Signal Complexity Reduction Tested on Recurrent Neural Networks with Applications on Stock Market Series

Henry Chacón [1] **Emre Kesici** [2] **Matthew Forey** [3] **Paul Rad** [4]

Abstract

Recurrent neural networks have gained a lot of attention in forecasting signals due to their flexibility in capturing dependencies on different time scales. However, as in most of the classical forecasting methods, its accuracy is tied to the degree of signal complexity. Stock market prices, on the other hand, are commonly classified to be non-linear, non-stationary and chaotic signals, since they exhibit erratic behavior that conducts a poor performance in the long short term memory or LSTM. In this paper, we propose a methodology to improve the accuracy in forecasting the trend of a stock closing price by using an LSTM network, the empirical mode decomposition and the intrinsic sample entropy. Our tests show a dependency between the decomposed signal entropy and the direction accuracy. This suggests that in those cases where the short term noise is small compared to the series amplitude, the forecasting capabilities are improved after the decomposed highest frequency is removed. Furthermore, our tests also show an improvement in forecasting the direction of the price in 100% using the classical LSTM model.

1. Introduction

In the stock market literature, it is common to find methods for explaining observed behaviors of a stock based on related variables to induce causality. For instance, in technical analysis the focus is on generating an excess of returns based on past information rather than forecasting future values and patterns of time series, exhibiting less success due to a high noise-to-signal ratio (Wu et al., 2019). While in fundamental analyses, the use of real economic factors including, but not limited to, interest rates, inflation, capital expenditure, industrial production index change in risk premium correspond to latent variables used to estimate the excess in returns.

Recent studies have considered economic news data as proxy variables to examine their impact on stock market returns and predicting earnings (Tetlock, 2007; Tetlock et al., 2008; Dougal et al., 2012; Dzielinski & Hasseltoft, 2013; Garcia, 2013). Other authors found a forecasting accuracy improvement by including news components in their models. Wu et al. (2019) analyzed the usefulness of news for predicting stock returns in the Taiwan stock market by converting textual news into a numerical variable in a regression model with microeconomic variables. Tetlock et al. (2008) found that high media pessimism in a popular Wall Street Journal column could predict downward pressure on market prices, and very high or low pessimism could predict high trading volume using vector auto-regressions.

Mok et al. (2004) proposed an interesting approach to quantify the news impact on the stock index prediction. They suggest a methodology based on *independent component analysis* (ICA) to decompose non-Gaussian signals into orthogonal elements. Each decomposed unit is associated with "extracted" news from intraday stock data assuming market *efficiency*. Where "efficient" is a hypothesis proposed by Fama in his seminal work (Fama, 1970) that defines a market in which prices always *fully reflect* all available information. The signal separation returned by the ICA method corresponds to the influence of statistically independent news observed in the stock price mixed by an unknown process. In the model, orthogonal elements de-mixed by ICA are used as input in an artificial neural network to improve the forecast through a reduction in the mean error prediction.

The frequency domain is also explored by researchers to detect components that affect a series. The *empirical mode decomposition* (EMD) proposed by Huang et al. (1998a), also called Hilbert-Huang EMD, is based in the Fourier

[1] Department of Management Science and Statistics, University of Texas at San Antonio, San Antonio, USA [2] Department of Finances, University of Texas at San Antonio, San Antonio, USA [3] Department of Data Analytics in the College of Business Graduate Studies, University of Texas at San Antonio, San Antonio, USA [4] Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, USA. Correspondence to: Henry Chacón <henry.chacon@utsa.edu>, Paul Rad <peyman.najafirad@utsa.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

transformation that decomposed the signal according to the time scale (Cao et al., 2019). Each filtered sequence represents a different quality of the series in a particular scale with a physical representation, unlike projection-based methods (Looney et al., 2015). Therefore, if a decomposed element dominantly captures features behaviors from the signal, any similarity metric between it and the original sequence should be larger than the comparison between other decomposed elements.

In the forecasting horizon, recent news plays an important role with an expected correlation in successive headlines. However, model predictability could be affected by the volume of news and its selection (Mok et al., 2004). In this case, quantity is not the only factor, but also the quality is important since signals can be confounding by noise. This consideration is relevant for the Long Short-Term Memory (LSTM) due to its forget mechanism. It is not robust to noisy components forcing the model to retain irrelevant information (Baddar & Ro, 2019). This is the reason because hybrid models between LSTM architectures and other procedures used to reduce the noisy effect are found in the literature. See for instance the model proposed by Cao et al. (2019) where the signal is decomposed in factors used as input of the LSTM with outputs feeding connected layer of artificial neurons to isolate the permanent effect of invaluable information saved by the memory unit. Other similar approaches are found in (Wang et al., 2015b; Anish & Majhi, 2016).

The problem with any time series forecasting method, including the LSTM, is their capability of detecting the pattern presented in the sequence, where classical statistical methods fail in detecting non-linear relations. Financial time series are commonly associated with deterministic chaos (Tiwari & Gupta, 2019). However, they are not the only one with that description. Clinical cardiovascular and other biological series shared a chaotic behavior too (Pincus, 1991). So, it is common to find in the medical literature different metrics developed to determine the complexity of a system. Richman and Moorman (2000) suggested the *sample entropy* (SampEn) as a metric closely related to the entropy and based on the *approximate entropy* (ApEn) proposed by Pincus. It reflects the degree of regularity and predictability found in a signal. The SampEn is based in "*the likelihood that similar patterns within a time series will remain similar when pattern lengths are increased*" (Looney et al., 2015). This means an increment in the SampEn is a consequence of more irregularities and fewer patterns predictability in the series.

Quasiperiodicity HERE

Contributions:

- The IMF are not statistically independent

2. Background and Related work

2.1. The empirical mode decomposition

Most of the classical time series methods are based on finding regular patterns or correlations in the sequence's behavior. In ARMA methods, the lag dependency contribution is defined by the model parameters based on the autocorrelation function (Nyoni, 2019). This metric measures a linear dependency considering, in general, a pairwise-joint uniform distribution, which limit the inclusion of more complex types of dependencies (Guhathakurta et al., 2008).

Frequency domain is an alternative commonly used in signal analysis to study the contribution of different harmonics in the series. Popular methods such as Fourier transformation, spectral analysis, etc. (Warner, 1998) have been applied in several fields. However, in the presence of noise, it is difficult to filter the signal from the noise in cases where the noise shares fundamental harmonics with the series (Wu & Huang, 2004). In nonlinear and statistically non-stationary series, Fourier-based metrics are not convenient (Guhathakurta et al., 2008). In this regards, the *empirical mode decomposition* (EMD) proposed by Huang et al. (1998b) arises as an elegant alternative in the Hilbert space to extract the periodicity of a series. EMD is appropriate for non-stationary sequences since it decomposes the signal based on a local time scale feature without a pre-setting function (Cao et al., 2019) nor the use of any harmonic.

It is developed as an iterative algorithm that decomposes the series in a set of orthogonal signals called *intrinsic mode functions*, or IMFs. The collection of break down elements explains the time variation dynamic presented in the signal at different scales and "*at the highest level of resolution–instantaneous frequency and instantaneous amplitude–via the Hilbert transformation*" (Looney et al., 2015). Since the basis of IMFs is not constant and locally determined, each function have physical meaning of the frequency per time in the underlying process. Imposed constrains in the algorithm for each IMF are the following (Huang et al., 1998b):

1. As global condition, the number of extreme points, positive or negative in each decomposed signal should differ maximum in one unit. Which means, every IMF is symmetric.
2. As local criteria, the mean of each IMF must be approximately zero.

Let's consider $x(t) \in \mathbf{R}$ decomposed in a finite number $N \in \mathbf{N}$ of IMFs $C_n(t) : n \in N$, given by the expression (1) and the algorithm (1) found in (Looney et al., 2015).

$$x(t) = \sum_{i=1}^{N} C_i(t) + r(t) \tag{1}$$

Algorithm 1 shifting process, EMD algorithm

```
Input: data $x(t)$
Define the proto-IMF as $\tilde{x} = x(t)$, and $N = 1$
repeat
  $k = 0$
  while satisfies IMF $SC$ defined in (2) do
    Identify all local maxima and minima of $\tilde{x}(t)$
    Find an "envelope" $e_{min}(t)$ that interpolates all local minima
    Do the same for the local maxima $e_{max}(t)$
    Compute a temporal IMF by
      $\tilde{c}_{n,k}(t) = \tilde{x}(t) - \left[e_{min}(t) + e_{max}(t)\right]/2$
    Let $\tilde{x}(t) = \tilde{c}_{n,k}(t)$
    $k = k + 1$
  end while
  $N = N + 1$
  Save the resulting IMF $C_n(t) = \tilde{c}_{n,k}(t)$
  Let $\tilde{x}(t) = x(t) - \sum_{j=1}^{N} C_j(t)$
until $\tilde{x}(t)$ is a monotonic residue $r(t)$ or trend
```

Notice the algorithm stops when the remaining section of the signal $\tilde{x}(t)$ is monotonic or a trend and not a normal distribution like in the ARIMA model. Besides the global and local constrains imposed to the IMF and mentioned above, the numerical stopping criteria is the following (Wu & Huang, 2005):

$$SC = \frac{\sum_{i=1}^{T} \left[\tilde{c}_{n,k}(i) - \tilde{c}_{n,k-1}(i)\right]^2}{\sum_{i=1}^{T} \tilde{c}_{n,k}^2(i)} < 2 \times 10^{-4} \qquad (2)$$

According to Wu & Huang (2004; 2005), the EMD is a *dyadic filter* in noise cancellation procedures since the frequency ratio of consecutive IMFs is almost equal to two. Also, each IMF shares the same area in terms of the Fourier spectra in a semi-logarithmic period scale. Besides, authors suggest that decomposed signals IMF follows a normal distribution. However, it is important to point out here that they are not statistically independent, as it will be demonstrated in the Markov network exhibited in the experimental section and also supported in the *envelope* formula included in the algorithm (1), computed from the locally dependent sequence of maximum and minimum values.

Some modifications of the original EMD algorithm are found in the literature. One is called *the ensemble empirical mode decomposition* or EEMD. It was proposed by Wu & Huang (2009) and is intended to solve the potential mode-mixing problem in which the IMFs share similar frequencies (Cao et al., 2019). It could limit the representation of some signal characteristics in the EMD method (Wang et al., 2015a). Another improvement was suggested by Torres et al. (2011) focused on completely eliminate the Gaussion noise added in the EEMD model to solve the mixing problem known as *the complete ensemble empirical mode decomposition with adaptive noise* or CEEMDAN.

2.2. The sample entropy

In the signal analysis, the common approach is to try to remove the noise from the signal assuming a model $x(t) = s(t) + n(t)$ where $s(t)$ and $n(t)$ are the signal and the noise respectively. However, a more interesting question was proposed by Wu & Huang (2005) - is it possible to determine if a signal or its components contain useful information?. The reason is simple - to determine the noise, a prior knowledge of both the signal and noise are required, which is a condition that is unlikely in the analysis of real series.

Entropy is usually associated to information, and in dynamic systems is linked to its information rate (Looney et al., 2015) or a metric to determine the complexity changing in a system (Pincus, 1991). The *sample entropy* (SampEn) proposed by Richman & Moorman (2000) and based in *the approximate entropy* (ApEn) suggested by Pincus (1991). It estimates how likely similar patterns in the series will be found in the sequence if patterns length increases, meaning it could be considered as a proxy of irregularities in the signal. Therefore, small values of SampEn are associated with more predictability of patterns in the sequence. The sample entropy procedure found in (Looney et al., 2015) is exhibited in algorithm (2).

3. Noise consideration

Complex temporal dependencies are one of the most important challenges in recurrent neural networks (Tabor, 2002) since the memory unit has no infinite capacity to store and process all the patterns found in the signal. Noise is an important source of disturbances that could mislead the detection of main patterns from the series. In an EEG test, for instance, noise can distort the signal through sensors, or even other signals generated by the body.

Noise removal methods have been developed, and most of them consider it as a Gaussian distribution with parameters $(\mu = 0, \sigma)$. In general, a threshold is defined to discriminate the signal from the noise, see for instance (Wu & Huang, 2004; 2005; Mert & Akan, 2014). However, as it was mentioned above, a prior understanding of both components is required for a successful implementation. In some situations, the noise is not added by any external source, and the erratic behavior is part of the sequence. Therefore, discriminating between signal and noise is not only unsuitable but also undesired, since important elements from the signal could be removed. A good example of erratic signals that

Algorithm 2 Sample entropy algorithm

```
Input: data $x(t)$ and threshold $r$
repeat
    For lag $\tau$ and window $m$ get vectors $X_m(t)$ of length $\tau(m-1)$
    Compute the maximum norm $d_j$ between vectors $X_m(t_{j-1})$ and $X_m(t_j)$ for $t_j \in \{1, \cdots T - \tau(m-1) + 1\}$ by the equation:
        $d_j = \max\{|X_m(t_{j-1})|, |X_m(t_j)|\}$
    For a given time $t$ compute
        $\phi_{t,m}(t,r) = \frac{1}{|X_m(t)|} \sum_{i=1}^{|X_m(t)|} \mathcal{I}_{d_i \leq r}(i)$
    Define $T^* = T - \tau(m-1) + 1$ and get
        $\phi_m(r) = \frac{1}{T^*} \sum_{t=1}^{T^*} \phi_{t,m}(t,r)$
until $m \geq T$
Compute the $SampEn = -\ln\left[\frac{\phi_{m+1}(r)}{\phi_m(r)}\right]$
where $|X_m(t)|$ denotes the number of elements in the embedding vector and $\mathcal{I}_{d_i \leq r}(i)$ the indicator function of maximum norms less than the threshold.
```

are sensible to several sources are the stock market prices, despite the fact that some economists such as Eugine Fama have considered them as a random walk. Its price is used by investors to produce a profit, regardless of the unpredictable number of patterns that are driving it and without considering a separation between signal and noise.

In this document, we follow the same approach proposed by Guhathakurta et al. (2008). As they suggest, financial time series are *quasi-periodic* without a predominant signal periodicity. Therefore, instead of handling the problem as a pure series and random noise, the signal is considered as a sequence of stable and irregular patterns. Where, depending on the last one's amplitude, the series could be more predictable or not. The following notation will be used in this paper to represent the signal composition:

$$x(t) = \Phi(t) + \Psi(t) : \quad (\Phi(t), \Psi(t)) \in \mathbf{R} \tag{3}$$

where $\Phi(t)$ and $\Psi(t)$ corresponds to periodic and quasi-periodic components respectively.

4. Methodology

5. Experiments

6. Summary and discussion

References

Anish, C. and Majhi, B. Hybrid nonlinear adaptive scheme for stock market prediction using feedback flann and factor analysis. *Journal of the Korean Statistical Society*, 45(1):64–76, 2016.

Baddar, W. J. and Ro, Y. M. Mode variational lstm robust to unseen modes of variation: Application to facial expression recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3215–3223, 2019.

Cao, J., Li, Z., and Li, J. Financial time series forecasting model based on ceemdan and lstm. *Physica A: Statistical Mechanics and its Applications*, 519:127–139, 2019.

Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A. Journalists and the stock market. *The Review of Financial Studies*, 25(3):639–679, 2012.

Dzielinski, M. and Hasseltoft, H. Aggregate news tone, stock returns, and volatility. *Unpublished Working Paper*, 2013.

Fama, E. F. Papers and proceedings of the twenty-eight annual meeting of the american finance association new york, ny december 28–30, 1969. *J. Financ*, 25:383–417, 1970.

Garcia, D. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.

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

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971): 903–995, 1998a.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971): 903–995, 1998b.

Looney, D., Hemakom, A., and Mandic, D. P. Intrinsic multi-scale analysis: a multi-variate empirical mode decomposition framework. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2173):20140709, 2015.

Mert, A. and Akan, A. Detrended fluctuation thresholding for empirical mode decomposition based denoising. *Digital signal processing*, 32:48–56, 2014.

Mok, P., Lam, K., and Ng, H. An ica design of intraday stock prediction models with automatic variable selection. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, volume 3, pp. 2135–2140. IEEE, 2004.

Nyoni, T. Forecasting australian cpi using arima models. 2019.

Pincus, S. M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

Richman, J. S. and Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.

Tabor, W. Fractal learning neural networks. *Unpublished Manuscript, University of Connecticut*, 2002.

Tetlock, P. C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168, 2007.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467, 2008.

Tiwari, A. K. and Gupta, R. Reprint of: Chaos in g7 stock markets using over one century of data: A note. *Research in International Business and Finance*, 49:315–321, 2019.

Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 4144–4147. IEEE, 2011.

Wang, W.-c., Chau, K.-w., Qiu, L., and Chen, Y.-b. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on eemd decomposition. *Environmental research*, 139:46–54, 2015a.

Wang, W.-c., Chau, K.-w., Qiu, L., and Chen, Y.-b. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on eemd decomposition. *Environmental research*, 139:46–54, 2015b.

Warner, R. M. *Spectral analysis of time-series data*. Guilford Press, 1998.

Wu, G. G.-R., Hou, T. C.-T., and Lin, J.-L. Can economic news predict taiwan stock market returns? *Asia Pacific Management Review*, 24(1):54–59, 2019.

Wu, Z. and Huang, N. E. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611, 2004.

Wu, Z. and Huang, N. E. Statistical significance test of intrinsic mode functions. In *Hilbert-Huang Transform and its applications*, pp. 107–127. World Scientific, 2005.

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