Reducing Uncertainty in Stock Market Forecasting

Matthew Forey

The University of Texas at San Antonio

Fall 2019

Introduction:

To accurately forecast future changes in stock prices and limit the amount of observed variance in prices, it is important to understand the extent to which market forces such as public information play in price evolution. Price changes in the stock market are dynamic and very difficult to predict, and they are reflective of outside forces ranging from favorable or unfavorable earnings reports, potential mergers, or changes in demand from investors. Fama and Malkiel [1] hypothesized that all stock share price variations are driven by all publicly available information, including past information. Called the efficient market hypothesis, this theory would mean it is impossible to beat the market, much less predict what will happen in the future.

However, it is a difficult task to quantify the extent that any number of events or news affects market prices because there is no observable metric to correlate with the movement observed in the stock price. The unpredictable nature of these events will yield the potential of high amounts of error in any forecast, therefore the proposed resolution in this paper is to reduce as much uncertainty as possible in the forecast, rather than create an outright prediction output. This study will propose the use of LSTM to use historical past information in the forecast model, in combination with empirical mode decomposition filters that decomposes the original stock signal to improve accuracy. The final prediction is obtained by reconstructing the filtered signals explained by the LSTM, limiting the amount of uncertainty in the forecast model.

Related Work:

Other authors have discussed various measures of quantifying uncertainty as a latent variable, and applied these measures to time-series forecasting. A proposed index developed by Baker et al. [2] searched economic policy uncertainty based on the frequency of terms in newspaper coverage related to economic uncertainty. Box and Jenkins [11] proposed a model based on moving averages to forecast data that is more linear and stationary. However, for this research, the selected data will represent different-sized companies across many sectors from Yahoo finance, and is both nonlinear and nonstationary.

Rather than selecting individual companies, Huarng et al. [13] used the Dow Jones and NASDAQ indices to create their own forecast of the Taiwan stock index using a multi-variable approach to improve their forecast. Furthermore, another time-series prediction method proposed by Song and Chissom [12] analyzes fuzzy historical university enrollment time series data to develop a model for forecasting future enrollment. Our approach will differ from these authors in that we propose the use of an artificial neural network to transform our data into an accurate forecast.

A forecasting technique proposed by Cao, Li, and Li [3] suggests that a supplemental approach to a traditional artificial neural network is required to predict stock prices because the stock market is not only related to presently available data, but also past data. The previous price information can be lost if only the latest and current data is used, leading to significant model accuracy error. To solve this issue, they proposed the use of filters in an LSTM model architecture with multilayer perceptrons to reduce uncertainty in forecasting [3]. The LSTM model has become one of the most popular models [9] used for short and long dependency in the data by keeping the error constant while solving the vanishing gradient problem.

Because the selected financial time series data contains the presence of noise, traditional models like LSTM cannot create the most accurate forecast on these nonlinear and nonstationary signals on their own. Other authors have proposed hybrid models to solve this limitation in other fields. Hadavandi et al. [14] used health care data to develop a hybrid artificial intelligence model to try to forecast monthly outpatient visits for hospitals. They created their model using a data clustering technique to improve upon their algorithm [14]. Yu et al. [17] addressed the same type of patient forecasting problem. In their research, they proposed a hybrid approach by combining wavelet decomposition with an artificial neural network. To address the complexity and noisiness of the selected data signals, the use of a wavelet filtering approach is one that this paper will seek to address through the principle of "decomposition and ensemble" [17].

First proposed by Huang et al. [6] empirical mode decomposition (EMD) is used to transform the original complex time-series signal into smaller sub-signals of different frequencies to better analyze the time-series data. As a Fourier transform-based method of filtering, EMD is useful on nonlinear and nonstationary stock market data to decompose these signals into simpler frequency components and stronger correlations, allowing them to be more predictive.

Other papers have found EMD to be one of the best models for reducing prediction error in time series analysis [10][15][16]. Wei [10] found that the best way to address the drawbacks of other models was to implement EMD to decompose the raw noisy stock market data into much simpler frequency components. Guhathakurta et al. [15] used EMD to analyze two separate financial time series to compare the two resulting behaviors of the decomposed signals. Yu and Liu [16] also used decomposition and ensemble as a two-stage combination of EMD with Support Vector Machine to predict the prices of the stock market. In their approach, EMD first decomposed the signal into intrinsic mode functions, and SVMs then used different learning parameters on several IMFs to create the predictions. This paper will propose the use of decomposition filtering in conjunction with LSTM to create a model that tries to reduce as much uncertainty as possible in stock market forecasting.

Background:

LSTM

Because the financial time series data is ever changing, previous historical data must be maintained and included in the prediction model. Hochreiter and Schmidhuber [4] first proposed LSTM in 1997 as a way to solve problems involving long minimal time lags between input signals and corresponding error signals. LSTM is a deep and recurrent model of neural networks that contains neurons which pass data between layers through gates. LSTM solves the issue of gradient disappearance by introducing memory cells that store information through three gates - the input gate, the forget gate, and the output gate. To allow the model to remember past data, we will include the forget gate that was first introduced and added to LSTM by F. Gers et al. [18]. These gates help LSTM to store, reset, and update past stock price information as new time-series data is inputted. Its schematic definition uses the same notation and definitions considered by Bianchi et al. [9].

However, with data containing a significant amount of noise, LSTM alone cannot create an accurate prediction model. Nelson et al. observed in their research that their LSTM model was able to predict with almost 56% accuracy whether the direction of a particular stock price would go up or down within the next 15 minutes [5].

Filtering

To reduce the impact level that noise has on the LSTM prediction, a series of filters are proposed in combination with LSTM. Empirical mode decomposition will transform the original time-series signal into smaller IMFs of different frequencies to better analyze the time-series stock price data. The decomposed frequencies that produce high amounts of reconstructed error can be more easily calculated and removed from the model. The original data set can be decomposed into six smaller intrinsic mode functions (IMF) that admit well-behaved Hilbert transforms, meaning they have frequency time-distributions that can be calculated. [6]. Once the decomposed time series are transformed into IMFs and the energy-frequency-time distributions are constructed, they can be analyzed and selected for reconstructing a more accurate forecast.

(EMD notation and figure here)

More advanced versions of filtering can also be used, including ensemble empirical mode decomposition (EEMD) or complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). EMD will fall short when frequent mode mixing occurs, which occurs when one IMF consists of oscillations in scales that are incapable of being compared with each other, preventing it from having any useful features to analyze. EEMD compliments EMD by adding a white noise to the original signal to overcome the scale separation problem [7]. Wu and Huang [19] showed in their research that EEMD can serve as major improvement of EMD when it uses the capability of enabling EMD to be a dyadic signal filter by adding noise to preserve the original signal's physical attributes.

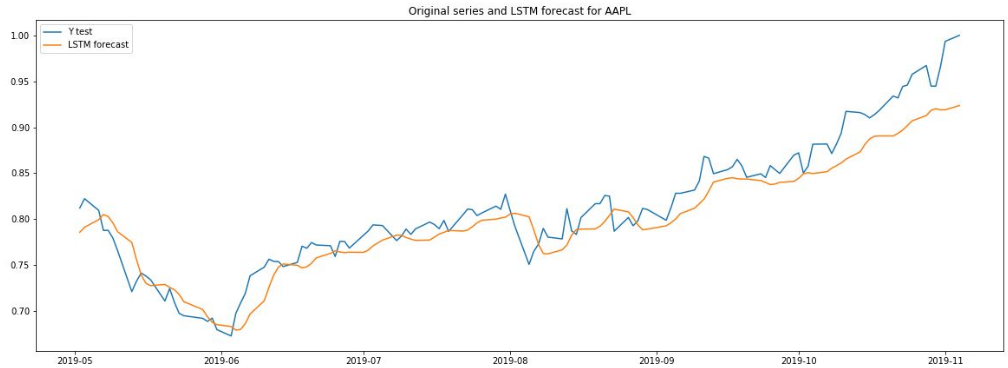
CEEMDAN is a variation of EMD that replicates the signal and better separates the IMFs [3]. It does so in a similar way to EEMD, but instead of adding noise independently distributed, it adds a pair of noises that are perfectly anti-correlated to cancel out the residual noise when the original signal is reconstructed. [20]. Yeh and Shieh [20] found that CEEMDAN is the best approach on data where the goal is to keep the noise in the reconstructed signal.

We will improve upon a methodology used by Cao, Li, and Li [3] in which they propose a hybrid time-series forecasting methodology by combining a CEEMDAN algorithm with LSTM in their research.

Experiment:

To achieve a suitable level of prediction accuracy, we use a recurrent neural network that connects hidden units to keep memory stored. The neural network best suited is LSTM, which will maintain the presence of historical data in the final prediction result.

The stocks used for testing our proposed LSTM architecture were selected from 20 companies belonging to 5 different sectors, using as criteria the highest and lowest values for market capitalization, as well as the highest and lowest values for trade volume. Choosing individual stocks ensured a suitable sample of different sized companies for our testing. LSTM used past closing prices to forecast future closing prices for each company, as seen in the following figure for Apple.



To solve the problem of the amount of random fluctuations of observed noise in the price that prevents LSTM model from capturing total behavior of signal, we propose the use of latent variables to help describe the behavior observed in the stock price. The EMD decomposed filters that correspond to the frequency decomposition of the price will serve as latent variables in the model. Through factoring analysis, we will attempt to recover the signal by adding each of the individual extracted contributions [8].

$$p = Wh + b + noise$$

(Where p is the price of a stock as a vector, W the weight, h the latent variable IMF filter, and b a set constant value)

The selected company's signal is filtered in the frequency domain into latent variables (h), which contributes to the value of the magnitude of the stock (W), plus some added level (b), and noise. This noise corresponds with the maximum prediction error that is not able to be expressed by our model. Because the data is normalized, (b) holds a value of 0, and (W) can be expressed as a matrix that only has magnitude value of 1. The resulting equation is:

$$p = h_{[IMF\ 1-6]} + noise_{[IMF\ 0]}$$

	Company	Filter	Source	MAPE	WAPE
0	AAPL	EMD	Original	1.981235	2.006934
1	AAPL	EMD	IMF 0	110.298565	1847.570724
2	AAPL	EMD	IMF 1	75.194315	1739.584209
3	AAPL	EMD	IMF 2	18.400771	-85.108009
4	AAPL	EMD	IMF 3	9.724974	-47.129555
5	AAPL	EMD	IMF 4	70.643007	133.223997
6	AAPL	EMD	IMF 5	3.388198	3.458215

Some features appear to have small variabilities in magnitude and high amounts of reconstructed residual error, or noise, - namely IMF0. These latent variables occur because the LSTM model is not fully capable of understanding the full pattern of the past information to make an accurate forecast. Thus, IMF0 is considered as noise in the model. The magnitude of this noise is selected with the uncertainty level in the LSTM forecast. The final model includes the previously described resulting equation, with only the five latent variables that are well expressed by LSTM model (IMF 1 to 5), because they contain some patterns in the past that can be learned by the LSTM model. The added random error is the latent variable in this frequency decomposition that does not express the observed variability well enough. The amplitude of the reconstructed signal corresponds with uncertainty level in terms of MAPE and WAPE.

Independent Component Analysis (ICA) is another approach used to decompose signals into statistically independent signals [21]. In this case, independence would refer to information from the value of one signal would not affect any information on the value of another signal, and vice versa [22]. Once the signals are separated into independent components, any factors that contribute to the overall time-series can be observed in the underlying signals. With stock market data, the changes in stock prices can be explained by a variety of factors, so ICA seeks to separate any noisy components such as news that were mixed into the original signal [21].

Our ICA model could be described as a generative model, meaning that it describes how the observed signals are generated through the unknown mixing matrix [22]. The independent signal components are represented by the unobserved latent variables. Our original approach is to take the original stock market signal sample from the selected test company and separate at the frequency domain into independent latent variables. Instead of filtering the signal at the frequency domain, we can recreate a similar approach using ICA.

The selected time is divided up into 2 sub-samples: in-sample, which makes up 66% used for model estimation and training, and out-sample, which makes up 33% used for model testing. A prediction error is output, and biases are checked. If the amount of error is close to zero, the model is unbiased. If the error is far from zero, the model is biased. Bias indicates the predictor has deficiencies that still need to be fixed. Once all predictors are fixed to unbiased, we can compare model prediction accuracies for both in-samples and out-samples. The smaller the value of the prediction error variance, the better the prediction.

References:

[1] Fama E. (1970) « Efficient Capital Markets: a Review of Theory and Empirical Work » *Journal of Finance*, Vol.25, No.2.pp. 383-417.

[2] Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593-1636.

[3] Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications*, *519*, 127-139.

[4] Hochreiter, S., & Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. In *Advances in neural information processing systems* (pp. 473-479).

[5] Nelson, D. M., Pereira, A. C., & de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 1419-1426). IEEE.

[6] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, E. H., Zheng, Q., Yen, N.-C., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis. Proc. R. Soc. Lond. A 454, 903-99

[7] Wang, T., Zhang, M., Yu, Q., & Zhang, H. (2012). Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal. *Journal of Applied Geophysics*, *83*, 29-34.

[8] Heaton, Jeffrey. (2017). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618. Genetic Programming and Evolvable Machines. 19. 10.1007/s10710-017-9314-z.

[9] Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). *Recurrent neural networks for short-term load forecasting: an overview and comparative analysis*. Springer.

[10] Wei, L. Y. (2016). A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. *Applied Soft Computing*, *42*, 368-376.

[11] G. Box, G. Jenkins. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1976)

[12] Q. Song, B.S. Chissom. Forecasting enrollments with fuzzy time-series Part I. Fuzzy Sets Syst., 54 (1993), pp. 1-10

[13] K.H. Huarng, H.K. Yu, Y.W. Hsu. A multivariate heuristic model for fuzzy. IEEE Trans. Syst. Man Cybern. B: Cybern., 37 (4) (2007), pp. 836-846

[14] E. Hadavandi, H. Shavandi, A.Ghanbari, S. Abbasian-Naghneh Developing a hybrid artificial intelligence model for outpatient visits forecasting in hospitals.

[15] Guhathakurta, K., Mukherjee, I., & Chowdhury, A. R. (2008). Empirical mode decomposition analysis of two different financial time series and their comparison. *Chaos, Solitons & Fractals*, *37*(4), 1214-1227.

[16] Yu, H., & Liu, H. (2012, October). Improved stock market prediction by combining support vector machine and empirical mode decomposition. In *2012 Fifth International Symposium on Computational Intelligence and Design* (Vol. 1, pp. 531-534). IEEE.

[17] Yu, L., Hang, G., Tang, L., Zhao, Y., & Lai, K. K. (2017). Forecasting Patient Visits to Hospitals using a WD & ANN-based Decomposition and Ensemble Model. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(12), 7615-7627.

[18] F. Cummins F.A. Gers, J. Schmidhuber. Learning to forget: Continual prediction with lstm. Technical Report, IDSIA-01-99, 1999.

[19] Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, *1*(01), 1-41.

[20] Yeh, J. R., Shieh, J. S., & Huang, N. E. (2010). Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Advances in adaptive data analysis*, *2*(02), 135-156.

[21] PY MOK, Lam - "An ICA Design of Intraday Stock Prediction Models with Automatic… Selection"

[22] Aapo Hyvärinen and Erkki Oja "Independent Component Analysis: Algorithms and Applications"