Economics
The Open-Access, Open-Assessment E-Journal

Vol. 10, 2016-9 | March 24, 2016 | http://dx.doi.org/10.5018/economics-ejournal.ja.2016-9

A Historical Analysis of the US Stock Price Index Using Empirical Mode Decomposition over 1791–2015

Aviral K. Tiwari, Arif B. Dar, Niyati Bhanja, and Rangan Gupta

Abstract
In this paper, the dynamics of Standard and Poor's 500 (S&P 500) stock price index is analysed within a time-frequency framework over a monthly period 1791:08–2015:05. Using the Empirical Mode Decomposition technique, the S&P 500 stock price index is divided into different frequencies known as intrinsic mode functions (IMFs) and one residual. The IMFs and the residual are then reconstructed into high frequency, low frequency and trend components using the hierarchical clustering method. Using different measures, it is shown that the low frequency and trend components of stock prices are relatively important drivers of the S&P 500 index. These results are also robust across various subsamples identified based on structural break tests. Therefore, US stock prices have been driven mostly by fundamental laws rooted in economic growth and long-term returns on investment.

JEL C22 G10
Keywords Empirical Mode Decomposition; stock prices, S&P 500 Index; United States

Authors
Aviral K. Tiwari, Faculty of Management, IBS Hyderabad, IFHE University, Donthanapally Shankarapalli Road, Hyderabad, Andhra Pradesh 501203, India, aviral.eco@gmail.com
Arif B. Dar, Institute of Management Technology, Rajnagar, Ghaziabad, Delhi, 201001, India, billaharif0@gmail.com or abillah@imt.edu
Niyati Bhanja, MICA, Ahmedabad Gujarat, India, niyati.eco@gmail.com
Rangan Gupta, ✉ Department of Economics, University of Pretoria, Pretoria, 0002, South Africa, rangan.gupta@up.ac.za

Citation Aviral K. Tiwari, Arif B. Dar, Niyati Bhanja, and Rangan Gupta (2016). A Historical Analysis of the US Stock Price Index Using Empirical Mode Decomposition over 1791–2015. *Economics: The Open-Access, Open-Assessment E-Journal*, 10 (2016-9): 1—15. http://dx.doi.org/10.5018/economics-ejournal.ja.2016-9

Received January 16, 2016 **Published as Economics Discussion Paper** February 24, 2016
Revised March 14, 2016 **Accepted** March 17, 2016 **Published** March 24, 2016

1 Introduction

In recent years, analyses of stock prices within the time-frequency framework have attracted a lot of attention from academicians and market practitioners. The intrinsic complexities of the stock markets have made them least worthy of analysis using the conventional time-domain tools. The obvious reason for this is that stock prices are determined by traders, who deal at different frequencies. While institutional investors and central banks constitute the low-frequency traders, speculators and market makers fall into the category of high-frequency traders in stock markets. Price formation in the stock markets can be attributed to trading by heterogeneous traders within different frequencies. Therefore, some appealing events may remain hidden under different frequencies when stock prices are analysed within the time-domain framework.

In the literature of financial economics, a number of frequency-based approaches have been used to unravel the hidden characteristics of financial time series. Zhang et al. (2008) used the Empirical Mode Decomposition (EMD) to unravel the price characteristics of crude oil at different frequencies. Zhu et al. (2015) analysed price formation in the carbon markets by using the EMD. Using wavelet-based approaches, studies like Tiwari et al. (2013a, 2013b), and Chang et al. (2015), and references cited therein, used wavelet decompositions to study the behaviour of financial variables like oil prices, exchange rates, inflation and stock prices at different frequencies. Wavelets nevertheless have certain disadvantages which EMDs can easily overcome. For example, one cannot chose an objective wavelet function from the set of wavelet functions. In wavelets choice of basis function and decompositions levels is subjective and this subjectivity can lead to the extraction of false cycles form the time series (Wang et al. 2014; Percival and Walden 2000). In EMD however the decomposition is based on local characteristic time scale of the data, hence, doesn’t need the subjective tests. Nevertheless, EMDs do not find much use in the analysis of stock prices. Some of the recent studies using EMDs for the analysis of stock prices are Yu and Liu (2012) Sun and Sheng (2010), Chang and Wei (2014).

Therefore, for this paper we conducted, for the first time, a time-frequency analysis using EMDs for Standard & Poor's 500 (S&P 500) index, covering the long monthly sample of 1791:08–2015:05. This allowed us to determine the various frequency components that have driven stock prices in the US over a prolonged historical sample. We attempted to identify the frequencies that have a substantial impact on stock prices. But why is the identification of these frequencies important? According to "Supply side" models equity returns have their roots in the productivity of underlying real economy. The GDP growth of the underlying economy flows to shareholders in the number of steps. First, it induces corporate growth followed by the earning per share growth and stock price increase. GDP growth and stock price increase being a lengthy process, one can safely assume that the long term trend or the low frequency components of the stock prices represent the underlying real growth of the economy.

The Ensemble EMD (EEMD), introduced by Huang et al. (1998), was used to decompose the stock price data into different intrinsic modes. The IMFs and the residual extracted were then reconstructed into high-frequency, low-frequency and trend components using the hierarchical clustering method. Different measures were then used to assess the importance of each frequency for the overall stock price series. Logically, if one finds the substantial impact of low frequency or trend components of stock prices on the aggregate stock prices, it can be inferred that stock prices are driven by the real growth of the economy.

The rest of the scheme according to which this paper is organized is as follows. Section 2 provides the information about the methodology followed in the paper. Section 3 provides a discussion on the data and results, and section 4 concludes, with the main findings.

2 Methodology

An EMD algorithm for extracting Intrinsic Mode Functions (IMFs) was followed as: [1]

In the first step, the minima and maxima of a time series $x(t)$ were identified. Then with the cubic spline interpolation upper $e_{min}(t)$ and lower $e_{max}(t)$ envelopes were generated. In the third step, the point-by-point mean ($m(t)$) was calculated from the lower and upper envelopes as: $m(t) = (e_{min}(t) + e_{max}(t)) / 2$. The mean form time series was calculated in step 4, and d(t) as the difference of $x(t)$ and $m(t)$ was calculated as $d(t) = x(t) - m(t)$. The properties of $d(t)$ were checked in step 5. If, for

[1] For more on this methodology, please refer to Zhu et al. (2015).

example, it was an IMF, the ith IMF was denoted by $d(t)$. The $x(t)$ was replaced by the residual, given as: $r(t) = x(t) - d(t)$. Often the ith IMF was denoted by $c_i(t)$, where I was interpreted as index. If $d(t)$ was not an IMF, it was replaced by $d(t)$. These five steps were repeated until the residuals satisfied some conditions known as stopping criteria.[2]

Contrary to the EMD, the Ensemble EMD proposed by Wu and Huang (2009) avoids the limitation of the mode mixing associated with EMD. The procedure involves an additional step of adding white noise series to targeted data, followed by the decomposition to generate the IMFs. The procedure was repeated by adding different white noise series each time to generate the Ensemble IMFs from the decompositions as an end product.

3 Results and Discussion

Our analysis is based on a historical data set of US stock prices. The monthly data on the S&P 500, covering the period 1791:08 to 2015:05 was obtained from the Global Financial Database (GFD). The natural logarithmic values of the data have been plotted in Figure A1, in the Appendix.

Through EEMD, four data samples of the US stock prices were decomposed into (IMFs) and residuals. The data sets include the full sample ranging from 1791:08 to 2015:05, and three subsamples ranging from 1791:08 to 1862:12, 1863:01 to 1940:04 and 1940:05 to 2015:05. The subsamples were identified by applying the Bai and Perron (2003) test of structural breaks in both mean and trend to the natural logarithms of the S&P 500 stock index. The division of the data into three subsamples gives a better idea of how the dynamics of the US stock market have evolved over time, and added to the robustness of the results. The IMFs along with the residual are shown in Figures A2, A3, A4 and A5, in the Appendix. The IMFs were generated in the order of highest to lowest frequency. The IMFs were then analysed by three measures. First, the mean period of each IMF – defined as the value extracted by dividing the total number of points by the number of peaks in the dataset – was calculated. Second, the pairwise correlation between the original data series and the IMFs was estimated by using a Pearson and Kendall

[2] For the stopping criteria, please refer to Zhang et al. (2008).

rank correlation. Third, the variance and variance percentage of each IMF were calculated. These results are shown in Tables 1, 2, 3 and 4.

Both the Pearson and Kendall coefficients between the original and high-frequency IMFs are low. However, the correlation is higher between the low-frequency IMFs and the original series. It can also be seen that the variances between lower (higher) frequencies contribute substantially (less) to the total variability.

Within these decompositions, however, the residues are the dominant modes. Their contribution to the total variability is highest, and the correlation with the original data series is also highest. The residue referred to as the deterministic long-term trend by Huang et al. (1998) indicates a very high correlation and accounts for a very high variability in the original series. A noteworthy observation here is that the correlation of the long-term trend with the data and the

Table 1: Measures of IMFs and residuals with the full sample, 1791:08–2015:05

	Mean	Pearson	Kendall	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
Original Series	2.630			3.9060		
IMF1	–0.00	0.0004	0.00095	0.0096	0.2467	0.2775
IMF2	–0.000	–0.001	0.00028	0.003	0.0875	0.0984
IMF3	–8.76E-05	0.011	0.01366	0.0022	0.0585	0.0658
IMF4	–0.000	0.032*	0.04***	0.0030	0.0784	0.0882
IMF5	0.0028	0.091***	0.07***	0.0055	0.1432	0.1610
IMF6	–0.002	–0.05***	0.03***	0.0067	0.17328	0.1948
IMF7	0.0115	0.267***	0.13***	0.025	0.64185	0.7218
IMF8	0.0197	0.543***	0.22***	0.0167	0.42926	0.4827
IMF9	–0.033	0.613***	0.24***	0.0204	0.52310	0.5882
IMF10	–0.016	–0.59***	–0.55***	0.0029	0.07659	0.0861
Residual	2.649	0.98***	0.84***	3.3771	86.46059	97.234
SUM					88.9192	100

Table 2: Measures of IMFs and residuals for the subsample 1791:08–1862:12

	Mean	Pearson	Kendall	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
Original Series	0.9094	1	1	2.74E-02		
IMF1	6.09E-05	0.073**	0.042*	0.000312	1.135786	1.763126
IMF2	0.0003	0.137***	0.072***	0.000327	1.19354	1.85278
IMF3	–0.002	0.342***	0.198***	0.001429	5.208142	8.084805
IMF4	0.0001	0.287***	0.1938***	0.001237	4.508703	6.999039
IMF5	0.0014	0.535***	0.313***	2.97E-03	10.83975	16.82697
IMF6	–6.35E-05	0.722***	0.4495**	0.006696	24.40875	37.89066
IMF7	0.0030	0.442***	0.3627**	0.001552	5.655753	8.77965
IMF8	0.0003	0.477***	0.340***	6.47E-05	0.235886	0.366176
Residual	0.906	0.527***	0.337***	0.003081	11.23259	17.43679
SUM					64.41889	100

Table 3: Measures of IMFs and residuals for the subsample 1863:01–1940:04

	Mean	Pearson	Kendall	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
Original Series	1.924267	1	1	0.219461		
IMF1	–3.04E-06	0.0545*	0.0311	0.001021	0.465262	0.532097
IMF2	0.000412	0.0971**	0.0534**	0.000933	0.425158	0.486232
IMF3	1.76E-05	0.1496***	0.0902***	0.001964	0.894929	1.023486
IMF4	0.003251	0.3278***	0.2018***	0.006409	2.920237	3.339731
IMF5	0.008768	0.2621***	0.1285***	0.014668	6.683416	7.643492
IMF6	–0.00663	0.162***	0.1695***	0.013013	5.92947	6.781241
IMF7	–0.00093	0.1528***	0.1376***	0.004185	1.907059	2.181009
IMF8	–0.01321	0.7345	0.4504	0.003302	1.504771	1.720932
Residual	1.932461	0.8578	0.6827	0.146401	66.709	76.29178
SUM					87.4393	100

Table 4: Measures of IMFs and residuals for the subsample 1940:05–2015:05

	Mean	Pearson	Kendall	Variance	Variance as % of observed	Variance as % of (ΣIMFs + residual)
Original Series	5.022752	1	1	2.402625		
IMF1	–0.00118	–0.0168	-0.008	0.006386	0.265791	0.273297
IMF2	0.001165	0.0185	0.0059	0.002049	0.085274	0.087682
IMF3	–0.00149	–0.029	–0.014	0.001729	0.071971	0.074003
IMF4	0.002962	0.095***	0.082***	0.002563	0.106674	0.109687
IMF5	–0.00268	0.1187***	0.106***	0.005094	0.212012	0.217999
IMF6	–0.00681	–0.205***	–0.122***	0.007476	0.311159	0.319946
IMF7	0.052445	0.369***	0.220***	0.054572	2.27136	2.335505
IMF8	0.002928	–0.66***	–0.469***	0.000204	0.008507	0.008747
Residual	4.979491	0.983***	0.916***	2.256563	93.92072	96.57313
SUM					97.25347	100

variability contribution increases for the more recent samples. Based on supply side models, since the continuing increasing trend of the US stock market is consistent with the development of the US economy over the decades, it can be said that the long-term price behaviour of US stocks has been determined by the long-term growth of the US MSCI (2010).

We then used a hierarchical clustering analysis, and subsequently the Euclidean distance to group the IMFs and residuals into their high-frequency, low-frequency and trend components.[3] The extracted components for all the time series are shown in Figure 1.

[3] We have followed Zhu et al. (2013) to extract the different time series components. For the sake of brevity, we do not show the results here; however, they can be produced on request.

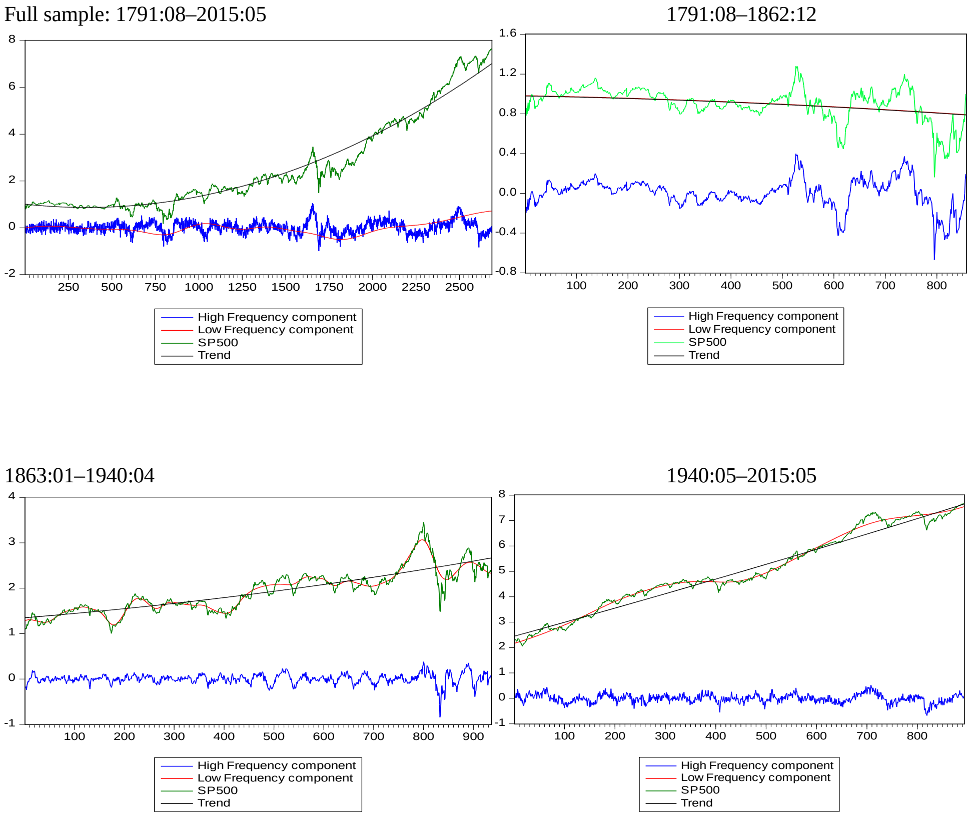

Figure 1: Three components of the S&P 500

Each component in these diagrams shows the distinct features. For example, the residuals show the slow variation around the long-term trend. Hence, it is considered as a long-term trend of a time series. The effect of medium to high frequencies was captured by two other frequencies, with the high frequency components reflecting the effect of short-term market fluctuations. For the moment of observed stock price series, the most important components are the low-frequency component and the trend. The Pearson and Kendall correlation between the different frequency components and the original series shown in Table 5 vary between samples. For example, they are comparatively higher for the lower

Table 5: Correlation and variance of components for the S&P 500 index

Full Sample: 1791:08–2015:05

	Mean	Pearson Correlation	Kendall Correlation	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
ORIGINAL_SERIES	2.63			3.901		
HFRQ	0.009	0.180***	0.109***	0.067	1.7168	0.904
LFRQ	–0.01	0.630***	0.282***	0.062	1.61	0.848
RESIDUAL	2.6499	0.983***	0.840***	3.377	86.48	45.556

1791:08–1862:12

	Pearson correlation	Kendall correlation	Mean	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
ORIGINAL_SERIES			0.909	0.0274		
HFRQ	0.94159***	0.7575***	0.003	0.0209	76.5	38.44
LFRQ	0.52703***	0.3373***	0.906	0.003	11.2325	5.645
RESIDUAL	0.52705***	0.3373***	0.906	0.003	11.232	5.645

1863:01–1940:04

1863:01–1940:04

	Mean	Pearson Correlation	Kendall Correlation	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
ORIGINAL_SERIES			1.92	0.2194		
HFRQ	0.3340***	0.2219***	0.003	0.0126	5.766	2.202
LFRQ	0.9724***	0.853***	1.92	0.196	89.312	34.11
RESIDUAL	0.8577***	0.6826***	1.932	0.1464	66.708	25.48

1940:05–2015:05

	Pearson correlation	Kendall correlation	Mean	Variance	Variance as % of observed	Variance as % of ΣIMFs + residual
ORIGINAL_SERIES			5.022	2.402		
HFRQ	-0.035	-0.002	-0.003	0.029	1.22	0.41
LFRQ	0.99597***	0.9180***	5.034	2.431	101.36	34.18
RESIDUAL	0.9835***	0.91596***	4.979	2.256	93.92	31.67

1863:01–1940:04

frequency and trend components of a time series, especially during the recent periods. This holds for the variance contribution too. The variance contribution is relatively greater from the low frequency and trend components of the time series. This is especially true for the more recent periods. The results obtained are robust to the subsamples. In nutshell, we did not find any evidence of US stock prices having been driven by short-term irrational behaviour. Our results support the view that the US stock market is driven mostly by fundamentals, which, in turn, are most likely rooted in economic growth and long-term returns on investment (Rapach and Zhou, 2013).

4 Conclusion

In this paper, the data of the S&P 500 index was decomposed into a number of IMFs and residuals, using the EEMD. The monthly data sets include the full sample ranging from 1791:08 to 2015:05 and three subsamples for the US stock prices: 1791:08 to 1862:12, 1863:01 to 1940:04 and 1940:05 to 2015:05. The division of the data into three subsamples gave a better idea of how the dynamics of the US stock market evolved over time, as well as the robustness of the results. The IMFs were generated in the order of highest to lowest frequency. The IMFs were analysed by three measures: mean, correlation with the original series and the contribution to the variability of the original series. It is shown that the residuals and low frequency IMFs indicate a very high correlation and account for very high variability in the original series. Also, it was found that the correlation of the long-term trend with the data and the variability contribution increased for the more recent samples. The IMFs and residuals were reconstructed into their high-frequency, low-frequency and trend components for the same full and subsamples. Again, it was found that the Pearson and Kendall correlation is comparatively higher for the lower-frequency and trend components of a time series, especially during the recent periods. The variance contribution was also found to be relatively greater from the low-frequency and trend components of the time series. The subsample results were found to corroborate the full-sample results. Therefore, it is concluded that, in general, US stock prices are not driven by the short-term irrational behaviour of investors, but seem to be driven mostly by fundamentals

Heaton and Lucas (2000); though, it is true that there have been episodes of bubbles, as indicated by Phillips et al. (2015).

References

Bai, J., and Perron, P., (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18(1): 1–22. https://ideas.repec.org/a/jae/japmet/v18y2003i1p1-22.html

Chang, T., Li, X.-L., Miller, S.M., Balcilar, M., and Gupta, R. (2015). The co-movement and causality between the U.S. real estate and stock markets in the time and frequency domains. *International Review of Economics and Finance* 38(1): 220–233.

Cheng, C.-H., and Wei, L.-Y. (2014). A novel time-series model based on empirical mode decomposition for forecasting TAIEX. *Economic Modelling* 36(2014): 136–141. http://www.sciencedirect.com/science/article/pii/S0264999313003933

Yu, H., and Liu, H. (2012). Improved stock market prediction by combining support vector machine and empirical mode decomposition. Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on, Hangzhou, 2012, pp. 531–534. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6407038&tag=1

Heaton, J., and Lucas, D. (2000). Stock prices and fundamentals. Ben S. Bernanke and Julio J. Rotemberg, (eds.), *NBER Macroeconomics Annual 1999*, Vol. 14: 213–264. http://www.nber.org/chapters/c11048.pdf

Huang, N.E., Shen, Z., and Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., and Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proceedings A of the Royal Society of London* 454(1971): 903–995. http://rspa.royalsocietypublishing.org/content/454/1971/903

MSCI (2010). Is there a link between GDP growth and equity returns? MSCI Barra Research Paper 2010-18. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1707483

Percival, D.B., and Walden A.T. (2000). Wavelet methods for time series analysis. Cambridge, UK: Cambridge University Press.

Phillips, P.C.B., Shi, S-P., and Yu, J., (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* 56(4): 1043–1078. http://onlinelibrary.wiley.com/doi/10.1111/iere.12132/abstract

Rapach, D., and Zhou, G. (2013). Forecasting stock returns. In: Elliott, G., and Timmermann, A. (Eds). *Handbook of economic forecasting.* Amsterdam: Elsevier.

Sun, J., and Sheng, H. (2010). Applications of ensemble empirical mode decomposition to stock-futures basis analysis. Information and Financial Engineering (ICIFE), 2010 2nd IEEE International Conference on, Chongqing, 2010, pp. 396–399. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5609386&tag=1

Tiwari, A.K., Dar, A.B., and Bhanja, N. (2013a). Oil price and exchange rates: A wavelet based analysis for India. *Economic Modelling* 31(1): 414–422. http://www.sciencedirect.com/science/article/pii/S0264999312003999

Tiwari, A.K., Dar, A.B., and Bhanja, N. (2013b). Stock market integration in Asian countries: Evidence from wavelet multiple correlations. *Journal of Economic Integration 28*(3): 441–456. https://ideas.repec.org/a/ris/integr/0608.html

Wang, B., Hu, Y.M., Du, Y.D., Zhai, Z.H., and Wu, X.Y. (2014) Study on the difference between wavelet analysis and EEMD in multi-scale decomposition of temperature and precipitation of Guangzhou. *Journal of Tropical Meteorology* 30(4): 769–776

Wu, Z., and Huang, N.E., (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis* 01 (1): 1–41. http://www.worldscientific.com/doi/abs/10.1142/S1793536909000047

Zhang, X., Lai, K.K., and Wang, S.Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics* 30(3): 905–918. http://www.sciencedirect.com/science/article/pii/S0140988307000436

Zhu, B., Wang, P., Chevallier, J., and Wei, Y. (2015). Carbon price analysis using empirical mode decomposition. *Computational Economics* 45(2): 195–206. https://ideas.repec.org/a/kap/compec/v45y2015i2p195-206.html

Appendix

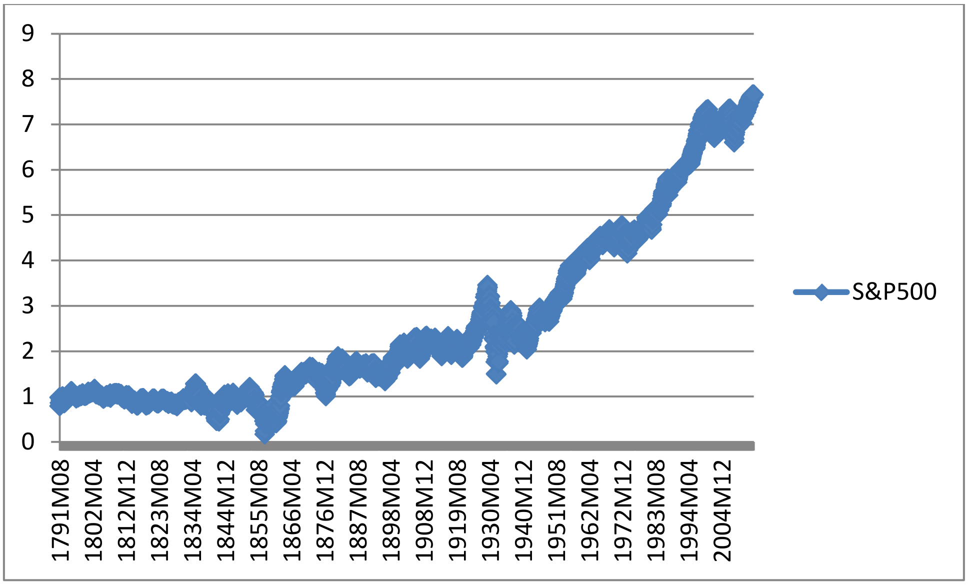

Figure A1: Natural Logarithms of S&P 500 Index (1791:08–2015:05)

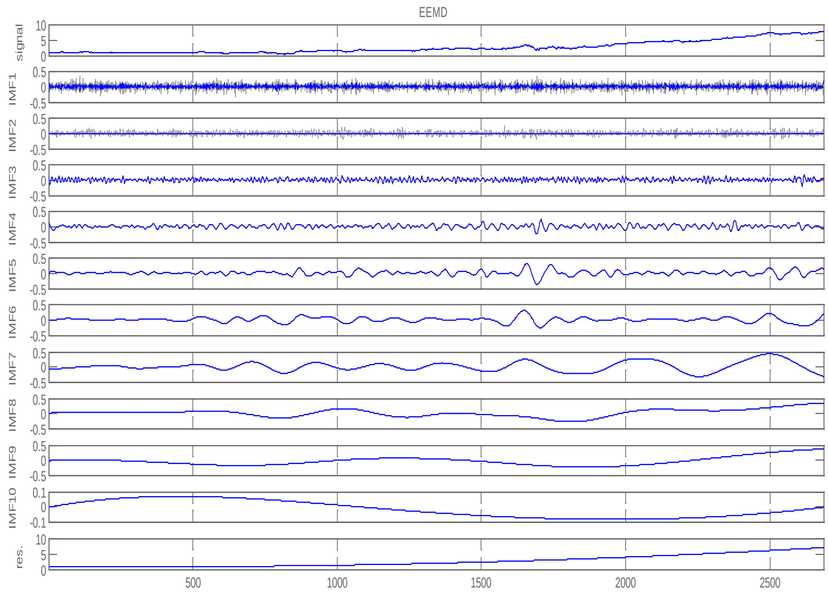

Figure A2: IMFs for the Full-Sample (1791:08–2015:05)

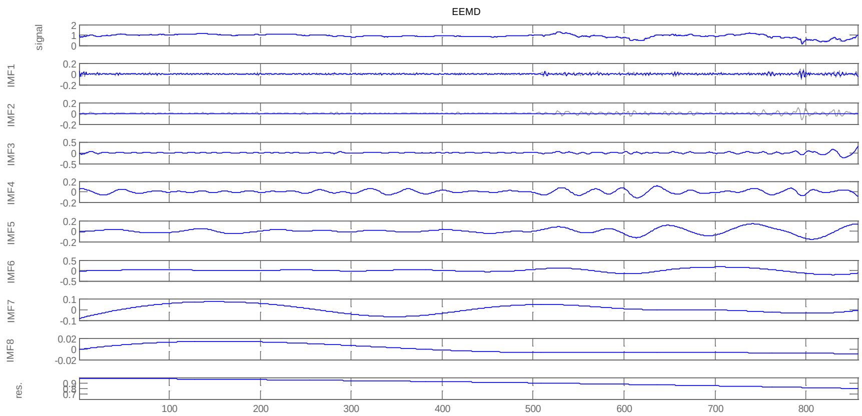

Figure A3: IMFs for the period 1791:05–1862:12

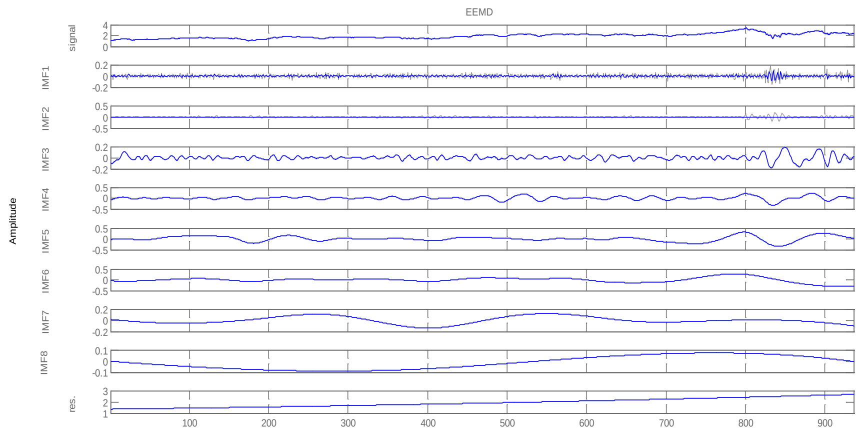

Figure A4: IMFs for the 1863:01–1940:04

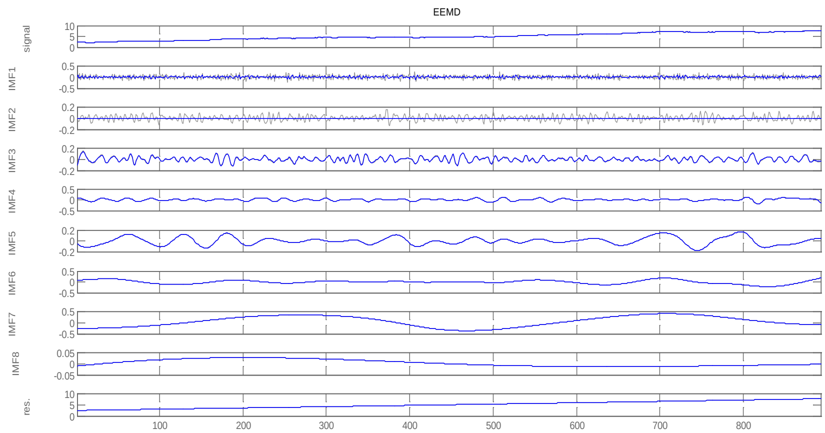

Figure A5: IMFs for the period 1940:05–2015:05

Economics
The Open-Access, Open-Assessment E-Journal

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

http://www.economics-ejournal.org/economics/discussionpapers/2016-9

The Editor