BANK OF GREECE
EUROSYSTEM

Working Paper

Disaggregating VIX

Stavros Degiannakis
Eleftheria Kafousaki

335

JANUARY 2025

BANK OF GREECE
Economic Analysis and Research Department – Special Studies Division
21, E. Venizelos Avenue
GR-102 50 Athens
Tel: +30210-320 3610
Fax: +30210-320 2432

www.bankofgreece.gr

Published by the Bank of Greece, Athens, Greece

ISSN: 2654-1912 (online)
DOI: https://doi.org/10.52903/wp2025335

DISAGGREGATING VIX

Stavros Degiannakis
Bank of Greece and Panteion University

Eleftheria Kafousaki
Panteion University

ABSTRACT

The present study highlights the economic profits of markets' participants, accumulated from the disaggregated forecasts of the stock market's implied volatility, generated from an ensemble modelling architecture. We incorporate six decomposition techniques, namely, the EMD, the EEMD, the SSA, the HVD, the EWT and the VMD and four different model frameworks that of AR, HAR, HW and LSTM, which are tested against a trading strategy. We diverge from quantifying forecast accuracy solely on statistical loss functions and report the cumulative returns of short or long exposure on roll adjusted VIX futures. The findings show that decomposing a time series into its intrinsic modes prior to modelling and forecasting, can result in generating forecast gains that are translated into improved profits for trading horizons of 1 to 22 days ahead. Important trading implications are drawn from these results.

Keywords: decomposition techniques, implied volatility forecasting, ensemble learning, objective-based evaluation criteria, VIX futures.

JEL classification: C22, C53, C61, G11, G17

Acknowledgements: The views expressed in this paper are those of the authors and not necessarily those of either the Bank of Greece or the Eurosystem.

Correspondence:
Stavros Degiannakis
Bank of Greece,
21 E. Venizelos Avenue, 10250
Athens, Greece
email: sdegiannakis@bankofgreece.gr

1. Introduction

Since the mid 00's when VIX futures and options were introduced as the tradable part of the VIX CBOE volatility index, trading volumes on these derivatives have reached tremendous heights, especially during turbulent, for the global economy, moments. The reason lies in the special features these products display, especially futures and the strong ties with VIX index and more generally the equity market (Szado, 2020; Fernandes *et al.*, 2014). These characteristics make them quite appealing especially for investors and risk managers, who use them daily to hedge uncertainty and diversify portfolios, transforming trading strategies based on short and long volatility exposure in options and futures contracts into top financial risk management tools, no matter if the use of derivatives encloses high risk of capital loss and demand careful manipulation. Therefore, engaging in derivatives markets' practices, requires the ability to accurately forecast the underlying implied volatility index, which is the key input for pricing them (Degiannakis *et al*., 2018). Forecasting implied volatility, which in our study is the VIX, the leading volatility index for the U.S stock market, was and will continue to be a demanding and of immense importance task, not only for those directly involving in markets but also for stakeholders, researchers and policy makers, who desire to navigate through past and present trends, but in the same moment desire to shape future regimes.

Literature has to provide an abundant of studies that model and forecast volatility, conditional, stochastic, implied or realized (Becker and Clements, 2008; Degiannakis and Filis, 2017; Kambouroudis *et al.*, 2021), especially realized that flourished the last 20 and so years through the studies of Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen *et al.* (2008) to name a few, due to the availability of ultra-high frequency data that completely altered the modeling landscape. The use of implied volatility primary involved in forecasting realized volatility, but steadily gained ground due to the information expected volatility encloses for the future path of a market, but also for macroeconomic conditions, energy commodities, speculative strategies, portfolio optimization etc. (Fernandes *et al.*, 2014; Bams *et al.*, 2017; Degiannakis and Filis, 2022). Various parametric frameworks have occasionally been used to assess the usefulness of volatility forecasts, such as ARIMA, ARFIMA, VAR (Konstantinidi *et al*., 2008), GARCH-type, ARMA-type

(Kampouroudis *et al*., 2016; Taylor, 2004) and HAR. HAR (Corsi, 2009) the state-of-the-art framework for modeling and forecasting volatility, was initially proposed for realized volatility (Bollerslev *et al.*, 2009; Degiannakis and Filis, 2017). Lately, it has been established for implied as well, since it proved to be efficient in producing meaningful forecasts and economic profits when incorporated in trading practices involving implied volatility derivatives (Delis *et al*., 2023).

Quite recently in the financial analysis, modeling theory and practice utilized non-parametric decomposition methodologies, paired either with parametric or non-parametric techniques, resulting in hybrid models (Risse, 2019). These hybrid frameworks, are the outcome of the evolution contacted in computer science that allowed the construction of powerful algorithms, liberating the scientific and engineering work processes. One of the major contributors of such algorithms by far is the signal processing field, followed by the machine learning, deep learning and artificial intelligence environments. Signal decomposition methodologies along with ensemble frameworks, suddenly flourished providing a more integrated analysis for time series data. These techniques, efficiently tackle with the restrictions raised from classic econometric models and effectively cope with the special characteristics, economic and financial data processes undergo, the non-linearity and non-stationarity (Vrontos *et al.,* 2021; Huang *et al.,* 1998; Civera and Surace, 2021; Prasad and Bakhshi 2022).

Decomposition frameworks, can reveal the peculiar features time series display and are capable of separating trend, high and low frequency components, periodicities, noise etc. Then, the retrieved components can be optimally analyzed and modeled separately, with the help of parametric models, if the nature of the retrieved component allows such, or with more advanced. Adaptive mode decomposition, variational mode decomposition, semi-variational mode decomposition, wavelet decomposition, convolution neural networks, recursive neural networks, deep neural networks are only few of the available algorithms combined together to conduct analysis, modelling and forecasting of complex systems, resulting in optimal statistical and economic gains (Yu *et al.,* 2008; Hewamalage *et al.*, 2021; Degiannakis *et al*., 2018; Rua, 2017).

We will not go into an extensive review of existing literature that exhibit the superiority of the above-mentioned techniques, as it exceeds the scope of this paper. Rather we will highlight some points of relevant studies that will better clarify the

contribution of this work amongst the bulk of adequate research. In particular, current studies incorrectly incorporate decomposition techniques in their modeling and forecasting frameworks, resulting in their disability to be utilized for real life applications (Liu *et al*., 2022; Wang *et al.* 2016, Guo *et al*., 2012). The reason lies in the boundary issue that stands for the inclusion of future info into the modeling system (Chen *et al*., 2022; Kaufman *et al*., 2012). This arises when one disaggregates entire time series data into distinctive components and then splits components into training and test sets in order to proceed with rolling out of sample forecast and evaluation techniques. As a result, the demonstrated low statistical loss functions and trading profits as the major gains of decomposition prior to modeling and forecasting, is the outcome of a rather biased procedure that suffers from future data leakage, data snooping and deterioration of the component selection process that each of the diverse decomposition frameworks, follows.

The difficult part is to split original dataset into smaller periods and observe how components and forecasts alter as new info gets updated iteratively on daily basis and new expectations of the future path are formed each consecutive day. At this very point lies the actual contribution of this study. More specifically, we disaggregate VIX index by taking original sample consisting of 2600 trading days and proceed by incorporating a rolling window approach, with a fixed window of 1000 observations that gets updated daily with new info. That said, we split original input into subsamples of 1000 observations. Each subsample ends at time t, which is the 1000 observation, in order to totally exclude the future info, t+1. Thus, in an iterative manner, each day we use 1000 observations as our training data set in order to decompose, utilizing diverse techniques, model and forecast for a completely unknown horizon of multiple days ahead, spanning from 1 to 22 days and for an out of sample period of 1600 trading days in place of our test data set. When above steps are completed, we proceed by adding all individual forecasted components, from each incorporated method, forming the final aggregated forecast.

Disaggregation of VIX is conducted via six signal decomposition techniques namely, the empirical mode decomposition (EMD) (Huang *et al.* 1998), the embedded empirical mode decomposition (EEMD) (Wu and Huang, 2009), the singular spectrum analysis (SSA) (Golyandina *et al*., 2001), the Hilbert vibration decomposition (HVD) (Feldman,2006), the empirical wavelet decomposition (EWD) (Gilles, 2013) and the

variational mode decomposition (VMD) (Dragomiretskiy and Zosso, 2014), where all exhibit different theoretical background, but all are successfully utilized in the analysis of real and complex signals. Furthermore, modeling takes place by employing four models, the Autoregressive model (AR), the Heterogeneous autoregressive framework (HAR), the Holt Winters (HW) and the Long short-term memory model (LSTM) (Chen *et al.*, 2021). Models were not randomly chosen but is the outcome of carefully inspecting the stochastic path of each and every component, their linear or non-linear dynamics. For comparison, the same models are also applied to the rolling samples of VIX as a unity, to form our benchmark models.

Additionally, another point we address is to assess how evaluation of forecasts should be conducted, especially when dealing with financial time series that are daily employed in real asset allocation practices. Evaluation criteria should be able to reflect the purpose for which forecasts are generated, in our case implied volatility forecasts, so as findings be comparable to pure market applications and able to be replicated.

There exists a literature strand that employs objective-based evaluation criteria developed to directly relate economic decisions with the generated forecasts (Elliot and Timmerman, 2008), contrasted to classical loss functions of MSE, MAE, RMSE, MAPE, etc. One of the first studies to make use of an economic loss function to evaluate the performance of volatility forecasts, is of Engle *et al*. (1993) who apply an options trading strategy, followed by Engle *et al*. (1996) who evaluate the ability of volatility forecasts to maximize trading profits. Since then, literature has recorded significantly contrasting findings of forecasts being evaluated through statistical loss functions along economic criteria (West *et al*., 1993, Degiannakis and Filis, 2022, Angelidis and Degiannakis, 2008, Becker *et al*., 2015). Patton and Sheppard (2015) consider these criteria as an indirect evaluation of volatility's forecast accuracy, since an economic criterion does not directly assess accuracy, rather measures the profit a market participant can extract from the forecast. The same states Taylor (2014) who maintains that indirect evaluation of volatility forecasts is a way to measure the value the end-user receives no matter if statistical loss functions are well defined. Furthermore, Mukherjee and Swanson (2021) point the usefulness of economic criteria such as trading strategies, are critical for highlighting the linkage between predictive accuracy and trading profitability.

Adding to this literature, we develop a simple yet effective objective-based evaluation framework in order to evaluate the generated volatility forecasts according to the purpose these forecasts serve. Since our study is centered to VIX index, that has an extensive network of derivatives products, we employ front month's VIX futures contracts as a fitted choice for our trading practice. Therefore, our forecasts are evaluated based on the profitability of naively trading the most liquid VIX futures. Depending on the signal the comparison between the aggregated forecast values generated daily and the actual VIX levels, sends, we take interchangeably short or long positions to VIX futures by reconstructing our position on a daily basis. Afterall, VIX futures should not be considered as long-term, buy-and-hold investments.

Moreover, we consider the roll of futures, that is the roll to the next month's contract. Roll takes place the 3rd week of each month, when front month's contract expires and an investor would have to decide whether to roll or not to the next month's contract since it is, its final settlement day. In this study, that day we close our position and roll to the following month's contract, no matter if that entails losses due to contango (roll cost) or gains due to backwardation (roll yield)[1]. At the end of the out of sample period, we report the cumulative returns, either positive (gains) or negative (losses). Obviously, there are many other trading strategies we could have enlisted especially because we employ VIX futures, a recognized portfolio diversification tool, but the extra complexity, disaggregation techniques may have imposed in our modeling architecture, dictated that we should keep things simple. Finally, in order to evaluate the risk-adjusted performance of our naïve trade in line with relative literature, we also adopt the Sharpe and Sortino ratios.

To the best of our knowledge, this is the first paper to compare six diverse decomposition frameworks in an out-of-sample rolling forecast practice for multiple horizons[2], that is tailored to the exact moment information becomes available. Our empirical findings suggest that decomposing prior to modelling can be multi beneficial. The evaluation of the out-of-sample forecasts shows that each technique performs differently and simply answers if decomposition techniques used that way, which is the

[1] Szado (2020), reports that a roll five days prior to maturity has lower carry over for VIX futures contracts and escapes the maturity effect, but here we have chosen to be risky and measure final outcome.
[2] The LSTM, part of the recursive neural networks has a detailed structure that for the case of the rolling scenery and for the multiple forecast horizons becomes computational heavy compared with other models incorporated in our study.

proper and the only unbiased way, can end with the significant outcomes, relative studies demonstrate. Even more, without focusing on the minimization of a statistical loss function[3], but instead on the maximization of the objective-based evaluation criterion employed here, we report significant economic gains for EEMD-based, EMD-based and SSA-based frameworks, which by far outperformed compared to other decomposition-based models or to the benchmark models. The fact that some outperformed while others failed to even compete in MSE terms is rather informative, as we noticed that the different modelling combinations of the components of each of the disaggregation techniques, can alter the final outcome and even generate a return of 44 times of invested capital[4]. Thus, if a trader invests $1 on VIX futures and decides daily of going long or short based on the VIX forecast, would earn multiple times her money during our out-of-sample period. Finally, we highlight the purpose, the generation of objectively accurate forecasts, should serve and stress out how critical it is, when producing forecasts of core financial indices, to assess the conditions under which markets operate and exposure is conducted, where there is a complete absence of any forward information. Thus, we maintain that decomposition enhances the economic outcome when outcome is evaluated based on objective-based criteria. Results remain robust against several robustness tests.

The remainder of this study is organized in the following manner. Sections 2, 3 and 4 present the six diverse decomposition techniques, the four forecast frameworks and the evaluation tests, providing all appropriate justification for their choice, the number of resulting components, the choice for the best fitted modelling frameworks, the final generated forecasts and the choice of the appropriate evaluation steps to verify robustness of empirical findings, respectively. Section 5, provides a description on the data used, their choice, their significance for this study and rεports their descriptive statistics. Section 6, presents the empirical findings of this study by presenting all relative tables and finally, section 7 sets the concluding remarks.

[3] Our study provides an indicative table where the classic mean square error is recorded for any of the proposed models, but only for highlighting of how illusive this turns out to be and how contradictive cumulative returns are for models with infinitesimally small divergence in this statistical measurement.
[4] Valid for EEMD-HW-AR-HAR model for the 5 trading days ahead horizon.

2. Decomposition techniques

This study incorporates 6 diverse decomposition techniques. All methods were carefully selected to be applicable to the implied volatility index. The fact that most were originally developed for application on continuous time series, earthquake motions, liquid motions, vibration mechanics, medical data, mechanical engineering, is not restrictive for a more general use, in our case on financial time series. Generally speaking, there are always restrictions, restrictions as for the length of the sample to be decomposed, the randomness, the aperiodicity and other relative issues, for the reason that methods will not have enough data to conduct analysis or will not be able to effectively capture all those features that would allow for a more thorough analysis. Huang *et al*. (1998), when introduced EMD as a general use method, specified that few are the datasets either from natural or artificial phenomena that satisfy all definitions imposed by these processes-methods and maintained that the retrieved components do not guarantee a well-defined physical meaning, something true for all decomposition techniques, however in most cases, IMFs do carry a physical significance. Feldman (2011) who provides an extensive description of Hilbert Transform (HT), points out that it can be applied to any oscillatory signal, since mathematically it is correct for any vibration, no matter if in practice is essential mostly for narrowband signals and states that from a general point of view, the analytic signal method is equally applicable to deterministic and random processes because it allows investigating any oscillating time function, no matter if it does not separate then at all. It is still a good method for solving problems of stationary and non-stationary vibrations. Thus, we deduce that the selection and application of any of these techniques that each has a differentiated theoretical and practical core, solely depends on the perspective desired (Feldman, 2011). Inevitably all these methods come with advantages and disadvantages[5] and that is the reason why occasionally new extensions come forth.

In this study we did not go into an extensive analysis of the components, that makes sense to applications from other scientific fields, since our analysis is targeted elsewhere. The analysis conducted on the individual components is only for assisting us in the effort to discover their data generating process in order proper models to be

[5] All techniques have pros and cons, but our intention is not to extent through the issues the methods face, either mathematical or technical, rather provide a basic understanding of the underlying theory. Those who desire a more thorough learning can always refer to the original papers of their initiators.

fitted, although, we did notice differences between EMD and EEMD performance. Thus, there must be indeed an impact on empirical findings by the choice of the methodologies. Moreover, as for the number of components that each method returns is dictated, for EMD, EEMD and HVD by the very own procedures as there are certain stopping criteria. For the cases of EWT and VMD it is user defined, while for SSA depends on the choice of specific parameters and auxiliary methods. Hence, for the methods that at least have a user dependency, the number of components, was dictated by the reason we originally engaged in this disaggregation practice, forecasting. Forecasting demands accuracy and as such the closer approximation of the summation of retrieved components to original input is the key for our decision and definitely can have an impact on the reported empirical findings.

2.1. The empirical mode decomposition method

It was the year 1998 when the study of Huang *et al*. (1998) published, proposing a new 2-step tool to cope with the difficulty faced when trying to analyze noisy, non-stationary and non-linear time series and partially replace the Fourier spectral analysis that due to its restrictions, would not fit for the analysis of such data. The first step was the EMD that could decompose any complex data through a spline algorithm into a finite and small number of intrinsic mode functions (imfs) that were complete and orthogonal (Huang *et al*., 2003) and would localize an event on time. The second step was the Hilbert transform of the imfs that would localize event in the frequency domain and would allow for a more efficient and meaningful analysis. Since then, EMD has been widely applied and paired with parametric and non-parametric models for decomposing and modeling wind-speed data, crude-oil price time series and other financial time series (Ali *et al*., 2023; Shu and Gao, 2020; Jin *et al*., 2022).

EMD is an adaptive and direct algorithmic method for decomposing a signal. The signal in order to be decomposed into a zero-mean amplitude modulated imfs and a residue, must comply with three assumptions (Huang *et al*., 1998), 1) have at least two extrema, one minimum and one maximum, 2) the characteristic time scale to be defined by the lapse of time between the extrema and 3) when there is a total absence of extrema and only the presence of inflection points, then differentiate one or more times in order to reveal the extrema. Furthermore, an imf must satisfy two conditions, 1) the number of extrema and the number of zero-crossings in the entire sample to be either equal or

differ by one[6] and 2) at any point the mean value of the envelope defined by the local maxima and the envelope defined by local minima to be zero. Thus, in order to identify an imf, EMD follows an iterative process that is better described in five steps:

1. For any signal $\mathrm{X}_t = \{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_t\}$, which in our case is the VIX_t, we identify all extrema.
2. Then, we connect all local maxima with a cubic spline and create the upper envelope. We do the same for the local minima to create the lower envelope. Upper and lower envelopes should cover all data points, afterall, the creation of the envelopes is the one that will help in separating components with close frequencies (Feldman, 2011).
3. We designate the mean of the upper and lower envelopes as $m_1 = (\mathrm{env}_{max} - \mathrm{env}_{min})/2$. We subtract m_1 from original signal and retrieve the first component $\mathrm{c}_1 \rightarrow \mathrm{X}_t - \mathrm{m}_1 = \mathrm{c}_1$
4. Does the component comply with imposed restrictions? If so, then it is an *imf*, if not we replace X_t with c_1 and iterate over *i* times until component equals an *imf*[7]. That is called the sifting process.
5. In the case the resulted component satisfies the characteristics of an *imf*, then $imf_1 = c_i$ (with $i = 1, 2, \dots$). And we get the first *imf*. After having the first *imf* we subtract it from original series $\mathrm{X}_t - imf_1 = r_1$. The r_1 now takes the place of signal and process starts over until all imfs are retrieved and only a residue is left, res_t, that is a monotonic function. At that point EMD terminates. The first imf is the one with the highest frequency, and progressively frequency of ongoing imfs lessens as gradually moves towards the final monotonic residue.

The summation of all *imf*s and residue returns original series. Among proposed and adopted decomposition techniques, EMD is the only complete method whom summation of components (imfs and residue) returns the exact same signal $\rightarrow \mathrm{X}_t = \sum_{k=1}^{t} imf_k + \mathrm{res}_t$.

[6] Huang *et al*. (2003) argue that imfs retrieved based on this condition are orthogonal and not over sifted. Basically, one of the cons that is being addressed to EMD is the lack of a mathematic formula and the fact that although imfs are indeed orthogonal, orthogonality itself cannot be proved.

[7] In the original work of Huang *et al*. (1998) the stopping criterion of the sifting process with which a component is recognized as an imf is given by a standard deviation function. Other studies use different criteria. In our study we use the standard deviation function $\sum_{t=0}^{T}\left[\frac{|(\left(c_{1(i-1)}(t) - c_{1i}(t)\right)|^2}{c_{1(i-1)}^2(t)} < 0.2\right.$.

There have been proposed various extensions to the above process regarding cubic spline end conditions or stopping criteria. In this study we follow a distinct literature branch that adopts a specific cubic spline end criterion, that is to place the end points of the slope of the cubic spline equal to 0. According to Peel *et al*. (2007) this condition returns fewer imfs and tends to be more efficient compared to other cubic spline rules proposed for EMD (Rilling *et al*., 2003; Flandrin *et al*., 2004). We evidence the same findings[8]. We decompose VIX on a daily basis for an out-of-sample period of 1600 trading days, each day the process ends on 5 or 6 components consequently. As such the number of components is dictated by the stopping criterion imposed that terminates process when only the long term trend component is left and finally by the very own structure of the input, thus it alters through successive iterations. As for the nature of components, each follows a different stochastic process that determines the choice of the proper model for the analysis that follows.

2.2. The Ensemble empirical mode decomposition method

EMD process although being an adaptive and fully data driven decomposition method, which is widely implemented, it has frequently been accused of mode mixing, an outcome of signal intermittency (Huang *et al*., 1999). This can raise annoying issues in the analysis conducted in the resulting components that may lack in any physical meaning through the second step that is the Hilbert transform. In order to cope with this issue Wu and Huang (2009) proposed the EEMD based on the study of Flandrin *et al*. (2004). This noise-assisted signal extraction method infuses signal with white noise of finite amplitude that leads to efficient frequency separation. Thus, the resulting imfs constitute the mean of an ensemble of repeated iterations (Yeh *et al*., 2010). The process of EEMD is in the same line as EMD and follows the same steps with the difference that:

1. We add a white noise series, wn_t to the original signal, s_t: $\mathrm{X}_t = \mathrm{s}_t + \mathrm{wn}_t$.
2. We proceed with the decomposition as conducted in EMD and retrieve *imf*s.
3. We iterate[9] above steps many times adding different white noise series.
4. We calculate the ensemble means of the *imf*s returned from every iteration. The means returned are the final *imf*s of entire process.

[8] Classic cubic spline conditions returned 10 and 11 components, the zero ends returned 5 and 6 components interchangeably.
[9] The number of ensembles is dictated by the data itself. For this study, we went through 30 trials.

According to Wu and Huang (2009) this repeated addition of different white noise series of finite amplitude not infinitesimal, in the original signal are cancelled out in the end result and the mode mixing problem is efficiently tackled without perturbating original signal. Moreover, the addition of finite amplitude white noise allows EMD to act as dyadic filter bank (Flandrin *et al.,* 2004), and accomplish meaningful *imf*s.

In this study, we thought the addition of EEMD would clarify if the *imf*s returned are more refined compared to those returned by EMD. We noticed that original EMD was efficient enough to decompose VIX and no mode mixing was optically inspected[10], but despite the fact that both methods in each of the 1600 iterations terminated at 5 or 6 components interchangeably, the reported economic gains of the implemented trade, at section 6 and Table 5, differ. Thus, although both methods are almost identical if not for the addition of white noise series, the infinitesimally differences between the resulted components of the EMD and EEMD processes seems to have a critical impact at the very end. Many are the studies who end with promising results when EEMD is incorporated in hybrid models or EEMD extensions are used (Dong *et al.,* 2019; Sun *et al.,* 2018; Tang *et al*., 2018).

2.3. The Singular Spectrum Analysis method

SSA, constitutes another non-parametric time series analysis and decomposition technique with various applications and many extensions (Golyandina and Zhigljavsky, 2013; Golyandina *et al.,* 2001; Hassani 2007) that approximately decomposes a time series into noise, periodicities, trend etc. Although SSA is demonstrated and primary used as a decomposition and filtering technique, it can also be used for forecasting purposes as in Degiannakis *et al.* (2018), Thomakos *et al.* (2002), Sulandari *et al*. (2020) to name a few, with remarkable results. In this study we will not test its forecasting potential, rather its efficiency to generate meaningful components. Afterall, the primary reason why one uses a decomposition technique is to separate the different harmonics, scales, frequencies, periodicities, trend, noise etc., who reveal the inner nature of a time series and allow for a more targeted analysis.

SSA is performed by following some distinctive steps of a robust mathematical path (Hassani and Thomakos, 2010). Process has two stages, the decomposition and the

[10] It is important though to mention that for series as earthquake data and climate data, where mode mixing issues are obvious, where analysis for bio sustainability is the main target followed by forecast of nature's elements, EEMD method has proved to be optimal.

reconstruction. The very first step is to construct the trajectory matrix and "transform" the time series into a multidimensional matrix:

Stage 1: Decomposition

1. Embedding. Let X_N be a time series of length *N*, $\mathrm{X}_N = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_N)$. We need two parameters, *L*, which is called embedding dimension and *K*, two integers that will constitute the number of rows and columns of the trajectory matrix respectively, with L being $(1 < L < N)$[11] and $K = N - L + 1$. We form a sequence of *L* and *K* lagged vectors out of the original sample. So, the X matrix of eq.1 constitutes a Hankel matrix and has equal anti-diagonal elements:

$$\mathrm{X} = [\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{K}] = (\mathrm{x}_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \mathrm{x}_3 & \cdots & \mathrm{x}_K \\ \mathrm{x}_2 & \mathrm{x}_3 & \mathrm{x}_4 & \dots & \mathrm{x}_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_L & \mathrm{x}_{L+1} & \mathrm{x}_{L+2} & \cdots & \mathrm{x}_N \end{pmatrix}. \quad (1)$$

2. Decomposition. In the core of SSA lies the singular value decomposition (SVD). Having constructed the trajectory matrix, we perform SVD on X and decompose it into a sum of rank-one matrices:

$$X = X_1 + X_2 + \cdots + X_d, \quad (2)$$

where d denotes the rank of X and $\mathrm{X}_i = \Sigma_i \sqrt{\lambda_i} \mathrm{U}_i \mathrm{V}_i^{\mathrm{T}}$ $(i = 1, 2, \dots, d)$. SVD process returns the eigenvalues λ_i of XX^{T} in decreasing order of magnitude, $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0)$, the left singular vectors of X, U_i and V_i the right singular vectors. The assortment of $(\sqrt{\lambda_i} \mathrm{U}_i \mathrm{V}_i)$ is called the i^{th} eigentriple of SVD (Hassani *et al.* 2021).

Stage 2: Reconstruction

1. Grouping. Grouping is the one that defines the way the individual eigentriples of Eq. 2, (1, 2, …,*d*), are going to be grouped with each other into *m* disjoint subsets $\mathrm{I}_1, \mathrm{I}_2, \dots, \mathrm{I}_m$ called eigentriple grouping[12]:

$$X = \mathrm{X}_{\mathrm{I}1} + \mathrm{X}_{\mathrm{I}2} + \cdots + \mathrm{X}_{\mathrm{Im}}. \quad (3)$$

[11] According to Hassani *et al.* (2021) *L* plays an important role during the reconstruction phase and sets how well reconstructed time series approximates original one and how well efficient separation of components is dealt. Values of *L* varying in between ($2 < L < N/2$) have higher resolution. In this study we have chosen *L* to be 400, dictated by the length of VIX subsamples that consist of 1000 observations. Parameter *m* is 4, since we get the trend, two elementary components and noise.

[12] For the case where $m = d$ and $I_j = \{j\}, j = 1, 2, \dots, d$, the corresponding group is called elementary.

2. Diagonal averaging. Having grouped the eigentriples, the final step is to transform these X_{Ij} matrices into new series/components of length *N*, the length of original series. In order to achieve that we first transform these matrices into Hankel matrices Y, of $L \times K$ dimensions and elements y_{ij}, for $1 \leq i \leq L$ and $1 \leq j \leq K$. Then, via performing diagonal averaging on Y, we transform it into the desired series $(y_1, y_2, \dots, y_N)$:

$$\widetilde{y_s} = \sum_{(l,k)\in A_s} y_{l,k} \,/|A_s|, \tag{4}$$

where $A_s = \{(l,k): l + k = s + 1, 1 \leq l \leq L, 1 \leq k \leq K\}$ that is the averaging process of the antidiagonals and $|A_s|$ the number of elements in the set of A_s. Thus, from the original series of $(x_1, x_2, \dots, x_N)$, the entire process returns the reconstructed series of $\widetilde{X}^{(k)} = (\tilde{x}_1^{(k)}, \tilde{x}_2^{(k)}, \dots, \tilde{x}_N^{(k)})$ that are the SSA *m* components and their summation approximates back original series:

$$x_n = \sum_{k=1}^{m} \tilde{x}_n^{(k)}, \qquad n = 1, 2, \dots, N. \tag{5}$$

Diagonal averaging completes the SSA decomposition process but still the step of grouping remains the most intrinsic one. That is why different frameworks have been proposed part of the separability issue faced in SSA that signals the way distinct components are going to be grouped together. Golyandina *et al.* (2001) propose the weighted correlation concept along with the graphs of eigenvectors that can reveal components who strongly interact. Hassani *et al.* (2021) against the classic weighted correlation propose hierarchical clustering methods and there is also the relative entropy[13] that we adopt. Thus, based on this criterion of the most efficient way to group eigentriples that terminates process, we ended generating 4 components in each iteration.

2.4. The Hilbert Vibration Decomposition method

In the same spirit to EMD, that forms the first step for Hilbert Huang transform, Feldman (2006) introduced the HVD, as part of the evolution conducted in signal analysis. HVD is a decomposition method used especially for the analysis of

[13] Relative entropy or else the Kullback–Leibler divergence, part of probability and information theory, is a type of statistical distance and measures the similarity between two probability density functions (Theodoridis, 2020).

mechanical vibrations, earthquake motions (Huang *et al.*, 2012), and for the analysis of Electroencephalograms (EEG), Electrocardiograms (ECG) or Seismocardiograms (SCG) (Shankar *et al.*, 2021; Singh *et al.*, 2022) among others. HVD constitutes an iterative algorithm, where in each iteration the component with the highest energy is subtracted from original data relying on the synchronous demodulation process (Feldman and Braun, 2017). HVD follows some distinctive rules and must comply with three assumptions, 1) the underlying vibration to be the outcome of the superposition of quasi-harmonics functions, 2) the envelopes of each vibration component to differ, and 3) the total length of the vibration to include several longest periods of the corresponding slowest component (Feldman, 2006). In the core of HVD lies the Hilbert transform (HT), a linear operator designated for the analysis of complex signals with varying amplitude and frequency through the course of time. For a vibration process**, x(t)**, the HT takes the form of an integral transform considered a Cauchy Principal Value:

$$\mathrm{H}[\mathrm{x}(t)] = \tilde{\mathrm{x}}(t) = \pi^{-1} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} dt, \tag{6}$$

where $\tilde{\mathrm{x}}(t)$ is the HT of x(t). Let us consider x(t) being a multicomponent signal, then it can be denoted as the sum of monocomponents of slow varying instantaneous amplitudes and frequencies:

$$\mathrm{x}(t) = \sum_{\mathrm{l}} \mathrm{a_l}(t)\cos\left(\int \omega_l(t)dt\right), \tag{7}$$

where $a_l(t)$ represents the instantaneous amplitude (envelope) and $\omega_l(t)$ denotes the instantaneous frequency of the i^{th} component. For estimating these parameters each time, in every successive iteration, we have to obtain the analytical signal, X(t), expressed by the summation of the real part along the imaginary part and by the exponential form (Feldman, 2011) both appearing in Eq. 7:

$$\mathrm{X}(t) = \mathrm{x}(t) + \mathrm{j}\tilde{\mathrm{x}}(t) = \mathrm{A}(t)\mathrm{e}^{j\varphi(t)}. \tag{8}$$

The instantaneous amplitude (envelope, magnitude) is then estimated as:

$$\mathrm{A}(t) = |\mathrm{X}(t)| = \sqrt{\mathrm{x}^2(t) + \tilde{\mathrm{x}}^2(t)} = \mathrm{e}^{\mathrm{Re}[\log(\mathrm{X}(t))]}, \tag{9}$$

and its instantaneous phase as:

$$\varphi(t) = \arctan\left(\frac{\tilde{\mathrm{x}}(t)}{\mathrm{x}(t)}\right) = \mathrm{lm}\big[\log\big(\mathrm{X}(t)\big)\big]. \quad (10)$$

From the first derivative of the instantaneous phase, we get the instantaneous angular frequency $\omega(t) = \dot{\varphi}(t)$. Now, the slow varying vibration component can be extracted by imposing a low pass filter on the instantaneous envelope and frequency respectively that will help subtract the largest vibration component out of initial signal $\mathrm{x}_{l-1}(t) = \mathrm{x}(t) - \mathrm{x}_{\mathrm{l}}(t)$ and then treat $\mathrm{x}_{l-1}(t)$ as the initial and repeat process until its termination. The cutoff frequency of the low pass filtering[14] is the one responsible for the frequency resolution of the HVD process and should be of a small value because on every successive iteration, after having subtracted the previous frequency, the frequency of the next to be extracted component dominates and so components with close frequencies must be able to be efficiently separated. The HVD process terminates when the standard deviation difference between two successive components become ≤ 0.01. Overall, HVD constitutes an unpretentious process, easily implemented and computationally fast. In this study HVD terminated at 5 components in each iteration as was dictated by the standard deviation stopping criterion.

2.5. The Empirical Wavelet Transform method

Among the concepts that seem to dominate the signal processing field, lies the wavelet theory with a vast literature dedicated on wavelets' features and their use (Meyer, 1997; Daubechies, 1992). Inspired by the wavelet theory and the continuous and discrete wavelet transforms, Gilles (2013) proposed a new adaptive method for decomposing non-stationary and non-linear signals into amplitude modulated-frequency modulated (AM-FM) components, the EWT. Since then, EWT appeared in the analysis of power system signals, wind data, medical disease diagnosis, seismic data, machine fault diagnosis, image processing etc. (Liu *et al.*, 2016; Singh and Sunkaria, 2016; Beoula *et al*, 2017), but also, in combination with parametric and non-parametric frameworks for forecasting short-term wind speed, drought etc. (Hu and Wong, 2015; Shaari *et al.*, 2018).

EWT builds a family of wavelets that can easily be adapted to the signal to be decomposed. The very starting point and the most crucial one in the entire process, is

[14] In our study we use the Elliptic low pass filter that fits perfect in order to efficiently manage the randomness that financial time series experience. The choice of filter is critical for the efficient disaggregation.

how to segment the Fourier spectrum of the signal. Each segment denotes a mode, so there are equivalent segments corresponding to the inspected components. This step seems to be and the most controversial as one has to predefine the modes the moment that any prior relevant information of the analyzed to be signal is absent. Thus, setting the boundaries of Fourier segments becomes demanding and various algorithms have been proposed to efficiently cope with this issue (Shi *et al*. 2017). The next step is to define the empirical scaling function and the empirical wavelets that will act as band pass filters on each predefined segment. Let us assume that we segment Fourier support [0, π] into N points. Let ω_n be the boundaries of the different segments with $\omega_0 = 0$ and $\omega_N = \pi$. Each segment lies in $[\omega_{n-1}, \omega_n]$. Around every ω_n a transition phase is defined, T_n, with width $2\tau_n$ and thus the empirical scaling function is defined by:

$$\widehat{\varphi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \le \omega_n - \tau_n\,, \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n + \tau_n)\right)\right] & \text{if } \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

And the empirical wavelets by:

$$\widehat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } \omega_n + \tau_n \le |\omega| \le \omega_{n+1} - \tau_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}(|\omega| - \omega_{n+1} + \tau_{n+1})\right)\right] & \text{if } \omega_{n+1} - \tau_{n+1} \le |\omega| \le \omega_{n+1} + \tau_r \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n + \tau_n)\right)\right] & \text{if } \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Assuming $\beta(x)$ being an arbitrary function $C^k([0,1])$, where:

$$\beta(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 & \text{if } x \ge 1 \end{cases} \text{and } \beta(x) + \beta(1-x) = 1 \quad \forall x \in [0,1]. \tag{13}$$

Setting τ_n proportional to ω_n, $\tau_n = \gamma\omega_n$, with $0 < \gamma < 1$, then above equations can be rewritten as:

$$\widehat{\varphi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \le (1-\gamma)\omega_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & \text{if } (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

And

$$\widehat{\psi}_{\mathrm{n}}(\omega) = \begin{cases} 1 & \text{if } (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_{n+1}}(|\omega| - (1-\gamma)\omega_{n+1})\right)\right] & \text{if } (1-\gamma)\omega_{n+1} \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & \text{if } (1-\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_n \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Now, in order to construct the EWT, denoted as $W_f^{\varepsilon}(n,t)$, the detail and approximation coefficients have to be estimated given by the inner products of empirical wavelets and scaling function respectively:

$$\mathrm{W}_{\mathrm{f}}^{\varepsilon}(n,t) = \langle \mathrm{f}, \psi_n \rangle = \int \mathrm{f}(\tau)\overline{\psi_{\mathrm{n}}(\tau - \mathrm{t})}\mathrm{d}\tau = (\hat{\mathrm{f}}(\omega)\overline{\widehat{\psi}_n(\omega)}^{\mathrm{V}}, \quad (16)$$

$$\mathrm{W}_{\mathrm{f}}^{\varepsilon}(0,t) = \langle \mathrm{f}, \varphi_1 \rangle = \int \mathrm{f}(\tau)\overline{\varphi_1(\tau - t)}d\tau = (\hat{\mathrm{f}}(\omega)\overline{\widehat{\varphi}_1(\omega)}^{\mathrm{V}}, \quad (17)$$

where $\widehat{\psi}_n(\omega)$ and $\widehat{\varphi}_1(\omega)$ are estimated above through Eq. 14 and Eq. 15 and the reconstruction:

$$\begin{aligned} \mathrm{f}(t) &= \mathrm{W}_{\mathrm{f}}^{\varepsilon}(0,t) * \varphi_1(t) + \sum_{n=1}^{N} \mathrm{W}_{\mathrm{f}}^{\varepsilon}(n,t) * \psi_n(t) \\ &= (\widehat{\mathrm{W}_{\mathrm{f}}^{\varepsilon}}(n,\omega)\widehat{\psi}_n(\omega) + \sum_{n=1}^{N} \widehat{\mathrm{W}_{\mathrm{f}}^{\varepsilon}}(n,\omega) * \psi_n(\omega))^{\mathrm{V}}. \end{aligned} \quad (18)$$

Finally, the empirical mode, f_k returned by the EWT is expressed as:

$$\mathrm{f}_0(t) = \mathrm{W}_{\mathrm{f}}^{\varepsilon}(0,t) * \varphi_1(t), \quad (19)$$

$$\mathrm{f}_k(t) = \mathrm{W}_{\mathrm{f}}^{\varepsilon}(k,t) * \psi_\kappa(t). \quad (20)$$

The empirical modes returned out of EWT come in ascending frequency order. Process starts from separating the component with the lowest frequency and terminates to the one with the highest. Actually, the number of modes and the point that terminates process is user defined, after inspecting the Fourier spectrum. For this method we defined 4 components for each iteration.

2.6. The Variational Mode Decomposition method

The final decomposition method applied in our study is attributed to Dragomiretskiy and Zosso (2014), who proposed a non-recursive variational model that could address limitations rising from corresponding decomposition models such as

EMD, HVD or EWT etc., the VMD. The use of this method has accelerated and many are the studies to incorporate VMD along neural networks, reconstruction techniques or optimization techniques in order to forecast oil price, short-term power load, financial data etc. (Lahmiri, 2016; Li *et al*., 2020; Ping Yu *et al*., 2021).

Dragomiretskiy and Zosso (2014), utilize three concepts, the Wiener filtering, the HT and the frequency mixing and heterodyne demodulation to synthesize the VMD process that decomposes a real signal f(x) into distinct modes. Each mode is required to be compact around a central frequency, ω_k determined through the process, so, they determine the bandwidth of the modes by following three steps, 1) they compute analytic signal of each mode through HT and obtain a unilateral frequency spectrum, 2) they shift mode's frequency spectrum to baseband, by mixing it with an exponential tuned to the respective estimated center frequency, and 3) they estimate the bandwidth through the H1 Gaussian smoothness of the demodulated signal. So, they end up in a constrained variational problem:

$$\min_{u_k,\,\omega_k,} \left\{ \sum_{\mathrm{k}} \left\| \partial_{\mathrm{t}} \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mathrm{u}_k(t) \right] \mathrm{e}^{-j\omega_k t} \right\|_2^2 \right\} \text{ s.t. } \sum_k \mathrm{u}_k = \mathrm{f}(x), \tag{21}$$

where u_k denotes the modes, ω_k denotes their central frequencies, * denotes convolution and the summation of all u_k approximates original signal. The solution to above problem is given through Lagrangian multipliers, λ and a quadratic penalty term with the assistance of the alternate direction method of multipliers (ADMM) (Gabay and Mercier, 1976; Eckstein and Bertsekas, 1992) an iterative sub-optimization method, so the Lagrangian takes the form:

$$\begin{aligned} \mathcal{L}(\mathrm{u}_k, \omega_k, \lambda) = \mathrm{a} \sum_{\mathrm{k}} \left\| \partial_{\mathrm{t}} \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mathrm{u}_k(t) \right] \mathrm{e}^{-j\omega_k t} \right\|_2^2 \\ + \left\| \mathrm{f}(t) - \sum_k \mathrm{u}_k(t) \right\|_2^2 + \langle \lambda(t), \mathrm{f}(t) - \sum_k \mathrm{u}_k(t) \rangle . \end{aligned} \tag{22}$$

The process returns the predefined from the user modes. The Lagrangian multiplier when used along with the penalty term ensures that the summation of all returned components will reconstruct actual signal not perfect but approximately close. The inclusion of only the penalty term will slightly vary the final outcome but will deal

more sufficiently with noisy data. After experimentation on the number of components that best approximates original series, we ended that 5 components were the most resolute approximation and thus we defined 5 components in each of the 1600 iterations, who consist the out-of-sample period.

3. Modelling and forecasting frameworks

Modeling and forecasting consist the second step of this study. Our goal is to fit optimal models to the individual components. By optimal we mean models that best fit to the stochastic process components undergo. Therefore, we applied all appropriate statistical tests[15] that would clarify the one that would capture the linear and the non-linear patterns. Forecasts are then generated in fully alignment to the fitted model specifications. Consequently, the four models proposed here, were selected carefully since each component holds a distinctive characteristic of original time series that we managed to isolate. Models are also applied to original series for comparison reasons to investigate whether the disaggregation step as conducted here, by following the only unbiased way to prohibit data leakage and data snooping, allows for forecast accuracy apart from added complexity. Therefore, entire process is designed to warrant a robust and unbiased outcome.

Moreover, it is critical to proceed by incorporating a rolling window approach with a fixed window of 1000 observations. Initial sample consists of 2600 trading days ($T_{total} = 2600$). Thus, we split sample into rolling subsamples of 1000 observations that are used daily for training ($T_{train} = 1000$) that is for decomposing, in-sample estimation and forecasting. The remaining observations are then used for testing, that is for the out of sample validation ($T_{test} = 1600$). The first rolling sample spans from 20th of August 2012 up to 12th of August 2016. Each day in our rolling sample, we drop the oldest observation, include a new one and proceed over and over with decomposition, model fitting and forecasting for a completely unknown future path of 1, 5, 10 and 22 trading days ahead[16]. Then by aggregating the individual forecasts, we form the final forecast for each model combination, for each of the proposed techniques

[15] For the parametric frameworks, we did consider the fit of the model, the significance of model's parameters, the R^2, used the available information criteria that would specify the order and test for the validity of the conditions that should hold, for a process to be unbiased, that said we did check for heteroskedasticity, normality, etc.

[16] For the importance of considering multi step ahead forecast horizon one can resort to Degiannakis and Filis (2022).

and for each of the chosen horizons. In every iteration, forecasts are generated based on data that is available to forecaster at time $t = 1000$.

Subsections that follow, present the proposed frameworks. We end up with 28 model combinations and 4 benchmark models for comparison reasons, in total 32 models. Table 1A at the Appendix section, presents a detailed list of the decomposition-based models and informs of the forecasting framework that was applied to the individual components that are enclosed in each model. Models are distinguished by disaggregation methods' abbreviations. For components that constitute a white noise process, no modelling takes place. Figure 1 illustrates the modeling architecture.

[FIGURE 1 HERE]

3.1. The Autoregressive model

The first model incorporated is the AR model. The inclusion of an intrinsic econometric model along with the more advanced ones, helps to model and forecast components that bare a first or higher order autoregressive pattern. For brevity, we present the model's specification for the case of the first order, for modelling the logarithm of VIX components, for each of the decomposition techniques[17], that is:

$$\log\left(\mathrm{VIX}_{i,j,t}\right) = \mathrm{w}_0^{(t)} + \mathrm{w}_1^{(t)} \log\left(\mathrm{VIX}_{i,j,t-1}\right) + \varepsilon_t, \tag{23}$$

where, $\mathrm{w}_0^{(t)}$ and $\mathrm{w}_1^{(t)}$ represent the rolling estimated coefficients, i is the number of the component, $j = \{\mathrm{EMD}, \mathrm{EEMD}, \mathrm{SSA}, \mathrm{HVD}, \mathrm{EWT}, \mathrm{VMD}\}$ and ε_t are the residuals that thought to be normally distributed, $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$. Forecasting for the $t+h$ horizon, with h = {1,5,10,22} is conducted through:

$$\mathrm{VIX}_{i,j,t+h|t} = \exp\left(\mathrm{w}_0^{(t)} + \mathrm{w}_1^{(t)} \log\left(\mathrm{VIX}_{i,j,t+h-1}\right)\right). \tag{24}$$

3.2. The Heterogeneous Autoregressive model

The HAR, the state-of-the-art framework for volatility modelling of Corsi (2009), is an additive linear combination of indicators of volatility components at different time horizons, fully aligned with markets' fractal structure. HAR has proven to be one of the best performing models for generating forecasts, especially when these forecasts entail asset allocation aspects (Degiannakis and Filis, 2022). The simple HAR for the

[17] Here equation is specified for VIX components. When modelling and forecasting the VIX as a unity, then specification becomes $\log(\mathrm{VIX_t})$. The same applies for the rest of the benchmark models, but for brevity we refrain from complete presentation in the relevant sections.

logarithm of VIX components, for each of the decomposition techniques, takes the form:

$$\log\left(\mathrm{VIX}_{i,j,t}\right) = \mathrm{w}_0^{(t)} + \mathrm{w}_1^{(t)}\log\left(\mathrm{VIX}_{i,j,t-1}\right) + \mathrm{w}_2^{(t)}\left(5^{-1}\sum_{\mathrm{k}=1}^{5}\log\left(\mathrm{VIX}_{i,j,t-k}\right)\right) + \mathrm{w}_3^{(t)}\left(22^{-1}\sum_{\mathrm{k}=1}^{22}\log\left(\mathrm{VIX}_{i,j,t-k}\right)\right) + \varepsilon_t, \tag{25}$$

where, $\mathrm{w}_0^{(t)}, \mathrm{w}_1^{(t)}, \mathrm{w}_2^{(t)}$ and $\mathrm{w}_3^{(t)}$ denote the rolling estimated coefficients and ε_t a normally distributed process, $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$. Since we have used the log form for model estimation, forecasts for the h-days-ahead horizon, are given by:

$$\begin{aligned}\mathrm{VIX}_{i,j,t+h|t} = \exp\Big(&\widehat{\mathrm{w}}_0 + \widehat{\mathrm{w}}_1\log\left(\mathrm{VIX}_{i,j,t+h-1|t}\right) \\ &+ \widehat{\mathrm{w}}_2\left(\mathrm{s}^{-1}\sum_{\mathrm{k}=1}^{\mathrm{s}-1}\log(\mathrm{VIX}_{i,j,t-k+h|t}) + (5-h)^{-1}\sum_{\mathrm{k}=\mathrm{s}}^{5}\log\left(\mathrm{VIX}_{i,j,t-k+h}\right)\right) \\ &+ \widehat{\mathrm{w}}_3\left(\mathrm{s}^{-1}\sum_{\mathrm{k}=1}^{\mathrm{s}-1}\log(\mathrm{VIX}_{i,j,t-k+h|t}) + (22-h)^{-1}\sum_{\mathrm{k}=\mathrm{s}}^{22}\log\left(\mathrm{VIX}_{i,j,t-k+h}\right)\right) + {}^{1}\!/_{2}\,\widehat{\sigma}_\varepsilon^2\Big).\end{aligned} \tag{26}$$

3.3. The Holt Winters model

Apart from HAR and AR models, we further employ the HW framework as some components present features that best apply to HW specification. HW is a simple univariate procedure for producing forecasts based on the past and current values of a time series that utilizes triple exponential smoothing and allows to deal with both seasonal variation and trend (Winters, 1960; Chatfield, 1978). HW unfolds in two versions[18], an additive and a multiplicative one where seasonal effects are thought to be of constant size or proportional to the local mean respectively (Chatfield, 1978).

[18] Additive and multiplicative versions apply depending on the data and their values. For running the HW model we made use of the statsmodels package of the python programming language.

There are three smoothing constants α, β, γ. The model specifications for the components of VIX, for the two distinctives versions have the following form[19]:

Additive

$$\begin{aligned} \widehat{\mathrm{M}}_t &= \hat{\mathrm{a}}_t\left(\log\left(\mathrm{VIX}_{i,j,t}\right) - \widehat{\mathrm{T}}_{t-c}\right) + (1-\hat{\mathrm{a}})\left(\widehat{\mathrm{M}}_{t-1} + \hat{\mathrm{F}}_{t-1}\right), \\ \hat{\mathrm{F}}_t &= \hat{\beta}\left(\widehat{\mathrm{M}}_t - \widehat{\mathrm{M}}_{t-1}\right) + (1-\hat{\beta})\hat{\mathrm{F}}_{t-1} \\ \widehat{\mathrm{T}}_t &= \hat{\gamma}_t\left(\log\left(\mathrm{VIX}_{i,j,t}\right) - \widehat{\mathrm{M}}_{t-1} - \hat{\mathrm{F}}_{t-1}\right) + (1+\hat{\gamma})\widehat{\mathrm{T}}_{t-c}. \end{aligned} \tag{27}$$

Multiplicative

$$\begin{aligned} \widehat{\mathrm{M}}_{\mathrm{t}} &= \hat{\mathrm{a}}\left(\frac{\log(\mathrm{VIX}_{i,j,t})}{\widehat{\mathrm{T}}_{t-c}}\right) + (1-\hat{\mathrm{a}})\left(\widehat{\mathrm{M}}_{t-1} + \hat{\mathrm{F}}_{t-1}\right), \\ \hat{\mathrm{F}}_t &= \hat{\beta}\left(\widehat{\mathrm{M}}_t - \widehat{\mathrm{M}}_{t-1}\right) + (1-\hat{\beta})\hat{\mathrm{F}}_{t-1} \\ \widehat{\mathrm{T}}_t &= \hat{\gamma}\left(\frac{\log\left(\mathrm{VIX}_{i,j,t}\right)}{\widehat{\mathrm{M}}_t}\right) + (1-\hat{\gamma})\widehat{\mathrm{T}}_{t-c}, \end{aligned} \tag{28}$$

where, $\widehat{\mathrm{M}}_t$, $\hat{\mathrm{F}}_t$, $\widehat{\mathrm{T}}_t$ stand for the estimate of the de-seasonalized mean for time *t*, the estimated seasonal factor at time *t* and the estimated trend term at time *t*, respectively. The c term that appears as subindex in above system of equations denotes the number of observations included in a seasonal cycle (Chatfield, 1978). Now, forecasts from the HW procedure can be generated for any horizon. Hence, forecasts of *h* days ahead of components in logarithmic form, for the additive and multiplicative versions respectively, are given by the formulas:

Additive

$$\mathrm{VIX}_{i,j,t+h|t} = \exp\left(\widehat{\mathrm{M}}_t + \mathrm{h}\hat{\mathrm{F}}_t + \widehat{\mathrm{T}}_{t+h-c}\right). \tag{29}$$

Multiplicative

$$\mathrm{VIX}_{i,j,t+h|t} = \exp\left(\widehat{\mathrm{M}}_t + \mathrm{h}\hat{\mathrm{F}}_t\right) * \widehat{\mathrm{T}}_{t+h-c}. \tag{30}$$

Someone will argue that a seasonal model is no proper for financial time series, where there is a total absence of seasonality. That depends, because financial time series exceed periodicity and volatility clustering. Volatility clustering involves fluctuations in asset prices that form clusters. Large (small) asset price changes tend to be followed by large (small) changes of either sign. Thus, clustering is the outcome of extreme volatility movements in response to crises of different sources recorded on the global landscape. The observed clusters seem to persist over specific periods and periods of high volatility are replaced by periods of low volatility. This pattern is closely studied

[19] Again log $(VIX_{i,j,t})$ can be rewritten as log (VIX_t) to account for VIX as a unity.

in volatility literature (Patton and Sheppard, 2011) and perhaps this is to blame why HW model seems to fit that well when modeling and forecasting volatility.

3.4. The Long Short-Term memory model

In the field of deep learning algorithms there stands out the LSTM (Hochreiter and Schmidhuber, 1997). Part of recursive neural networks (RNNs), LSTM came as a solution in a major drawback of RNNs, their inability to produce accurate predictions attributed to the explosive gradient descent when trying to learn the long-term dependencies of a series (Bengio *et al*., 1994). LSTM ever since has proven to be a powerful enough tool for time series forecast scenarios and many are the studies to include it (Chen *et al.*, 2021; Michańków *et al*., 2022; Shu and Gao, 2020; Liu *et al*., 2022). Its main advantage lies in the addition of extra memory cells, extra layers, consisting of three gates, the input gate, ig_t, the output gate, og_t and the forget get, fg_t. Memory cells with the help of activation functions[20], such as sigmoid and hyperbolic tangent function, act as filters concerning the information that will eventually reach the next cell state. The mathematical expression of how information flows inside each memory cell and the role each gates plays is presented in the following equations:

$$\begin{aligned} ig_t &= \text{sigm}(\text{weight}_{ig} \cdot [h_{t-1}, x_t] + \text{bias}_{ig}), \\ fg_t &= \text{sigm}(\text{weight}_{fg} \cdot [h_{t-1}, x_t] + \text{bias}_{fg}), \\ \tilde{\text{m}}_t &= \text{hyptan}(\text{weight}_m \cdot [h_{t-1}, x_t] + \text{bias}_m), \end{aligned} \tag{31}$$

where, $\tilde{\text{m}}$ is the value of the memory cell in order to update the cell, x_t is the input data at time *t*, h_{t-1} is the output of the previous layer that enters as input to present state, the weight with subindices of the different gates denote the weights set every time by forward and back propagation (Hecht-Nielsen, 1992), that try to minimize the error condition[21] through successive efforts and finally, the bias term stands for the error term. The sigm and hyptan symbols, appearing in Eq. 31 denote the sigmoid and the hyperbolic tangent functions respectively. The new value of the cell becomes:

$$\text{c}_\text{t} = \text{fg}_\text{t} * \text{c}_{\text{t}-1} + \text{ig}_\text{t} * \tilde{\text{m}}_\text{t}, \tag{32}$$

[20] For setting the model and its parameters (layers, optimizers, epochs, etc.), we use the keras deep learning API in python programming language (keras.io) that is built on top of TensorFlow, a machine learning platform.

[21] LSTM utilizes different types of loss functions. A frequently used loss function is the "mean squared error" and is the one we also incorporate for constructing our model.

where * denotes the convolution and c_{t-1} denotes the value of the previous cell state. Now, the value of the output gate, $\mathrm{og_t}$, is given by:

$$\mathrm{og}_t = \mathrm{sigm}(\mathrm{weight}_{og}[h_{t-1}, \mathrm{x}_t] + \mathrm{bias}_{og}, \tag{33}$$

Finally, the above steps result in the final layer's filtered output that is of the form:

$$\mathrm{h}_t = \mathrm{og}_t * \mathrm{hyptan}(\mathrm{c}_t). \tag{34}$$

The above procedure provides the steps occurring in each memory cell and the way they control information, progressively over and over again among layers until the final estimated value that minimizes the error term is eventually reached. Perhaps it sounds a little complicated, but that is how RNNs work and LSTM being the improved version of RNNs, holds their core architecture. Now, LSTM contrariwise to RNNs thanks to its lookback period window and its ability to learn both the short-term and the long-term data characteristics, can generate multiple days ahead forecasts. Having produced the 1 day ahead forecast, incorporates forecasted value, updates data set and moves on to the next and so on, until all values for the specified horizon are gathered:

$$\mathrm{h}_{t+h} = \mathrm{og}_{t+h} * \mathrm{hyptan}(\mathrm{c}_{t+h}). \tag{35}$$

In this study the inclusion of LSTM is used for modelling and forecasting components that their nature excludes more elementary frameworks.

4. Criteria for model selection and forecast evaluation

4.1. Forecast evaluation criteria

The aim of this study is twofold. The first is to preserve the robustness of decomposition process when employed in modeling financial time series, while the second is to figure out whether the proposed modelling architecture, can actually result in optimal forecast performance. We specify optimality through the production of meaningful and profitable forecasts for market participants, who engage dynamically in derivatives markets and seek to encompass futures contracts in their portfolios. Thus, we develop an objective-based forecast evaluation criterion, an economic criterion, incorporated along the classic statistical criterion of mean squared error, MSE, which is estimated by:

$$MSE_{i,h} = \widetilde{\mathrm{T}}^{-1} \sum_{t=1}^{\widetilde{\mathrm{T}}} (\mathrm{VIX}_{i,t+h|t} - \mathrm{VIX}_{t+h})^2 \tag{36}$$

where, $\mathrm{VIX}_{i,t+h|t}$ is the forecast from model i for day $t+h$ and VIX_{t+h} is the actual value of VIX volatility index at day $t+h$. $\widetilde{\mathrm{T}}$ is the out-of-sample forecast period and the subindex i denotes the i=1,2,…,32 different models. The reported MSE on the section of empirical findings, when for the decomposition-based models, denotes the aggregated forecast error, thus, tables report the total error. It may appear that we rest upon the empirical findings of a single study, to validate that total error from aggregating forecasts of components is less or equal to the ones generated without primary decomposing, but there is ground evidence that total error can be smaller by forecasting the decomposed variables. Degiannakis (2023) pointed out through a series of Monte Carlo simulations, that MSE of aggregated predictions of GDP[22] subcomponents was lower to the one of the predicted GDP in total.

Now, as for the economic criterion we follow a simple, yet powerful trading rule best expressed by two conditions:

1. If $\mathrm{VIX}_{i,t+h|t} > \mathrm{VIX}_t$ then we go long on VIX futures.
2. If $\mathrm{VIX}_{i,t+h|t} < \mathrm{VIX}_t$ then we go short on VIX futures.

Depending on the forecasted values, each day for the following day and up to 22 days ahead, for the entire out-of-sample period, we reconstruct daily our portfolio by holding long or short positions on VIX futures. Short or long positions are translated into selling or purchasing VIX futures, respectively. Therefore, at the end of the entire out-of-sample we calculate the cumulative returns (CR), provided through:

$$\mathrm{CR}_{i,h} = \sum_{t=1}^{\widetilde{\mathrm{T}}} \left(\mathrm{I}_{i,t} \times \frac{\left(\mathrm{VIX}_{i,t+h} - \mathrm{VIX}_t\right)}{\mathrm{VIX}_t} \right) \text{ and } \mathrm{I}_{i,t} = \begin{cases} 1 & \text{if} \quad \mathrm{VIX}_{i,t+h|t} > \mathrm{VIX}_t \\ -1 & \text{if} \quad \mathrm{VIX}_{i,t+h|t} \leq \mathrm{VIX}_t. \end{cases} \tag{37}$$

4.2. Model confidence set

A study that faces a multiple comparisons problem, deals with 32 models and diverse forecast horizons, would be diminished, if it did not employ a model selection procedure like the one proposed by Hansen *et al*. (2011), the model confidence set (MCS) test. MCS test, has turned out to be a valuable tool with several advances over

[22] GDP refers to the Gros Domestic Product.

other identical tests[23]. MCS test identifies the set of the best models, in terms of the evaluation criterion applied in a study and treats entire set of models equally without testing against the value of a benchmark model. Here we have the statistical loss function of MSE and the economic criterion. MCS test will provide the p-values that will tell of the set of the best performing models or the ones that inevitably will be excluded, depending on the prespecified significance level. Thus, in the tables of findings presented in section 6, we also report the p-values of the MCS test.

4.3. Direction-of-Change

In this study we employ two contradictive evaluation measures, a classic one and an economic criterion directly involving profits and losses. Therefore, the inclusion of an extra forecast evaluation technique, is essential. Direction-of-Change, (DoC), is a substantial feature, especially for trading exercises, in order to test the ability of a forecast to generate economic profit. Economic profits from long or short positions accelerate only when correctly predicting market's direction. Sometimes what is critical is the ability to predict directional accuracy and not the exact forecast accuracy (Degiannakis and Filis, 2018). DoC reports a proportion, PR_i, denoting the percentage of days a forecast correctly predicted the direction of the actual implied volatility index:

$$P_{i,t,h} = \begin{cases} 1 \text{ if } \mathrm{VIX}_{i,t+h|t} > \mathrm{VIX}_t \text{ and } \mathrm{VIX}_{t+h} > \mathrm{VIX}_t \\ 1 \text{ if } \mathrm{VIX}_{i,t+h|t} < \mathrm{VIX}_t \text{ and } \mathrm{VIX}_{t+h} < \mathrm{VIX}_t \\ 0 \text{ otherwise,} \end{cases} \tag{38}$$

and

$$\mathrm{PR}_{i,h} = \widetilde{\mathrm{T}}^{-1} \sum_{i=1}^{\widetilde{\mathrm{T}}} \mathrm{P}_{i,t,h}, \tag{39}$$

where $P_{i,t,h}$, is a dummy variable that reports whether the forecast of model i ($i=1, 2\ldots,32$) correctly predicted the volatility's upward or downward movement. Now, in order for the reported DoC rate to be significant, we need to be able to evaluate it. Thus, we also employ the non-parametric test of Pesaran and Timmerman (2009), which answers to the null hypothesis of no directional accuracy.

[23] Here we refer to the superior predictive ability test (SPA) of Hansen (2006) that again is used in identical studies, but it proceeds with its estimations for model selection, based on the values a benchmark model returned for a statistical or economic loss function.

4.4. Adjusting for the portfolio risk

Trading practices, which claim excess returns, should be able to measure the risk they undertake and account for whether they are indeed profitable or not. There are two widely applied metrics to measure the risk-adjusted performance of a trade, amongst others, the Sharpe ratio that was introduced by William Sharpe (Sharpe, 1963) and the Sortino ratio named after Frank Sortino. Sharpe Ratio measures excess return of a portfolio per unit of volatility and is denoted by:

$$ShR_{i,h} = \frac{R_{i,h} - r_f}{\sigma_{i,h}}, \tag{40}$$

where, $R_{i,h}$ are the annualized daily returns of model *i*, r_f is the risk-free rate that is conceived to be better represented by the 3-month T-bill rate and $\sigma_{i,h}$ is the standard deviation of the returns of model *i*. The $R_{i,h}$ can be computed as $\sqrt{252}\widetilde{\mathrm{T}}^{-1}\mathrm{CR}_{i,h}$ or even better based on the log-returns as $R_{i,h} = \sqrt{252}\widetilde{\mathrm{T}}^{-1}\sum_{t=1}^{\widetilde{\mathrm{T}}}\mathrm{I}_{i,t}\left(log\left(\mathrm{VIX}_{i,t+h}\right) - log(\mathrm{VIX}_t)\right)$, for $\mathrm{I}_{i,t} = 1$ if $\mathrm{VIX}_{i,t+h|t} > \mathrm{VIX}_t$ and $\mathrm{I}_{i,t} = -1$ if $\mathrm{VIX}_{i,t+h|t} \leq \mathrm{VIX}_t$. We select the second formulation although there are no qualitatively differences. Sortino and Price (1994) suggest instead of using the risk-free rate to use the average annual or monthly return of a market index (*MAR*).

The Sortino ratio, on the other hand, is an extension of Sharpe ratio without though being a complete measure of risk, rather it was proposed to face some recorded limitations of the standard deviation metric (Sortino and Forsey, 1996) Thus, it was created guided by the realization that large positive performance deviations, should not be penalized in the same manner to negative deviations. That is the reason why there are different versions for measuring the ratio, here we use the following form:

$$SorR_{i,h} = \frac{R_{i,h} - MAR}{\sigma_{i,h}^{(-)}}, \tag{41}$$

where MAR denotes the minimum accepted return; so, we stick to the risk-free rate and $\sigma_{i,h}^{(-)}$ is the standard deviation of portfolio's negative returns. A Sharpe or Sortino ratio over 2 are considered to be good for the trading strategy employed, but generally, common practise holds that the higher the rate the better, although it holds true that implied volatility indices and the linked assets come with rather low ratios, lower from unity.

5. Data description

Study is focused on implied volatility. We chose VIX index, the trademark of US stock market's implied volatility. The index itself is not tradable but is an index that forms an expectation[24] of markets' volatility for the following 30-days and is the key factor for pricing VIX derivatives. It constitutes a major informative tool for implementing trading strategies with options and futures, while having a distinctive structure differentiated by other volatility measures. As a result, is widely exploited by investors, policy makers, market makers etc. Here, we decompose, model and forecast VIX in order to invest in VIX futures, one of the major tradable instruments of VIX. No matter if index itself is more sensitive to market movements, by modelling VIX we avoid the frictions imposed by the constitution of VIX futures time series. Futures are characterized by jumps and discontinuities, since every month the tracked product alters, so it is not a proper instrument for modeling. Thus, VIX is more appropriate for the proposed framework and futures more fitted for volatility allocation practices.

But, we did not naively go for the VIX futures series, nor did calculations that would result in misleading profits or losses. VIX futures are characterized of having unique return drivers (Moran and Dash, 2007) and unique properties (Szado, 2018) attributed to the fact that their returns are highly correlated to VIX, but negatively correlated with equites not necessarily the SP500 equites, but other equites of major stock indices as well. This fact makes then extremely alluring for investors and risk managers, who incorporate them daily for hedging, capitalization, or portfolio diversification in order to get protection against extreme negative movements. Moreover, there also exists the expiration and the roll, that raises the challenging bar, The third Tuesday of each month, the expiration of the present month contract takes place. The Wednesday that follows, the front month's contract is set in action. There is when the roll cost or the roll yield occurs and can be a source of losses or gains, due to the difference in the price levels between the two contracts and the spot price of VIX. When the price of the contract that expires is lower than the next month's contract (contango), then rolling comes with the cost of paying more in order to swap in the next month's contract. When the price of the contract that expires is higher than the next

[24] VIX is a widely recognized index and the way it is calculated has been widely covered and analyzed in numerous studies and for practical reasons it is omitted from the present study. For the not so informed user, one can always resort to https://cdn.cboe.com/resources/vix/VIX_Methodology.pdf where there is extensive analysis on the methodology is being applied for index's construction.

month's contract (backwardation), then rolling incurs gains due to the roll yield. Since VIX futures most of the times are in contango and only few in backwardation, investors entail a significant roll cost[25].

Table 1 presents the price of the active contract and the price of next month's contract at the expiration date of the randomly chosen third week of the October of 2022. The price of October's contract is 30.86, while the one for November is 30.15. The following day the price of October's contract appears in the official records, while November is the active one. So, no matter what the final outcome shall be, for an investor that chooses as a settlement day the final trading day for the asset, she/he should close her/his position on October's contract and go for the November's on Tuesday.

[TABLE 1 HERE]

In this tailored trading exercise, we consider the roll for the entire out of sample period, report the cumulative returns, account for the total losses or profits and manage to exploit the exact conditions under which VIX futures market operates and exposure is conducted. Daily data for both VIX and VIX futures were retrieved from CBOE[26]. The sample of VIX, consists of 2600 trading days spanning from 21st of August 2012 up to 30th of November 2022, while for VIX futures, sample consists of 1600 trading days spanning from 12th of August 2016 up to 30th of November 2022, that is for the out-of-sample period.

5.1. Descriptive statistics

Table 2 reports the descriptive statistics of VIX, VIX futures, the logarithm of VIX and the correlation of VIX to VIX futures, which discloses the strong underlying relation (Daigler and Rossi, 2006; Szado 2009; Alexander *et al*., 2016), vital for the trade we engage in. Table 1 confirms the non-normal distribution of VIX, the positive skewness and the leptokurtic condition mainly attributed to extreme volatility movements (Degiannakis and Filis, 2022). Inspection also confirms another highlighted feature, the mean reverting property (Szado, 2020). But still there is another one recognizable feature this time concerning the logarithm of VIX. The logarithmic form

[25] An investor in order to get protection against extreme movements during the final settlement day and minimize the possible roll cost, has the ability to roll 5 or 6 days earlier.
[26] Data for VIX futures concerning the values of the first month contract at the day and the following day when settlement and roll takes place were retrieved from www.vixcentral.com.

comes with better statistical properties compared to VIX, more suitable for modeling; hence, in our study we incorporate the logarithmic form of VIX to construct the frameworks, an action dictated not only from the improved features, but also from the fact that some decomposition methods operate optimally under the transformed version.

[TABLE 2 HERE]

6. Empirical Findings

In a study, whose major aim and contribution lies in displaying the proper use of decomposition techniques, when forecasting and trading financial time series, analysis could not but start from the inspection of retrieved components. Figures 2 to 4, depict the components of two randomly chosen rolling samples of VIX for the methods of EMD, HVD and EWT, respectively. It is important to investigate, how the nature and trend of processes, completely alters through successive rolling samples[27], since they will be used as the daily input for the in-sample estimation and later on for forecast generation of a completely unknown future path. Although this aspect raises a little the complexity of entire process, it is also the one that validates the economic significance of our findings.

[FIGURE 2 HERE]

[FIGURE 3 HERE]

[FIGURE 4 HERE]

At this point it would be helpful to reiterate how individual components were modeled so as there will not be any inconsistencies. All components are modeled according to the stochastic process they undergone. Each component holds a distinctive nature, a part of the original input and its inner characteristics. These are the ones to be unmasked during the decomposition process. We distinguish their nature by applying, at least for the parametric models, all statistical tests that would specify whether a component was simply an AR (1), AR (3), AR (4) and so on. We test all conditions that apply, in order to hold and recure to all relevant information criteria, Akaike information criterion (AIC), Schwarz information criterion (SIC) etc. For components

[27] The number of components for techniques that are not user specified alters between successive samplings. For the EMD and EEMD techniques, there were 5 components, while in others were 6. This of course adds a bit in computational difficulty but in that way, we replicate the exact way information arrives.

that are not stationary and generally exceed a more distinctive pattern, the LSTM or HW are employed. There are also components that exceed a white noise process thus, no modeling takes place on them. We have 28 decomposition-based models. In order to refrain from an extensive representation that would make narration rather tiresome, we have gathered all 28 models in Table A1, in the Appendix section. The fact that we experimented especially with the HW and went on to also model all components from all techniques with HW as well, it was something dictated by modern literature who find it be really efficient in generating profitable forecasts (Degiannakis *et al.*, 2018). For a researcher that would desire to replicate proposed procedure a thing to bear in mind is that all depends firstly on the variable of interest, since decomposition unmasks all its inner characteristics and secondly on the appropriate tests that have to be conducted.

The secondary scope of this study is to show that objective-based evaluation criteria are more informative rich compared to statistical loss functions, when trading profits, need to be evaluated. Table 3 reports the values of the MSE along the p-values of its respective MCS, for all forecast horizons. The reported MSE is the aggregated one, as emerges from the summation of the individual forecasts of the individual components for each of the 6 disaggregation techniques. MCS informs that amongst the best performing models are included, apart from some of the proposed frameworks, 3 out of the 4 benchmark models. Outcome shows that especially for horizons of 5 to 22 trading days ahead, there are infinitesimally small differences between benchmark models and the more sophisticated ones. Thus, we cannot claim any significant gain in predictive accuracy. The fact that the higher p-values of MCS are reported for the more sophisticated frameworks, remains only an indication that decomposition-based models from a statistical perspective, are efficient enough. But the verdict remains that simple models can compete equally well with the more advanced ones.

[TABLE 3 HERE]

Now, when considering the economic criterion, the statistically imposed equality amongst forecast models, suddenly debases. Table 4 reports the cumulative returns of models[28] followed by the MCS test p-values at the right column of each forecast horizon. The measure clearly highlights how asymmetric outcome between the

[28] Cumulative returns reported in Table 4 are not expressed on % points.

statistical and the economic loss functions, can actually be. It also highlights how accuracy should be quantified while evaluating trading strategies. The reported cumulative returns at least for three of the proposed techniques, by far exceed those generated by the AR1, the HAR, the HW and the LSTM, something also confirmed by the reported MCS test p-values. Most of the EEMD-based and EMD-based models outperformed for horizons of 5 to 22 trading days ahead, followed by SSA-based models for the 1 and 5 trading days ahead horizon and HVD-HAR-AR, HVD-AR1-AR and HVD-LSTM-AR for the 5 days ahead. EEMD-HW-AR-HAR generated the highest cumulative returns that reach the value of 44.17 for the 5 days and the 41.53 for the 10 days ahead horizon, that is 44 and 41 times the initial invested capital, respectively. EMD-based models follow, with significant cumulative returns ranging between 12.44 to 39.93 for horizons of 5, 10 and 22 days ahead[29].

[TABLE 4 HERE]

Tables 3 and 4 for the same models, demonstrate two contradictive outcomes, an outcome of no predictive gain and an outcome of accelerated economic significance. Thus, we raise an important issue, that statistic loss functions, who measure the distance of forecasts from the actual price level can possibly direct towards incremental losses, if they cannot successfully point out towards the most profitable model that correctly predicts the direction of the underlying asset, especially in cases when an investor or risk manager follows a model-based trading strategy. Table 5 that follows, reports the rates of the DoC evaluation test and validates recordings of Table 4, with which it is aligned. The EEMD-HW-AR-HAR model that recorded the highest value in cumulative returns, according to DoC, correctly predicted the direction of VIX index for the 82% of times in the out-of-sample period of the 1600 trading days for the 5 days ahead horizon and about 78% of the times for the 10 days ahead horizon. Significant rates, also, are reported for the models that performed a little lower but correctly predicted the directional movement of the implied volatility index in a range spanning from 55% to 77%. We further validate the significance of all overperforming models, by applying the Pesaran and Timmermann (2009) test. We find that for all models that

[29] We do not place much attention on the 1 day ahead forecast horizon for the cases of the EEMD and EMD-based models due to the way the two techniques are implemented. At sections 3.1 and 3.2 we specified that we used the zero-end cubic spline condition, so literally the last observation in every rolling sample of the components, is excluded as no decomposition takes place.

record rates 55% and over, for all forecasting horizons, the null hypothesis is rejected at 1% level of significance.

[TABLE 5 HERE]

Furthermore, DoC rates that high are unique in financial literature. Relative studies as Degiannakis and Filis (2018) and Delis *et al*. (2022) have reached DoC rates up to 68%. Generally, rates over 55% denote gains, so no matter if it is 55% or 65%, the models are profitable because they generate extra returns. When models do not correctly predict the direction of the market index, we end up with losses (DoC<50%).

Moreover, there are also the two metrics the Sharpe and the Sortino ratios that we have also included in our study to account for the risk undertaken by reconstructing our position on a daily basis. Results for the monthly Sharpe and Sortino ratios are reported on Tables 6 and 7, respectively. We end with ratios way over 2 for all those models that outperformed, a number that signifies a balanced portfolio.

[TABLE 6 HERE]

[TABLE 7 HERE]

Empirical findings of this study, as they are reported in this final section, conclude by justifying the multiple steps of this modeling architecture. Decomposition techniques boost the performance of even classic models, when incorporated in the various modelling combinations of the separate modes and enhance the final outcome of implemented strategy, as the diverse tests implemented, verify. Perhaps not all decomposition-based models generated remarkable excess returns, but the ones that did, were high enough. Thus, there are three points to be highlighted. The one is that from all the proposed decomposition techniques, EMD framework is the only one to be characterized complete (Huang *et al*., 1998). EEMD follows in resolution as is the one with the lowest MSE[30] that of 0.02. For SSA this rate is 0.42, while for HVD it is 0.49. The MSE for the EWT is 1.2 and finally, for VMD, by using 5 components, it ends being 2.4. Thus, we comprehend especially for the user defined methods that the chosen number of components can seriously deteriorate process. A rate of 0.02 to a rate of 2.4 of how close, the addition of individual components, approximates original input, is

[30] Here MSE is the same statistical loss function used in our data only here it measures the divergence each decomposing method had from actual values of original input signal. That is how close the aggregated values of the components, approximate the values of the rolling samples of VIX index.

significant for the forecasting procedure and especially when forecasts have to be evaluated compared to actual future values. Perhaps that is the reason for the mediocre or bad performance of EWT-based and VMD-based models in the diverse forecast horizons. These declines between methods, probably shape optimality of final forecasts. Figures 5 to 8, visualize the cumulative returns of the MCS models, along the benchmark models for the 1, 5, 10 and 22 trading days ahead forecast horizon, while Figure A1 at the appendix section, depict a mixture of most sophisticated models for the same horizons, where also the performance of EWT-based and VMD-based, is illustrated. Another point that could be highlighted and possibly plays a critical role, is the modelling combinations of the forecasted components of each of the six decomposition techniques. Let us consider Figures 6 and 8 for the 5 and 22 days ahead horizon, respectively, and concentrate on EMD-HW, EMD-LSTM-AR or EMD-HW-AR models. For the case of EMD process, where all components are being modelled via HW framework, we notice cumulative returns to skyrocket at the rate of 38.78 for the 5 days ahead horizon. When the same components are modelled through a combination of the HW or LSTM and AR[31], cumulative returns decrease to 21.71. Furthermore, these model combinations apart from resulting in decreased cumulative returns, seem to be affected during the COVID-19 pandemic outburst and steadily recover through the remaining modeling period. This could be an indication that trading strategy followed could be altered during the turmoil, but our intension was to keep process as simple as possible and naively trade VIX futures purely guided by the forecasted value of VIX index out of the presented models. Additionally, it could be an indication that the way, components of this disaggregation method, were modeled, although statistically correct was not efficient for trade in period covered by extreme turbulence. It could also simply imply that there was a wrong combination of the modeled components.

[FIGURE 5 HERE]

[FIGURE 6 HERE]

[FIGURE 7 HERE]

[FIGURE 8 HERE]

[31] Not AR1 but for convenience we simply refer to as an AR process, since components whose process was a pure autoregressive one were found to exceed higher orders or combinations of it.

Finally, the 3rd point is the very own trading instrument, the VIX futures. VIX futures have some distinctive characteristics that allow for a close replication of a markets' operation but at the same time allow for a profitable trade (Moran and Dash, 2007; Szado, 2018). Despite the roll and despite the fact that for most of out-of-sample period, contract prices where in contango, indicating a loss by rolling to next month's contract, process was not negatively deteriorated. We ended recording increased cumulative returns.

7. Conclusion

The original scope of this study is to showcase how decomposition techniques should be properly incorporated and utilized when engaging in trading practices of financial time series. In such cases efficient replication of real-life operations is desired and thus, refraining from data leakage and data snooping is the most critical step in order to successfully validate robustness of empirical findings. In this way we manage to contribute to this literature strand and propose a process that can be utilized for other financial time series as well. Afterall, financial time series are not plain data, they are core instruments of financial markets and therefore of the global financial system. They hold incremental information for traders, risk and portfolio managers, policy agencies, so research must be targeted to their utility, especially when forecast is involved. Hence, study also focusses on a secondary aim, which is to show that when forecasts serve a realistic economic purpose, their accuracy and performance, should be evaluated based to this purpose.

Our study manages to serve this purpose and certify their profitability under economic standards. We produce useful for the market participant forecasts by pairing six decomposition techniques, the EMD, the EEMD, the SSA, the HVD, the EWT and the VMD, with parametric and non-parametric modelling frameworks, namely the AR, the HAR, the HW and the LSTM. On a daily basis in our rolling training sample, we decompose, model and forecast VIX index for 1, 5, 10 and 22 trading days ahead horizon and for an out of sample period of 1600 trading days. We record the profitability of forecasts by engaging in trading VIX futures, a dynamic tool in the hands of investors, used massively for hedging, diversification and even capitalization aims. There by reconstructing daily our position, solely depending on the forecasted value of

VIX we consider the cumulative returns recorded at the end of the out-of-sample period. Robustness of findings was verified by implementing a series of empirical tests.

The EMD-based and the EEMD-based followed by the SSA-based models were amongst the best performing models, with EEMD-HW-AR-HAR by far exceeding expectations, returning 44 times the invested capital for the 5 days ahead horizon and 41 times for the 10 days ahead. The fact that not all decomposition-based models performed equally well, only intensifies the different theoretical and empirical backgrounds that could take a toll when modelling financial data and reveal shortcomings of techniques that are disclosed only when properly incorporated in similar practices.

Concluding, we point out that ensemble methods can be effectively utilized in a process of forecasting and trading practices when objective-based evaluation criteria are applied. We also stress that the modelling combinations of the produced modes out of the decomposition techniques, does play a critical role in the generated cumulative returns, but we definitely highlight the significance of taking into consideration the actual way markets operate and exposure is conducted. Future research can shed new light on above findings by incorporating different financial time series, different decomposition techniques, different forecasting frameworks or different objective-based evaluation criteria.

Tables

Table 1 . VIX futures contracts prices

18/10/2022	**Tuesday:** Expiration date of October's contract	**October's contract:** 30.86 **November's contract:** 30.15
19/10/2022	**Wednesday:** First day of November's contract	**October's contract:** 31.77 (**expired**) **November's contract:** 30.40 (**active**)

Note: Table reports the prices of VIX futures contracts at the expiration and the following day of the expiration. In most studies October's price is reported for both days, but especially for Wednesday where October's have ceased to exist is more appropriate November's to be officially reported since is the one traded. The values reported here, were retrieved from vixcentral.com.

Table 2. Descriptive Statistics of VIX, the logarithm of VIX and VIX futures

	VIX	Log. VIX	VIX futures
Mean	19.27	2.88	19.91
Median	17.28	2.84	18.17
Max	82.69	4.41	72.63
Min	9.14	2.21	9.88
St. Dev	8.52	0.38	7.44
Coef. Of Variation	0.44	0.13	0.37
Skewness	2.21	0.52	1.71
Kurtosis	9.17	0.10	5.94
J-Bera	6889.52	73.29	3135.72
ADF	-4.438*	-4.769*	-3.943*
Corr. with VIX	-	-	0.95

Note: Descriptive statistics for both VIX and VIX futures agree with the stylized facts of volatility. The correlation between VIX and VIX futures for the period under investigation also confirms their strong connection, crucial for trading exercises. The * denotes significance at the 1% level for the ADF, so both series do not exceed unit root issues.

Table 3. MSE values of forecasting models

Model	FORECASTING HORIZON							
	1 day		5 days		10 days		22 days	
	MSE	MCS	MSE	MCS	MSE	MCS	MSE	MCS
AR1							53.12	**0.836***
HAR							51.28	**0.859***
HW							2553.5	0.000
LSTM							9	**0.715***
EMD-HW	4.342	0.000	16.960	**0.815***	30.91	**0.422***	59.150	0.000
EMD-HW-AR	4.317	0.000	16.775	**0.845***	30.37	**0.422***	297.84	0.000
EMD-LSTM-AR	4.257	0.000	111.609	0.000	1042.20	0.000	4	**0.461***
EMD-LSTM-HW-AR	14.039	0.000	15.978	**0.901***	29.523	**0.773***	119.97	0.004
EEMD-HW	3.144	**1.000***	32.803	0.000	185.492	0.000	3	0.000
EEMD-HW-AR	4.145	0.000	18.618	**0.607***	45.175	0.001	71.294	0.000
EEMD-HW-AR-HAR	6.703	0.000	16.197	**0.887***	33.393	**0.156***	96.308	0.000
EEMD-LSTM-AR	5.146	0.000	20.675	0.004	50.346	0.000	197.16	**0.617***
EEMD-LSTM-AR-HAR	3.145	**0.652***	33.892	0.000	94.862	0.000	5	**0.371***
SSA-LSTM-AR	3.269	0.000	16.791	0.845	47.860	0.000	139.84	**0.901***
SSA-LSTM-HAR	3.468	0.000	28.476	0.000	88.862	0.000	9	**0.836***
SSA-LSTM-HW	7.094	0.000	20.948	0.004	38.448	0.075	150.10	**0.598***
SSA-HW	8.145	0.000	27.667	0.000	81.486	0.000	4	**0.598***
SSA-HW-HAR	4.121	0.000	15.671	**0.901***	28.007	**0.773***	63.780	**0.859***
SSA-HW-AR	4.106	0.000	16.683	**0.843***	30.676	**0.422***	82.054	**1.000***
HVD-HW	4.139	0.000	16.045	**0.901***	29.098	**0.773***	49.665	0.000
HVD-HW-AR	4.047	0.000	16.413	**0.867***	30.547	**0.422***	53.296	0.000
HVD-HAR-AR	4.063	0.000	16.229	**0.891***	29.025	**0.773***	65.104	**0.371***
HVD-AR1-AR	3.971	0.000	15.455	**1.000***	27.722	**1.000***	66.653	**0.715***
HVD-LSTM-AR	6.781	0.000	83.256	0.000	779.544	0.000	52.227	0.000
EWT-HW	6.615	0.000	32.464	0.000	558.121	0.003	49.555	**0.461***
EWT-HW-AR	5.658	0.000	18.946	**0.580***	39.863	0.060	4818.4	0.063
EWT-LSTM-HW	4.676	0.000	16.187	**0.901***	35.712	**0.103***	52	**0.371***
EWT-LSTM-AR	13.826	0.000	26.867	0.000	244.755	0.000	2461.5	0.063
VMD-HW	6.802	0.000	20.778	0.004	37.577	0.089	66	**0.598***
VMD-HW-AR	6.445	0.000	25.071	0.000	44.505	0.003	82.434	**0.461***
VMD-LSTM-AR	6.583	0.000	20.636	0.004	37.778	0.089	99.441	0.063
VMD-LSTM-HW	4.728	0.000	25.813	0.000	46.461	0.001	272.67	**0.483***
	9.553	0.000	19.876	0.045	32.956	**0.231***	7	
	9.934	0.000	22.630	0.004	39.041	0.060	76.215	
	11.155	0.000	24.055	0.000	42.385	0.003	89.045	
	10.659	0.000	21.282	0.004	34.560	**0.156***	81.528	
							89.040	
							66.874	
							78.114	
							89.704	
							69.392	

Note: Table reports the MSE. Forecast horizon spans from 1, 5 10 to 22 days ahead, respectively. With * and bold writing we denote the models that belong to the MCS. The choice of the models for the MCS test, depends on the threshold significant level we choose, as well other parameters involving MCS test framework. In this study we use the 0.15 level. The MSEs reported are for the aggregated forecasts that were generated after the summation of the individual forecasts of the components of each decomposition technique. Each model in the table denotes a different aggregated forecast combination. For instance, EMD-HW model denotes an aggregated forecast composed of forecasts of components of EMD, where all were forecasted via HW framework.

We deduce, that based on MSE in the MCS of the best performing models do belong the AR1, the HAR and the LSTM for the 5, 10 and 22 days ahead horizons.

Table 4. Cumulative returns of forecasting models

Model	FORECAST HORIZON							
	1 day		5 days		10 days		22 days	
	Cum. Returns	MCS	Cum. Returns	MCS	Cum. Returns	MCS	Cum. Returns	MCS
AR1	2.691	0.005	3.915	0.090	3.969	0.000	4.319	0.006
HAR	3.249	0.005	4.692	0.099	5.446	0.000	5.558	0.021
HW	0.247	0.000	3.505	0.085	-2.462	0.000	3.511	0.000
LSTM	3.238	0.005	0.188	0.000	2.140	0.000	-2.372	0.000
EMD-HW	33.038	**0.973***	38.783	**0.615***	38.616	**0.145***	37.028	**1.000***
EMD-HW-AR	32.593	**0.953***	38.347	**0.615***	36.935	**0.145***	31.839	**0.427***
EMD-LSTM-AR	-10.921	0.000	12.442	**0.488***	13.342	0.013	18.634	**0.144***
EMD-LSTM-HW-AR	4.029	0.011	21.707	**0.507***	27.202	**0.145***	28.065	**0.217***
EEMD-HW	34.400	**1.000***	35.990	**0.615***	38.043	**0.145***	36.264	**0.463***
EEMD-HW-AR	27.528	**0.937***	24.686	**0.507***	32.704	**0.145***	31.281	**0.427***
EEMD-HW-AR-HAR	30.171	**0.950***	44.417	**1.000***	41.530	**1.000***	34.360	**0.463***
EEMD-LSTM-AR	1.461	0.001	9.252	**0.470***	11.527	0.013	19.608	**0.144***
EEMD-LSTM-AR-HAR	0.647	0.000	38.731	**0.615***	37.077	**0.145***	22.734	**0.186***
SSA-LSTM-AR	14.247	**0.223***	5.241	**0.162***	8.594	0.004	6.225	0.060
SSA-LSTM-HAR	12.409	**0.223***	3.295	0.087	6.725	0.004	5.076	0.021
SSA-LSTM-HW	10.486	**0.223***	1.097	0.036	6.265	0.004	5.787	0.021
SSA-HW	8.219	0.053	7.077	**0.162***	6.191	0.004	4.174	0.015
SSA-HW-HAR	9.149	0.053	6.940	**0.150***	9.488	0.004	9.105	0.060
SSA-HW-AR	10.143	**0.233***	9.660	**0.470***	8.830	0.004	12.233	0.098
HVD-HW	1.316	0.001	0.160	0.000	-0.591	0.000	0.602	0.000
HVD-HW-AR	4.099	0.011	1.856	0.036	0.104	0.000	0.579	0.000
HVD-HAR-AR	4.789	0.011	5.603	**0.162***	4.652	0.000	3.069	0.000
HVD-AR1-AR	3.221	0.005	9.881	**0.470***	3.854	0.000	1.103	0.000
HVD-LSTM-AR	7.603	0.053	22.198	**0.507***	10.623	0.004	6.876	0.021
EWT-HW	2.115	0.005	1.448	0.036	0.812	0.000	0.110	0.000
EWT-HW-AR	5.134	0.011	0.761	0.000	3.040	0.000	-0.217	0.000
EWT-LSTM-HW	3.034	0.005	1.247	0.036	2.081	0.000	-0.792	0.000
EWT-LSTM-AR	3.124	0.005	3.724	0.090	2.847	0.000	1.377	0.000
VMD-HW	1.432	0.001	0.544	0.000	-0.160	0.000	-0.115	0.000
VMD-HW-AR	2.349	0.005	0.296	0.000	1.396	0.000	1.473	0.000
VMD-LSTM-AR	2.641	0.005	2.025	0.085	2.636	0.000	2.415	0.000
VMD-LSTM-HW	3.437	0.005	1.099	0.036	0.487	0.000	-0.722	0.000

Note: Table reports the cumulative returns of trading VIX futures (are **not** expressed in % points) for 1, 5, 10 and 22 days ahead forecast horizon along with the p-values of the MCS test. With * and bold writing we denote the models that belong to the MCS. The choice of the models for the MCS test, depends on the threshold significance level we choose, as well other parameters involving MCS test framework. In this study we use the 0.15 level. According to the objective-based evaluation criterion, none of the AR1, HAR, HW and LSTM, considered to be the benchmark models, do belong to the set of the best-performing models.

Table 5. Direction of Change rates of forecasting models

	Direction of Change			
Model	1 day	5 days	10 days	22 days
AR1	0.474	0.491	0.488	0.509
HAR	0.486	0.502	0.510	0.512
HW	0.450	0.479	0.405	0.489
LSTM	0.485	0.446	0.476	0.387
EMD-HW	**0.743***	**0.768***	**0.755***	**0.766***
EMD-HW-AR	**0.616***	**0.703***	**0.670***	**0.680***
EMD-LSTM-AR	0.397	**0.553***	0.564	**0.602***
EMD-LSTM-HW-AR	0.493	**0.642***	**0.688***	**0.710***
EEMD-HW	**0.749***	**0.754***	**0.769***	**0.756***
EEMD-HW-AR	**0.704***	**0.670***	**0.712***	**0.704***
EEMD-HW-AR-HAR	**0.723***	**0.821***	**0.778***	**0.731***
EEMD-LSTM-AR	0.445	0.534	0.538	**0.608***
EEMD-LSTM-AR-	0.427	**0.766***	**0.722***	**0.667***
HAR	**0.608***	0.505	0.520	0.523
SSA-LSTM-AR	**0.581***	0.488	0.512	0.511
SSA-LSTM-HAR	**0.575***	0.469	0.508	0.517
SSA-LSTM-HW	0.544	0.536	0.501	0.508
SSA-HW	**0.551***	0.513	0.528	0.530
SSA-HW-HAR	**0.572***	0.541	0.520	0.541
SSA-HW-AR	0.441	0.455	0.411	0.456
HVD-HW	0.497	0.474	0.451	0.466
HVD-HW-AR	0.499	0.518	0.487	0.479
HVD-HAR-AR	0.485	0.542	0.479	0.461
HVD-AR1-AR	0.511	**0.651***	0.516	0.499
HVD-LSTM-AR	0.469	0.473	0.430	0.440
EWT-HW	0.489	0.450	0.465	0.411
EWT-HW-AR	0.482	0.469	0.453	0.402
EWT-LSTM-HW	0.482	0.479	0.459	0.453
EWT-LSTM-AR	0.459	0.444	0.423	0.404
VMD-HW	0.463	0.441	0.442	0.458
VMD-HW-AR	0.468	0.467	0.455	0.470
VMD-LSTM-AR	0.491	0.455	0.411	0.399
VMD-LSTM-HW				

Note: The DoC is the extra forecast evaluation test we adopted for our study, since the MCS test of cumulative returns, highlighted the outperformance of some ensemble frameworks. DoC rates do confirm the superiority and the ability of some techniques to effectively predict VIX's correct directional movement. With bold writing and * we specify the models with rates over 55%. Their significance is gauged by employing the Pesaran and Timmermann (2009) test, under the null hypothesis of no directional accuracy. We find that for all models of proportion of 55% and above and for the relevant forecast horizons, the null hypothesis is rejected at 1% level of significance.

Table 6. Trading profitability: The Sharpe Ratio

	Sharpe Ratios			
Model	1 day	5 days	10 days	22 days
AR1	0.833	0.659	0.643	0.721
HAR	0.653	0.835	0.796	0.703
HW	0.036	0.027	0.314	-0.376
LSTM	0.479	0.554	-0.390	0.634
EMD-HW	5.738	5.780	5.711	5.817
EMD-HW AR	4.344	4.714	4.132	3.585
EMD-LSTM AR	-1.600	1.816	1.981	4.287
EMD-LSTM HW AR	0.896	3.227	4.125	4.088
EEMD HW	5.306	7.135	1.702	5.681
EEMD HW AR HAR	4.595	5.590	5.793	4.822
EEMD HW AR	4.163	1.361	5.965	5.348
EEMD LSTM AR	0.186	6.069	5.036	2.940
HAR	0.378	3.655	6.602	3.430
EEMD LSTM AR	0.315	0.212	0.119	0.016
EWT HW	0.758	0.111	0.446	-0.032
EWT HW AR	0.463	0.546	0.418	0.203
EWT LSTM AR	0.438	0.183	0.306	-0.117
EWT LSTM HW	1.121	3.323	1.268	0.680
HVD LSTM AR	0.707	0.822	0.683	0.453
HVD HAR AR	0.477	1.454	0.566	0.149
HVD AR1 AR	0.280	0.023	-0.087	0.089
HVD HW	0.420	0.272	0.162	0.233
HVD HW AR	0.344	0.043	0.205	-0.017
VMD HW AR	0.210	0.080	-0.024	0.356
VMD HW	0.387	0.297	0.387	0.217
VMD LSTM AR	0.503	0.161	0.071	-0.107
VMD LSTM HW	2.103	0.776	1.267	0.918
SSA LSTM AR	1.344	1.022	1.299	1.348
SSA HW HAR	1.491	1.425	1.396	1.819
SSA HW AR	1.828	0.490	0.987	0.748
SSA LSTM HAR	1.542	0.164	0.921	0.803
SSA LSTM HW	1.206	1.043	0.908	0.615
SSA HW				

Note: Table reports the Sharpe ratios of each model. Generally, a ratio above 2 is considered to be good signal for the profitability of an investment. The Sharpe ratio actually adjusts portfolio's returns based on the risk they undertake and evaluates them, Here each model denotes a univariate portfolio constructed by naïvely following trading VIX futures. We calculate ratio by subtracting the risk-free rate from the annualized portfolio returns and then dividing by portfolio's standard deviation. Risk free rate is set to be the 3-month T-Bill. Here, we observe ratios to follow the performance of models as it was recorded at Table 4.

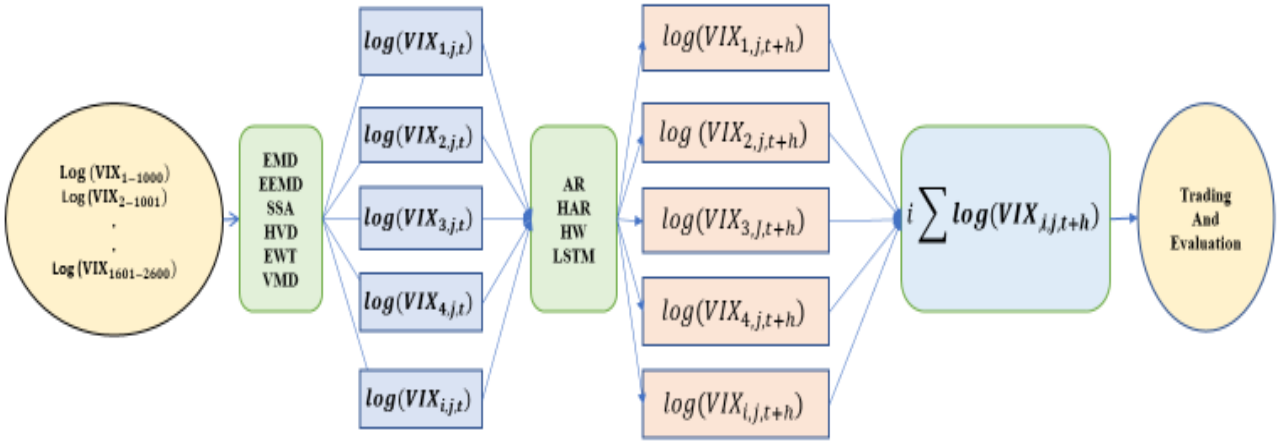
Table 7. Trading profitability: The Sortino Ratio

	Sortino Ratios			
Model	1 day	5 days	10 days	22 days
AR1	1.254	2.321	1.175	1.194
HAR	1.683	3.950	1.452	1.655
HW	-0.012	0.534	-0.476	-0.567
LSTM	0.787	1.618	1.006	1.361
EMD-HW	7.356	8.504	9.703	10.869
EMD-HW AR	6.418	6.347	6.032	6.875
EMD-LSTM AR	-2.324	3.323	4.243	6.212
EMD-LSTM HW AR	1.250	5.125	5.335	5.473
EEMD HW	8.135	14.419	8.697	8.240
EEMD HW AR HAR	7.000	16.719	9.181	7.221
EEMD HW AR	5.647	12.438	7.733	6.895
EEMD LSTM AR	-0.198	19.283	9.804	6.391
HAR	0.381	3.763	2.387	4.827
EEMD LSTM AR	0.794	2.076	0.443	0.443
EWT HW	2.188	0.935	0.513	0.513
EWT HW AR	0.923	5.230	0.831	0.831
EWT LSTM AR	1.031	2.001	0.749	0.749
EWT LSTM HW	2.015	6.030	1.999	0.543
HVD LSTM AR	1.858	1.422	1.011	0.963
HVD HAR AR	0.872	3.684	1.336	0.686
HVD AR1 AR	0.589	0.434	0.048	0.374
HVD HW	0.691	0.534	0.350	1.064
HVD HW AR	0.624	0.993	0.398	-0.231
VMD HW AR	0.506	0.671	0.007	0.726
VMD HW	0.701	3.633	1.041	0.044
VMD LSTM AR	1.029	1.530	0.119	-0.455
VMD LSTM HW	4.017	3.771	2.680	1.825
SSA LSTM AR	2.001	1.909	-1.408	3.846
SSA HW HAR	2.283	4.375	2.787	2.870
SSA HW AR	3.469	3.308	2.027	1.503
SSA LSTM HAR	2.884	1.690	1.468	1.231
SSA LSTM HW	1.934	3.042	1.368	2.616
SSA HW				

Note: The Sortino ratio although is not a complete measure of risk, it considers some aspects of behavioral finance, that argues that large negative shocks do not produce the same feeling as large positive shocks. Thus, ratio being an extension of Sharpe ratio, replaces the portfolio's standard deviation at the denominator of Sharpe ratio to the standard deviation of portfolio's negative returns, penalizing the negative returns that are the ones that seriously deteriorate a portfolio's performance. We notice that Sortino follows the Sharpe ratios recorded at Table 7. Ratios above 2 are a good indication for our portfolio profitability. That said, models that outperformed based on the recordings of cumulative returns of Table 4, were the ones to record higher Sortino ratios.

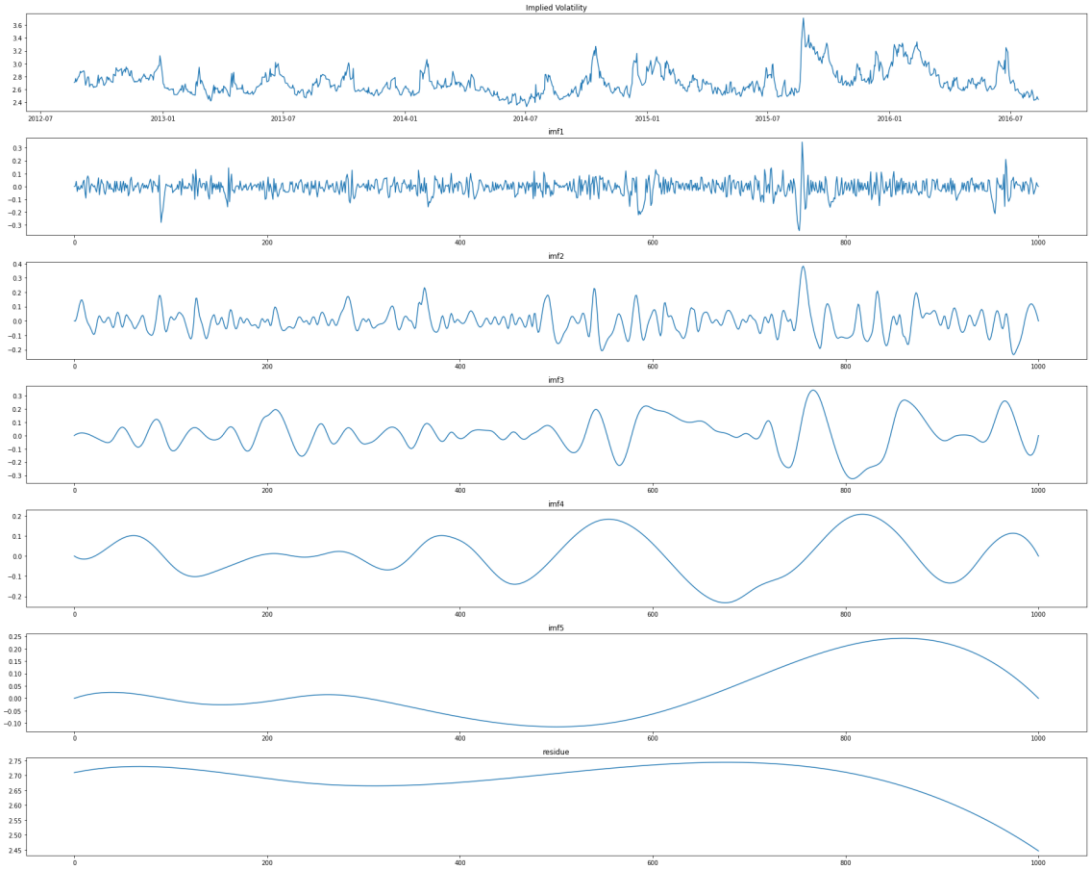
Figures

Figure 1. Decomposition, modelling and forecasting process



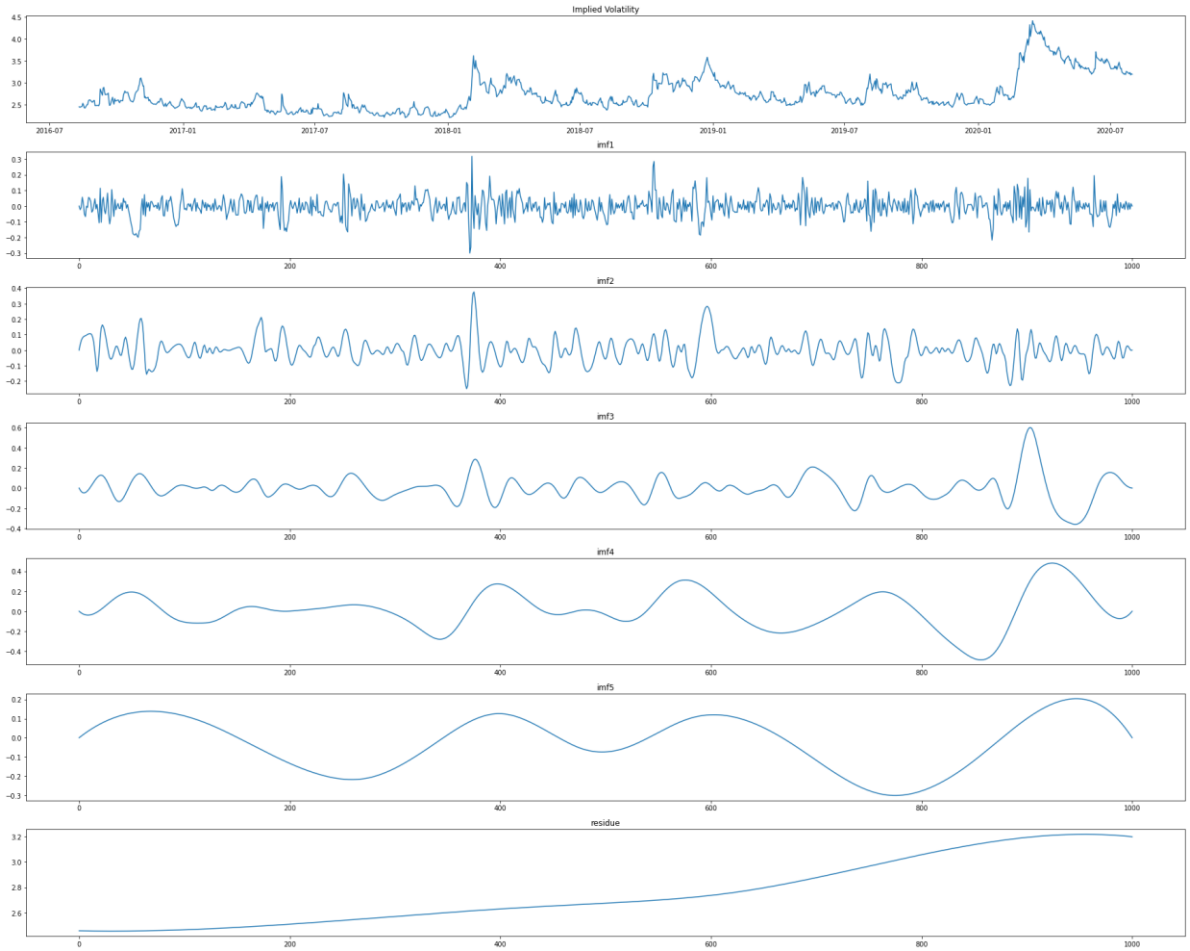
Note: Figure illustrates the rolling samples of the logarithm of VIX that consist of a fixed window of 1000 observations. On a daily basis the 1000 observations are used as the training data set that is decomposed, modelled and forecasted in order to get the aggregate forecasts of VIX for forecast horizons spanning from 1, 5, 10 and 22 days ahead for the out of sample period of the 1600 trading days, used as the test data set. Evaluation is then conducted based on statistical and economic criteria. From the above, i stands for the number of components, j = {EMD, EEMD, SSA, HVD, EWT, VMD} and h is the forecast horizon, with h = {1,5,10,22} days ahead.

Figure 2a. Components of EMD from two different samples (sample number 1)



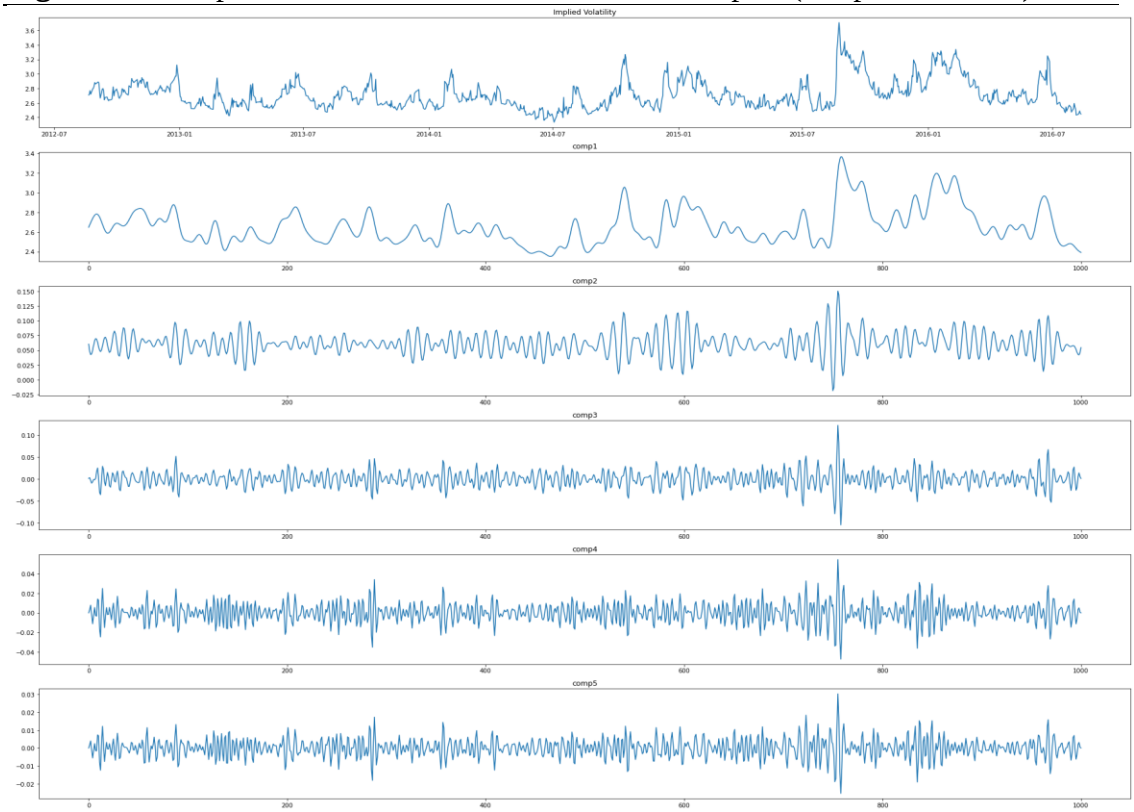
Note: At the top of the figure the VIX process of the chosen sample is depicted in order to present how radically components alter through successive samples and how unknown the future process becomes, something challenging for the forecasting frameworks that follow decomposition step. Here are the 6 components of the 1st rolling sample of the EMD process.

Figure 2b. Components of EMD from two different samples (sample number1000)



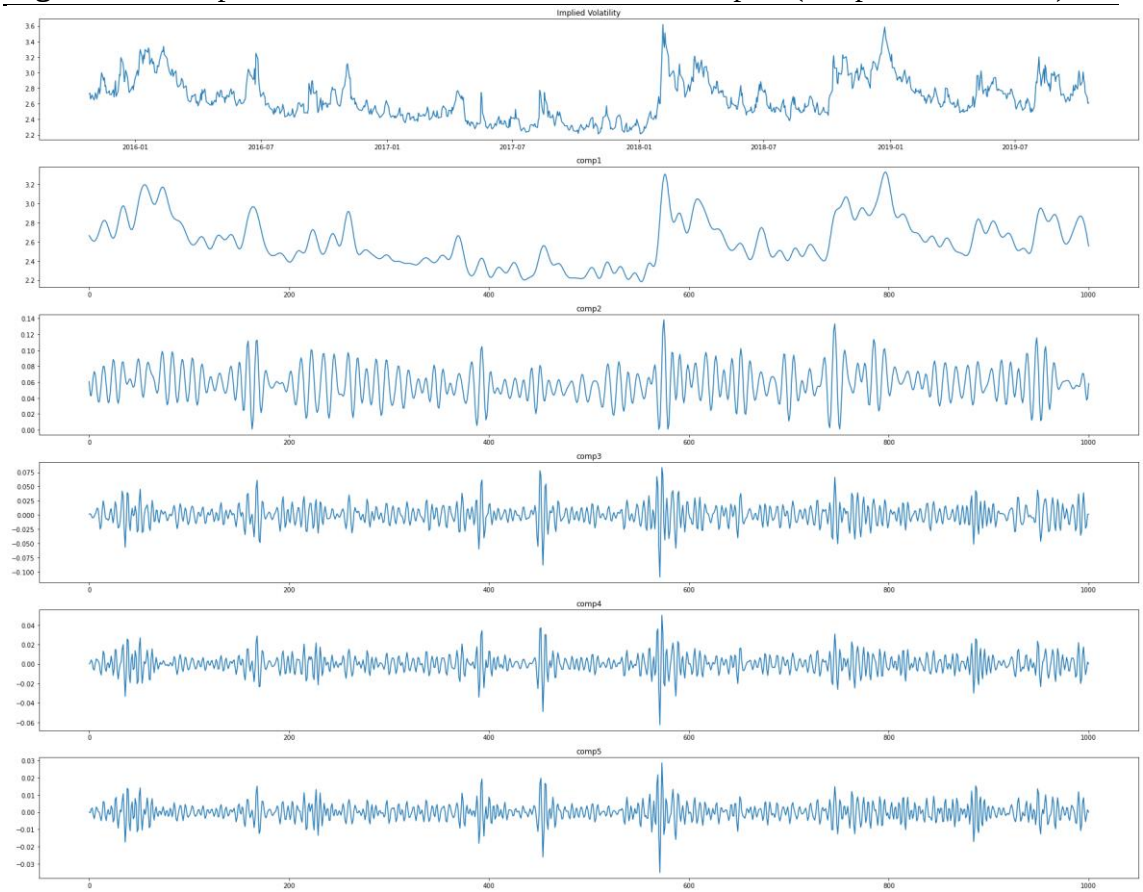
Note: At the top of each figure the VIX process of the chosen sample is depicted in order to present how radically components alter through successive samples and how unknown the future process becomes, something challenging for the forecasting frameworks that follow decomposition step. Above are presented the 6 components of the 1100th rolling sample of the EMD method.

Figure 3a. Components of HVD from two different samples (sample number 1)



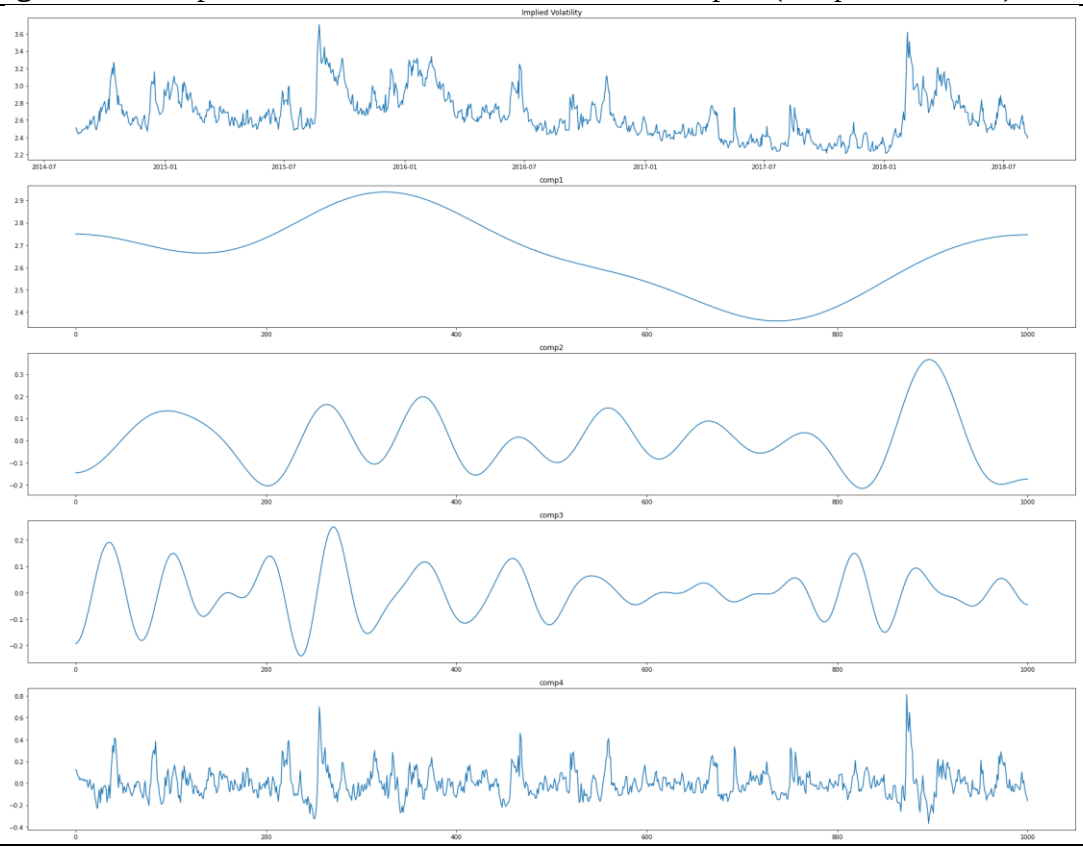
Note: HVD process terminates when the difference of the standard deviation between two iterations becomes minimum and that plays significant role during the approximate reconstruction of original time series. At the top is the VIX process. Above are the 5 components of the 1st rolling sample.

Figure 3b. Components of HVD from two different samples (sample number 800)



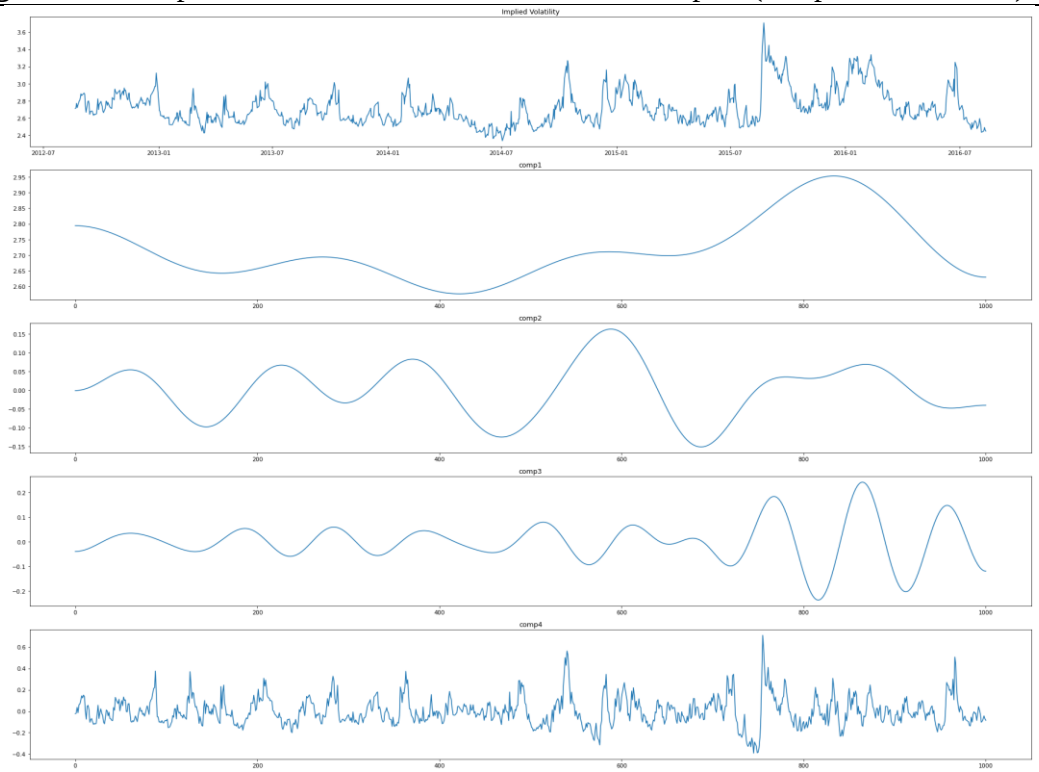
Note: HVD process terminates when the difference of the standard deviation between two iterations becomes limited and that plays significant role during the approximate reconstruction of original time series. At the top is the VIX process and below VIX are presented the 5 components of the HVD process of the 800th rolling sample.

Figure 4a. Components of EWT from two different samples (sample number 1)



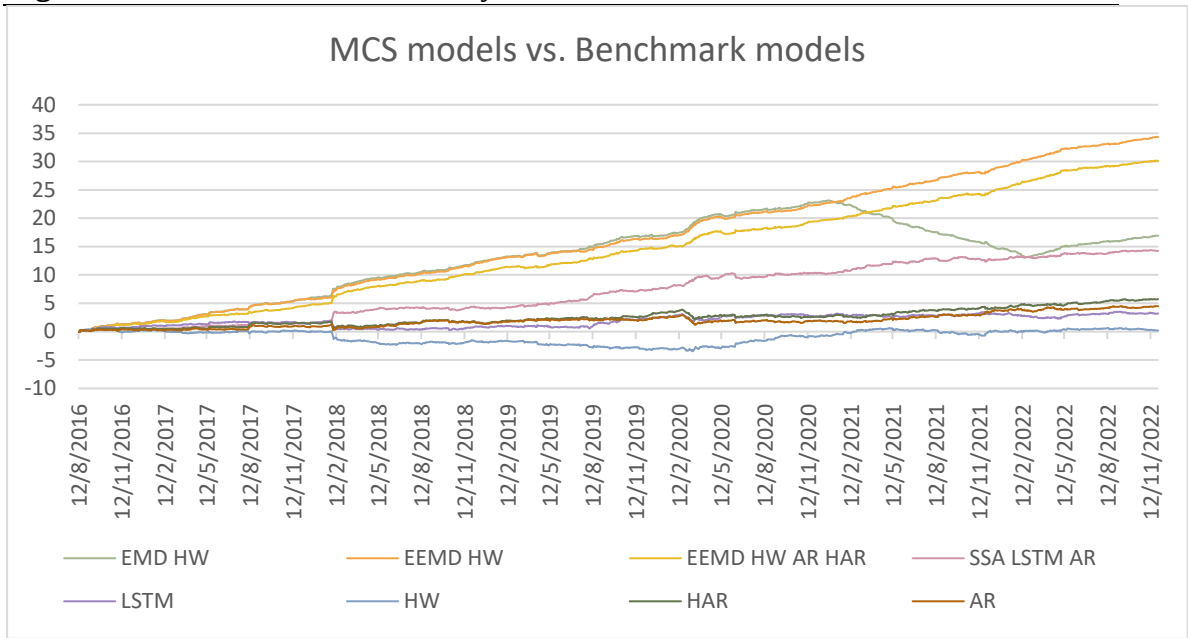
Note: EWT is the method that the user must specify the segments of the Fourier spectrum. The correct choice of boundaries will return components that, when reconstructed give an optimum approximation of initial data. Components follow the wavelet theory and decomposition begins from the lowest energy component towards the highest. The top graph is the VIX process and underneath VICX follow the 4 components of the EWT for the 1st rolling sample

Figure 4b. Components of EWT from two different samples (sample number 500)



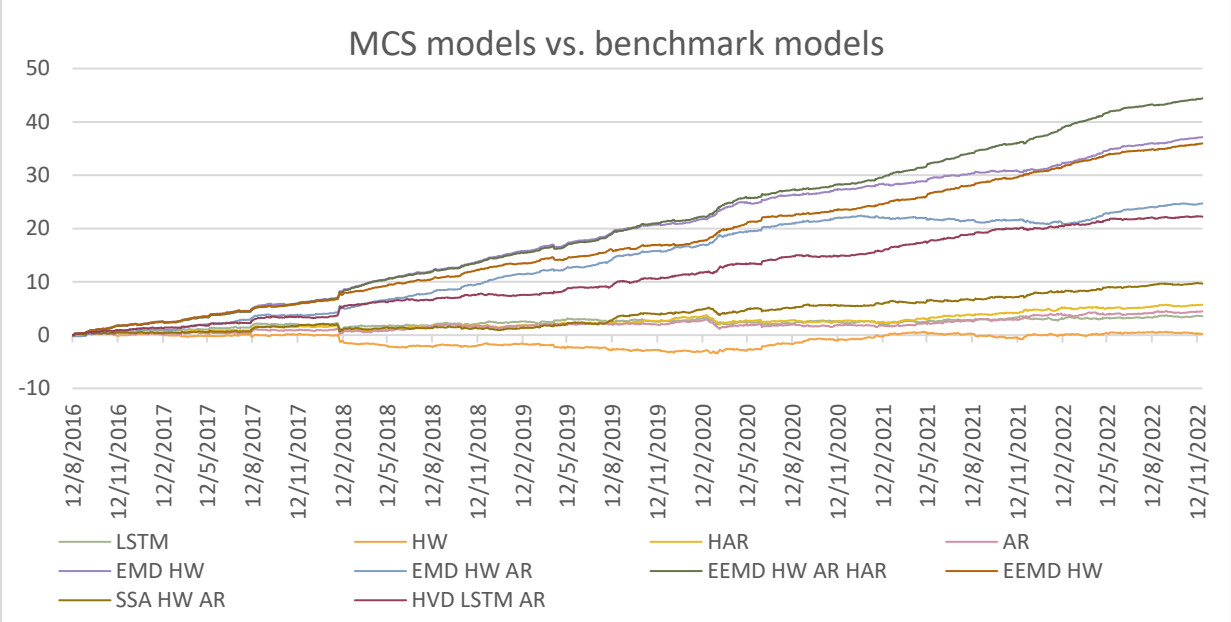
Note: EWT is the method that the user must specify the segments of the Fourier spectrum. The correct choice of boundaries will return components that, when reconstructed give an optimum approximation of initial data. Components follow the wavelet theory and decomposition begins from the lowest energy component towards the highest. At the top is the VIX process followed by the 4 components of the EWT method of the 500th rolling sample.

Figure 5. Cumulative returns of 1 day ahead horizon



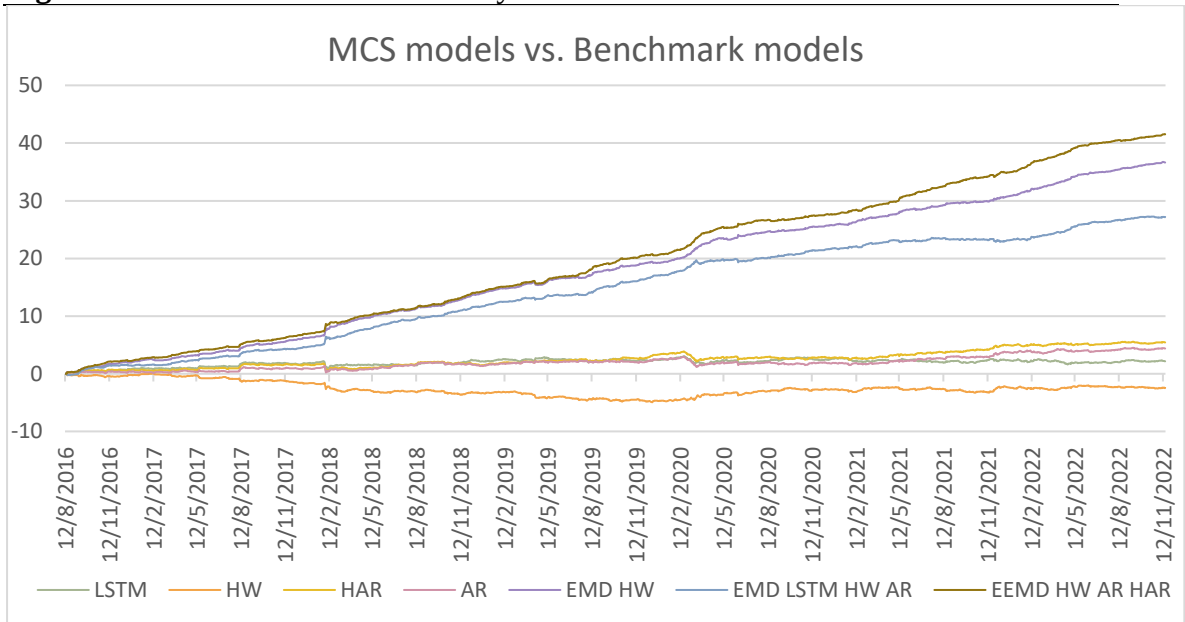
Note: Cumulative returns of the best performing models, according to MCS, along with the ones of the 4 benchmark models, for the 1 day ahead forecast horizon. All 4 benchmark models have mediocre or negative returns, compared to EMD-based, EEMD-based and SSA-based models, which moved with increasing rates during the period under investigation. The effect on cumulative returns based on the model combination is noticeable, but more noticeable is in the Figure 1a of Appendix section. The values reported in the y-axis are not in % points.

Figure 6. Cumulative returns of 5 days ahead horizon



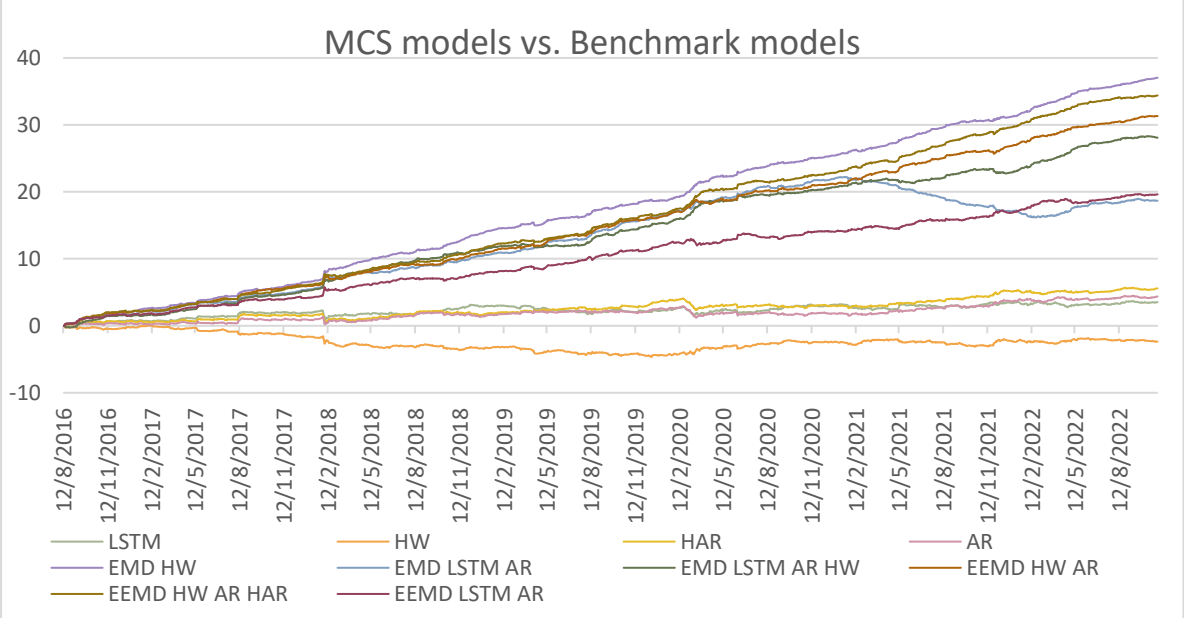
Note: For five days ahead forecast horizon, the cumulative returns earned by the trading strategy we implemented, do seem to diverge substantially between the ensemble models and the benchmark models. Of course, there were a few more models included in the MCS, but their cumulative returns were closely enough with others and their presentation would be rather messy, so we refrained from their inclusion. The values reported in the y-axis are not in % points.

Figure 7. Cumulative returns of 10 days ahead horizon



Note: For the forecast horizon of ten days ahead fewer models were included in the MCS, but the ones that did, come with significant gains compared to benchmark models. The values reported in the y-axis are not in % points.

Figure 8. Cumulative returns of 22 days ahead horizon



Note: All EMD-based and EEMD-based models were the only ones to be included in the MCS for the forecast horizon of the 22 days ahead. Generally, most of the EMD-based and EEMD-based models were amongst the best performing models for most horizons with significant increased cumulative returns, while benchmark models performed mediocre with HW being the one with systematically negative returns for entire period under investigation. On the other hand, EMD-HW and EEMD-HW, whose all components were modelled and forecasted via the HW framework, were the best performing. The values reported in the y-axis are not in % points.

Appendix

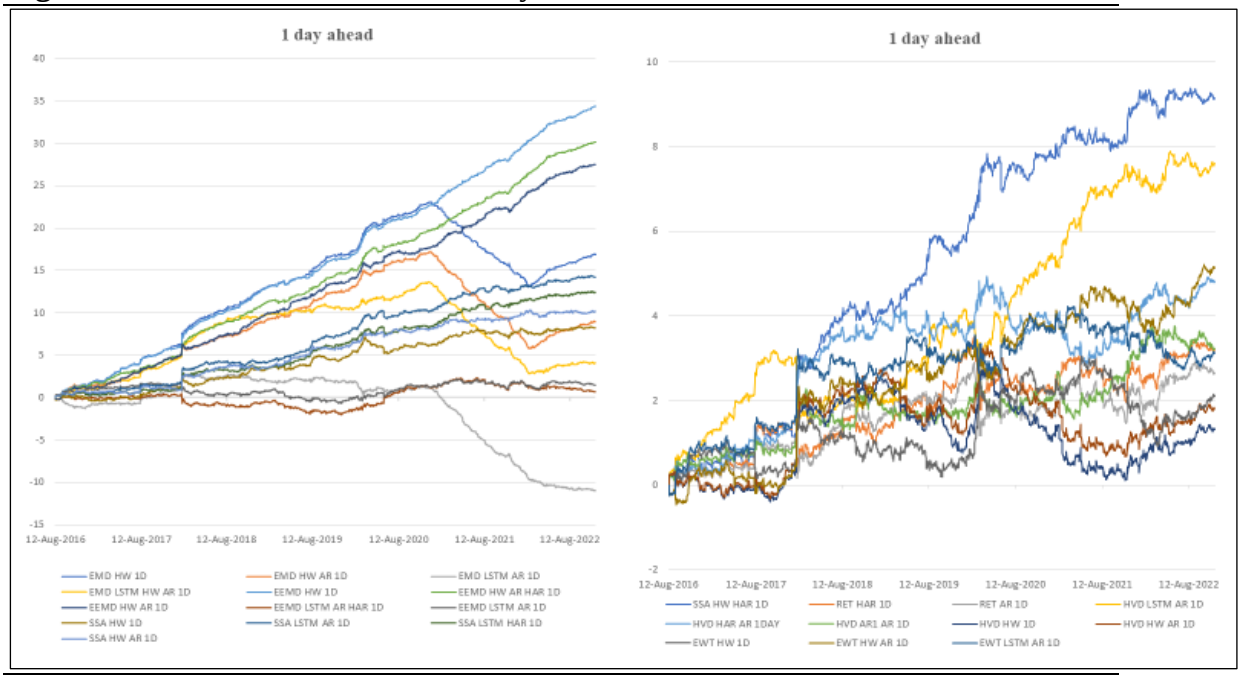
Table A1. Decomposition-based models, their components, and their forecasting techniques.

Model	# Components	Forecasting Technique
EMD HW	5-6 Components	Component 1: HW Component 2: HW Component 3: HW Component 4: HW Component 5: HW
EMD HW AR	5-6 Components	Component 1: HW Component 2: HW Component 3: AR (3) Component 4: AR (2) Component 5: AR (1)
EMD LSTM AR	5-6 Components	Component 1: LSTM Component 2: LSTM Component 3: AR (3) Component 4: AR (2) Component 5: AR (1)
EMD-LSTM HW AR	5-6 Components	Component 1: LSTM Component 2: HW Component 3: HW Component 4: AR (2) Component 5: AR (1)
EEMD HW	5-6 Components	Component 1: HW Component 2: HW Component 3: HW Component 4: HW Component 5: HW
EEMD HW AR HAR	5-6 Components	Component 1: HW Component 2: HW Component 3: HAR Component 4: HAR Component 5: AR (1)
EEMD HW AR	5-6 Components	Component 1: HW Component 2: HW Component 3: AR (3) Component 4: AR (3) Component 5: AR (1)
EEMD LSTM AR HAR	5-6 Components	Component 1: LSTM Component 2: LSTM Component 3: HAR Component 4: HAR Component 5: AR (1)

EEMD LSTM AR	5-6 Components	Component 1: LSTM Component 2: LSTM Component 3: AR (3) Component 4: AR (2) Component 5: AR (1)
SSA-LSTM-AR	4 Components	Component 1: LSTM Component 2: AR (2) Component 3: AR (1)
SSA-LSTM-HAR	4 Components	Component 1: LSTM Component 2: HAR Component 3: HAR
SSA-LSTM-HW	4 Components	Component 1: LSTM Component 2: HW Component 3: HW
SSA-HW	4 Components	Component 1: HW Component 2: HW Component 3: HW
SSA-HW-HAR	4 Components	Component 1: HW Component 2: HAR Component 3: HAR
SSA-HW-AR	4 Components	Component 1: HW Component 2: AR (2) Component 3: AR (1)
HVD LSTM AR	5 Components	Component 1: LSTM Component 2: AR (1) Component 3: AR (2) Component 4: AR (5) Component 5: AR (5)
HVD HAR AR	5 Components	Component 1: HW Component 2: AR (1) Component 3: AR (2) Component 4: AR (5) Component 5: AR (5)
HVD AR1 AR	5 Components	Component 1: AR (1) Component 2: AR (1) Component 3: AR (2) Component 4: AR (5) Component 5: AR (5)
HVD HW	5 Components	Component 1: HW Component 2: HW Component 3: HW Component 4: HW
HVD HW AR	5 Components	Component 1: HW Component 2: AR (1) Component 3: AR (2) Component 4: AR (5) Component 5: AR (5)

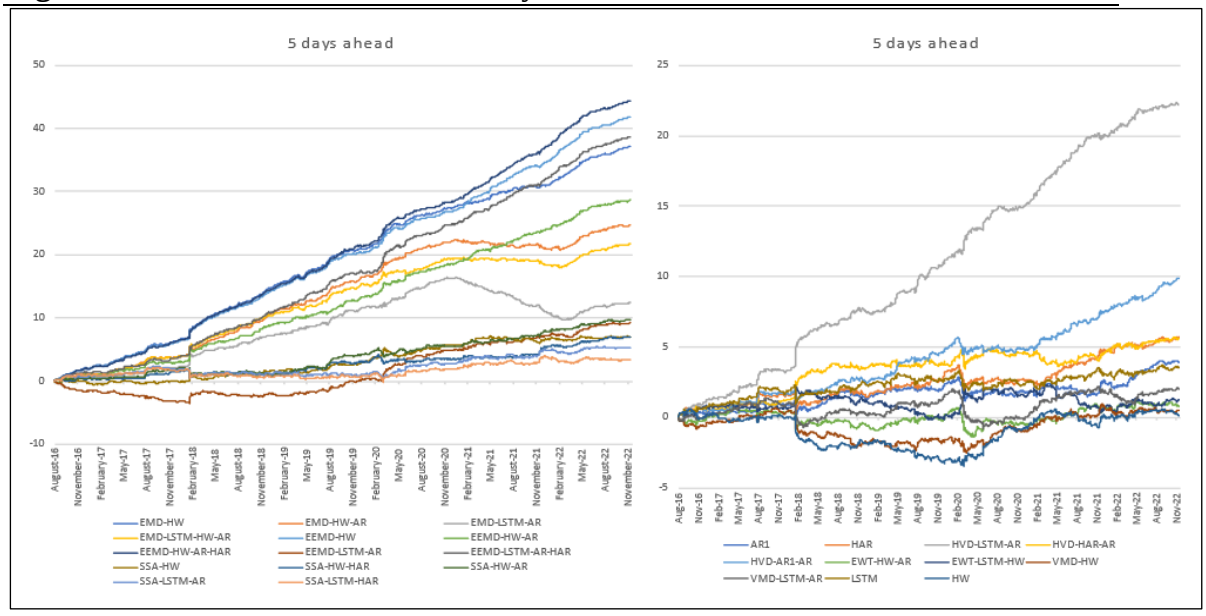
EWT HW	4 Components	Component 1: HW Component 2: HW Component 3: HW Component 4: HW
EWT HW AR	4 Components	Component 1: HW Component 2: AR (1) Component 3: AR (1) Component 4: HAR
EWT LSTM AR	4 Components	Component 1: LSTM Component 2: AR (1) Component 3: AR (1) Component 4: AR (1)
EWT LSTM HW	4 Components	Component 1: LSTM Component 2: HW Component 3: HW Component 4: HW
VMD-HW	5 Components	Component 1: HW Component 2: HW Component 3: HW Component 4: HW Component 5: HW
VMD-HW-AR	5 components	Component 1: HW Component 2: AR (1) Component 3: AR (1) Component 4: AR (2) Component 5: AR (2)
VMD-LSTM-AR	5 components	Component 1: LSTM Component 2: AR (1) Component 3: AR (1) Component 4: AR (2) Component 5: AR (2)
VMD-LSTM-HW	5 components	Component 1: LSTM Component 2: HW Component 3: HW Component 4: HW Component 5: HW

Figure A1a. Cumulative returns of 1 day ahead



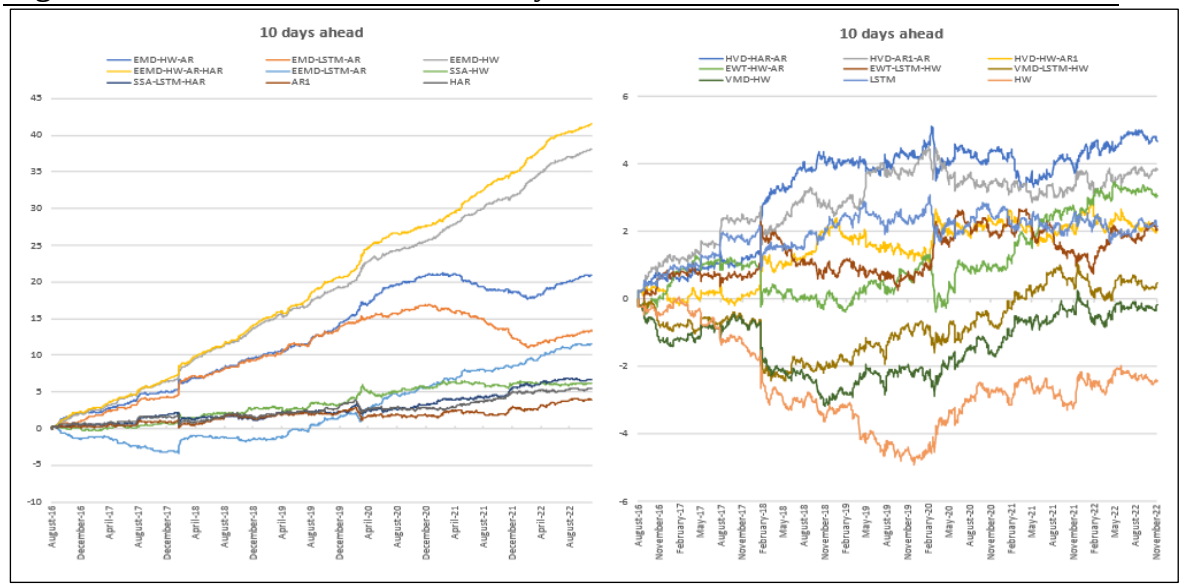
Note: Figure presents in two groups the cumulative returns of some of the various forecasting models for the 1 day ahead forecast horizon. Here we can see clearly how modelling the separate components of the six diverse techniques, with alternating combinations of AR, HAR, HW and LSTM frameworks, results in completely different outcomes, even for the same technique. The values reported in the y-axis are not in % points.

Figure A1b. Cumulative returns of 5 days ahead



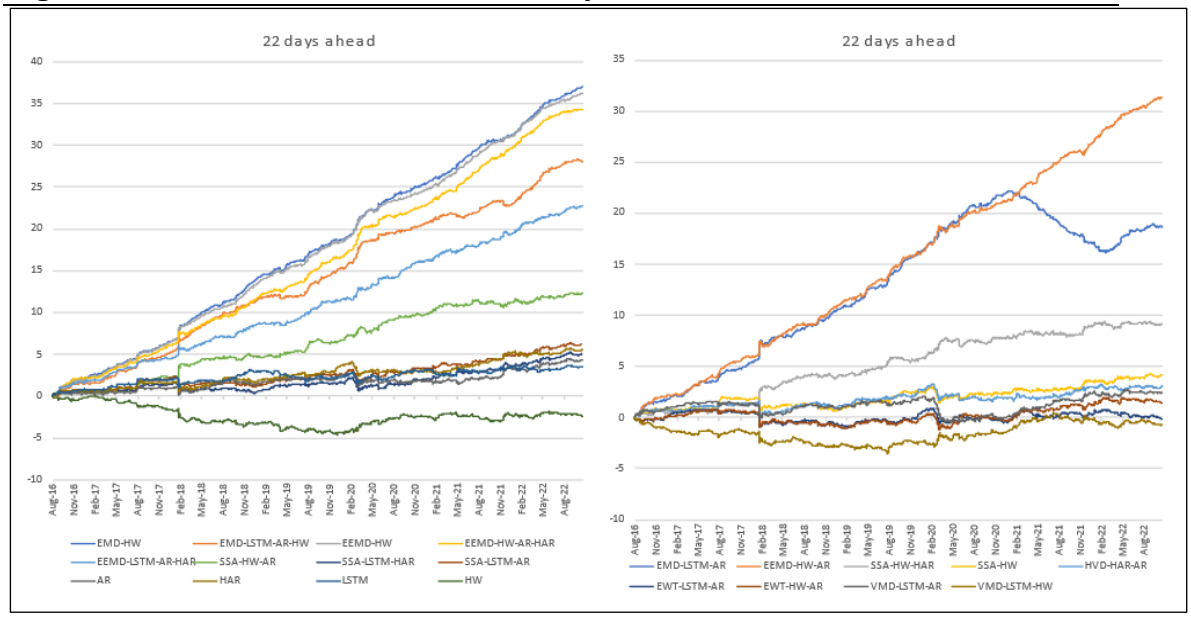
Note: The cumulative returns of some of the proposed models, for the 5 days ahead forecast horizon, appear in two groups. We definitely observe how outcome alters among different combinations and perhaps that is the purpose of experimentation, to find the optimal modelling combinations of components, belonging to the same decomposition technique. The optimal combination is the one that will boost returns. The values reported in the y-axis are not in % points.

Figure A1c. Cumulative returns of 10 days ahead



Note: Different model combinations, result in different cumulative returns. Here the model combinations with higher cumulative returns are more restricted. For the 10 days ahead horizon as for 5 days ahead, EEMD-HW-AR-HAR was the best performing model. Due to space and presentations limitations, we include only few of the 32 models in total. The values reported in the y-axis are not in % points.

Figure A1d. Cumulative returns of the 22 days ahead



Note: Even for the 22 days ahead horizon the group of the best performing models consists of the EEMD-based and EMD-based models. The models that belong to each of the two groups presented here, are randomly chosen between the 32 models. It was impossible to include all proposed models in above figures. The values reported in the y-axis are not in % points**.**

References

Alexander, C., Korovilas, D. and Kapraun, J. (2016). Diversification with volatility products, *Journal of International Money and Finance*, 65, 213-235.

Ali, M., Dost M. K., Alshanbari, H. M. and AAAH El-Bagoury. (2023). Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model. *Applied Sciences,* 13(3).

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review*, 39, 885–90.

Angelidis, T., and Degiannakis, S. (2008). Volatility forecasting: Intra-day versus inter-day models. *Journal of International Financial Markets, Institutions and Money*, 18(5), 449-465.

Bams, D., Blanchard G. and Lehnert, T. (2017). Volatility measures and Value-at-Risk. *International Journal of Forecasting,* 33, 848–863.

Barndorff-Nielsen, O.E. and Shephard, N. (2002). Estimating Quadratic Variation Using Realized Variance, *Journal of Applied Econometrics*, 17, 457-477.

Barndorff-Nielsen, O. E., Hansen, P.R., Lunde, A. and Shephard, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica,* 76 (6), 1481–1536.

Becker, R., and Clements A.E. (2008). Are combination forecasts of SandP 500 volatility statistically superior? *International Journal of Forecasting*, 24, 122–133.

Becker, R., Clements, A.E., Doolan, M.B. and Hurn, A.S. (2015). Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting*, 31 (3), 849-861.

Bengio, Y., Simard, P. and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult, *IEEE Transactions on Neural Networks*, 5(2), 157-166.

Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected stock returns and variance risk premia, *Review of Financial Studies,* 22(11*),* 4463–4492.

Chatfield, C. (1978). The Holt-Winters Forecasting Procedure. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 27(3), 264–279.

Chen, W., Xu, H., Jia, L. and Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting,* 37, 28–43.

Chen, Y., Yu, S., Islam, I., Peng Li, C. and Muyeen, S. M. (2022). Decomposition-based wind power forecasting models and their boundary issue: An in-depth review and comprehensive discussion on the potential solutions. *Energy Reports*, 8, 8805-8820.

Civera, M. and Surace, C. (2021). A Comparative Analysis of Signal Decomposition Techniques for Structural Health Monitoring on an Experimental Benchmark. *Sensors*, 21, 1825.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of. Financial Econometrics*, 7(2), 174–196.

Corsi, F. and Reno, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling, *Journal of Business and Economic Statistics*, 30(3), 368–380.

Daigler, R.T. and Rossi, L. (2006). A Portfolio of Stocks and Volatility. *Journal of Investing*, 99–106.

Daubechies, I. (1992). Ten Lectures on Wavelets, *Society for Industrial and Applied Mathematics*, Regional Conference Series in Applied Mathematics.

Degiannakis, S. (2023). The D-model for GDP nowcasting. *Swiss Journal of Economics and Statistics*, 159(1), 7.

Degiannakis, S. and Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, 76(C), 28–49.

Degiannakis, S. and Filis, G. (2018). Forecasting oil prices: High-frequency financial data are indeed useful. *Energy Economics*, 76, 388-402.

Degiannakis, S. and Filis, G. (2022). Oil price volatility forecasts: What do investors need to know? *Journal of International Money and Finance*, 18(5), 449–465.

Degiannakis, S., Filis, G. and Hassani, H. (2018). Forecasting global stock market implied volatility indices. *Journal of Empirical Finance*, 46, 111–129.

Delis, P., Degiannakis, S., and Filis, G. (2022). What matters when developing oil price volatility forecasting frameworks? *Journal of Forecasting*, 41(2), 361–382.

Delis, P., Degiannakis, S. and Giannopoulos, C. (2023). What should be taken into consideration when Forecasting Oil Implied Volatility Index? *Energy Journal*, 44(5).

Dong, J, Dai, W., Tang, L. and Yu, L. (2019). Why do EMD-based methods improve prediction? A multiscale complexity perspective. *Journal of Forecasting*, 38, 714– 731.

Dragomiretskiy, K. and Zosso, D. (2014). Variational Mode Decomposition. *IEEE Transitions on Signal Processing*, 62, 531–544.

Eckstein, J. and Bertsekas. P.D. (1992). On the Douglas Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3), 293–318.

Elliott, G. and Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*. 46. 3-56.

Engle, R.F., Hong, C.H., Kane, A., and Noh, J. (1993). Arbitrage valuation of variance forecasts with simulated options. *Advances in Futures and Options Research*, 6, 393–415.

Engle, R.F., Kane, A. and Noh, J. (1996). Index-Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts. Review of Derivatives Research. 1. 139-157.

Feldman, M. (2006). Time-varying vibration decomposition and analysis based on the Hilbert transform. *Journal of Sound Vibrations*, 295, 518– 530.

Feldman, M. (2011) Hilbert Transform Applications in Mechanical Vibration; John Wiley and Sons: Toronto, Canada.

Feldman, M. and Braun, S. (2017). Nonlinear vibrating system identification via Hilbert decomposition. *Mechanical Systems and Signal Processing*, 84(B), 65-96.

Fernandes, M., Medeiros, M.C. and Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking Finance* 40, 1–10.

Flandrin, P., Rilling, G. and Goncalves, P. (2004). Empirical Mode Decomposition as a Filter Bank. *IEEE Signal Processing Letters*, 11, 112– 114.

Gabay D. and Mercier, B. (1976). A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1), 17–40.

Gilles, J. (2013). Empirical Wavelet Transform. *IEEE Transactions on Signal Processing*, 61, 3999-4010.

Golyandina, N. and Zhigljavsky, A. (2013). Singular Spectrum Analysis for Time Series. *Springer Briefs in Statistics.*

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). Analysis of Time Series Structure: SSA and Related Techniques. *Monographs on Statistics and Applied Probability*, 90.

Guo, Z., Zhao, W., Lu, H. and Wang, J. (2012). Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renewable Energy*, 37 (1), 241-249.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence Set. *Econometrica*, 79(2), 453-497.

Hansen, P.R. (2005). A test for superior predictive ability. *Journal of Business Economic Statistics*, 23, 365–380.

Hassani, H., Kalantari, M. and Beneki, C. (2021). Comparative Assessment of Hierarchical Clustering Methods for Grouping in Singular Spectrum Analysis. *Applied Math*, 1, 18–36.

Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5,239–257.

Hassani, H. and Thomakos, D. (2010). A review on singular spectrum analysis for economic and financial time series. *Statistics and Its Interface*,3, 377–397.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural networks for perception*, 65-93, Academic Press.

Hewamalage, H., Bergmeir, C. and Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, 37, 388-427.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-80.

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

Huang, N.E., Shen, Z. and Long, S.R. (1999). A New View of Nonlinear Water Waves: The Hilbert Spectrum. *Annual Review of Fluid Mechanics*., 31, 417–457.

Huang, N.E., Wu, M.-L.C., Long, S.R., Shen, S.S.P., Qu, W., Gloersen, P. and Fan, K.L. (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society A, Mathematical, Physical and Engineering Sci*ences, 459, 2317–2345.

Huang, Y., Yan, C.J. and Xu, Q. (2012). On the difference between empirical mode decomposition and Hilbert vibration decomposition for earthquake motion records. *In Proceedings of the 15th World Conference on Earthquake Engineering*, Lisbon, Portugal.

Hu, J. and Wang, J. (2015). Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression, *Energy*, 93(2), 1456-1466.

Jin, Z., Jin Y. and Chen, Z. (2022). Empirical mode decomposition using deep learning model for financial market forecasting. *PeerJ Computer Science*, 8.

Kambouroudis, D. S., McMillan, D. G. and Tsakou, K. (2021). Forecasting realized volatility: The role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility. *Futures markets*, 41(10), 1618-1639.

Kambouroudis, D. S., McMillan D. G. and Tsakou, K. (2016). Forecasting stock return volatility: A comparison of GARCH implied volatility and realized volatility models. *Journal of Futures Markets*, 36(12), 1127-1163.

Kaufman, S., Rasset, S., Perlich, C. and Stitelman, O. (2012). Leakage in data mining: Formulation, detection and avoidance. *ACM Transactions on Knowledge Discovery from Data*, 6 (4), 1-21.

Konstantinidi, E., Skiadopoulos, G. and Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking and Finance,* 32(11), 2401-2411.

Lahmiri, S. (2016). A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Systems with Applications*, 55, 268-273.

Li, Z., Wang, M., Wang, X., Liu, Z. and Shi, A. (2020) Oil Price Forecasting Based on Variational Mode Decomposition, Relative Entropy and LSTM Neural Network. *IOP Conference Series: Materials Science and Engineering*, 750.

Liu, M., Sun, X., Wang, Q. and Deng, S. (2022). Short-Term Load Forecasting Using EMD with Feature Selection and TCN-Based Deep Learning Model. *Energies*, 15, 7170.

Liu, W., Cao, S. and Chen, Y. (2016) Applications of variational mode decomposition in seismic time-frequency analysis. *Geophysics*, 81, 365–378.

Meyer, Y. (1997). Wavelets, Vibrations and Scalings. *American Mathematical Society, Centre de Recherches Mathematiques*, 9, 133.

Michańków, J., Sakowski, P. and Ślepaczuk R. (2022). LSTM in Algorithmic Investment Strategies on BTC and SandP500 Index. *Sensors (Basel)*, 22(3), 917.

Moran, M. T. and Dash, S., (2007). VIX Futures and Options: Pricing and Using Volatility Products to Manage Downside Risk and Improve Efficiency in Equity Portfolios. *Journal of Trading*, 2 (3),96–105.

Mukherjee, A. and Swanson, N.R. (2021). Evidence on the Importance of Volatility Density Forecasting for Financial Risk Management. *Available at SSRN 3964200*.

Patton, A.J. and Sheppard, K. (2015). Good volatility, bad volatility: signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, 97(3): 683-697.

Peel, M. C., Pegram, G. S. and McMahon, T. A. (2007). Empirical mode decomposition: improvement and application. *In Proceedings of International Congress Modelling Simulation*, 1, 2996 – 3002. Modelling and Simulation Society of Australia, Canberra.

Pesaran, M. H., and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485), 325–337.

Ping Yu, Fang, J., Xu, Y. and Shi, Q. (2021). Application of Variational Mode Decomposition and Deep Learning in Short-Term Power Load Forecasting. *Journal of Physics: Conference Series, 1883.*

Prasad, A. and Bakhshi, P. (2022). Forecasting the Direction of Daily Changes in the India VIX Index Using Machine Learning. *Journal of Risk and Financial Management,* 15.

Rilling, G. Flandrin, P. and Goncalves P. (2003). On empirical mode decomposition and its algorithms. *In Proceedings 6th IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing.*

Risse, M. (2019). Combining wavelet decomposition with machine learning to forecast gold returns. *International Journal of Forecasting*, 35, 601-615.

Rua, A. (2014). A wavelet-based multivariate multiscale approach for forecasting. *International Journal of Forecasting,* 33, 581–590.

Shankar, S., Dandapat S. and Barma, S. (2021). Classification of Epileptic Seizure from EEG Signal Based on Hilbert Vibration Decomposition and Deep Learning, *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Mexico, 2802-2805.

Shaari, M., Samsudin, R. and Shabri, A. (2018). Forecasting Drought Using Modified Empirical Wavelet Transform-ARIMA with Fuzzy C-Means Clustering. *Indonesian Journal of Electrical Engineering and Computer Science*, 11, 1152-1161.

Sharpe, F.W. (1963). A Simplified Model for Portfolio Analysis. *Management Science*. 9 (2): 277–293.

Shi, P., Yang, W., Sheng, M. and Wang, M. (2017). An Enhanced Empirical Wavelet Transform for Features Extraction from Wind Turbine Condition Monitoring Signals. *Energies,* 10(7), 972.

Shu, W. and Gao, Q. (2020). Forecasting Stock Price Based on Frequency Components by EMD and Neural Networks. *In IEEE Access*, 8, 206388-206395.

Singh, O. and Sunkaria, R. (2016). ECG signal denoising via empirical wavelet transform. *Australasian Physical and Engineering Sciences in Medicine*, 40, 1-11.

Singh, M. J., Sharma L. N. and Dandapat, S. (2022). Hilbert Vibration Decomposition of Seismocardiogram for HR and HRV Estimation. *2022 IEEE International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, India,1-5.

Sortino, F.A. and Forsey, H.J. (1996). On the Use and Misuse of Downside Risk. *Journal of Portfolio Management*, 22, 35-42

Sortino, F.A. and Price, L.N. (1994). Performance Measurement in a Downside Risk Framework. *Journal of Investing*, 3, 59-64.

Sulandari, W., Subanar, M., Lee, H. and Rodrigues, P.C. (2020). Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks, *Energy*, 190.

Sun, S., Wang, S., Zhang, G., and Zheng, J. (2018). A decomposition–clustering–ensemble learning approach for solar radiation forecasting. *Solar Energy*, 163, 189–199.

Szado, E. (2009). VIX Futures and Options: A Case Study of Portfolio Diversification during the 2008 Financial Crisis. *Journal of Alternative Investments,* 12 (2), 68–85.

Szado, E. (2018). The Distinctive Characteristics of VIX Futures and Options. *Working Paper*.

Szado, E. (2020). Selling VIX® Futures and Options for Portfolio Return Enhancement. *Working paper.*

Tang, L., Wu, Y. and Yu, L. (2018). A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting. *Energy*, 157, 526–538.

Taylor, J. M. (2004). Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting*, 20, 273-286.

Taylor, N. (2014). The Economic Value of Volatility Forecasts: A Conditional Approach. *Journal of Financial Econometrics*. 12.

Thomakos, D., Wang, T. and Wille, L. (2002). Modeling daily realized futures volatility using singular spectrum analysis. *Physica A: Statistical Mechanics and its Applications*, 312(3), 505–519.

Theodoridis, S. (2020). Machine Learning: A Bayesian and Optimization Perspective (Second Edition). Academic Press.

Vrontos, S. D., Galakis J. and Vrontos, I. D. (2021). Implied volatility directional forecasting: a machine learning approach. *Quantitative Finance*, 21(10), 1687-1706.

Wang, C., Zhang, H. and Ma, P. (2016). Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renewable Energy, 94, 629-636.

West, K. D., Edison, H. J. and Cho, D. (1993). A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics*, 35 (1–2), 23-45.

Wu, Z., and Huang, N.E. (2009). Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 1, 1–41.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342.

Yeh, J. R., Shieh, J. S. and 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.

Yu, L., Wang, S. and Lai, K. (2008). Forecasting crude oil price with an emd-based neural network ensemble learning paradigm. *Energy Economics*, 30(5), 2623–2635.

BANK OF GREECE WORKING PAPERS

316. Kotidis, A. M. MacDonald, D. Malliaropulos, "Guaranteeing trade in a severe crisis: cash collateral over bank guarantees", March 2023.

317. Degiannakis, S. "The D-model for GDP nowcasting", April 2023.

318. Degiannakis, S., G. Filis, G. Siourounis, L. Trapani, "Superkurtosis", April 2023.

319. Dixon, H. T. Kosma, and P. Petroulas, "Explaining the endurance of price level differences in the euro area", May 2023.

320. Kollintzas, T. and V. Vassilatos, "Implications of market and political power interactions for growth and the business cycle II: politico-economic equilibrium", May 2023.

321. Bragoudakis, Z. and I. Krompas "Greek GDP forecasting using Bayesian multivariate models", June 2023.

322. Degiannakis, S. and E. Kafousaki "Forecasting VIX: The illusion of forecast evaluation criteria", June 2023.

323. Andreou C. P., S. Anyfantaki, C. Cabolis and K. Dellis, "Exploring country characteristics that encourage emissions reduction", July 2023.

324. Dimakopoulou, V., Economides, G., Philippopoulos, A., and V. Vassilatos, "Can central banks do the unpleasant job that governments should do?", December 2023.

325. Chrysanthakopoulos, C. and A. Tagkalakis, "The medium-term effects of fiscal policy rules", January 2024.

326. Manou, K. and E. Papapetrou, "Does uncertainty matter for household consumption? A mean and a two tails approach", February 2024.

327. Kakridis, A., "War, mobilization, and fiscal capacity: testing the bellicist theory in Greece, 1833-1939", March 2024.

328. Mavrogiannis, C. and A. Tagkalakis, "From policy to capital: assessing the impact of structural reforms on gross capital inflows", April 2024

329. Delis, P., S. Degiannakis, G. Filis, T. Palaskas and C. Stoforos, "Determinants of regional business cycle synchronization in Greece", May 2024.

330. Sideris, D. and G. Pavlou, "Market power and profit margins in the Euro area countries in the post-pandemic period", June 2024.

331. Kasimati, E. and N. Veraros, "The dry-bulk shipping market: a small econometric model", September 2024.

332. Mermelas, G. and A. Tagkalakis, "Monetary policy transmission: the role of banking sector characteristics in the euro area", November 2024.

333. Anastasiou, D., Pasiouras, F., Rizos, A., and A. Stratopoulou, "Do macroprudential policies make SMEs more-or-less discouraged to apply for a bank loan?", December 2024.

334. Malliaropulos, D., Passari, E., and F. Petroulakis, "Unpacking commodity price fluctuations: reading the news to understand inflation", December 2024