reaction of traders. Journal of Financial Economics. 1986; 17(1):5–26. https://doi.org/10.1016/0304-405X(86)90004-8

2. Lo AW, MacKinlay AC. Stock market prices do not follow random walks: Evidence from a simple specification test. The Review of Financial Studies. 1988; 1(1):41–66. https://doi.org/10.1093/rfs/1.1.41

3. Fama EF, French KR. Permanent and temporary components of stock prices. Journal of Political Economy. 1988; 96(2):246–273. https://doi.org/10.1086/261535

4. Cutler DM, Poterba JM, Summers LH. Speculative dynamics. The Review of Economic Studies. 1991; 58(3):529–546. https://doi.org/10.2307/2298010

5. Chopra N, Lakonishok J, Ritter JR, et al. Performance measurement methodology and the question of whether stocks overreact/1991: 130. BEBR faculty working paper; no 91-0130. 1991.

6. Bondt WF, Thaler R. Does the stock market overreact? The Journal of Finance. 1985; 40(3):793–805. https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

7. Bogousslavsky V. Infrequent rebalancing, return autocorrelation, and seasonality. The Journal of Finance. 2016; 71(6):2967–3006. https://doi.org/10.1111/jofi.12436

8. Chowdhury SSH, Rahman MA, Sadique MS. Behaviour of stock return autocorrelation in the GCC stock markets. Global Business Review. 2015; 16(5):737–746. https://doi.org/10.1177/0972150915591420

9. Li L, Leng S, Yang J, Yu M. Stock Market Autoregressive Dynamics: A Multinational Comparative Study with Quantile Regression. Mathematical Problems in Engineering. 2016; 2016.

10. Chowdhury SSH, Rahman MA, Sadique MS. Stock return autocorrelation, day of the week and volatility. Review of Accounting and Finance. 2017. https://doi.org/10.1108/RAF-12-2014-0146

11. Xue WJ, Zhang LW. Stock return autocorrelations and predictability in the Chinese stock market—Evidence from threshold quantile autoregressive models. Economic Modelling. 2017; 60:391–401. https://doi.org/10.1016/j.econmod.2016.09.024

12. Fama EF, French KR. Dividend yields and expected stock returns. Journal of Financial Economics. 1988; 22(1):3–25. https://doi.org/10.1016/0304-405X(88)90020-7

13. Jaffe J, Keim DB, Westerfield R. Earnings yields, market values, and stock returns. The Journal of Finance. 1989; 44(1):135–148. https://doi.org/10.1111/j.1540-6261.1989.tb02408.x

14. Fama EF, French KR. The cross-section of expected stock returns. The Journal of Finance. 1992; 47 (2):427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

15. Silvennoinen A, Teräsvirta T. Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations. SSE/EFI Working Paper Series in Economics and Finance; 2005.

16. Dennis P, Mayhew S, Stivers C. Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon. Journal of Financial and Quantitative Analysis. 2006; 41(2):381–406. https://doi.org/10.1017/S0022109000002118

17. Eom C, Park JW. Effects of common factors on stock correlation networks and portfolio diversification. International Review of Financial Analysis. 2017; 49:1–11. https://doi.org/10.1016/j.irfa.2016.11.007

18. Dimic N, Kiviaho J, Piljak V, Äijö J. Impact of financial market uncertainty and macroeconomic factors on stock–bond correlation in emerging markets. Research in International Business and Finance. 2016; 36:41–51. https://doi.org/10.1016/j.ribaf.2015.09.001

19. Han L, Lv Q, Yin L. The effect of oil returns on the stock markets network. Physica A: Statistical Mechanics and its Applications. 2019; 533:122044. https://doi.org/10.1016/j.physa.2019.122044

20. Augustyniak M. Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics & Data Analysis. 2014; 76:61–75. https://doi.org/10.1016/j.csda.2013.01.026

21. Stein C. Estimation of a covariance matrix, Rietz Lecture. In: 39th Annual Meeting IMS, Atlanta, GA, 1975; 1975.

22. Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis. 2004; 88(2):365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

23. Ohno S, Ando T. Stock return predictability: A factor-augmented predictive regression system with shrinkage method. Econometric Reviews. 2018; 37(1):29–60. https://doi.org/10.1080/07474938.2014.977086

24. Yang L, Couillet R, McKay MR. Minimum variance portfolio optimization with robust shrinkage covariance estimation. In: 2014 48th Asilomar Conference on Signals, Systems and Computers. IEEE; 2014. 1326–1330.

25. Ledoit O, Wolf M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance. 2003; 10(5):603–621. https://doi.org/10.1016/S0927-5398(03)00007-0

26. Wilder JW. New concepts in technical trading systems. Trend Research; 1978.

27. Rougier J. Ensemble averaging and mean squared error. Journal of Climate. 2016; 29(24):8865–8870. https://doi.org/10.1175/JCLI-D-16-0012.1

28. Scharf LL, Demeure C. Statistical signal processing: detection, estimation, and time series analysis. vol. 63. Addison-Wesley Reading, MA; 1991.

29. Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology. 1933; 24(6):417. https://doi.org/10.1037/h0071325

30. Jolliffe I. Principal component analysis. In: International Encyclopedia of Statistical Science. Springer; 2011. 1094–1096.

31. Ince H, Trafalis TB. Kernel principal component analysis and support vector machines for stock price prediction. IIE Transactions. 2007; 39(6):629–637. https://doi.org/10.1080/07408170600897486

32. Fung W, Hsieh DA. Empirical characteristics of dynamic trading strategies: The case of hedge funds. The Review of Financial Studies. 1997; 10(2):275–302. https://doi.org/10.1093/rfs/10.2.275

33. Alexander C. Market risk analysis, value at risk models. vol. 4. John Wiley & Sons; 2009.

34. Shukla R, Trzcinka C. Sequential tests of the arbitrage pricing theory: a comparison of principal components and maximum likelihood factors. The Journal of Finance. 1990; 45(5):1541–1564. https://doi.org/10.1111/j.1540-6261.1990.tb03727.x

35. Yu H, Chen R, Zhang G. A SVM stock selection model within PCA. Procedia computer science. 2014; 31:406–412. https://doi.org/10.1016/j.procs.2014.05.284

36. Zhong X, Enke D. Forecasting daily stock market return using dimensionality reduction. Expert Systems with Applications. 2017; 67:126–139. https://doi.org/10.1016/j.eswa.2016.09.027

37. Wang J, Wang J. Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. Neurocomputing. 2015; 156:68–78. https://doi.org/10.1016/j.neucom.2014.12.084

38. Pasini G. Principal component analysis for stock portfolio management. International Journal of Pure and Applied Mathematics. 2017; 115(1):153–167. https://doi.org/10.12732/ijpam.v115i1.12

39. Narayan PK, Narayan S, Thuraisamy KS. Can institutions and macroeconomic factors predict stock returns in emerging markets? Emerging Markets Review. 2014; 19:77–95. https://doi.org/10.1016/j.ememar.2014.04.005

40. Velicer WF, Jackson DN. Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. Multivariate behavioral research. 1990; 25(1):1–28. https://doi.org/10.1207/s15327906mbr2501_1 PMID: 26741964

41. Bartholomew DJ, Steele F, Moustaki I. Analysis of multivariate social science data. Chapman and Hall/CRC; 2008.

42. Meglen RR. Examining large databases: a chemometric approach using principal component analysis. Marine Chemistry. 1992; 39(1-3):217–237. https://doi.org/10.1016/0304-4203(92)90103-H

43. Hermus K, Wambacq P, et al. A review of signal subspace speech enhancement and its application to noise robust speech recognition. EURASIP Journal on Advances in Signal Processing. 2006; 2007 (1):045821. https://doi.org/10.1155/2007/45821

44. Tufts DW, Kumaresan R, Kirsteins I. Data adaptive signal estimation by singular value decomposition of a data matrix. Proceedings of the IEEE. 1982; 70(6):684–685. https://doi.org/10.1109/PROC.1982.12367

45. Ballings M, Van den Poel D, Hespeels N, Gryp R. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications. 2015; 42(20):7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

46. Yao J, Tan CL. A case study on using neural networks to perform technical forecasting of forex. Neuro-computing. 2000; 34(1-4):79–98. https://doi.org/10.1016/S0925-2312(00)00300-3

47. Torun MU, Akansu AN, Avellaneda M. Portfolio risk in multiple frequencies. IEEE Signal Processing Magazine. 2011; 28(5):61–71. https://doi.org/10.1109/MSP.2011.941552

48. Byrnes CI, Lindquist A, Zhou Y. On the nonlinear dynamics of fast filtering algorithms. SIAM Journal on Control and Optimization. 1994; 32(3):744–789. https://doi.org/10.1137/S0363012991202422

49. Brock W, Lakonishok J, LeBaron B. Simple technical trading rules and the stochastic properties of stock returns. The Journal of finance. 1992; 47(5):1731–1764. https://doi.org/10.1111/j.1540-6261.1992.tb04681.x

50. Taylor MP, Allen H. The use of technical analysis in the foreign exchange market. Journal of international Money and Finance. 1992; 11(3):304–314. https://doi.org/10.1016/0261-5606(92)90048-3

51. Hong K, Satchell S. Time series momentum trading strategy and autocorrelation amplification. Quantitative Finance. 2015; 15(9):1471–1487. https://doi.org/10.1080/14697688.2014.1000951

52. Pafka S, Potters M, Kondor I. Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization. arXiv preprint cond-mat/0402573. 2004.