IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 July 2024, accepted 8 August 2024, date of publication 15 August 2024, date of current version 4 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444479

RESEARCH ARTICLE

Long/Short Equity Risk Premia Parity Portfolios via Implicit Factors in Regularized Covariance Regression

COLE VAN JAARSVELDT[1], GARETH W. PETERS[2], (Senior Member, IEEE), MATTHEW AMES[3], AND MIKE CHANTLER[4]

[1]Department of Actuarial Mathematics and Statistics, Heriot-Watt University, EH14 4AS Edinburgh, U.K.
[2]Department of Statistics and Applied Probability, University of California at Santa Barbara, Santa Barbara, CA 93106, USA
[3]ResilientML, Melbourne, VIC 3000, Australia
[4]Department of Computer Science, Heriot-Watt University, EH14 4AS Edinburgh, U.K.

Corresponding author: Cole van Jaarsveldt (colevj0303@gmail.com)

ABSTRACT A robust time series basis decomposition and non-stationary trend extraction technique, known as Empirical Mode Decomposition (EMD), will be combined with Regularised Covariance Regression (RCR) to produce a novel covariance forecasting technique. EMD is designed for multiscale and adaptive time-frequency decomposition in nonstationary time series. EMD-RCR generates multi-time-frequency resolution adaptive forecasting models of predictive covariance forecasts for a universe of selected asset returns. This provides a unique method to obtain predictive covariance regression structures for the short-, mid-, and long-time-scale portfolio dynamics. EMD isolates structures in a frequency-hierarchical fashion (with automated sorting of structures through EMD-MDLP available) which allows this multi-time-frequency covariance forecasting framework that uses the structures isolated using EMD (referred to as IMFs: Intrinsic Mode Functions) as the explanatory variables in the RCR framework. Having developed these techniques, a case study is used for exposition for active portfolio asset management. The case study is based on a dynamic long/short equity and risk premia parity (or risk parity) portfolio-of-portfolios investment strategy using the 11 sectors dividing the 505 stocks of the S&P 500. Each of the 11 sector indices is constructed using a market capitalisation ratio of the companies within the respective sector. The portfolio will be reweighted monthly based on the covariance structure forecast using covariance regression, in which covariance regression factors will be obtained at multiple time-frequency scales endogenously from the ETF asset price returns from each sector. At the end of each month, the covariance is forecast for the next month or investment horizon. This is done using low-, mid-, and high-frequency IMFs isolated using EMD from the 11 sector indices over the previous year. The IMFs isolated from the 11 sector indices over the previous year are fitted against the daily logarithmic returns in the RCR model to make multi-frequency covariance forecasts. We construct long/short equity and risk premia parity portfolios using each different covariance forecast and review the results. The performance of the portfolios will be measured using multiple performance measures (the most relevant being risk-related measures with risk premia parity in focus) and contrasted against multiple benchmark portfolios using several well-known portfolio optimisation techniques such as PCA and multivariate GARCH extensions. This paper promotes what we term ''implicit factor'' extraction, empirical market factors, and RCR in portfolio optimisation, horizon-specific active portfolio optimisation, long/short equity portfolios, and risk parity portfolios.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

INDEX TERMS Risk parity, long/short equity, active fund management, portfolio optimisation, empirical mode decomposition (EMD), singular spectrum analysis (SSA), singular spectrum decomposition (SSD), regularised covariance regression (RCR), expectation maximisation.

I. INTRODUCTION

Traditionally, factor models have been developed in the context in which the factors have a direct association with a particular component of market risk, such as:

- a market factor;
- a leverage factor;
- an ESG (environmental, social, and governance) factor;
- an emerging markets factor;
- a convexity or carry factor;
- a currency or interest rate basis risk factor; or
- or combinations of the above (or proxies) and others.

The above are a few of the common explicit factor approaches. In such applications factors (used as time series regression covariates) are constructed with an explicit interpretation. These are used within time series regression models for financial asset price dynamic forecasting applications.

In this work, the factors explored (implicit factors as opposed to the above explicit factors), build upon concepts from works such as [1]. Whilst they don't admit the same direct interpretation as the explicit factors mentioned above, they are still obtained from asset price returns time series. The contribution developed in this manuscript in this regard is to create such implicit factors to produce frequency-specific non-stationary adaptive trend information. This allows one to develop time series forecasting methods using various time resolutions within time series factor models.

To position the context of this ''implicit factor'' development approach, we first provide a clear purpose for this paper in the next section, Section I-A. The contribution and structure are summarised in Section I-B with a diagrammatic representation of the structure of this paper given in Figure 1. This is followed by a brief review of modern portfolio theory in Section I-C. How such factor models have been used in financial econometrics to produce forecasts that act as inputs to dynamic portfolio allocation strategy design is summarised in Section I-C. This is the focus of the problems addressed in this manuscript.

A. PURPOSE OF PAPER

This paper suggests using implicit or empirical (as distinct from explicit factors such as US GDP, SOFR rates, mortgage capital, etc.) as factors in forecasting the covariance of a different set of financial time series. The decomposition algorithms we review (to produce the factors used as inputs for covariance forecasting) are EMD, SSA, and SSD. For further information on EMD see [1], [2], [3], [4], and [5] (among many others); for SSA see [6], and for SSD see [7]. There are numerous algorithm extensions available for EMD, SSA, and SSD - see [1], [8], and [9].

The extensions are not the focus of this work. The factors isolated using these techniques, or others, or the explicit factors directly can be used in the covariance regression framework from [10] and reviewed in Section II-B. Minor extensions are also proposed that allow for the regularisation of the covariance regression through different well-known techniques such as LASSO, ridge, elastic-net, group-LASSO, and subgradient descent with modifications being made to the definition rather than estimation of the mean in [10] algorithm.

This paper further gives the proposed technique exposition with a real-world case study using this technique against well-known benchmarks presented in Section V. The methods in this paper and others can be used in portfolio optimisation whereby risk and return can be optimised with small correlation couplings being captured. See Figure 1 for a diagrammatic summary of this paper and the steps involved. Figure 1 is a gross over-simplification with many other potential additions such as lagging versus formal forecasting of independent variables. All these techniques are used in a real-world case study explained in Section IV.

B. CONTRIBUTIONS AND STRUCTURE

This work begins with Section II outlining the methodology developed for constructing risk parity portfolios using implicit factor extraction techniques and regularised covariance regression. Section II-A outlines the factor extraction techniques. This includes EMD, SSA, and SSD explained in Sections II-A1, II-A2, and II-A3, respectively. The outputs of the implicit factor extraction techniques are used to isolate different time-frequency trend structures from price processes to differentiate the attributable volatility to different time scales. In Section II-B, the original formulation of covariance regression is outlined and uses the framework of [10] before outlining the extension presented here as RCR. The performance of the proposed covariance regression model will be compared with the class of reference models selected based on DCC-MGARCH which is an MGARCH technique within the family of nonlinear combinations of univariate GARCH models.

Finally, the methodology section concludes by describing the risk parity portfolio weighting technique in Section III to calculate the required weighting once all assumptions are made about the covariance structure during the coming period. The risk parity portfolio construction framework is developed based on the covariance forecasts from the proposed models, and the portfolio framework is based on the approach of [11] and [12]. This is subsequently applied to a portfolio-of-portfolios which is constructed from sector indexes. All methodology is carefully analysed on a set of real data case studies based on the S&P500 in order to illustrate the proposed methods and their ability to capture instantaneous correlation and correlation coupling during significant periods of upward and downward momentum. It is

demonstrated how such features from the model translate into the equal risk-parity portfolio structuring and risk-return performance. In this regard, various performance measures will be used to compare the long/short equity risk parity portfolios with the benchmarks detailed in Section V-A.

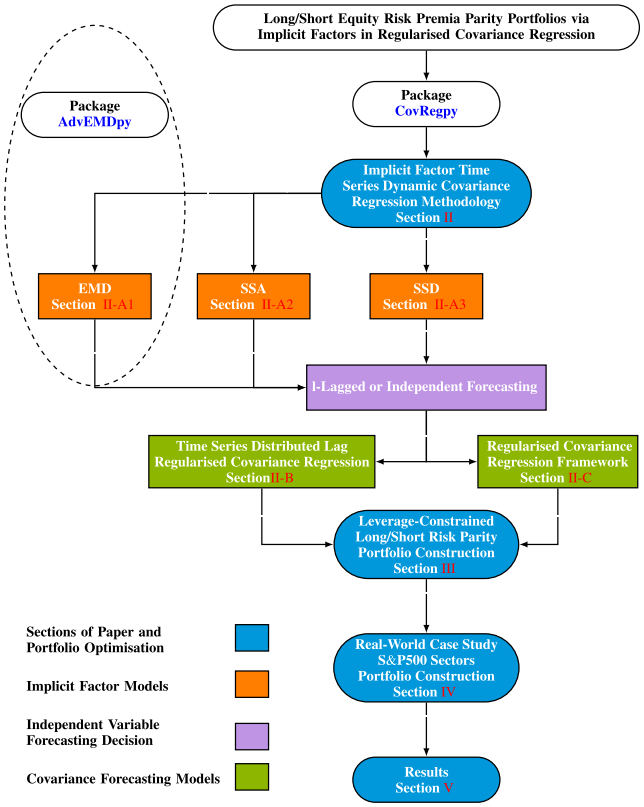

FIGURE 1. Diagrammatic summary of the sections of this paper and stages in designing leveraged risk premia parity portfolios with implicit factor models (such as EMD, SSA, SSD, and many others) in an *l*-lagged and regularised covariance regression framework.

C. LITERATURE REVIEW

The modern portfolio selection criterion of mean variance established in [13] and [14], and developed further in [15], would continue to influence theories such as the Capital Asset Pricing Model (CAPM) developed in [16], [17], and [18] and later works. Extensions were made in [19] where short sales are restricted or in [20] where different interest rates are assumed for lending and borrowing of the risk-free asset. Later work would use the entire distribution in the formulation of the optimised portfolio, as in [21]. The mean-variance criterion is formalised to produce the efficient frontier and mutual fund theorem as in [22].

Arbitrage pricing theory was subsequently introduced to such modelling settings in [23] and looks at relating the price of an asset linearly to an underlying market indicator as an extension of the Capital Asset Pricing Model introduced in [16] and [17]. This covariate, known as a market factor, paved the way for further extensions, as in the market factor models of [24] and [25]. These explicit market factors explained some of the underlying discrepancies seen in the original CAPM model. In [8], the concept of "implicit market factors" is introduced as an explanation of returns above the risk-free rate, where such factors are termed implicit as whilst they are derived from asset return decompositions they are not easy to directly associate to a unique identifiable source of risk and in fact will often comprise multiple sources of risk. As such, implicit market factors offer a treatment of mixed risk drivers that can accommodate both linear, nonlinear, and nonstationary return process structures, not easily facilitated in classical factor methods. Such implicit factors are specifically derived from asset price or macroeconomic time series dynamics that may have explanatory power in explaining dynamics of first- or second-order aspects of return processes in collections of risky assets that comprise the universe of investible assets under consideration.

In this work, implicit factors will be obtained based on S&P500 sector portfolios return dynamics using non-standard time series basis decomposition methods. As such, they are therefore not characteristic of the standard directly observed factors, which are more commonly used in the aforementioned factor model frameworks. This work will demonstrate how to construct such implicit factors, which will be based upon time series decomposition methods that use combinations of empirical mode decomposition (EMD), singular spectrum analysis (SSA), and singular spectrum decomposition (SSD). Such methods are developed to produce trend decompositions, which are capable of non-parametrically obtaining representations of factor structures that isolate different frequency information adaptively over time from the returns time series and ultimately produce a collection of different factor regression time series that can be used to develop targeted time series forecasting methods with a particular frequency resolution of interest.

To illustrate this concept, this manuscript will develop a class of models that utilises these implicit factors as covariates in a covariance regression to optimise a risk-parity portfolio with weighting restrictions. As such, this paper serves to promote the use of implicit factor extraction and time series forecasting methods based on regularised covariance regression (RCR) in the interrelated fields of portfolio optimisation, horizon-specific portfolio optimisation, long/short equity portfolios, and dynamic active risk parity portfolio management.

Over the last few decades, in parallel with developments in such asset return factor models, there has also been a significant advance in modelling of time series stochastic volatility. The most commonly used approaches is to utilise a class of Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models either in univariate or multivariate settings. The autoregressive conditional heteroskedasticity (ARCH) models of [26], [27], and [28] can be seen as direct applications of autoregressive (AR) and autoregressive moving average (ARMA) models to modelling univariate variance. In [29] the ARCH modelled is extended to the GARCH model by incorporating a random error component. All multivariate GARCH (MGARCH) models can be seen as various

multivariate extensions of univariate GARCH models; a thorough review of these extensions can be found in [30].

If one seeks to develop a factor regression model to forecast and explain the dynamics of multivariate volatility, in order to dynamically model the covariance structures for a vector of asset returns, alternative approaches to classical MGARCH models can also be developed based on covariance regression. In this work, we consider such alternative time series-based covariance regression methods for forecasting and will combine these with implicit factors. This work uses an extension of the methodology for covariance regression of [10] and [31]. The extensions proposed in this work include a variety of regularised estimation methods that extend the original covariance regression setting of [10] while also incorporating multiple time-frequency resolutions obtained from the implicit factor extraction methods introduced, which are then subsequently used as distributed lag time series factor covariates in such a covariance regression estimation and covariance regression forecasting framework, which is termed the EMD-RCR portfolio framework herein.

Several methods have been developed recently in the time series literature to treat feature extraction for nonlinear and nonstationary time series. This work explores three important examples: the first is based on EMD which was introduced in [2], [3], and [4]; secondly, Singular Spectrum Analysis (SSA) will be considered; and thirdly, Singular Spectrum Decomposition (SSD) is assessed in the context of an extension of SSA. These implicit factors are then used as distributed lag time series covariates in regularised versions of the covariance regression developed in [10]. It will be noted that the regularisation stages are not completely standard as they take place within an EM iterative estimation framework of the covariance regression.

Once the covariance structure has been forecast (id est, assumptions have been made about the forthcoming performance of the assets according to the portfolio optimisation dichotomy first proposed in [13]), the portfolio weighting, according to the individual's risk appetite, would need to be calculated. Risk premia parity portfolio weighting, first proposed in [11], and later used in [12], [32], and [33], is used to minimise the total variance of the portfolio while maintaining a predefined risk distribution amongst the available assets. The equal weighting is not necessary, with other possible weighting strategies being allowed based on the parameterisation of the problem. It is this final step that concerns itself with the second component of the observations originally made in [13]. All the previous work was concerned with estimating and forecasting the covariance structure, but it is the risk premia parity weighting that concerns itself with the second aspect of portfolio optimisation, which is the weighting of the portfolio once all assumptions about forthcoming returns and covariance have been made.

II. COVARIANCE FORECASTING METHODOLOGY

This section begins with the discussion of the construction of implicit factor extraction (IFE) techniques for EMD, SSA and SSD time series basis decomposition methods, see Section II-A. Additional details on various aspects of SSA and SSD as well as their implementation are also available in [9]. Then in Section II-B details of the time series regularised covariance regression (RCR) model incorporating the implicit factor covariate time series are provided. This involves detailing a dynamic random-effects covariance regression model representation in Section II-B1 before moving on to parameter estimation and inference in Section II-B2. Finally, the EM algorithm is presented in Section II-B3, which details the efficient estimation steps in the estimation algorithm. With this framework in place, the RCR is introduced in Section II-C with several adjustments to the original EM algorithm made to introduce regularisation steps of various types.

A. IMPLICIT FACTOR EXTRACTION

Implicit Factor Extraction (IFE) methods build upon the work of previous factor and feature extraction techniques. Constructing portfolios using Principal Component Analysis (PCA), such as in [34] and [35], is a form of compression with the attributes quantified in components based on variance, and this can result in components being mixed in the projections utilised. In [13] the global minimum variance portfolio and the efficient frontier are constructed using the realised mean and covariance, but all this was based exclusively on the past performance of assets with little to no factor or feature extraction. The CAPM model by [16] and [17] related assets prices to the risk-free proxy, with the arbitrage pricing model and the multifactor models as [24] and [25] being extensions thereof. The models all rely on asset data being used as observed with minimal pre-processing, in addition possibly to some standardisation and return transformations.

The Fama-French 3 Factor Model from [24] is compared against the Carhart 4 Factor Model from [36] in [37]. The Carhart 4 Factor Model from [36] is the same as the Fama-French 3 Factor Model with a fourth factor that tries to capture the salient risk present in the market with a momentum factor. The Carhart 4 Factor Model is found to be more accurate than the Fama-French 3 Factor Model in terms of the root mean square error when applied to eight years of the Hanoi Stock Exchange, but both models were found to be equally significant using their t-tests.

Formal model comparisons are conducted by [38] where the performance of a single index model, the Fama-French 3 Factor Model, and the Carhart 4 Factor Model are assessed in their ability to forecast the returns of the 20 largest stocks in the NASDAQ and NYSE exchanges. Much like [37] and [38] found that the Carhart 4 Factor Model performed the best in terms of the largest adjusted R^2 values. These three-factor models were, however, found to perform poorly when compared against the Black-Litterman model in constructing portfolios. The portfolio constructed using Black-Litterman technique outperformed the portfolios constructed using the

three-factor models by having relatively higher returns with relatively lower variance and having the highest Sharpe ratio.

A different approach, by [39], performs feature selection techniques on several candidate micro- and macro-economic factors from a potential pool of 44 to select a set of 10 relatively independent factors. These factors are used to construct a multifactor model (explicit factors with minimal preprocessing as opposed to implicit factor models explored herein) to measure the portfolio performance and backtest the model by adjusting the factors based on varying the retracement periods. Finally, this explicit multifactor model is then combined with machine learning techniques to perform stock selection. The machine learning-augmented explicit multifactor stock selection method was shown to outperform the traditional explicit multifactor stock selection model.

In another portfolio optimisation technique begun with feature selection preprocessing, [40] uses double-selection LASSO feature selection before using PCA on the selected features. Double-selection LASSO feature selection uses traditional LASSO feature selection first between the covariates and the dependent variables before, secondly, traditional LASSO feature selection is then performed between the covariates and the independent variables. PCA is then used to isolate the principal components from the remaining features. These factors are then used to predict asset prices using both support vector regression (SVR) and vanilla multivariate regression. SVR is shown to exceed multivariate regression in the accumulated returns from the case studies conducted using US and China stock market data.

There is an opportunity, therefore, to consider the development of robust multiple-time-frequency factor extraction techniques to be used in finance, especially as will be illustrated in the examples in this manuscript; they can be beneficial to the study of multiple-time-scale investment decision-making.

In Sections II-A1, II-A2, and II-A3 this work gives an exposition of EMD, SSA, and SSD, respectively. These three methods are not exhaustive with many other potential decomposition algorithms capable of providing implicit market factors for later use in covariance forecasting in Section II-B. As an example, variational mode decomposition (VMD) was developed in [41] to address some of the possible limitations of EMD, such as sensitivity to noise and sampling. This technique is described as a generalisation of classic Wiener filters from [42]. VMD extends classic Wiener filters into multiple and adaptive frequency bands. VMD is shown to perform better than EMD in certain specific synthetic scenarios such as the summation or different frequency sawtooth waves, but in the one real-world example in [41] (an electrocardiogram signal) the methods are not compared, but rather the results for VMD are presented. Admittedly, these results are promising.

EMD and VMD are compared by [43] in their ability to extract the salient features from grayscale images before machine-learning classification algorithms are used to identify and classify objects. Both support vector machines (a kernel-based classifier) and random forest (an ensemble classifier technique which is an extension of decision-tree algorithms) are used in the classification of the features extracted using VMD and EMD. References [44] and [45] use VMD and EMD, respectively, to improve speech recognition through adaptive denoising and filtering of online speech signals. To improve upon VMD from [41] for the specific task of speech signal decomposition, [44] develops adaptive frequency-shifting VMD (VMD-FS) to shift speech to lower-frequency bandwidths for improved decomposition.

Wavelet decomposition, an extension of wavelet analysis, was also considered as a potential decomposition algorithm, but it, as well as its inspiration from Fourier analysis, were not considered appropriate and disregarded based on their constructive nature (using predefined bases). Both wavelet decomposition and Fourier decomposition have been extended from their original exclusively stationary time series decomposition algorithms to their non-stationary variations of empirical wavelet transforms (developed in [46]) and empirical Fourier decomposition (developed in [47]). Despite these extensions and developments (noted in [46] and [47] as having been inspired by EMD), these techniques are still constructive and [48] found that EMD is superior to wavelet decomposition in both accuracy and computation complexity.

Lower-order tensor decomposition was also considered and is discussed in [49] in the context of signal denoising through factor analysis and decomposition. In [49] Tucker decomposition is modified via truncation based on the signal-to-noise ratio and subsequently paired with PCA and tested against other techniques such as wavelet transforms augmented with PCA in their decomposition ability. This technique is concluded to be more effective than the other techniques surveyed in [49], but it is noted that it is computationally expensive and not appropriate for online or time-sensitive forecasting and prediction. Many other decomposition techniques exist with their advantages and disadvantages, but their review is beyond the scope of this work.

1) EMPIRICAL MODE DECOMPOSITION

This section begins with the EMD method which extracts basis functions from a signal, in our case price or return time series, known as Intrinsic Mode Functions (IMFs) through an algorithmic procedure known as Sifting (see details and a python package to perform variations of sifting in [1]) to produce the following decomposition:

$$x_{IMF}(t) = \sum_{j=1}^{K} \gamma_j(t) + r_K(t), \tag{1}$$

with $\gamma_j(t)$ being the j^{th} IMF decomposed from the time series and $r_K(t)$ being the non-stationary residual after K IMFs have been decomposed, which has the special property that it has, at the most, a single convexity change over the time-domain on which the EMD basis extraction is performed.

There are numerous algorithmic variations designed to address specific shortcomings of the original algorithm. The most ubiquitous problem in EMD (and possibly in time series analysis) is the edge of the time series that, because of the iterative nature of the sifting algorithm, results in the proliferation of errors throughout the decomposed IMFs. Reference [1] investigates the edge effect and the numerous boundary conditions (amongst algorithmic variations of other subroutines within the algorithm) used to address this proliferation of errors. Reference [50] investigates a few of these edge effects and some others along with algorithmic extensions to EMD in their ability to minimise the proliferation of these errors and further begins to formalise the decay rate of the errors in the components as a function of the locations of extrema within the decomposed time series.

If the edge effect is not properly managed with an appropriate static or dynamic boundary effect, the errors will proliferate throughout the IMFs and will result in a related problem in non-stationary time series decomposition of mode-mixing which is the mixing of frequencies between components - see [51]. This mixing of components based on their frequency content is difficult to assess when there are no ground truth target components (for example, the earthquake data in [2]) such as in synthetic experiments. Reference [52] compares ensemble EMD (EEMD) and masking EMD (M-EMD) against EMD in their ability to minimise mode mixing whilst decomposing a series of synthetically constructed time series. Reference [52] found that M-EMD performed the best with EEMD performing better than EMD, but not as well as M-EMD.

Another useful characteristic of EMD is its function as a robust non-stationary filtering technique. This is particularly useful when short-time-interval artefacts interfere with signals and instead of truncating the signal or clipping the corrupted intervals where valuable diagnostic or predictive information would be discarded. To this end, [53] investigates the ability of EMD Interval Thresholding (EMD-IT) and Iterative EMD-IT (EMD-IIT) to specifically isolate artefacts from time series whilst retaining as much of the original information in the signal. It is concluded by [53] that EMD-IIT is the most accurate in its isolation and extraction of the corrupting artefacts, but it is computationally expensive and difficult to use in real-time applications. EMD-IT is less accurate but also less computationally expensive and, therefore, more amenable to online applications.

In a similar application, [54] uses EMD and EEMD as frequency-adaptive filters to denoise signals before using them for interpretation and prediction. A shortcoming of this technique noted in [54] is the automation of the stopping of the EMD algorithm and individual IMF sifting subroutines as well as the appropriate splines to use to best isolate and extract the salient features. Reference [1] assesses a number of stopping criteria for EMD whilst [8] assesses different spline techniques used to approximate the IMFs. Reference [55] addressed the same problem as [54], but instead of using EEMD, an amalgamation of EMD and wavelet analysis which was found in [53] to be effective in removing artefacts with [55] also finding it effective at denoising time series. The are numerous algorithmic variations, [1], and extensions, [8], of EMD, but these will not be assessed herein.

A decomposition of type Equation (1) is applied to each asset price series in a given asset universe that is used to construct a portfolio. The resulting IMFs provide different time-frequency scale basis functions that are adapted to nonstationary price or return signal time series dynamics. The sifting procedure results in a collection of IMF basis function which each satisfy the following two conditions, see discussions in [45]:

Condition 1 $\text{abs}\left(\left|\left\{\frac{d\gamma_k(t)}{dt} = 0 : t \in (0,T)\right\}\right| - \left|\left\{\gamma_k(t) = 0 : t \in (0,T)\right\}\right|\right) \leq 1$, and

Condition 2 $\tilde{\gamma}_k^{\mu}(t) = \left(\frac{\tilde{\gamma}_k^M(t)+\tilde{\gamma}_k^m(t)}{2}\right) = 0 \; \forall \, t \in [0,T]$ with,
$\gamma_k(t) = \tilde{\gamma}_k^M(t)$ if $\frac{d\gamma_k(t)}{dt} = 0$ and $\frac{d^2\gamma_k(t)}{dt^2} < 0$,
$\gamma_k(t) \leq \tilde{\gamma}_k^M(t) \; \forall \, t \in [0,T]$,
$\gamma_k(t) = \tilde{\gamma}_k^m(t)$ if $\frac{d\gamma_k(t)}{dt} = 0$ and $\frac{d^2\gamma_k(t)}{dt^2} > 0$, and
$\gamma_k(t) \geq \tilde{\gamma}_k^m(t) \; \forall \, t \in [0,T]$,

with $\tilde{\gamma}_k^M(t)$ being the spline fitted through the maxima of the time series and $\tilde{\gamma}_k^m(t)$ being the spline fitted through the minima that produce an envelope of the IMF basis. In the application of EMD in this work, cubic B-splines are used to fit and smooth the IMFs. The mean squared error optimality of cubic splines has already been shown in [56], [57], and [58]. Numerous packages have been developed to implement the EMD basis extraction method, see detailed discussion in [1] and the accompanying Python package at:

https://github.com/Cole-vJ/AdvEMDpy.

In Section V we present a case study using IMFs from different frequency bandwidths, isolated using EMD described above, to make different frequency covariance forecasts about the possible investible assets in the portfolio reweighted monthly. The observations of the different performances of the portfolios based on different frequency bandwidth forecasts follow the conclusions of [45] that the separation of structures using in EMD in non-stationary time series benefits later analysis and further predictive capabilities. The sorting of these IMFs based on bandwidths that are not fixed, owing to the robust nature of the EMD algorithm, presents a unique problem that is addressed in Section IV-C1 that references [8] where it was first proposed to be used in conjunction with EMD in sorting structures with temporally varying bandwidths.

2) SINGULAR SPECTRUM ANALYSIS

The second intrinsic factor extraction method proposed to be utilised on each assets return time series is a method proposed in [6] that consists of four core stages, namely Embedding, Singular Value Decomposition, Grouping, and

Trend Estimation through Diagonal Averaging. Only a single factor is extracted using the original SSA technique, and the frequency of this component is determined by the predetermined window width, which is discussed below. This component can be extracted from each asset in a portfolio and used to forecast the covariance of the portfolio or to forecast the price of individual assets.

a: STAGE 1 OF SSA: EMBEDDING

Given a time series, $x(t)$, with T samples, the embedding stage consists of converting the univariate time series into a multi-variate time series by embedding lagged increments of the time series into a matrix $\mathbf{X}$ such that:

$$\mathbf{X} = [\mathbf{X}_1(t), \ldots, \mathbf{X}_K(t)], \tag{2}$$

with $\mathbf{X}_j(t) = [x(t_j), x(t_{j+1}), \ldots, x(t_{j+L-1})]^T$, L being the window length, and $K = T - L + 1$.

b: STAGE 2 OF SSA: SINGULAR VALUE DECOMPOSITION

Let λ denote the set of eigenvalues, $[\lambda_1, \ldots, \lambda_L]^T$, arranged in descending order of magnitude such that $\lambda_1 \geq \cdots \geq \lambda_L$ with the associated set of orthonormal eigenvectors denoted by $\mathbf{U}$ such that $\mathbf{U} = [\mathbf{U}_1, \ldots, \mathbf{U}_L]^T$, with $\mathbf{U}_j$ being the eigenvector associated with λ_j. Then by denoting $\mathbf{V}_j = \mathbf{X}^T\mathbf{U}_j/\sqrt{\lambda_j}$, the Singular Value Decomposition of the trajectory matrix can be written as:

$$\mathbf{X} = \sum_{k=1}^{K} \mathbf{X}_k = \sum_{k=1}^{K} \sqrt{\lambda_k}\mathbf{U}_k\mathbf{V}_k^T. \tag{3}$$

c: STAGE 3 OF SSA: GROUPING

For some integer $p < L$, denote by I the subset of indices such that $I = \{i_1, \ldots, i_p\} \subset \{i_1, \ldots, i_L\}$ that allows for the specification of partial components:

$$\mathbf{X}_I = \sum_{j=i_1}^{i_p} \mathbf{X}_j. \tag{4}$$

d: STAGE 4 OF SSA: TREND ESTIMATION & DIAGONAL AVERAGING

The final step of Singular Spectrum Decomposition is essentially the reversal of the embedding step performed on the chosen approximation of $\mathbf{X}$, now denoted by $\mathbf{X}_I$. Unlike the other techniques listed in Section II-A, the final output of SSA decomposition is a single trend estimate based on a summation of some subset of the diagonally averaged set such that:

$$x_{\text{SSA}}(t) = \sum_{i=1}^{p-1} \frac{(p+1)-i}{p}\mathbf{X}_{I,i} + \sum_{i=p}^{L-p} \frac{1}{p}\mathbf{X}_{I,i} + \sum_{i=L-p+1}^{L} \frac{(L-i)}{p}\mathbf{X}_{I,i}. \tag{5}$$

3) SINGULAR SPECTRUM DECOMPOSITION

This extension of Singular Spectrum Analysis was formalised in [7] to automate some steps such as the bandwidth of each chosen decomposition structure and to formalise the technique as a decomposition algorithm with many bandwidth-specific components rather than simply a trend estimation technique such as outlined in the original SSA method. This can be interpreted as a modification of down-sampling to multiple, possibly disjoint, bandwidths without the decimation step. In [1] down-sampling without decimation is introduced as a preprocessing technique. SSD automates the isolation of multiple significant (power spectral density nodes) structures in a time series - these structures can be used in covariance modelling and forecasting of financial data. The stages of SSD that generalise those of SSA are presented below.

a: STAGE 1 OF SSD: SIGNIFICANT TREND

The first step of the SSD algorithm is to test for a significant trend. A time series is deemed transient or nonstationary if the normalised dominant frequency peak (f_{max}/F_s) is below a prespecified, user-defined threshold. In [7] a threshold of $f_{\text{threshold}} = 10^{-3}$ is recommended. If the dominant peak is below this threshold, then the window width or embedding dimension, L, is set to $T/3$ as recommended in [59]. If this frequency threshold is not met, the embedding dimensions are set to $1.2 \times \frac{F_s}{f_{\max}}$, which is dynamically adjusted throughout the algorithm. This step automates and formalises the identification of a significant trend, which is often present in financial data. The threshold recommended in [59] does not need to be strictly adhered to, but can be adjusted to user's desired resolution or time scale of focus in the feature extraction.

b: STAGE 2 OF SSD: DOWN-SAMPLING

Three Gaussian functions are then fitted to the subsequent power spectral densities to estimate the next iteration of the extraction algorithm. The following function is fit to the power-spectral density:

$$\gamma(f, \theta) = \sum_{i=1}^{3} A_i e^{-\frac{(f-\mu_i)^2}{2\sigma_i^2}}. \tag{6}$$

where the μ parameters for each Gaussian density basis are specified as follows:

$$\mu_1 = f_{\max},\ \mu_2 = f_2,\ \mu_3 = \frac{f_{\max} + f_2}{2}, \tag{7}$$

and where $f_{\max}$ is the dominant spectral peak and f_2 is the second spectral peak. The initial estimates for the other parameters are:

$$\begin{aligned} A_1^{(0)} &= \frac{1}{2}\text{PSD}(f_{\max}), \quad \sigma_1^{(0)} = f : \text{PSD}(f) = \frac{2}{3}\text{PSD}(f_{\max}), \\ A_2^{(0)} &= \frac{1}{2}\text{PSD}(f_2), \quad \sigma_2^{(0)} = f : \text{PSD}(f) = \frac{2}{3}\text{PSD}(f_2), \\ A_3^{(0)} &= \frac{1}{4}\text{PSD}(f_3), \quad \sigma_3^{(0)} = 4|f_{\max} - f_2|, \end{aligned} \tag{8}$$

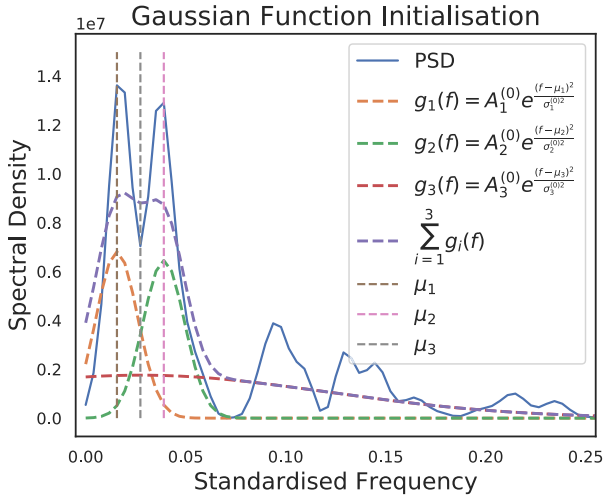

FIGURE 2. SSD power spectral density Gaussian function initialisation.

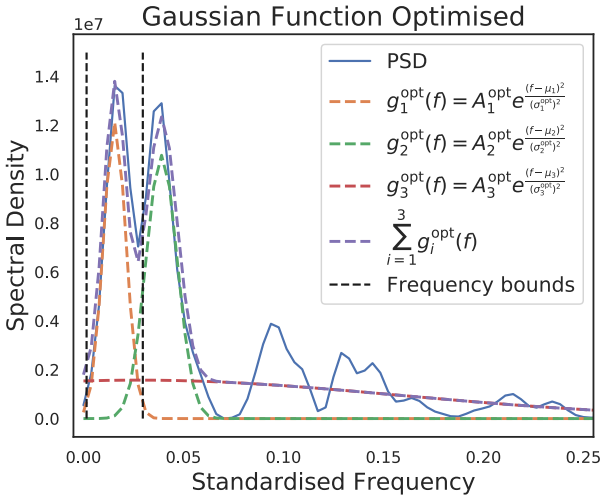

FIGURE 3. SSD power spectral density Gaussian function optimisation.

where $f : \text{PSD}(f) = \frac{2}{3}\text{PSD}(f_*)$ denotes the frequency, f, such that the power-spectral density at this frequency, f, is two thirds the power-spectral density at f_*. With these initial conditions, Equation (6) is fitted to the complete power spectral density as can be seen in Figure 2 in some sample data from a subset of the S&P 500 after the significant trend has been extracted.

The initialised parameters in Equation (8), and plotted in Figure 2 for reference, are mean squared error optimised to produce the parameters to fit the power-spectral density as in Figure 3. Once σ_1^{opt} has been estimated, the frequency bound for the down-sampling, $\delta f = 2.5\sigma_1^{\text{opt}}$, such that it captures approximately 99% of the area underneath the primary power-spectral density mode. The frequency bounds can be seen as the black dashed lines in Figure 3. By formally acknowledging the presence of a second significant frequency node in the power spectral density, this technique is able to separate two structures that would otherwise confound modelling and forecasting attempts in the time domain for financial signals. The original SSA technique would not be able to separate the two significant structures observable in Figures 2 and 3.

The SSA technique with a modified embedding step is then applied to this component estimated through down-sampling. Down-sampling is not formally acknowledged as the process, but it is clear from the application that this initial estimate is as a result of down-sampling.

c: STAGE 3 OF SSD: MODIFIED EMBEDDING

This modified embedding step is similar to that presented in Equation (2) with a slight adjustment. The new matrix, $\boldsymbol{X}^{mod}$, is calculated as:

$$\mathbf{X}^{\text{mod}} = [\mathbf{X}_1^{\text{mod}}(t), \ldots, \mathbf{X}_K^{\text{mod}}(t)], \tag{9}$$

where

$$\mathbf{X}_j^{\text{mod}}(t) = [x(t_j), x(t_{j+1}), \ldots, x(t_{j+L-1}), x(t_1), \ldots, x(t_{j-1})]^T, \tag{10}$$

and with L being the window width, and $K = T - L + 1$ as before. This will result in a modified trend estimate that is more cyclical as a result of what is referred to in [7] as wrapping the matrix around. This 'wrapping' results in the isolated structures being more cyclical, which prevents large deviations, particularly at the edges of the isolated components, and which in the applications in this manuscript will prevent extreme estimations of the covariance and assets prices.

d: STAGE 4 OF SSD: SCALING

Once the final component estimate has been made, a scaling factor is introduced to best fit the component to the underlying structure using mean squared error optimisation. This scaling factor is calculated as in Equation (11):

$$\hat{a}_i = \min_{a_i} \left|\left| v_i(t) - a_i g_i(t) \right|\right|_2^2, \tag{11}$$

with $v_i(t)$ being the remainder of the time series after the first $i-1$ components have been removed and $g_i(t)$ is the unscaled i^{th} component. The i^{th} scaled component is then given by $\tilde{g}_i(t) = \hat{a}_i g_i(t)$.

Figure 4 demonstrates the remaining time series ($v_i(t)$), the unscaled i^{th} component ($g_i(t)$) and the scaled i^{th} component ($\tilde{g}_i(t)$). This allows one to adjust the financial factor to be extracted to better fit (mean squared error) the residual time series.

e: STOPPING CRITERION

The final result of the SSD is a decomposition of the underlying time series, $x(t)$, such that:

$$x_{\text{SSD}}(t) = \sum_{i=1}^{M} \tilde{g}_i(t) + v_{M+1}(t), \tag{12}$$

where $\tilde{g}_i(t)$ is the i^{th} component extracted and $v_{M+1}(t)$ is the residual. To stop the 'sifting' (word borrowed from EMD, but accurately describes what is happening) procedure, a stopping criterion must be introduced. A normalised mean

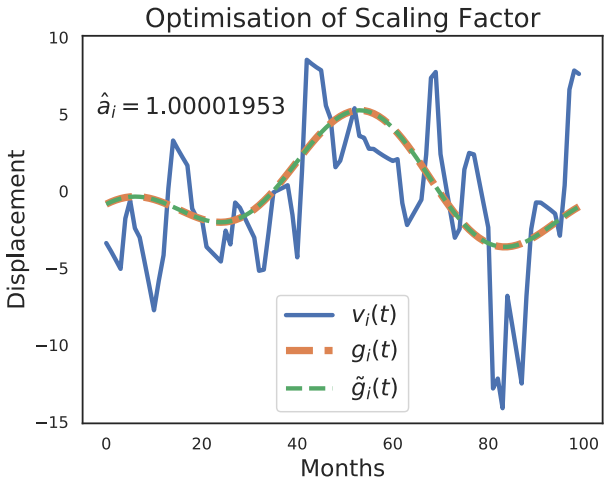

FIGURE 4. Singular Spectrum Decomposition scaling factor mean squared error optimisation.

squared error (NMSE) is introduced with some threshold value so that the sifting stops when the NMSE:

$$\text{NMSE}_i = \frac{\sum_{i=0}^{N} v_{M+1}^2(t_i)}{\sum_{i=0}^{N} x^2(t_i)}, \tag{13}$$

is below a user-specified threshold value and $x(t)$ is the original unsifted time series. In [7], it is suggested that a threshold value of 1% be used - this stops the algorithm when 99% of the energy of the time series has been decomposed. This prevents over-sifting and the proliferation of nonsensical components for further modelling.

B. LAGGED REGULARISED COVARIANCE REGRESSION

Sample covariance matrix (SCM) estimation, which assumes the covariance stationarity of the time series, is commonly used when predicting portfolio risk-return trade-offs such as the classic efficient frontier originally formulated in [13]. A natural and popular regularised extension of assuming covariance stationarity (see [60]) is shrinkage SCM which has the same eigenvectors as the original SCM, but the eigenvalues are regularised (via a shrinkage parameter) towards the average of all the eigenvalues of the unregularised SCM. Refernce [60] proposes an extension of this shrinkage SCM estimation by using the maximum likelihood estimator or "M-estimator" of the scatter matrix with an automated and adaptive technique for calculating the shrinkage parameter. This approach also allows a more generalised weighting function to be used which is shown in [60] to be better suited to real-world problems where error distributions do not follow the Gaussian distribution, but come from more generalised heavy-tailed symmetric distributions.

The adjustment of SCM estimation to robustify the estimation procedure for errors of heavy-tailed distribution is the focus of [61]. Reference [61] uses Tyler's cost function when using the M-estimator in calculating the regularised SCM estimation. Reference [61] uses the natural structure inherent in covariance matrices to demonstrate that the structure-constrained (constraints based on observed structure) Tyler's estimator of the SCM exhibits lower error rates than the unconstrained optimisation in the simulated case studies. Reference [62] approaches the robust estimation of a structured covariance matrix by using the Kronecker products of matrices of low rank. To this end, [62] derives the constrained Tyler's estimator which is the minimiser of Tyler's estimator cost function with the additional constraint of being structurally a Kronecker product.

Heavy-tailed sampling experiments are conducted by [63] who assess the regularised estimation of the mean and covariance when samples are drawn from generalised elliptical distributions. In this work by [63], the developed regularised estimation is found to be accurate and robust when synthetic samples are drawn from heavy-tailed distributions or when there are numerous outliers. The proposed method outperforms the Cauchy maximum likelihood estimate and the student t-distribution estimate when measured using the Kullback–Leibler divergences of the predicted covariance versus the distribution from which the samples are drawn. Reference [64] specifically develops a robust mean and covariance forecasting framework for financial applications where there is missing data. This mean and covariance estimation assumes the missing data is monotone which has applications in financial data as well as using Student-t distributions to better capture the heavy-tailed distribution of the errors in finance. It is shown that robust estimation using the minorisation-maximisation framework converges more quickly than expectation-maximisation and parameter-expanded expectation-maximisation.

Drawbacks of the above-mentioned methods and techniques of covariance estimation are the difficulties in estimating the mean or expected returns of the distributions when applied in financial applications - see [65], [66], [67], and [68]. Further, the dual mean-covariance estimation problem is very sensitive to small errors in the expected returns - see [65], [66], [67], [68], and [69]. It is also generally agreed (see [70] and [71]) that modelling the covariance is easier than modelling the expected returns, given the historical data. Also, the above frameworks do not permit exogenous variables of factors to be used in the covariance forecasting. [72] partly addresses this by using factor analysis in forecasting the covariance, but with these concerns, we propose using implicit market factors derived in II-A within the [10] framework to make regularised financial covariance forecasts.

With the multi-resolution basis extraction techniques, notation, and framework established in Section II-A, one can now use these different basis extraction methods on financial price return time series, either for individual assets or on indices for portfolios of assets, to produce a collection of factors that are termed here Implicit Factors (IFs) at multiple time-frequency resolutions. These IFs then form the factors that will be subsequently used in Regularised Covariance Regression (RCR) to forecast the covariance. The concept of covariance regression methods for dynamic portfolio design

has been used in financial applications previously; see [73]. This work extends these concepts to multi-time resolution settings as well as the addition of various regularisation constraints in the estimation procedures.

The reason that such time-frequency Intrinsic Factors (IFs) (used interchangeably with implicit factors) are so useful is that they will allow us to use these multiresolution IF components within the RCR, to construct horizon-specific portfolios that can then accommodate covariance forecasting optimised in its accuracy when forecasting for specific time horizons or on specific sampling rates. In this setting, the IFs that form the time series covariates can be obtained as the intrinsic mode functions (IMFs) extracted by the EMD, the SSA or the SSD methodologies when applied to past of the time series of asset returns considered in the regression model construction. For instance, if one seeks to forecast the covariance for a vector of asset returns, one could obtain the IFs endogenously by extracting these implicit factors from the historical price time series, and then from a given forecast origin use the estimated covariance regression to forecast the covariance at some future forecast horizon. In this manner, the extracted IFs are introduced to the covariance regression as a distributive lag time series covariance regression model.

The proposed regularised covariance regression (RCR) model is developed under a formulation that can be expressed as a type of time series distributed lag random-effects model. In this section, one begins by outlining the results underlying the random-effects formulation of covariance regression in Section II-B1. This is followed by parameter estimation and inference in Section II-B2 which concludes with results concerning the pseudoinverse of random errors in the model. In Section II-B3 the expectation maximisation algorithm is shown which allows the translation to pseudocode. This section is concluded with Section II-C which outlines the various regularised covariance regression techniques presented in this paper. The MGARCH alternative that is commonly used can be found in the Supplement to: Long/Short Equity Risk Premia Parity Portfolios via Implicit Factors in Regularised Covariance Regression, Appendix D.

1) RANDOM EFFECTS REPRESENTATION

Let $\mathbf{y}_t \in \mathbb{R}^p$ denote the multivariate response, corresponding to the daily log returns of a set of p assets at time t, denote by $\widetilde{\mathbf{x}}_{t-i,1} \in \mathbb{R}^q$ the column vector of q covariates at lag i of factors used to approximate the trend dynamic at time t, denote by $\widetilde{\mathbf{x}}_{t-j,2} \in \mathbb{R}^r$ the column vector of r covariates at lag j used to explain the covariance of the log returns about the trend at time t. Denote the model parameter matrices at lag i by $\mathbf{A}_i \in \mathbb{R}^{p\times q}$ being the coefficients of the mean and $\mathbf{B}_j \in \mathbb{R}^{p\times r}$ being the matrix of coefficients for the random-effects terms that incidentally describe the relationship of the regression structure covariance between the lagged covariates $\mathbf{x}_{t-j,2}$ and the covariance of the returns for $\mathbf{y}_t$ at time t. We denote the latent variables for random effects by process $\{\gamma_t\}$ and the uncorrelated regression errors in time, but potentially cross-correlated by process $\{\boldsymbol{\epsilon}_t\}$. The random effects covariance regression model may therefore be specified in the following model structure, for n_1 lags of covariates in the trend and n_2 lags of covariates in the covariance dynamic:

$$\begin{aligned}\mathbf{y}_t &= \sum_{i=0}^{n_1}\mathbf{A}_i\widetilde{\mathbf{x}}_{t-i,1} + \gamma_t\sum_{j=0}^{n_2}\mathbf{B}_j\widetilde{\mathbf{x}}_{t-j,2} + \boldsymbol{\epsilon}_t,\\ &= \mathbf{A}\mathbf{x}_{t,1} + \gamma_t\mathbf{B}\mathbf{x}_{t,2} + \boldsymbol{\epsilon}_t,\end{aligned} \tag{14}$$

where the following stacked matrices and covariate vectors are denoted for a compact notation used in the remainder of the manuscript:

$$\begin{aligned}\mathbf{x}_{t,1} &= \mathtt{vec}\left[\widetilde{\mathbf{x}}_{t,1}^T, \widetilde{\mathbf{x}}_{t-1,1}^T, \widetilde{\mathbf{x}}_{t-2,1}^T, \ldots, \widetilde{\mathbf{x}}_{t-n_1,1}^T\right]^T,\\ &\quad \text{lagged covariate vectors of dimension } (q(n_1+1)\times 1)\\ \mathbf{x}_{t,2} &= \mathtt{vec}\left[\widetilde{\mathbf{x}}_{t,2}^T, \widetilde{\mathbf{x}}_{t-1,2}^T, \widetilde{\mathbf{x}}_{t-2,2}^T, \ldots, \widetilde{\mathbf{x}}_{t-n_2,2}^T\right]^T,\\ &\quad \text{lagged covariate vectors of dimension } (r(n_2+1)\times 1)\\ \mathbf{A} &= \left[\mathbf{A}_0 : \mathbf{A}_1 : \mathbf{A}_2 : \cdots : \mathbf{A}_{n_1}\right],\\ &\quad \text{a matrix of dimension } (p\times q(n_1+1))\\ \mathbf{B} &= \left[\mathbf{B}_0 : \mathbf{B}_1 : \mathbf{B}_2 : \cdots : \mathbf{B}_{n_2}\right],\\ &\quad \text{a matrix of dimension } (p\times r(n_2+1)).\end{aligned} \tag{15}$$

Furthermore, if one makes the following statistical assumptions regarding this model,

$$\begin{aligned}&\mathbb{E}[\boldsymbol{\epsilon}_t] = \mathbf{0},\ \mathrm{Cov}[\boldsymbol{\epsilon}_t] = \boldsymbol{\Psi},\\ &\mathbb{E}[\gamma_t] = 0,\ \mathrm{Var}[\gamma_t] = 1,\ \text{and } \mathbb{E}[\gamma_t\boldsymbol{\epsilon}_t] = \mathbf{0},\end{aligned} \tag{16}$$

then the resulting covariance matrix of the asset returns $\mathbf{y}_t$ given lagged covariates $\mathbf{x}_{t,1}$, $\mathbf{x}_{t,2}$ (and with $\boldsymbol{\mu}_{\mathbf{x}_{t,1}} = \mathbf{A}\mathbf{x}_{t,1}$) is expressed as follows:

$$\begin{aligned}&\mathbb{E}\left[(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x}_{t,1}})(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x}_{t,1}})^T\right]\\ &= \mathbb{E}\Big[{\gamma_t}^2\mathbf{B}\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T\mathbf{B}^T\\ &\qquad + \gamma_t\left(\mathbf{B}\mathbf{x}_{t,2}\boldsymbol{\epsilon}_t^T + \boldsymbol{\epsilon}_t\mathbf{x}_{t,2}^T\mathbf{B}^T\right) + \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\Big]\\ &= \mathbf{B}\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T\mathbf{B}^T + \boldsymbol{\Psi}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}}\\ &= \boldsymbol{\Sigma}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}}.\end{aligned} \tag{17}$$

In this manuscript, the Implicit Factors (IFs), extracted using the previously described EMD, SSA, or SSD methods, will be used for the covariates $\mathbf{x}_{t,1}$ and $\mathbf{x}_{t,2}$ and will be added to the regression with a lag structure. However, just as easily, it is clear that such a model also accommodates a hybrid model of Farma-French factors in the trend and IFs in the covariance, respectively.

With regard to estimation, such a random effects covariance regression representation is advantageous as it admits an Expectation Maximisation (EM) estimation framework which has both E and M steps obtainable in closed form, at least in some versions of the regularised covariance regression (RCR) model specification. In non-differentiable regularisation versions, one will be shown how to properly treat the M-step via proximal gradient methods.

2) LIKELIHOOD SPECIFICATION

The calibration of the aforementioned random effects covariance regression model proceeds as follows. With the random effects representation and errors defined as above with:

$$\gamma_1, \ldots, \gamma_n \overset{\text{iid}}{\sim} \mathcal{N}(0, 1), \tag{18}$$

$$\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n \overset{\text{iid}}{\sim} \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Psi}), \tag{19}$$

$$\widetilde{\mathbf{e}}_t = \mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x}_{t,1}}, \text{ and} \tag{20}$$

$$\mathbf{E} = \mathtt{vec}[\mathbf{e}_1^T, \ldots, \mathbf{e}_n^T]^T \tag{21}$$

where $\mathcal{MVN}(\mathbf{0}, \boldsymbol{\Psi})$ is the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Psi}$, and with the residual matrix defined as $\mathbf{E}$ and dependent on the choice of $\mathbf{x}_{t,1}$ and mean coefficient matrix, $\mathbf{A}$, (through $\boldsymbol{\mu}_{\mathbf{x}_{t,1}}$) the log-likelihood of parameters $(\mathbf{B}, \boldsymbol{\Psi})$ given $\mathbf{E}$ and $\mathbf{X}$ is:

$$l(\mathbf{B}, \boldsymbol{\Psi} : \mathbf{E}, \mathbf{X}) = c - \frac{1}{2}\sum_{t=1}^{n}\log\left|\boldsymbol{\Psi} + \mathbf{B}\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T\mathbf{B}^T\right| - \frac{1}{2}\sum_{t}\mathrm{tr}\left[\left(\boldsymbol{\Psi} + \mathbf{B}\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T\mathbf{B}^T\right)^{-1}\mathbf{e}_t\mathbf{e}_t^T\right]. \tag{22}$$

It is shown in the Supplement to: Long/Short Equity Risk Premia Parity Portfolios via Implicit Factors in Regularized Covariance Regression, Appendix A, that:

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1} &= \boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1}\mathbf{e}_t\mathbf{e}_t^T\boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1} \text{ and} \\ \boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1}\mathbf{B}\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T &= \boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1}\mathbf{e}_t\mathbf{e}_t^T\boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1}\mathbf{B}\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T,\end{aligned} \tag{23}$$

from which it follows that the MLE of $\boldsymbol{\Sigma}_{\mathbf{x}_{t,2}}^{-1}$ acts as a pseudoinverse for $\mathbf{e}_t\mathbf{e}_t^T$.

3) ESTIMATION VIA EXPECTATION MAXIMIZATION ALGORITHM

To perform the maximisation of the likelihood to perform the parameter estimation, it will be convenient to utilise an auxiliary variable method procedure known as the EM algorithm. This is convenient since Equation (22) is difficult to maximise directly, but the random effects representation allows it to be maximised using iterative methods. The iterative Expectation Maximisation (EM) Algorithm depends on the conditional distribution of $\{\gamma_1, \ldots, \gamma_n\}$ given $\{\mathbf{Y}, \mathbf{X}, \boldsymbol{\Psi}, \mathbf{B}\}$. One can refer to the Supplement to: Long/Short Equity Risk Premia Parity Portfolios via Implicit Factors in Regularized Covariance Regression, Appendix B, for a detailed derivation of Equation (24):

$$\begin{aligned}&\mathbb{P}(\gamma_t|\mathbf{y}_t, \mathbf{x}_{t,1}, \mathbf{x}_{t,2}, \boldsymbol{\Psi}, \mathbf{B}) \\ &= (2\pi)^{-\frac{1}{2}}\left(\frac{1}{1 + \mathbf{x}_{t,2}^T\mathbf{B}^T\boldsymbol{\Psi}^{-1}\mathbf{B}\mathbf{x}_{t,2}}\right)^{-\frac{1}{2}} \\ &\quad\times\exp\left(-\frac{1}{2}\frac{\left(\gamma_t - \frac{\left(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x}_{t,1}}\right)^T\boldsymbol{\Psi}^{-1}\mathbf{B}\mathbf{x}_{t,2}}{(1+\mathbf{x}_{t,2}^T\mathbf{B}^T\boldsymbol{\Psi}^{-1}\mathbf{B}\mathbf{x}_{t,2})}\right)^2}{(1 + \mathbf{x}_{t,2}^T\mathbf{B}^T\boldsymbol{\Psi}^{-1}\mathbf{B}\mathbf{x}_{t,2})^{-1}}\right).\end{aligned} \tag{24}$$

From Equation (24), it follows that,

$$\{\gamma_t|\mathbf{Y}, \mathbf{X}, \boldsymbol{\Psi}, \mathbf{B}\} \sim \mathcal{N}(m_t, v_t), \tag{25}$$

where,

$$\begin{aligned}v_t &= (1 + \mathbf{x}_{t,2}^T\mathbf{B}^T\boldsymbol{\Psi}^{-1}\mathbf{B}\mathbf{x}_{t,2})^{-1} \text{ and} \\ m_t &= v_t\left(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x}_{t,1}}\right)^T\boldsymbol{\Psi}^{-1}\mathbf{B}\mathbf{x}_{t,2}.\end{aligned} \tag{26}$$

The EM algorithm progresses by maximising the expectation of the complete data log-likelihood iteratively. The complete data log-likelihood can be seen in Equation (27):

$$l(\mathbf{A}, \mathbf{B}, \boldsymbol{\Psi}) = \log p(\mathbf{Y}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Psi}, \boldsymbol{X}, \boldsymbol{\gamma}). \tag{27}$$

This is obtained from the multivariate normal distribution of $\mathbf{y}_t$ such that $\mathbf{y}_t \sim \mathcal{MVN}(\mathbf{A}\mathbf{x}_{t,1} + \gamma_t\mathbf{B}\mathbf{x}_{t,2}, \boldsymbol{\Psi})$ from which Equation (28) follows:

$$\begin{aligned}-2l(\mathbf{A}, \mathbf{B}, \boldsymbol{\Psi}) &= np\log(2\pi) + n\log|\boldsymbol{\Psi}| \\ &\quad+ \sum_{t=1}^{n}\left(\mathbf{y}_t - \left[\mathbf{A}\mathbf{x}_{t,1} + \gamma_t\mathbf{B}\mathbf{x}_{t,2}\right]\right)^T\boldsymbol{\Psi}^{-1} \\ &\quad\times\left(\mathbf{y}_t - \left[\mathbf{A}\mathbf{x}_{t,1} + \gamma_t\mathbf{B}\mathbf{x}_{t,2}\right]\right).\end{aligned} \tag{28}$$

Given the current estimates of $(\mathbf{A}, \mathbf{B}, \boldsymbol{\Psi})$, $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}})$, with $\hat{\mathbf{e}}_t = \mathbf{y}_t - \hat{\mathbf{A}}\mathbf{x}_{t,1}$, and taking expectations one has:

$$\begin{aligned}-2\mathbb{E}\left[l(\mathbf{A}, \mathbf{B}, \boldsymbol{\Psi})|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}\right] &= np\log(2\pi) + n\log\left|\hat{\boldsymbol{\Psi}}\right| \\ &\quad+ \sum_{t=1}^{n}\mathbb{E}\left[\left(\hat{\mathbf{e}}_t - \gamma_t\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)^T\hat{\boldsymbol{\Psi}}^{-1}\right. \\ &\quad\left.\times\left(\hat{\mathbf{e}}_t - \gamma_t\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)\middle|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}\right],\end{aligned} \tag{29}$$

and noting that $m_t = \mathbb{E}[\gamma_t|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}, \mathbf{y}_t]$ and $v_t = \mathrm{Var}[\gamma_t|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}, \mathbf{y}_t]$ with $s_t = v_t^{\frac{1}{2}}$ one has:

$$\begin{aligned}&\mathbb{E}\left[\left(\hat{\mathbf{e}}_t - \gamma_t\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)^T\hat{\boldsymbol{\Psi}}^{-1}\left(\hat{\mathbf{e}}_t - \gamma\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)\middle|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}\right] \\ &\quad= \left(\hat{\mathbf{e}}_t - m_t\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)^T\hat{\boldsymbol{\Psi}}^{-1}\left(\hat{\mathbf{e}}_t - m_t\hat{\mathbf{B}}\mathbf{x}_{t,2}\right) \\ &\quad\quad+ s_t\mathbf{x}_{t,2}^T\hat{\mathbf{B}}^T\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{B}}\mathbf{x}_{t,2}s_t.\end{aligned} \tag{30}$$

With $\tilde{\mathbf{X}} \in \mathbb{R}^{2n\times(q+r)}$, the t^{th} row being $[\mathbf{x}_{t,1}^T, m_t\mathbf{x}_{t,2}^T]$, the $(n+t)^{th}$ row being $[\mathbf{0}_q^T, s_t\mathbf{x}_{t,2}^T]$, $\tilde{\mathbf{Y}} \in \mathbb{R}^{2n\times p}$ such that $\tilde{\mathbf{Y}} = [\mathbf{Y}^T, \mathbf{0}_{n\times p}^T]^T$, and $\hat{\mathbf{C}} = (\hat{\mathbf{A}}, \hat{\mathbf{B}})$, Equation (30) can be written as:

$$\begin{aligned}&\mathbb{E}\left[\left(\hat{\mathbf{e}}_t - \gamma_t\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)^T\hat{\boldsymbol{\Psi}}^{-1}\left(\hat{\mathbf{e}}_t - \gamma\hat{\mathbf{B}}\mathbf{x}_{t,2}\right)\middle|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}\right] \\ &\quad-np\log(2\pi) \\ &\quad= n\log\left|\hat{\boldsymbol{\Psi}}\right| + \mathrm{tr}\left(\left[\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\mathbf{C}}^T\right]\left[\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\mathbf{C}}^T\right]^T\hat{\boldsymbol{\Psi}}^{-1}\right).\end{aligned} \tag{31}$$

It is shown in the Supplement to: Long/Short Equity Risk Premia Parity Portfolios via Implicit Factors in

Regularized Covariance Regression, Appendix C, how to obtain:

$$(\check{\mathbf{A}}, \check{\mathbf{B}}) = \check{\mathbf{C}} = \tilde{\mathbf{Y}}^T \tilde{\mathbf{X}} (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}, \tag{32}$$

and

$$\check{\boldsymbol{\Psi}} = (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\check{\mathbf{C}}^T)(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\check{\mathbf{C}}^T)^T/n. \tag{33}$$

Under this formulation, one still needs to fit the model to each of these implicit factors, depending on the bandwidth(s) desired or the portfolio horizon(s) under investigation. Further to this point, the lag (n_2) used will have far-reaching consequences throughout the remainder of the iterative algorithm - in this work, a monthly window is used. In the following sections, the regularisation framework is introduced so as to ensure a smooth estimation of the dynamic covariance structure over time. Although covariance regression was introduced in [10] and used in the financial setting in [31], here a regularisation framework is introduced with an additional horizon-specific portfolio optimisation framework using implicit factors in a regularised lagged covariance regression and long/short equity risk parity weighting strategies.

C. REGULARISED COVARIANCE REGRESSION

In this section the framework of the covariance regression model, previously presented, is extended to develop a regularised version of the random effects formulation in which regularisation is applied iteratively within the EM algorithm. Consider the design matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{2n \times r}$, in which the t^{th} row is given by $[m_t \mathbf{x}_{t,2}^T]$, the $(n+t)^{th}$ row being $[s_t \mathbf{x}_{t,2}^T]$, $\tilde{\mathbf{Y}} \in \mathbb{R}^{2n \times p}$ such that $\tilde{\mathbf{Y}} = [[\mathbf{Y} - \mathbf{A}\mathbf{x}_{t,1}]^T, \mathbf{0}_{n \times p}^T]^T$ (note removal of 'mean' or detrending), Equation (31) can be written as:

$$\begin{aligned}\mathbb{E}\Big[\left(\hat{\mathbf{e}}_t - \gamma_t \hat{\mathbf{B}}\mathbf{x}_{t,2}\right)^T \hat{\boldsymbol{\Psi}}^{-1} \left(\hat{\mathbf{e}}_t - \gamma_t \hat{\mathbf{B}}\mathbf{x}_{t,2}\right) \Big| \hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}\Big] - np\log(2\pi) \\ = n\log|\hat{\boldsymbol{\Psi}}| + \mathrm{tr}\Big(\left[\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\mathbf{B}}^T\right]\left[\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\mathbf{B}}^T\right]^T \hat{\boldsymbol{\Psi}}^{-1}\Big).\end{aligned} \tag{34}$$

With this formulation, one can calculate, or rather, define, the trend to suit one's requirements. This formulation has been first described herein and named Independent Mean Regularised Covariance Regression (IM-RCR). One can proceed as before by appropriately adjusting Equation (32) such that:

$$\check{\mathbf{B}} = \tilde{\mathbf{Y}}^T \tilde{\mathbf{X}} (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}, \tag{35}$$

and by adjusting Equation (33) such that:

$$\check{\boldsymbol{\Psi}} = (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\check{\mathbf{B}}^T)(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\check{\mathbf{B}}^T)^T/n, \tag{36}$$

or $\hat{\mathbf{B}}$ can be estimated in a number of regularised means. Some of these regularised estimation techniques are described in the following sections. The key difference between covariance regression and RCR is the adjustment of the maximisation step in the EM algorithm.

1) RIDGE REGRESSION WITHIN EM

One can optimise $\boldsymbol{B}$ using Equation (37) which represents the regularised MSE (RegMSE):

$$\begin{aligned}\mathrm{RegMSE}(\boldsymbol{B}|\tilde{\mathbf{Y}}) &= (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T)^T(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T) + \lambda_2^2||\mathbf{B}||_2^2 \\ &= ||\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T + \lambda_2\mathbf{B}||_2^2,\end{aligned} \tag{37}$$

with λ_2 being the Ridge regularisation parameter, from which it follows:

$$\check{\mathbf{B}}_{\lambda_2} = \tilde{\mathbf{Y}}^T \tilde{\mathbf{X}} (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + \lambda_2 \mathbf{I})^{-1}, \tag{38}$$

and

$$\check{\boldsymbol{\Psi}}_{\lambda_2} = (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\check{\mathbf{B}}_{\lambda_2}^T)(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\check{\mathbf{B}}_{\lambda_2}^T)^T/n. \tag{39}$$

In this framework, the original covariance regression model from Equation (35) reduces to $\check{\mathbf{B}}_{\lambda_2=0}$. This technique has been named Ridge Regularised Covariance Regression (L2-RCR). The addition of the ridge penalty provides a bias variance trade-off that can improve performance of the covariance forecasting models forecasting and improve generalisation. However, it introduces a bias at each stage of the EM algorithm, as discussed in the following section.

2) HOMOGENEOUS AND HETEROGENEOUS RIDGE REGRESSION

The above description of Ridge regression is a member of the homogeneous local ridge regression class of regression models - this class of models uses the same penalty λ_2, throughout the iterative algorithm. There are numerous instances where a higher flexibility of this penalty term would show numerous structures previously unnoticed, as well as provide invaluable insight into the gradual sequential regularisation of components. A heterogeneous local Ridge regression model is described in the following section, with this family of models allowing appropriately decaying functions to redirect the solution space allowing for further flexibility and better resolution capabilities. In Section II-C3 an exponentially decaying penalty is introduced, but this is by no means exhaustive of all potential decaying functions.

3) EXPONENTIALLY DECAYING BIAS IN RIDGE REGRESSION

In this section is considered an approach that applies a progressive refinement of the regularisation in which the estimation starts with a stronger regularisation and greater bias in order to stabilise the initial solutions to the M-Step of the EM algorithm. After the iterations of the EM algorithm proceed, a progressive relaxation of the regularisation strength in the M-Step of the EM algorithm is performed. The decay rate of the regularisation can follow any monotonically decreasing function, we have opted to illustrate the method with the following functional decay in the regularisation parameter:

$$\lambda_j = \lambda e^{-j}. \tag{40}$$

This form of regularisation is found to produce additional stability in the EM iterative solution achieving more accurate

estimation in practise along with greater stability in the solution of the global optimisation objective of the EM method, and ultimately this in turn generally leads such a method to require fewer iterative steps in the EM algorithm. Furthermore, such regularisation can be particularly useful in higher-dimensional settings, i.e. large covariance matrix regressions with many covariates (IFs). To demonstrate this, let:

$$f(\beta_1, \beta_2; t) = \beta_1 x_1(t) + \beta_2 x_2(t) = \tilde{\mathbf{X}}_i \check{\mathbf{B}}, \tag{41}$$

with,

$$||\tilde{\mathbf{Y}} - f(\beta_1, \beta_2; t)||_2^2 = s^2 \tag{42}$$

it follows from Equation (37) that,

$$\sqrt{\beta_1^2 + \beta_2^2} \leq \frac{s}{\lambda_j}. \tag{43}$$

From this one can infer that the Ridge penalty or bias parameter, λ_2, functions as a limit on the parameters values with unique solutions at each penalisation level. If one were to use Equation (40), one can observe an example of these various Ridge regression solutions in Figure 5.

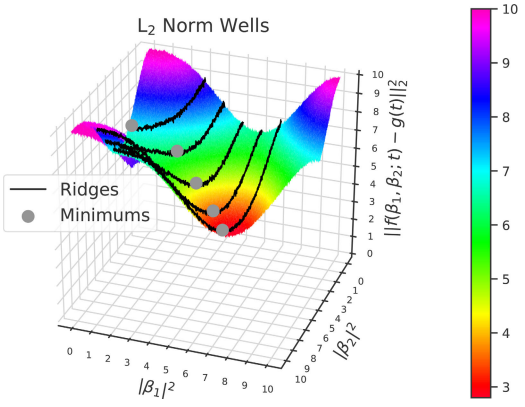

FIGURE 5. Exponentially decaying bias in ridge regression example.

4) REGULARISATION VERSUS SELECTION

Ridge regression, as detailed in Section II-C1, is a regularisation technique, whereas in Section II-C3 Ridge regression takes on the properties of the variable selection model, as can be seen in the example demonstrated in Figure 5. More complex methods that function as both regularisers and variable selectors in their vanilla forms are the least absolute shrinkage and selection operator (LASSO) regression as originally proposed in [74] and formalised in [75] and the least angle regression (LARS) as proposed in [76]. All of the further techniques discussed within this section are homogeneous in nature in that the penalisation parameter is held constant between iterations of the RCR algorithm. These further RCR extensions can easily be extended to heterogeneous regression by choosing an appropriately decaying monotonous function to decease the penalisation factor iteratively.

5) LASSO REGRESSION

By applying LASSO regression, first formalised in [75], at the M-step of the EM algorithm one now has LASSO Regularised Covariance Regression (LASSO-RCR) or L1-RCR. $\mathbf{B}$ can now be optimised iteratively in the M-step using Equation (44):

$$\text{RegMSE}\big(\mathbf{B}\big|\tilde{\mathbf{Y}}\big) = \big(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T\big)^T\big(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T\big) + \lambda_1||\mathbf{B}||_1, \tag{44}$$

with λ_1 being the LASSO or L1 regularisation parameter. In [77] LASSO regression and Ridge regression penalties are combined linearly to construct Elastic Net regression as an early comprise between Ridge or L2 regularisation and LASSO or L1 variable selection and is discussed in the following section.

6) ELASTIC NET REGRESSION

By applying Elastic Net regression, proposed in [77], to the M-step of the EM algorithm, one can optimise $\mathbf{B}$ using Equation (45):

$$\begin{aligned}\text{RegMSE}(\mathbf{B}|\tilde{\mathbf{Y}}) = (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T)^T(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T) \\ + \lambda_2||\mathbf{B}||_2^2 + \lambda_1||\mathbf{B}||_1,\end{aligned} \tag{45}$$

with Elastic Net regression reverting to Ridge regression when $\lambda_1 = 0$ and LASSO regression when $\lambda_2 = 0$. A common redefinition of Equation (45) allows a heterogeneous regression approach by varying the tuning parameter, $\alpha \in [1, 0]$, in:

$$\begin{aligned}\text{RegMSE}(\mathbf{B}|\tilde{\mathbf{Y}}) = (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T)^T(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T) \\ + \lambda_2(1-\alpha)||\mathbf{B}||_2^2 + \lambda_1\alpha||\mathbf{B}||_1.\end{aligned} \tag{46}$$

7) SUBGRADIENT DESCENT FOR LASSO AND ELASTIC NET

In the cases, listed above, in which the RegMSE objective function, in the M-Step of the EM algorithm in the covariance regression estimation, contains an L_1 penalty, then the standard gradient based optimization methods cannot be applied. Instead one may adopt a sub-gradient solution as follows:

$$\begin{aligned}&\text{minimize } ||\mathbf{B}||_1 \\ &\text{subject to } \tilde{\mathbf{X}}\mathbf{B}^T = \tilde{\mathbf{Y}},\end{aligned} \tag{47}$$

with $\mathbf{B}_1$ being initialised from solving Equation (37), the next iteration is calculated using:

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \alpha_k\Big(\mathbf{I}_p - \tilde{\mathbf{X}}^T\big(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\big)^{-1}\tilde{\mathbf{X}}\,\text{sign}\big(\mathbf{B}_k\big)\Big), \tag{48}$$

where α_k is the k^{th} step size, $\mathbf{I}_p \in \mathbb{R}^{p\times p}$ is the identity matrix, and sign$(\cdot)$ is defined as:

$$\text{sign}(x) = \left\{\begin{array}{r} +1, \quad \text{for } x > 0 \\ 0, \quad \text{for } x = 0 \\ -1, \quad \text{for } x < 0 \end{array}\right\}. \tag{49}$$

Numerous options are available for the step size, α_k. In this paper, a Polyak step size will be used will be used which is a

member of the family of square summable, but not summable step sizes that satisfy the following conditions:

$$\alpha_k > 0 \,\forall\, k > 0, \sum_{i=0}^{\infty} \alpha_k^2 < \infty, \text{ and } \sum_{i=0}^{\infty} \alpha_k = \infty, \quad (50)$$

for which convergence proofs exist - see [78]. An example used here is $\alpha_k = \frac{10^{-8}}{k}$.

III. PORTFOLIO CONSTRUCTION

In this work there are numerous simplifying assumptions about the temporal and practical problems that arise with portfolios rebalancing. In Figure 25 an unrealisable S&P 500 proxy is included as a general measure of market performance which partly addressed the simplifying assumptions of zero transaction costs. This proxy is rebalanced daily based on all 505 constituent stocks' market capitalisations with another proxy being included that assumes a daily rebalancing charge of 0.01% of the total value of the portfolio (the orange line in Figure 25) that demonstrates how quickly the inclusion of previously discarded simplifying constraints can radically alter the results of case studies demonstrating proposed portfolio optimisation techniques. In this work a further assumption of exclusively monthly (cost-free) rebalancing is made.

Several of these rather restrictive simplifying assumptions are loosened by [79] in the review of multi-period portfolio reweighting with constraints and transaction costs. [79] relaxes the restriction of regularly spaced (though monthly is not uniformly spaced) reweighting instances and generalises it to N reweighting intervals with a finite horizon. Further, the ultimate goal of the portfolio optimisation in [79] is to minimise the mean square deviation of the portfolio's return relative to some desired portfolio return. Additional constraints on portfolio rebalancing also considered in [79] are linearly increasing transaction costs and restrictions on shorting. In defence of the relaxation of the transaction costs in modelling portfolio optimisation strategies, [79] finds that despite the inclusion of transaction costs, the resulting suboptimal trading strategy very closely tracks the optimal portfolio with no transaction costs.

Among the first to extend portfolio optimisation to multiple periods was [80] - an apt analogy is made in [80] that relates the problem of investing over n horizons to taking a $\frac{1}{n}^{th}$ interest or risk in n independent shipping voyages. Reference [80] is also, possibly, the first to consider transaction costs in their portfolio optimisation. Reference [80] notes that if market imperfections cause loans to be costly, then investing in volatile stocks to achieve the same desired level of returns is rational.

The assumptions that all trading opportunities are available at all times and that they can be enacted costlessly are criticised by [81] as leading to unrealistic behaviour. Reference [81] finds that when transaction cost are included in models that investors will make trades at randomly spaced intervals (motivated by their knowledge) as opposed to continuously or uniformly which mimics real-world behaviour of investors. Further, [82] notes that in the absence of transactions costs that an investor's decisions in a single investment period with a multiple period horizon are indistinguishable from an investor's decisions in the same period over a single horizon with a convex utility function.

Now, considering the maturity-horizon(s) of interest to investors, one must select the implicit factors within a specific frequency band to forecast the covariance over the coming statically weighted forecast period. One could, for example, exclude all structures of higher frequency than monthly if one were reweighting monthly - this is one such contribution in this work. This dynamic framework affords the user choices such as the frequency of the implicit factors, maturity horizons, and leverage constraints of the investor.

Note that throughout this work, the dynamic time series implicit factor EMD-RCR framework will be used to produce the covariance dynamically over time that is used in the portfolio construction. In this section, the risk-parity portfolio framework will be described. To achieve this, denote the forecast portfolio covariance by $\mathbf{\Sigma}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}} \in \mathbb{R}^{p\times p}$ with the required restrictions such as symmetry and positive-definiteness and the weighting vector being $\boldsymbol{\omega} \in \mathbb{R}^{p\times 1}$, the portfolio standard deviation or risk (σ_P) is denoted:

$$\sigma_P = \sqrt{\boldsymbol{\omega}^T \mathbf{\Sigma}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}} \boldsymbol{\omega}}. \quad (51)$$

In practise, arbitrary long- and short-term positions are not feasible as this assumes incorrectly that a facility for unlimited borrowing is available. A commonly used long/short equity weighting limit is 130/30 where:

$$\sum_{\omega>0} \omega \leq 1.30 \text{ and } \sum_{\omega<0} \omega \geq -0.30, \quad (52)$$

must be satisfied in addition to the standard unity summation restriction. The risk contribution from each asset j, σ_j, to the total correlative risk of the portfolio, can be calculated as:

$$\sigma_j = w_j \frac{\partial \sigma_P}{\partial w_j} = w_j \frac{(\mathbf{\Sigma}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}} \boldsymbol{\omega})_j}{\sigma_P}. \quad (53)$$

The relative risk contribution of each asset, expressed as a fraction of the entire portfolio, is then calculated:

$$\frac{\sigma_j}{\sigma_P} = w_j \frac{(\mathbf{\Sigma}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}} \boldsymbol{\omega})_j}{\sigma_P^2} \quad (54)$$

Finally, with a given rule for the risk allocation - that is, how to construct the risk allocation vector ($\mathbf{b} = [b_1, \ldots, b_p]^T$) - a common rule being equal risk allocation - that is, $b_j = \frac{1}{p}\ \forall\, j \in \{1, \ldots, p\}$, one can calculate the portfolio weighting:

$$\boldsymbol{\omega}^* = \text{argmin} \sum_{j=1}^{p} \left(w_j \frac{(\mathbf{\Sigma}_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}} \boldsymbol{\omega})_j}{\sigma_P^2} - b_j \right). \quad (55)$$

The equal-risk premia parity weighting strategy is the most commonly used, but others could also be beneficial, such as weighting risk according to market capitalisations of the comprising assets.

IV. REAL-WORLD CASE STUDY: S&P500

In this section the practical aspects of constructing a case study to illustrate the portfolio methodology developed will be considered. First, Section IV-A provides an outline of the construction of the market cap weighted indices that will form the regression response variables of interest in this case study and will furthermore be used to construct the decomposed Implicit Factors used as lagged time series regression factors. These signals will be based on sectors of the S&P500 constituent companies. Then in Section IV-C an explanation of how these IMFs (intrinsic mode functions - implicit market factors) will be used in the covariance regression setting.

The implicit regression factors extracted from these sector indices for this case study will be based on IMFs extracted using EMD. The IMFs extracted from each of the eleven market indices are sorted using EMD-MDLP (first introduced in [8]). The Minimum Description Length Principle (discussed further in Section IV-C1) will be used to sort the decomposed components into qualitative groups representing a specific time scale (frequency bandwidth spectral groups) for the financial information they implicitly characterise. These will be used to assess which frequency components are best for estimating the forthcoming covariance for monthly portfolio rebalancing. This sets an important benchmark, as monthly portfolio rebalancing is tested based on different bandwidths of risk. Independent Mean RCR (IM-RCR) will be used to regress the covariance of the returns against the various extracted factors.

A. COVARIATE EXTRACTION: IMF INDEX FEATURES

The S&P 500 includes 505 large market cap stocks that serve as a proxy for the general state of the United States of America (and the global) economy. These 505 stocks are separated into eleven sectors (based on the relative market capitalisation of the assets within each of the sectors), namely: Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, and Utilities. Each index consists of all stocks in the related sector weighted by the market cap of the individual asset within the sector. Five years of daily observations are used from 31 December 2016 to 31 December 2021 in this case study.

Two notable features can be discerned from Figure 6. There is a significant drop in most of the indices in the last quarter of 2018. This is attributable to a number of factors such as the President of the United States imposing tariffs on industrial materials, interest rate increases leading to fears of a recession, the largest information technology stocks being under increased scrutiny, and inflated stock prices at the time being driven by tax cuts imposed by the President of the United States. The second notable feature is the significant and immediate drop in all indices at the beginning of 2020 as a result of the SARS-CoV-2 pandemic.

These two notable periods of marked market depreciation of differing magnitudes and intensities are of particular interest to this study, as two events occurred in the world resulting in significant increased positive correlation and depreciation amongst the assets. These two studies, however, are not identical because of the longer less-certain downturn period owing to a number of contributory economic factors and the correlation structure being different compared to the accelerated correlation coupling and increased rate of downturn. RCR allows for the study of these two different regions of significant market downturn and dynamic covariance.

With this framework, one can respond to either market-shock event by using selected implicit factors within specific bandwidths based on the forecasted period of downturn. In the following section, the realised covariances for the periods from 2 October 2018 until 23 December 2018 (extended low-intensity period of depreciation - 82 days) and 28 February 2020 until 1 April 2020 (short high-intensity period of depreciation - 32 days) will be calculated and compared against the forecasted covariance using selected implicit factors using an appropriate metric.

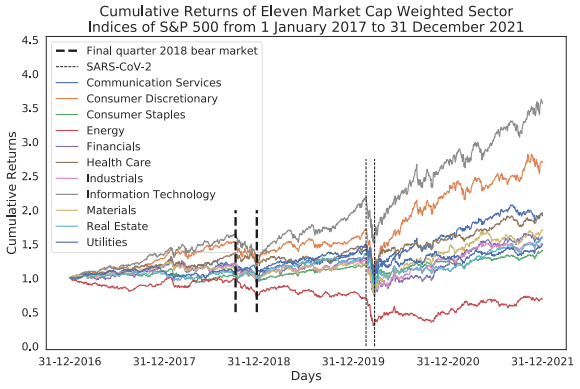

FIGURE 6. Eleven sector indices weighted by market cap from the S&P 500 with two periods of marked market depreciation of varying durations noted for comparison with Figure 19.

B. CORRELATION COUPLING

This section of the analysis will outline time periods of relevance in the S&P500 markets that one can explore as challenging times for studying the covariance regression based EMD-RCR framework. Challenges include rapid and slow market drawdowns, as well as rapid versus slow market appreciations. Such time-period selections will allow one to assess how well the EMD-RCR methodology captures the estimated correlation structures during periods of varying lengths of marked market depreciation.

Standard practise is to use the realised covariance over a fixed window, and it is assumed that this covariance is stationary over the forecast horizon - it is assumed to be static over the forecast period. The covariance regression model allows dynamic covariance regression using implicit factors; daily covariance can be forecast. In Figures 7, 8, 9, 10, 11, and 12 the covariance realised between the eleven

indices changes dramatically and almost instantaneously during periods of increased financial stress. In Figures 7, 8, and 9 a window of 82 days is used for consistency with the approximate duration of the first market downturn. In the second set of figures, namely Figures 10, 11, and 12 a window of 32 days is used to capture the compressed nature of the second market downturn, namely the SARS-CoV-2 pandemic.

The first period of market downturn is taken as being (for this study) from 2 October 2018 until 27 December 2018 (82 days), and the second period of market downturn is taken as being (for this study) from 28 February 2020 to 1 April 2020 (32 days). These windows are used as they capture the entire periods of near-continuous market depreciation. The realised covariance calculated in Figures 7, 8, and 9 uses the 82-day window, while in Figures 10, 11, and 12 the realised covariance is calculated using the 32-day window. For comparison, the best-fit model (Kullback–Leibler divergence or 'relative entropy') for each period is in this article, with numerous other factors and further analysis is provided in the supplement.

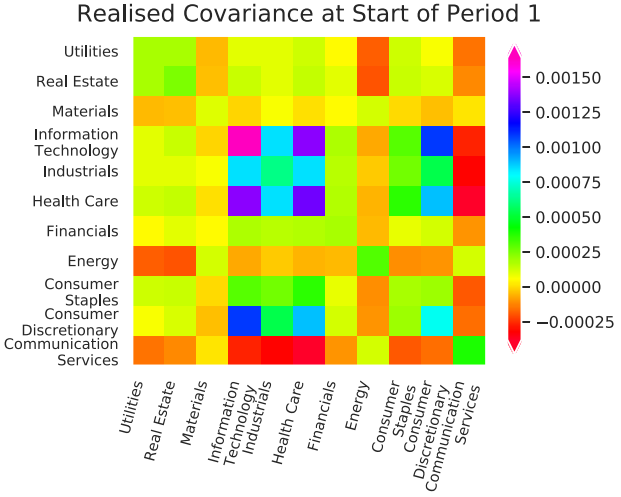

FIGURE 7. Realised covariance of 11 indices over 82-day window before first recession.

Direct comparisons should be made between Figures 8 and 13, and Figures 11 and 16, respectively. The Kullback-Leibler (KL) divergence between Figure 7 and 8 is: $D_{KL}^{real}(\mathbf{\Sigma}||\mathbf{\Sigma}_x) = 132.69$, while the KL divergence between Figure 13 (covariance forecast using the high frequency utilities component extracted using EMD) and Figure 8 is: $D_{KL}^{U,IMF}(\mathbf{\Sigma}||\mathbf{\Sigma}_x) = 119.50$. This is a significant (approximately 10%) decrease in 'distance' between the objects in this PSD hyper-cone space when compared against stationary model assumption where the previous covariance is taken as the best stationary estimate.

To a further point, this estimate is not only a significant improvement over standard practise, but also splits the covariance into a baseline unattributable covariance in Figure 14 and the attributable covariance in Figure 15. The attributable covariance in Figure 15 shows interesting and possibly unexpected correlation structures. It is revealed

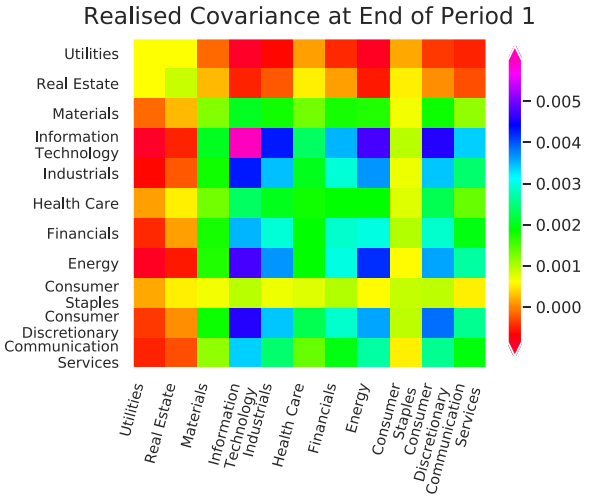

FIGURE 8. Realised covariance of 11 indices over 82-day window during first recession.

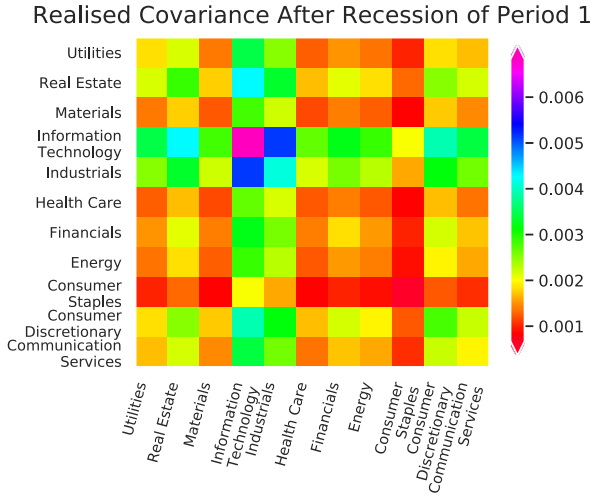

FIGURE 9. Realised covariance of 11 indices over 82-day window after first recession.

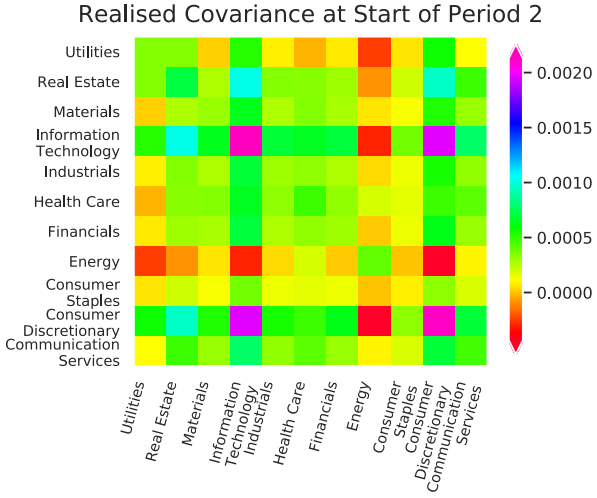

FIGURE 10. Realised covariance of 11 indices over 32-day window before second recession.

that structures in the utilities factor leading up to the marked market downturn have contributed to relatively large correlations between materials and information technology

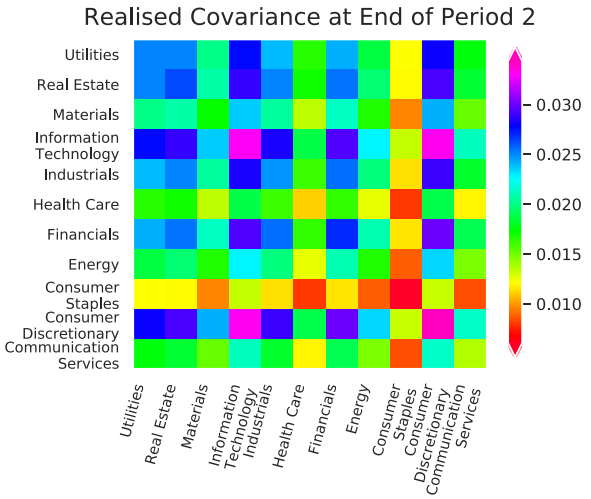

FIGURE 11. Realised covariance of 11 indices over 32-day window during second recession.

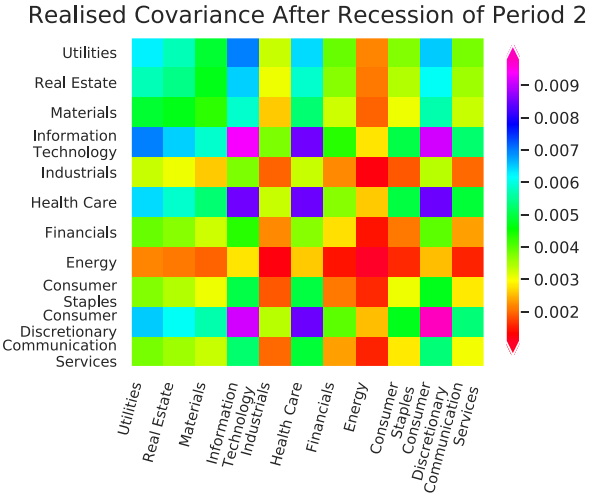

FIGURE 12. Realised covariance of 11 indices over 32-day window after second recession.

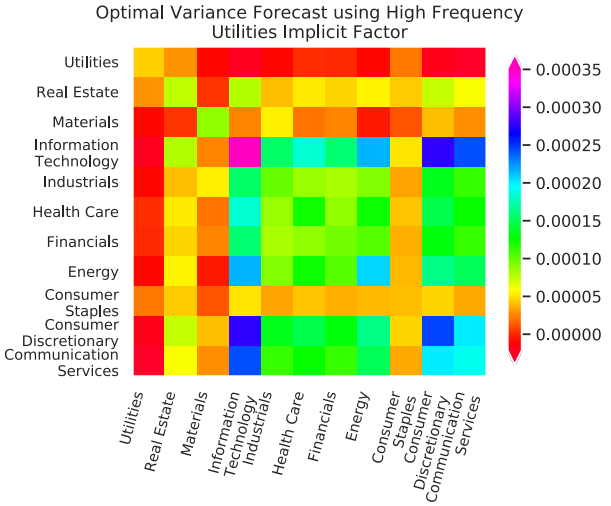

FIGURE 13. Best fitting model S&P500 portfolio covariance forecast ($\Sigma_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}}$) using 82-day lagged window and high frequency Utilities implicit factor.

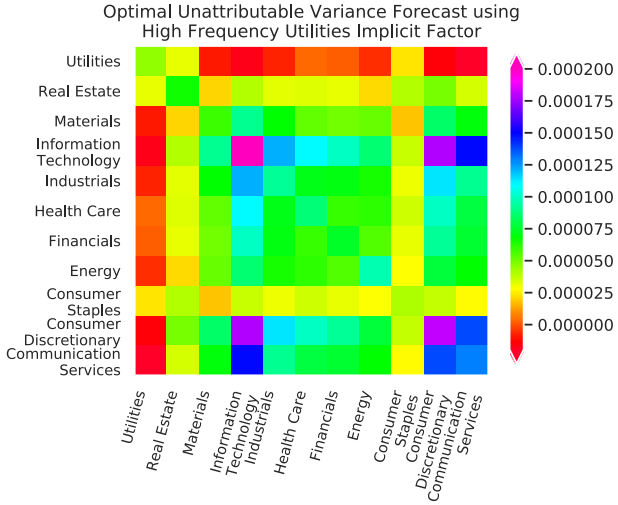

FIGURE 14. Best fitting model S&P500 portfolio unattributable covariance forecast ($\Psi_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}}$) using 82-day lagged window and high frequency Utilities implicit factor.

and other sectors with almost no attributable correlation between utilities and other sectors, which is expected during periods of stress - utilities are normally uncoupled from the others.

In Figure 18, similar inferences can be made. In Figure 18 it appears that the structures present in the low-frequency and high-frequency Information Technology sector leading up to the SARS-CoV-2 pandemic had predictive capabilities for the severe and accelerated economic downturn caused by the pandemic. Again, Information Technology and Materials sectors appear to be most affected by these structures, with all sectors being affected with a notable exception being utilities that, again, perform quite independently of the other sectors.

One has flexibility in the choice of the window length for measuring realised covariance amongst a group of variables. There is no ideal one-size-fits-all window in finance with weekly, monthly, quarterly, etc. being used for convenience - the model presented in this work attempts to formalise an instantaneous correlation structure which is window-independent and does not exist in the traditional covariance and correlation framework. The figures demonstrating the covariance structures of the eleven market indices before, during, and after the two stressed market events demonstrate the dynamics of covariance during periods of stress, covariance and correlation are usually presented as fixed or, at the very most, varies slightly and slowly.

In Figure 19 two notable downward momentum market draw down periods for the S&P 500 over the past four years. It is directly observable that during these periods one can see a sudden and strong positive coupling in correlation relationships between sectors. In the last quarter of 2018 a number of contributing factors led to a relatively prolonged period of marked market depreciation, whereas the short but extreme period of marked market depreciation in the first quarter of 2020 is solely attributable to the SARS-CoV-2 pandemic. A third notable feature, not explicitly highlighted

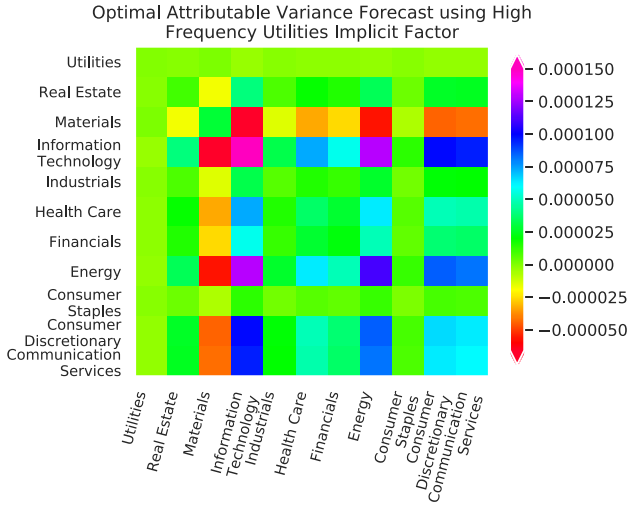

FIGURE 15. Best fitting model S&P500 portfolio attributable covariance forecast ($B\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T B^T$) using 82-day lagged window and high frequency Utilities implicit factor.

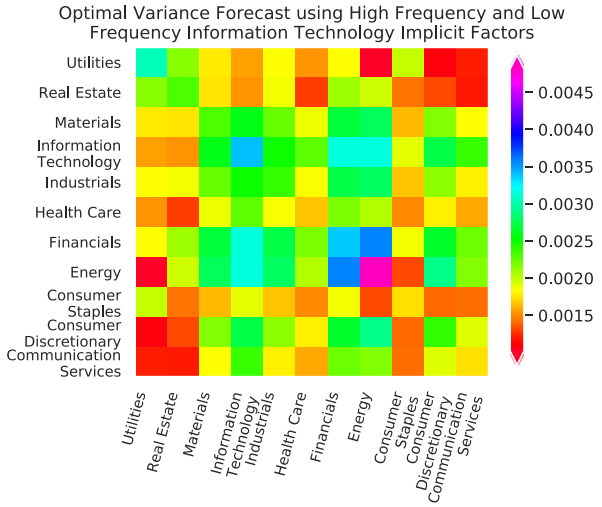

FIGURE 16. Best fitting model S&P500 portfolio covariance forecast ($\Sigma_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}}$) using 32-day lagged window and high and low frequency Information Technology IMFs.

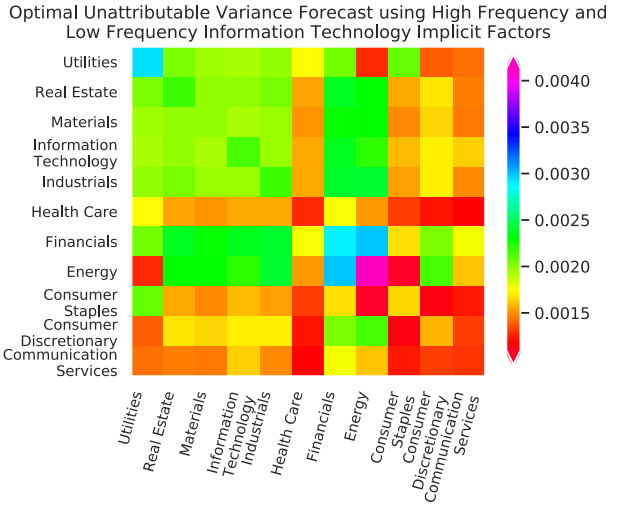

FIGURE 17. Best fitting model S&P500 portfolio unattributable covariance forecast ($\Psi_{\mathbf{y}_t|\mathbf{x}_{t,1},\mathbf{x}_{t,2}}$) using 32-day lagged window and high and low frequency Information Technology IMFs.

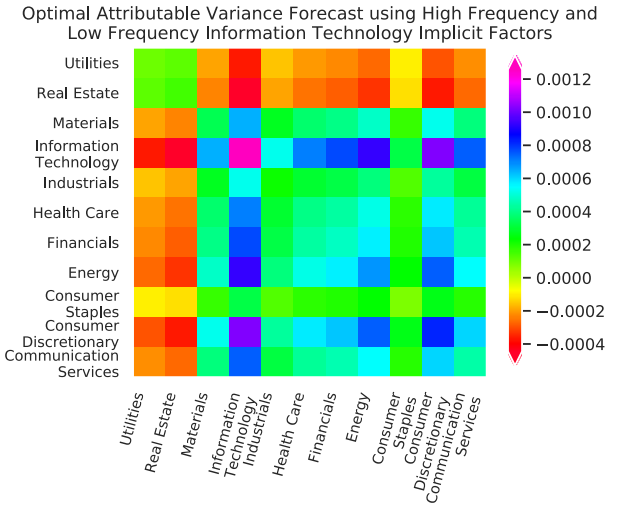

FIGURE 18. Best fitting model S&P500 portfolio attributable covariance forecast ($B\mathbf{x}_{t,2}\mathbf{x}_{t,2}^T B^T$) using 32-day lagged window and high and low frequency Information Technology IMFs.

in Figure 19, can be seen at the end of 2020 and at the beginning of 2021. This is as a result of a reflexively and stimulus-cheque driven period of marked market appreciation as a result of the extreme government intervention in the financial market.

Figures 14 and 15 are the decomposition of Figure 13 with Figures 17 and 18 being the decomposition of Figure 16 according to Equation (17). These are the best-fitting models during their respective periods of marked economic downturn, with further analysis using other components that showed significant forecasting abilities being in the supplement along with all KL divergence scores compared against stationarity assumption benchmark. From Figure 19 it is observable that the low and mid-frequency structures would be unsuitable during periods of economic stress owing to the rapid change in correlation structures - these would be more suited to periods of relative stability.

The above forecasts, Figures 13 and 16, of the periods of marked market downturns or recessions (depending on your definition), Figures 8 and 11, respectively, both show the time it takes for the lagged-effects covariance regression to respond to the in-sample covariance. Sudden, rapid shocks, external (in some sense) to the data (such as SARS-CoV-2) cause the model to take time to adjust. Further, EMD-RCR allows the separation of the frequency structures for forecasting and decision-making, such as passive managers focusing on the mid- to low-frequency components with more active managers (higher frequency trading) focusing on the highest frequency components.

C. COVARIANCE REGRESSION

With $\mathbf{X} \in \mathbb{R}^{T_X \times m}$ being the m covariates (of variable dimension owing to IMFs produced by sifting algorithm)

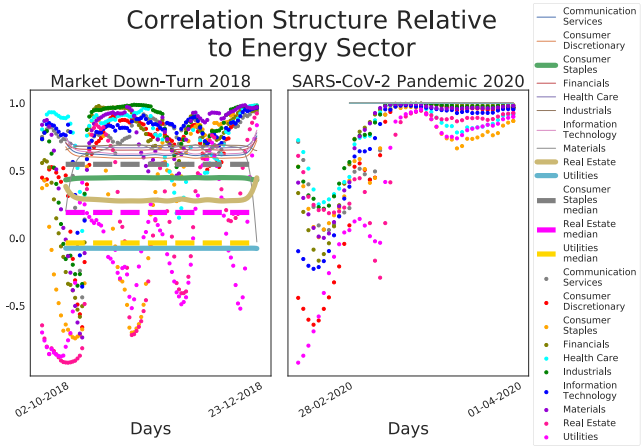

FIGURE 19. Forecasted instantaneous correlation structures versus 20-day rolling window of realised correlation of ten sector indices versus Energy weighted by market cap from the S&P 500.

over the previous 11 months (T_X of variable day length) and $\mathbf{Y} \in \mathbb{R}^{T_Y \times 11}$ being the log-returns of the 11 indices over the most recent 11 months (T_Y of variable day length), one can construct the optimisation problem as in Equation (34):

$$\begin{aligned}\mathbb{E}[(\mathbf{e}_t - \gamma_t \mathbf{B}\mathbf{x}_{t,2})^T \boldsymbol{\Psi}^{-1}(\mathbf{e}_t - \gamma_t \mathbf{B}\mathbf{x}_{t,2})|\hat{\mathbf{B}}, \hat{\boldsymbol{\Psi}}] - n p \log(2\pi) \\ = n \log|\boldsymbol{\Psi}| + \text{tr}([\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T][\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{B}^T]^T \boldsymbol{\Psi}^{-1}),\end{aligned} \tag{56}$$

with $\mathbf{e}_t$ being the error term vector at increment t, γ_t being the random effects variable, $\mathbf{B}$ the parameter matrix to be estimated, $\mathbf{x}_{t,2} \in \mathbb{R}^{1 \times m}$, and $\boldsymbol{\Psi}$ being the base or unattributable covariance structure. As T_X and T_Y may be different owing to the variable month lengths of throughout the year, a standard T is calculated such that:

$$T = \min(T_X, T_Y). \tag{57}$$

The intervals used for fitting and forecasting, respectively, are demonstrated more clearly in Figure 20. In this framework, there is a one month delay between the underlying structures in the price process and the resulting change in the covariance structure. There may also be a difference in length between the IMFs period used to for forecasting and the covariance of returns to be forecast. This is not as relevant as the fitting period discrepancy as the median of the forecast covariance is taken to calculate the portfolio weighting for the forthcoming month. In Figure 20, without loss of generality, the interval truncation demonstration is made assuming that $T_X > T_Y$.

1) MINIMUM DESCRIPTION LENGTH BINNING AND EMD

First proposed in [8], this technique is now used with modifications designed to cope with discrepancies often found in real world data. Unlike in [8], when dealing with real-world data, as opposed to well-behaved and easily separable synthetic data used first for the creation of this method, often the entire set of data is not easily classifiable using MDLP. An example of t_{31} of the first monthly iteration of the algorithm is shown in Figure 21 with two cut points. In this setting with potential cut-points, T_k, and the corresponding partitions of set $\boldsymbol{S}$ such that cut-point

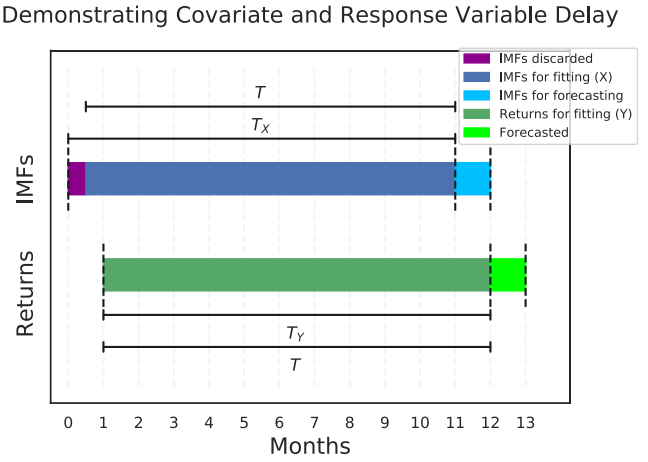

FIGURE 20. Plot demonstrating the delay between the covariates (IMFs or SSA trends extracted from price processes) and response variable (returns) and how period discrepancies are handled.

T_k partitions $\boldsymbol{S}$ into dichotomous subsets $\boldsymbol{S}_1^{T_k}$ and $\boldsymbol{S}_2^{T_k}$, the optimisation function is as follows:

$$\min_{T_k}(E(T_k; \boldsymbol{S})), \tag{58}$$

with $E(T_k; \boldsymbol{S})$ being the entropy or class information entropy that is induced by the partitioning by T_k of $\boldsymbol{S}$ such that:

$$E(T_k; \boldsymbol{S}) = \frac{|\boldsymbol{S}_1^{T_k}|}{|\boldsymbol{S}|}\text{Ent}(\boldsymbol{S}_1^{T_k}) + \frac{|\boldsymbol{S}_2^{T_k}|}{|\boldsymbol{S}|}\text{Ent}(\boldsymbol{S}_2^{T_k}), \tag{59}$$

with,

$$\text{Ent}(\boldsymbol{S}_j^{T_k}) = -\sum_{i=1}^{m} P(C_i, \boldsymbol{S}_j^{T_k}) \log(P(C_i, \boldsymbol{S}_j^{T_k})), \tag{60}$$

with $P(C_i, \boldsymbol{S}_j^{T_k}) = \frac{|\{C_i\}|}{\{\boldsymbol{S}_j^{T_k}\}}$, $|\{\cdot\}|$ being the cardinality measure, and with C_i being the class of objects that displays feature i within $\boldsymbol{S}_j^{T_k}$.

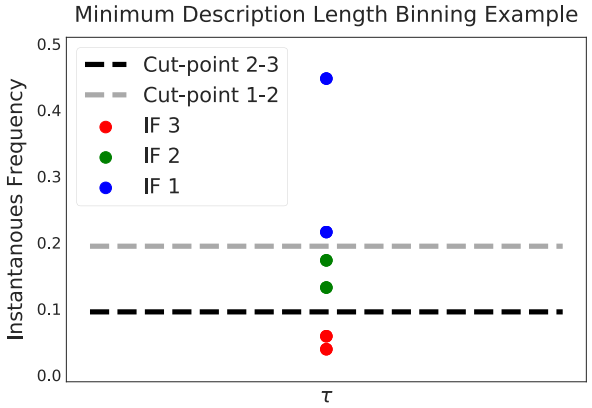

FIGURE 21. Plot demonstrating how an MDLP algorithm is applied iteratively.

D. LONG/SHORT EQUITY RISK PARITY PORTFOLIO

Once the covariance has been forecast for the month ahead, the median (more robust measure than mean) of the covariance is calculated. With the median of the forecast

covariance being denoted by $\boldsymbol{\Sigma}_{med}$, the long/short equity risk parity portfolio is calculated as:

$$\tilde{\boldsymbol{\omega}}_{med} = \text{argmin} \sum_{j=1}^{N} \left(\omega_j \frac{(\Sigma_{med}\boldsymbol{\omega})_j}{\boldsymbol{\omega}^T \boldsymbol{\Sigma}_{med} \boldsymbol{\omega}} - b_j \right), \tag{61}$$

$$\text{subject to,} \quad \sum_{j=1}^{N} \omega_j = 1, \text{ and} \tag{62}$$

$$\sum_{j=1}^{N} \omega_j \mathbb{I}_{\{\omega_j \geq 0\}} \geq k_{long} \text{ for } k_{long} \geq 1, \text{ and}$$

$$\sum_{j=1}^{N} \omega_j \mathbb{I}_{\{\omega_j \leq 0\}} \geq k_{short} \text{ for } k_{short} \leq 0, \tag{63}$$

with $k_{long} = 130\%$ and $k_{short} = -30\%$ being used throughout this paper as both a risk moderating technique as well as a real-world limitation proxy. If $k_{short} = 0$, this restricts the investor to only long positions. While the criticism of a long-only market goes back to at least [83], the formalisation of the long/short equity investment strategy with short summation limits can be attributed to [84], [85], and [86].

V. RESULTS

This section begins with Section V-A which outlines several benchmark portfolios and benchmark techniques for portfolio construction, against which the performance of the regularised covariance regression risk parity portfolios will be measured. In this section, five benchmark portfolios are introduced, namely: realised covariance risk parity portfolio, DCC-MGARCH risk parity portfolio, market cap weighting portfolio (as a proxy for the S&P 500), global minimum variance portfolio, and a principal component portfolio using three components.

The performance of these five benchmark portfolios will be measured using several well-known portfolio performance measures that are outlined in Section V-C. Classical measures such as cumulative returns (Section V-C1), mean returns (Section V-C2), and variance of returns or risk (Section V-C3 are discussed first. As risk parity portfolios are constructed to equally distribute risk, some more complex performance measures are required such as Value-at-Risk (Section V-C4), Maximum Drawdown (Section V-C5), Omega Ratio (Section V-C6), Sortino Ratio (Section V-C7) and Sharpe Ratio (Section V-C8).

A. BENCHMARKS

These benchmarks serve as viable and practised alternatives to the portfolio construction methods under investigation in the paper. The performance of these portfolios will be measured using the measures below.

1) REALISED COVARIANCE

This is the most common industry practise where the best estimate for the forthcoming portfolio balancing period's covariance is the current realised covariance structure looking backwards an agreed upon period. In the current paradigm, with this given covariance assumption looking ahead, one should use the same portfolio weighting technique, namely risk-parity portfolio weighting.

2) MULTIVARIATE GARCH

Multivariate GARCH models or MGARCH models are thoroughly reviewed in [30]. At their core, the various models try to find a concise, viable, and natural extension of univariate GARCH models. The most complex family of these models, the nonlinear combinations of univariate GARCH, includes the DCC-GARCH model where, as can be seen in the supplementary materials, the forecasted covariance simply becomes some bases forecasted covariance multiplied by the square root of the number of days into the future one is forecasting the variance. This very much mimics the uncertainty present in the Brownian motion. As a result, the median of the covariance structure forecasted over the next month will not differ significantly from the realised covariance above. Close monitoring of these two portfolios shows this.

3) MARKET CAPITILISATION WEIGHTING

This is a natural benchmark that functions as a proxy for the actual S&P 500. This is equivalent to weighting every one of the 505 assets by their individual market caps - the same result is obtained by weighting the eleven sector indices by their respective cumulative market caps.

4) GLOBAL MINIMUM VARIANCE

This portfolio strategy is a well-established means of minimising portfolio variance at the cost of returns. In the absence of any weighting restrictions (besides of course that the weights sum to 1), the weight vector that minimizes the variance of the portfolio is:

$$\boldsymbol{\omega}_{\text{MV}} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{64}$$

This solution is in the absence of any short-selling restrictions. To disregard short-selling entirely would view a solely one-sided unrealistic market; any theory that disregards what is seen in practise should be rejected outright. This is discussed in [83].

5) PRINCIPLE COMPONENT PORTFOLIO

Given a covariance matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{11\times 11}$, and a specified number of components, c, such that $c \leq 11$, one can construct a principal portfolio with a weighting vector:

$$\boldsymbol{\omega}_{PC} = (\text{diag}(\boldsymbol{\Delta})/\mathbf{1}^T\text{diag}(\boldsymbol{\Delta}))(\boldsymbol{\Gamma}/(\mathbf{1}^T\boldsymbol{\Gamma})) \tag{65}$$

with $\boldsymbol{\Delta}$ being the diagonal matrix containing the singular values of the PCA decomposition, $\boldsymbol{\Gamma}$ being the matrix containing the corresponding principle components of the PCA decomposition, and with $\mathbf{1} \in \mathbb{R}^{c\times 1}$ being the vector of ones where c is the number of desired components. $\boldsymbol{\Delta} \in \mathbb{R}^{c\times c}$

and $\mathbf{\Gamma} \in \mathbb{R}^{11\times c}$ are the solutions to the following optimisation problem:

$$\text{argmin}_{\mathbf{\Gamma},\mathbf{\Delta}}\text{MSE}\Big(\mathbf{\Sigma} - \mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Gamma}^T\Big). \tag{66}$$

B. FORECASTING PERFORMANCE MEASURES

The covariance forecasting ability of the model presented herein can be measured on a monthly basis by either comparing the forecasted monthly covariances directly with the realised monthly covariance or by using the resulting efficient frontiers which can also be used to contrast the resulting efficient frontiers with the forecasted efficient frontiers using low-frequency, mid-frequency, and high-frequency implicit factors.

1) EFFICIENT FRONTIERS

With reference to Figure 6, Figures 22, 23 and 24, one can observe the precision, as well as the shortcoming, of monthly efficient frontier forecasting using internal factors during periods of stability and periods of instability due to external factors. In Figures 22 and 24 the implicit factors of highest frequency forecast the efficient frontier with the best accuracy during these two periods. In Figure 23 all efficient frontier forecasts are inaccurate due to the market downturn due to external factors not exhibited in the implicit factors.

From Figures 22 and 24 the mid- and low-frequency structures overestimate the variance of the portfolios over the majority of efficient frontiers. In Figure 23 all efficient frontier estimates are relatively inaccurate as a result of external market factors. This analysis is the first of its kind and provides valuable information about different frequency implicit factors that are used to forecast the forthcoming covariance structure of assets.

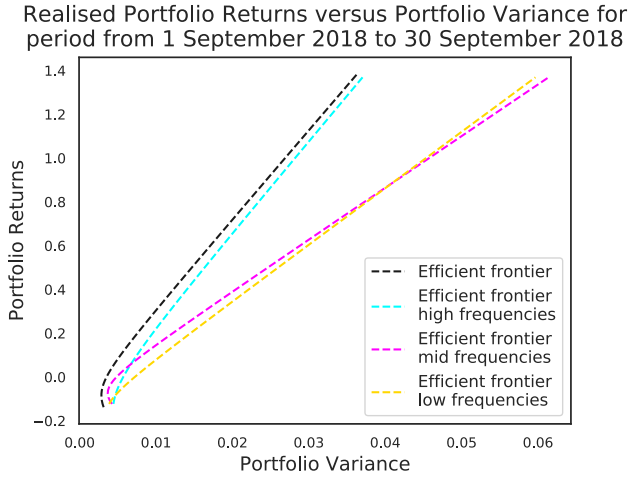

FIGURE 22. Plot comparing efficient frontiers using forecast covariance and realised covariance using realised returns for September 2018.

C. PORTFOLIO PERFORMANCE MEASURES

The following sections detail various performance measures that will be used to compare and contrast the IMF and SSA RCR portfolios against the five well-established techniques

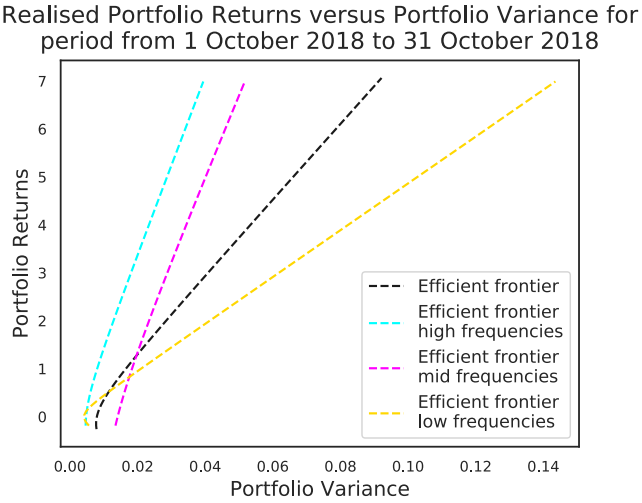

FIGURE 23. Plot comparing efficient frontiers using forecast covariance and realised covariance using realised returns for October 2018.

that function as benchmark portfolios mentioned above. Cumulative returns, mean returns, and variance of returns or risk are the most well-known and simple performance measures and are discussed in Section V-C1, Section V-C2, and Section V-C3, respectively. In Section V-C4, the value-at-risk measured for the various portfolios and benchmarks with a significance value of 5% such that $\alpha = 0.05$. This quantifies the 5% worst-case scenario of daily losses over the previous window. In Section V-C5, the maximum drawdown is the difference between the minimum daily returns and the maximum daily returns, normalised by the minimum daily returns.

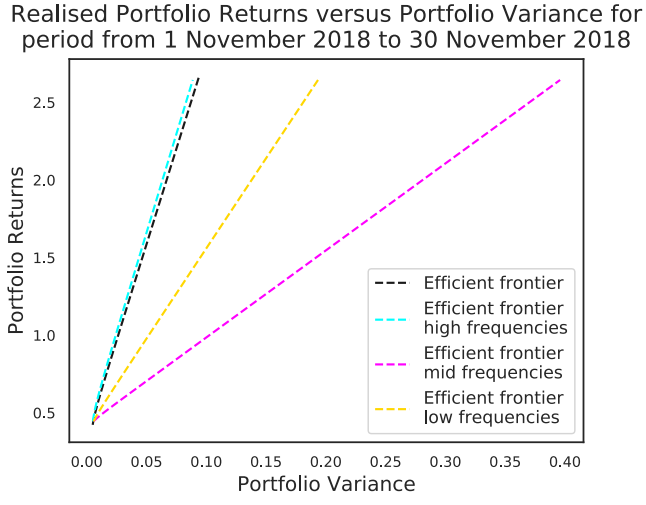

FIGURE 24. Plot comparing efficient frontiers using forecast covariance and realised covariance using realised returns for October 2018.

Section V-C6 details the omega ratio that calculates the ratio of the maximum profit from a call option on the asset during the previous window to the maximum profit from a put option on the asset during the previous window. Sections V-C7 and V-C8 outline the Sortino ratio and Sharpe ratio, respectively. These two ratios are similar, but

the Sortino ratio is a refinement of the Sharpe ratio. The Sharpe ratio measures the returns above the risk-free rate standardised by the risk over the previous period, but the Sortino ratio refines this by measuring the returns above the risk-free rate standardised by only the downside risk over the previous period; this is seen as a better measure of the risk.

1) CUMULATIVE RETURNS

At time t, the cumulative returns of a portfolio, CR_t, can be calculated as:

$$\mathrm{CR}_t = \prod_{s=1}^{t} e^{R_s} = e^{\sum_{s=1}^{t} R_s}, \tag{67}$$

with R_s being the daily return on day s. All implicit factor portfolios have final cumulative returns between those of the lowest benchmark portfolio (minimum variance portfolio) and the highest benchmark portfolio (the S&P 500 proxy or the maximum Sharpe portfolio). As cumulative returns (as well as all other performance measures in addition to annualized returns in Section V-C9) are for a single realisation of the implicit factor portfolios, the final evaluation of their performance will be reserved until Section 33. The S&P 500 proxy performs well being more heavily weighted towards Information Technology stocks that have performed well since the SARS-CoV-2 pandemic as a result of government assistance and paradigm-shifting lockdowns and work-from-home protocols.

A specific note can be taken of the PCA portfolio over the first two years. The portfolio has favourable returns above the other portfolios, but as a result of the weighting strategy (without any variance considerations), the portfolio is very responsive to drops in the economy. This happens during the 2018 lull, and one can note that the low frequency MDLP portfolio responds positively where the PCA responds negatively to essentially reverse positions. The PCA portfolio is no longer in contention from this point forward. Most implicit factor portfolios recover well from the SARS-CoV-2 pandemic-fuelled period of marked market depreciation, but as a result of the shorting restriction, the portfolios are susceptible to large fluctuations. The low-frequency MDLP portfolio exceeds its higher-frequency counterparts in cumulative returns.

2) MEAN RETURN

The mean return of a portfolio over a window width of length T, at time t can be calculated as:

$$\mu_t = \mathbb{E}[R_t] \approx \hat{\mu}_t := \frac{1}{T}\sum_{s=0}^{T-1} R_{t-s}. \tag{68}$$

Figure 26 displays the mean returns of the various portfolios over the previous 30 day window. Most portfolios follow each other quite closely, with a few exceptions. The two main outliers are the PCA and the maximum Sharpe ratio

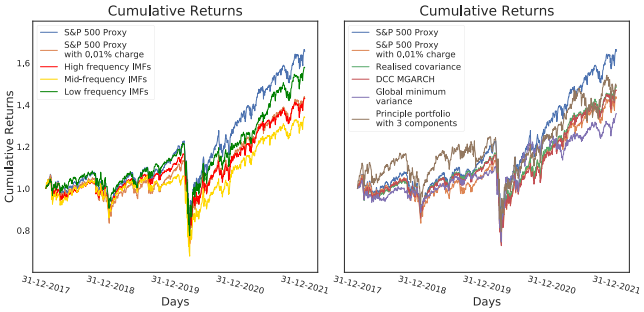

FIGURE 25. Cumulative returns of risk parity portfolios versus benchmarks.

portfolios - this increased variation in the returns has already been observed in the cumulative returns in Figure 25.

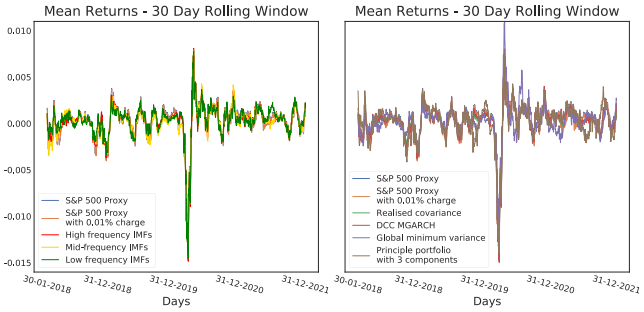

FIGURE 26. Mean returns of risk parity portfolio versus benchmark mean returns.

3) RISK

The unbiased variance estimate of a portfolio at time t with moving window of length T can be calculated as:

$$\sigma_t^2 = \mathbb{E}[(R_t - \mu_t)^2] \approx \hat{\sigma}_t^2 := \frac{1}{T-1}\sum_{s=0}^{T-1}(R_{t-s} - \mu_t)^2. \tag{69}$$

Figure 27 shows the variance over the moving window of the various portfolios. The largest variances reflect what has already been said about the cumulative returns and the mean returns of the PCA portfolio and the maximum Sharpe portfolio. The two major unanimous periods of increased variance have already been noted in the plot of the underlying indices and their respective correlations. During the SARS-CoV-2 pandemic, the variance is lowest in the minimum variance portfolio as is to be expected.

4) VALUE-AT-RISK (VaR)

Given a significance level, $\alpha = 0.05$, and a window, T, the Value-at-Risk of a portfolio gives the α percentile return over this window - this performance measure shows the worst-case scenario for the portfolio. The Value-at-Risk of a portfolio at time t over moving window T can be calculated as:

$$\mathrm{VaR}_{t_\alpha,T} = F^{-1}_{R_t,T}(\alpha) \approx \widehat{\mathrm{VaR}}_{t_\alpha,T} := \hat{F}^{-1}_{R_t,T}(\alpha), \tag{70}$$

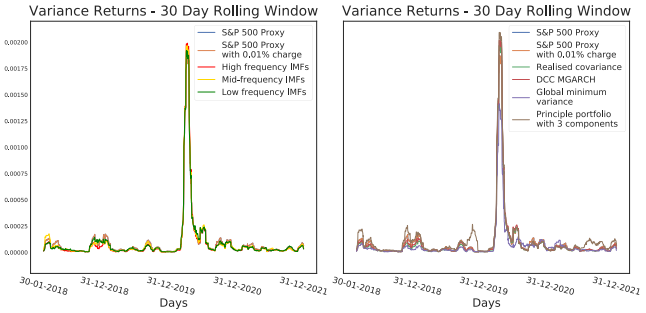

FIGURE 27. Variance of returns for risk parity portfolio versus benchmarks.

with $F_{R_t,T}(\alpha)$ being the cumulative distribution of the daily returns, R_t, over the period T at significance α. As can be noted in Figure 28, the PCA portfolio and the maximum Sharpe ratio portfolio have the largest Values-at-Risk. One may also note that all of the portfolios have similar Values-at-Risk during the SARS-CoV-2 Pandemic - this coupling of losses resulting in a significant correlation coupling that can be seen in the instantaneous covariance structures modelled by covariance regression and can be observed in Figure 19.

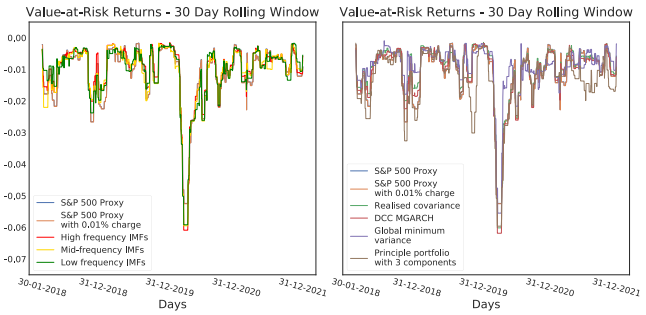

FIGURE 28. Value-at-risk of returns for risk parity portfolio versus benchmarks.

5) MAXIMUM DRAW-DOWN

The maximum drawdown is similar to the Value-at-Risk in that it uses the worst-case scenario as a measure of the performance of the portfolio. Maximum drawdown differs in that it measures the worst-case scenario when compared against the best-case scenario and then standardised using the best-case scenario. Given period T, the maximum drawdown of a portfolio at time t can be calculated as:

$$\text{MDD}_t = \frac{\min(R_{t-T+1}, \ldots, R_t) - \max(R_{t-T+1}, \ldots, R_t)}{\max(R_{t-T+1}, \ldots, R_t)}. \tag{71}$$

The maximum drawdown has an upper-limit of 0, where the Value-at-Risk has no theoretical upper-limit. The maximum drawdown is more erratic than the Value-at-Risk owing to the lack of the distribution in calculating the value. The maximum draw-down values that most break from the observable trend are the principle portfolio and the maximum Sharpe portfolio.

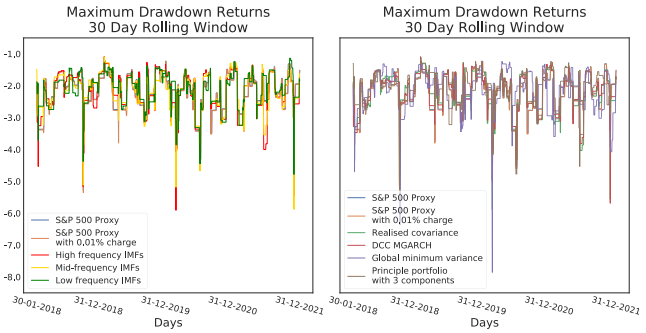

FIGURE 29. Maximum Draw-Down of returns for risk parity portfolio versus benchmarks.

6) OMEGA RATIO

The Omega ratio is the ratio of a hypothetical call option value to the hypothetical put option value over the previous window period. The Omega ratio, like the maximum drawdown, has a theoretical limit. Unlike the maximum drawdown, this is a theoretical lower limit where a Omega ratio can never be below zero - there is no theoretical upper limit. A larger Omega ratio is therefore desirable. Given a period, T, the Omega ratio of a portfolio can be calculated as:

$$\begin{aligned}\Omega_{t,T}(0) &= \frac{\int_0^\infty (1 - F_{R_t,T}(r))dr}{\int_{-\infty}^0 F_{R_t,T}(r)dr} \\ &\approx \widehat{\Omega}_t(0) := \frac{\int_0^\infty (1 - \hat{F}_{R_t,T}(r))dr}{\int_{-\infty}^0 \hat{F}_{R_t,T}(r)dr},\end{aligned} \tag{72}$$

with $F_{R_t,T}(r)$ being the cumulative distribution of the daily returns over the period T. All portfolios (even the minimum variance portfolio) appear to demonstrate significantly large omega ratios at various periods over the previous four years. This indicates that this is interesting period over which to apply portfolio optimisation techniques. One of the largest observable omega ratios are for the low frequency MDLP portfolio - this is a noteworthy feature as already observed in Section V-C1, the low frequency MDLP portfolio is able to respond more appropriately than its higher frequency counterparts.

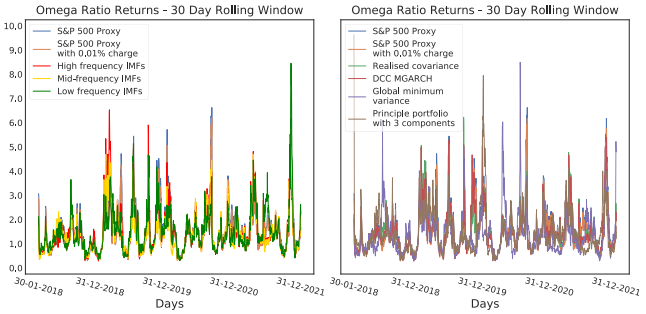

FIGURE 30. Omega Ratio of returns for risk parity portfolio versus benchmarks.

7) SORTINO RATIO

The Sortino ratio was introduced in [87] as a better alternative to the traditional Sharpe ratio. The returns above the risk-free

rate is standardised used only the downside risk. Given a period, T, the Sortino ratio at time t, can be calculate as:

$$\begin{aligned} \mathrm{SR}_t &= \frac{\mu_t - r_f}{\sqrt{\mathbb{E}[(R_t - \mu_t)^2 \mathbb{I}(R_t < 0)]}} \\ &\approx \widehat{\mathrm{SR}}_t := \frac{\hat{\mu}_t - r_f}{\sqrt{\frac{1}{T-1}\sum_{s=0}^{T-1}(R_{t-s} - \hat{\mu}_t)^2 \mathbb{I}(R_{t-s} < 0)}}. \end{aligned} \tag{73}$$

The Sortino ratios observable in Figure 31 are similar to the structures observable in Figure 30. This is attributable to both measures that track positive returns to some measure of the risk of negative returns. The maximum Sharpe ratio portfolio is the clear outlier over the vast majority of the 4 year period - this is as expected given this weighting strategy seeks to maximise the relative returns.

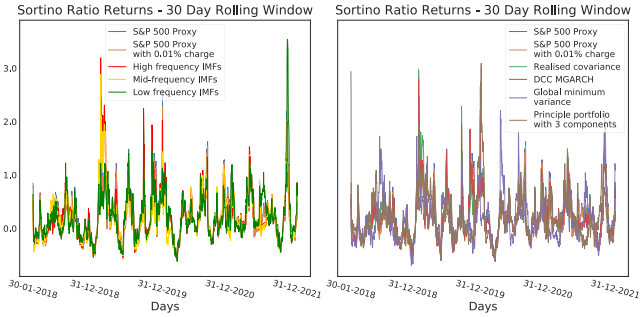

FIGURE 31. Sortino Ratio of returns for risk parity portfolio versus benchmarks.

8) SHARPE RATIO

The Sharpe ratio was introduced in [88] as a reliable measure of relative portfolio performance. Without the context of the level of variance of the returns about some underlying mean, a high return is obviously preferred to a slightly lower return, but should there be far less variance in the portfolio with a slightly lower return than in the portfolio with the higher return, many investors would prefer the near-guaranteed lower return to the less-guaranteed higher returns. The Sharpe ratio at time t with window length T can be calculated as:

$$\mathrm{Sharpe}_t = \frac{\mu_t - r_f}{\sigma_t} \approx \widehat{\mathrm{Sharpe}}_t := \frac{\hat{\mu}_t - r_f}{\hat{\sigma}}. \tag{74}$$

The Sortino ratio can be seen as a refinement of the Sharpe ratio, as the Sortino ratio measures the excess return above the risk free rate relative to only the variance of the negative returns compared to the Sharpe ratio measuring the excess returns above the risk free rate relative to the variance of all returns. This makes interpretation more difficult, as outliers become harder to detect and the interpretation opaque. All portfolios appear to have either the most negative or the most positive Sharpe ratio over the previous four-year period - this includes the minimum variance portfolio.

a: MAXIMUM SHARPE RATIO PORTFOLIO

With $\boldsymbol{\omega}_{\mathrm{MSR}} \in \mathbb{R}^{N\times 1}$ being the portfolio weighting vector, $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ being the vector of individual asset returns, r_f being the risk free rate which one can set to 1%, with $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$, being the covariance of the returns, and with $\mathbf{1} \in \mathbb{R}^{N\times 1}$ being the vector of ones, the maximum Sharpe ration portfolio weights be calculated as:

$$\boldsymbol{\omega}_{\mathrm{MSR}} = \frac{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1})}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1})} \tag{75}$$

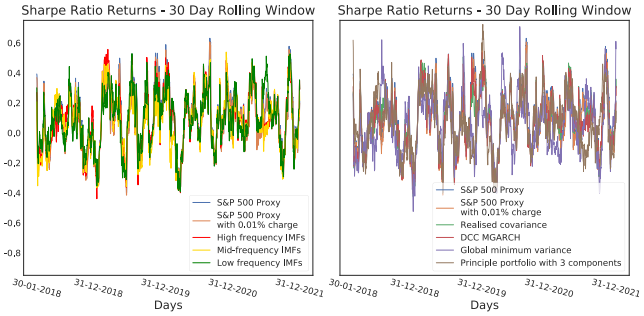

FIGURE 32. Sharpe Ratio of returns for risk parity portfolio versus benchmarks.

9) ANNUALISED RETURNS

To make a fair assessment of the efficacy of the technique presented here, the simulations were run a total of 250 times. The annualized returns of the five benchmark portfolios, as well as the distributions of the different frequency long/short equity portfolios, are plotted in Figure 33. The mean, variance, and skewness of each of the distributions are tabulated in Table 1 for additional context.

As expected, the minimum-variance portfolio has the lowest annualised returns of the five benchmark portfolios. The realised covariance, the DDC-GARCH portfolio, and the principal portfolio all have similar annualised returns. The greatest annualised returns among the benchmarks is clearly the S%P 500 proxy - this large annualised return is assisted by the larger weightings of the Information Technology sector, which had a stellar performance since the SARS-CoV-2 pandemic as a result of intensive government assistance and increased adoption of technology (AMZN, ZM, NFLX, et cetera) owing to the various stay and work from home protocols.

The means of the various long/short equity implicit factor portfolios differ by several points, the maximum difference being approximately 68 points between the high-frequency long/short equity portfolio and the low-frequency long/short restriction portfolio. The lowest frequency portfolio has the highest average returns compared to the other portfolios. This is as expected as a result of the relatively lower frequency implicit factor portfolios being able to adjust at a frequency most appropriate for monthly rebalancing. The mean returns of the portfolio using mid frequency structures are midway between the high frequency portfolio and the low frequency portfolio, as one would expect.

This analysis remains relevant when assessing the variances of the various portfolios. An important feature is that as the frequencies of the underlying implicit factors decrease,

the variances increase - this is counter to the analysis of the means in that the higher frequency structures allow the portfolios to adapt more rapidly to underlying structural changes, and as a result they have on average lower variances.

The skewness of the portfolios also have notable features.The least skewness value is the high frequency portfolio

TABLE 1. First three centred moments of different frequency risk parity portfolios with long/short equity weighting restrictions.

Frequency Content	Restrictions	Mean	SD	Skewness
High frequency	130/30	0.09511	0.02402	-0.04937
Mid frequency	130/30	0.09663	0.03144	0.08312
Low frequency	130/30	0.10196	0.04158	1.16445

using the implicit factor covariance forecasting; this means that it is the most positively skewed distribution meaning that on average the returns of the portfolios would be higher than the mean. In conjunction with reasonable returns, low relative variance, and large negative skewness, the best-performing portfolio is the long/short restricted implicit factor portfolio using high frequencies.

One should note the ubiquitous trend among mean, standard deviation, and skewness in that as the frequencies of the underlying implicit factors decrease, these values all increase. The returns for the low-frequency portfolio may be higher on average, but there is far more uncertainty in the consistency of these returns. Further study and analysis is needed before a conclusive technique is agreed upon as best for a monthly portfolio rebalancing.

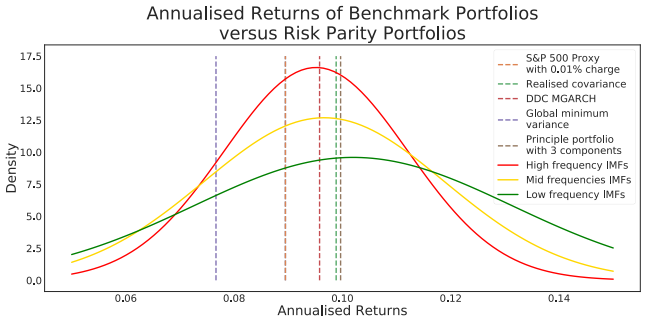

FIGURE 33. Annualised returns of benchmarks versus distribution of long/short equity risk parity portfolios.

VI. CONCLUSION

This paper describes several modern techniques rarely applied to financial data. The implicit factor extraction techniques discussed in Section II-A allow implicit factors to be used in modelling the covariance structure of financial instruments. The MGARCH techniques (given further exposition in the supplementary materials) are the current state of the art in covariance modelling in the financial setting (DCC MGARCH is used as a benchmark in our case study).

The results in Section V suggest that the performance of the long/short equity risk premia parity portfolio weighted based upon the covariance forecast using the implicit factors exceeds the analogous DCC MGARCH portfolios in a significant proportion of the simulations.The implicit factor portfolios appear to exceed more traditional methods (based on the measures used within this work), and beyond traditional models, allow one to directly model variance as attributable to specific market influences.

An exception is the cumulative returns of the S&P 500 proxy that grew significantly over the past four years (without considering the cost of daily trading expenses - this is used as a proxy for the general state of the economy and not tradeable). This is primarily due to the heavy weighting of technology companies within the index and the stellar adoption and performance of these companies after the SARS-CoV-2 pandemic and their subsequent elevated recovery (beyond the general market - see Figure 6) after the pandemic-induced sell-off.

The improved performance, as measured by several risk metrics, of the implicit factor models is due to the covariance structure being able to adjust to time-frequency changes in underlying implicit factors as proxies of trends in the indices and market. This paper promotes the development of implicit financial factor extraction to be used as factors or input variables in RCR to forecast the covariance structure of the underlying assets for inclusion and reweightings in portfolios.

RCR may be used without implicit factors, but rather using explicit factors such as unemployment rate or the federal funding interest rates, but are often then prone to discontinuities and stagnation in values - techniques would need to be developed within this framework to deal with the different frequency observations such as those developed in [89]. Without methods that monitor public opinion, such as natural language processing models, the daily movements in indices serve as a good proxy for the instantaneous opinion of the market.

Long/short equity weighting strategies and portfolio weighting techniques that promote risk parity deserve further study. We demonstrate these methods pair well with covariance regression methods in our case study with the high-frequency IMFs being most able to capture the evolving market conditions and construct a consistently low variance portfolio. All the techniques should be used more frequently and simultaneously in the pursuit of a portfolio optimisation framework.

VII. DISCUSSION

As discussed in Section II-A, VMD was discussed as a viable alternative to EMD and the other implicit factor extraction techniques discussed therein. [41] (who proposed VMD) found VMD to be more accurate in decomposing certain types of structures. VMD and EMD should be investigated further in the context of their accuracy in decomposing financial time series. Independent Component Analysis (ICA), which was proposed in [90], has been shown in [8] to be effective when paired with EMD (to form EMD-ICA or ICA-EMD) to isolate the underlying structures in simulated synthetic financial data.

The effectiveness of ICA versus PCA is examined by [91] in its ability to forecast the 28 largest Japanese stocks

using 3 years' of daily returns. ICA is shown to be more effective than PCA in capturing the dynamics of the stocks and to more accurately reconstruct the stocks. Reference [92] examines ICA's ability to better forecast the returns of stocks using the higher order mutual information (provided by ICA) with the intention of extending portfolio optimisation beyond traditional mean-variance optimisation. ICA shows promise in the analysis of financial time series with [93] demonstrating how ICA can be used to augment partial correlation coefficients (PCCs) - see [94] - to better isolate the correlation between structures by removing the correlations induced by other structures present.

The absence of the non-linear dependencies in the normalised elements of the precision matrix are addressed by [95] . The precision matrix is the inverse of the covariance matrix with the normalised elements thereof being the partial pairwise correlation coefficients. Reference [95] addresses this problem by proposing a generalised PCC model which is derived from the perspective of a multivariate mixture model.

Direct extensions of the framework described in Section IV-C are presented in [10] with these being given exposition in [31], [73], [96], and [97]. Reference [98] serves as an online appendix to [31]. Confidence bounds for the covariance values calculated using the framework described in Section IV-C are estimatable using the marginal distributions of each of the covariates used in calculating the covariance forecasts. Reference [73] uses this framework to estimate the confidence bounds for the individual skewnesses of the low-interest rate country currencies of the Japanese Yen, Swiss Franc, and Eurozone Euro.

The paper that proposed covariance regression through the random effects framework presented herein, [10] , proposes direct, higher-order, extensions of Equation (14) which is the random effects representation of our covariance regression such as:

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_{t,1} + \gamma_t \mathbf{B}\mathbf{x}_{t,2} + \phi_t \mathbf{C}\mathbf{x}_{t,3} + \cdots + \boldsymbol{\epsilon}_t, \tag{76}$$

with,

$$\begin{aligned} &\mathbb{E}[\gamma_t] = 0,\ \mathrm{Var}[\gamma_t] = 1, \mathbb{E}[\gamma_t \boldsymbol{\epsilon}_t] = \mathbf{0}, \\ &\mathbb{E}[\phi_t] = 0,\ \mathrm{Var}[\phi_t] = 1, \mathbb{E}[\phi_t \boldsymbol{\epsilon}_t] = \mathbf{0}, \\ &\quad \text{and}\ \ \mathrm{Cov}[\gamma_t \phi_t] = 0. \\ &\quad \vdots \end{aligned} \tag{77}$$

The interpretation of $\mathbf{C}\mathbf{x}_{t,3}$ is difficult in the presence of $\mathbf{B}\mathbf{x}_{t,2}$ and other additional orders of Equation (14). The extensions of [10] and [73] as well as other factor models are interesting fields related to the question of covariance forecasting and should be seriously considered when continuing this investigation.

REFERENCES

[1] C. van Jaarsveldt, M. Ames, G. W. Peters, and M. Chantler, "Package AdvEMDpy: Algorithmic variations of empirical mode decomposition in Python," *Ann. Actuarial Sci.*, vol. 17, no. 3, pp. 606–642, Nov. 2023.

[2] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[3] N. Huang, Z. Shen, and S. Long, "A new view of nonlinear water waves: The Hilbert spectrum," *Annu. Rev. Fluid Mech.*, vol. 31, no. 1, pp. 417–457, 1999.

[4] N. Huang, "Computer implemented empirical mode decomposition method, apparatus and article of manufacture," U.S. Patent 5 983 162, Nov. 9, 1999. [Online]. Available: https://patents.google.com/patent/US5983162/en

[5] Q. Chen, N. Huang, S. Riemenschneider, and Y. Xu, "A B-spline approach for empirical mode decompositions," *Adv. Comput. Math.*, vol. 24, nos. 1–4, pp. 171–195, Jan. 2006.

[6] H. Hassani, "Singular spectrum analysis: Methodology and comparison," *J. Data Sci.*, vol. 5, no. 2, pp. 239–257, 2007.

[7] P. Bonizzi, J. M. H. Karel, O. Meste, and R. L. M. Peeters, "Singular spectrum decomposition: A new method for time series decomposition," *Adv. Adapt. Data Anal.*, vol. 6, no. 4, 2014, Art. no. 1450011.

[8] C. van Jaarsveldt, G. W. Peters, M. Ames, and M. Chantler, "Tutorial on empirical mode decomposition: Basis decomposition and frequency adaptive graduation in non-stationary time series," *IEEE Access*, vol. 11, pp. 94442–94478, 2023.

[9] C. van Jaarsveldt, G. Peters, M. Ames, and M. Chantler, "Package CovRegpy: Regularized covariance regression and forecasting in Python," *Ann. Actuarial Sci.*, pp. 1–35, May 2024, doi: 10.1017/S1748499524000101.

[10] P. Hoff and X. Niu, "A covariance regression model," *Statistica Sinica*, vol. 22, no. 2, pp. 729–753, 2012.

[11] E. Qian, "Risk parity portfolios: Efficient portfolios through true diversification, Panagora Asset Manag., Boston, MA, USA, White Paper 270416 9/11, Sep. 2005. [Online]. Available: https://www.panagora.com/assets/PanAgora-Risk-Parity-Portfolios-Efficient-Portfolios-pdf

[12] S. Maillard, T. Roncalli, and J. Teïletche, "The properties of equally weighted risk contribution portfolios," *J. Portfolio Manage.*, vol. 36, no. 4, pp. 60–70, Jul. 2010.

[13] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[14] J. Tobin, "Liquidity preference as behavior towards risk," *Rev. Econ. Stud.*, vol. 25, no. 2, pp. 65–86, 1958.

[15] J. Tobin, "The theory of portfolio selection," in *The Theory of Interest Rates*, F. Hahn and F. Brechling, Eds., London, U.K.: Macmillan, 1965.

[16] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, 1964.

[17] J. Lintner, "Security prices, risk, and maximal gains from diversification," *J. Finance*, vol. 20, no. 4, pp. 587–615, 1965.

[18] J. Lintner, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets: A reply," *Rev. Econ. Statist.*, vol. 47, no. 1, pp. 13–37, 1965.

[19] F. Black, "Capital market equilibrium with restricted borrowing," *J. Bus.*, vol. 45, no. 3, pp. 444–455, 1972.

[20] W. Sharpe, *Portfolio Theory and Capital Markets*. New York, NY, USA: McGraw-Hill, 1970.

[21] H. Levy and M. Sarnat, *Investment and Portfolio Analysis*. Hoboken, NJ, USA: Wiley, 1972.

[22] R. Merton, "An analytic derivation of the efficient portfolio frontier," *J. Financial Quant. Anal.*, vol. 7, no. 4, pp. 1851–1872, 1972.

[23] S. A. Ross, "The arbitrage theory of capital asset pricing," *J. Econ. Theory*, vol. 13, no. 3, pp. 341–360, Dec. 1976.

[24] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *J. Financial Econ.*, vol. 33, no. 1, pp. 3–56, Feb. 1993.

[25] E. F. Fama and K. R. French, "A five-factor asset pricing model," *J. Financial Econ.*, vol. 116, no. 1, pp. 1–22, Apr. 2015.

[26] R. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: J. Econ. Soc.*, vol. 50, no. 4, pp. 987–1007, 1982.

[27] R. Engle, "Estimates of the variance of U.S. Inffation based on the ARCH model," *J. Money, Credit Banking*, vol. 15, no. 3, pp. 286–301, 1983.

[28] R. Engle and D. Kraft, "Multiperiod forecast error variances of inflation estimated from ARCH models," in *Applied Time Series Analysis of Economic Data*, A. Zellner, Ed. Washington, DC, USA: Bureau of the Census, 1983, pp. 293–302.

[29] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, Apr. 1986.
[30] L. Bauwens, S. Laurent, and J. Rombouts, "Multivariate GARCH models: A survey," *J. Appl. Econometrics*, vol. 21, no. 1, pp. 79–109, 2006.
[31] M. Ames, G. Bagnarosa, and G. W. Peters, "Violations of uncovered interest rate parity and international exchange rate dependences," *J. Int. Money Finance*, vol. 73, pp. 162–187, May 2017.
[32] Y. Choueifaty and Y. Coignard, "Toward maximum diversification," *J. Portfolio Manage.*, vol. 35, no. 1, pp. 40–51, Oct. 2008.
[33] L. Martellini, "Toward the design of better equity benchmarks: Rehabilitating the tangency portfolio from modern portfolio theory," *J. Portfolio Manage.*, vol. 34, no. 4, pp. 34–41, Jul. 2008.
[34] L. Yang, "An application of principal component analysis to stock portfolio management," M.S. thesis, Dept. Econ. Finance, Univ. Canterbury, Christchurch, New Zealand, Jan. 2015, doi: 10.26021/5235. [Online]. Available: http://hdl.handle.net/10092/10293
[35] G. Pasini, "Principal component analysis for stock portfolio management," *Int. J. Pure Appl. Math.*, vol. 115, no. 1, pp. 153–167, 2017.
[36] M. Carhart, "On persistence in mutual fund performance," *J. Finance*, vol. 52, no. 1, pp. 57–82, 1997.
[37] B. T. Khoa and T. Trong Huynh, "Comparing the predictive efficiency of fama-french three-factor model and carhart four-factor model: Financial market approach," in *Proc. IEEE 8th Int. Conf. for Converg. Technol. (I2CT)*, Lonavla, India, Apr. 2023, pp. 1–5, doi: 10.1109/I2CT57861.2023.10126181.
[38] L. Jia-long, L. Bo-wei, and L. Min, "Model contest and portfolio performance: Black-litterman versus factor models," in *Proc. Int. Conf. Manage. Sci. Eng. 20th Annu. Conf. Proc.*, Harbin, China, Jul. 2013, pp. 507–516.
[39] M. Yang, "Research on quantitative multi-factor stock selection model based on machine learning," in *Proc. 3rd Int. Academic Exchange Conf. Sci. Technol. Innov. (IAECST)*, Guangzhou, China, Dec. 2021, pp. 489–493.
[40] Q. Fan, F. Hu, and X.-P. Zhang, "Double-selection based high-dimensional factor model with application in asset pricing," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Ottawa, ON, Canada, Nov. 2019, pp. 1–5, doi: 10.1109/GlobalSIP45357.2019.8969175.
[41] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[42] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*, 1st ed. Cambridge, MA, USA: MIT Press, 1949.
[43] O. F. Karaaslan and G. Bilgin, "Comparison of variational mode decomposition and empirical mode decomposition features for cell segmentation in histopathological images," in *Proc. Med. Technol. Congr. (TIPTEKNO)*, Antalya, Turkey, Nov. 2020, pp. 1–4, doi: 10.1109/TIPTEKNO50054.2020.9299321.
[44] W. Liu, W. Hu, and D. Fu, "Frequency shifting-based variational mode decomposition method for speech signal decomposition," in *Proc. Int. Conf. Autom., Robot. Comput. Eng. (ICARCE)*, Wuhan, China, Dec. 2022, pp. 1–5, doi: 10.1109/ICARCE55724.2022.10046652.
[45] M. Campi, G. W. Peters, N. Azzaoui, and T. Matsui, "Machine learning mitigants for speech based cyber risk," *IEEE Access*, vol. 9, pp. 136831–136860, 2021.
[46] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
[47] W. Zhou, Z. Feng, X. Wang, and H. Lv, "Empirical Fourier decomposition," 2019, *arXiv:1912.00414*.
[48] R. Jayaparvathy, "Complexity comparison of empirical mode decomposition and wavelet decomposition methods in the detection of ventricular late potential," in *Proc. IEEE Conf. Interdiscipl. Approaches Technol. Manage. Social Innov. (IATMSI)*, Gwalior, India, Dec. 2022, pp. 1–5, doi: 10.1109/IATMSI56455.2022.10119348.
[49] S. Meng, L.-T. Huang, and W.-Q. Wang, "Tensor decomposition and PCA jointed algorithm for hyperspectral image denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 897–901, Jul. 2016.
[50] Y.-H. Wang and S.-H. Cheng, "Boundary effects for EMD-based algorithms," *IEEE Signal Process. Lett.*, vol. 29, pp. 1032–1036, 2022.
[51] T. Wang, M. Zhang, Q. Yu, and H. Zhang, "Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal," *J. Appl. Geophys.*, vol. 83, pp. 29–34, Aug. 2012.
[52] R. Ho and K. Hung, "A comparative investigation of mode mixing in EEG decomposition using EMD, EEMD and M-EMD," in *Proc. IEEE 10th Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Malaysia, Apr. 2020, pp. 203–210.
[53] A. Vijayasankar and P. R. Kumar, "Correction of blink artifacts from single channel EEG by EMD-IMF thresholding," in *Proc. Conf. Signal Process. Commun. Eng. Syst. (SPACES)*, Vijayawada, India, Jan. 2018, pp. 176–180.
[54] J. Jenitta and A. Rajeswari, "Denoising of ECG signal based on improved adaptive filter with EMD and EEMD," in *Proc. IEEE Conf. Inf. Commun. Technol.*, Thuckalay, India, Apr. 2013, pp. 957–962.
[55] N. Li and P. Li, "An improved algorithm based on EMD-wavelet for ECG signal de-noising," in *Proc. Int. Joint Conf. Comput. Sci. Optim.*, vol. 1, Sanya, China, Apr. 2009, pp. 825–827.
[56] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numerische Mathematik*, vol. 31, no. 4, pp. 377–403, 1978.
[57] G. Wahba, *Spline Models for Observational Data*. 6th ed. Philadelphia, PA, USA: SIAM, 1990.
[58] A. Bowman and L. Evers, "Nonparametric smoothing," Acad. Ph.D. Training Statist., Univ. Warwick, Lect. Notes, 2017. [Online]. Available: https://warwick.ac.uk/fac/sci/statistics/apts/students/resources-1314/ns_notes.pdf
[59] R. Vautard, P. Yiou, and M. Ghil, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signals," *Phys. D, Nonlinear Phenomena*, vol. 58, nos. 1–4, pp. 95–126, Sep. 1992.
[60] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of M-estimators of covariance matrix," *IEEE Trans. Signal Process.*, vol. 69, pp. 256–269, 2021.
[61] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed distributions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, South Brisbane, QLD, Australia, Apr. 2015, pp. 5693–5697.
[62] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, and D. P. Palomar, "Robust rank constrained Kronecker covariance matrix estimation," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 810–814.
[63] Y. Sun, P. Babu, and D. P. Palomar, "Regularized robust estimation of mean and covariance matrix under heavy tails and outliers," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, A Coruna, Spain, vol. 8, Jun. 2014, pp. 125–128.
[64] J. Liu and D. P. Palomar, "Robust estimation of mean and covariance matrix for incomplete data in financial applications," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Montreal, QC, Canada, Nov. 2017, pp. 908–912.
[65] R. Michaud, "The Markowitz optimization enigma: Is 'optimized' optimal?" *Financial Analysts J.*, vol. 45, no. 1, pp. 31–42, 1989.
[66] M. Best and R. Grauer, "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results," *Rev. Financial Stud.*, vol. 4, no. 2, pp. 315–342, 1991.
[67] V. Chopra and W. Ziemba, "The effect of errors in means, variances, and covariances on optimal portfolio choice," *J. Portfolio Manag.*, vol. 19, no. 2, pp. 6–11, 1993.
[68] K. Winston, "The 'efficient index' and prediction of portfolio variance," *J. Portfolio Manag.*, vol. 19, no. 3, pp. 27–34, 1993.
[69] L. Chan, J. Karceski, and J. Lakonishok, "On portfolio optimization: Forecasting covariances and choosing the risk model," *Rev. Financial Stud.*, vol. 12, no. 5, pp. 937–974, 1999.
[70] R. Merton, "On estimating the expected return on the market: An exploratory investigation," *J. Financial Econ.*, vol. 8, no. 4, pp. 323–361, 1980.
[71] D. Nelson, "Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model," *J. Econometrics*, vol. 52, nos. 1–2, pp. 61–90, 1992.
[72] R. Zhou, J. Ying, and D. P. Palomar, "Covariance matrix estimation under low-rank factor model with nonnegative correlations," *IEEE Trans. Signal Process.*, vol. 70, pp. 4020–4030, 2022.
[73] M. Ames, "Innovations in dependence modelling for financial applications," Ph.D. thesis, Dept. Stat. Sci., Univ. College London, London, U.K., 2017.
[74] F. Santosa and W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 4, pp. 1307–1330, 1986.
[75] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
[76] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[77] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.
[78] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, 1951.
[79] J. Skaf and S. Boyd. (2009). *Multi-Period Portfolio Optimization with Constraints and Transaction Costs*. Accessed: Apr. 20, 2009. [Online]. Available: https://web.stanford.edu/~boyd/papers/pdf/dyn_port_opt.pdf
[80] P. Samuelson, "Lifetime portfolio selection by dynamic stochastic programming," *Rev. Econ. Statist.*, vol. 51, no. 3, pp. 239–246, 1969.
[81] M. Magill and G. Constantinides, "Portfolio selection with transactions costs," *J. Econ. Theory*, vol. 13, no. 2, pp. 245–263, 1976.
[82] G. Constantinides, "Multiperiod consumption and investment behavior with convex transactions costs," *Manag. Sci.*, vol. 25, no. 11, pp. 1127–1137, 1979.
[83] G. Hoffman, "Short selling," in *The Security Markets*, A. Bernheim and M. Grant Schneiders, Eds., New York, NY, USA: Twentieth Century, 1936, pp. 356–401. [Online]. Available: https://www.jstor.org/stable/1807910
[84] B. Jacobs and K. Levy, "Long/short equity investing—Profit from both winners and losers," *J. Portfolio Manage.*, vol. 20, no. 1, pp. 52–63, 1993.
[85] R. O. Michaud, "Are long-short equity strategies superior?" *Financial Analysts J.*, vol. 49, no. 6, pp. 44–49, Nov. 1993.
[86] B. Jacobs and K. Levy, "More on long-short strategies," *Financial Analysts J.*, vol. 51, no. 2, pp. 88–90, 1995.
[87] F. A. Sortino and L. N. Price, "Performance measurement in a downside risk framework," *J. Investing*, vol. 3, no. 3, pp. 59–64, Aug. 1994.
[88] W. Sharpe, "Mutual fund performance," *J. Bus.*, vol. 39, no. 1, pp. 119–138, 1966.
[89] M. Armesto, K. Engemann, and M. Owyang, "Forecasting with mixed frequencies," *Federal Reserve Bank St. Louis Rev.*, vol. 92, no. 6, pp. 521–536, 2010.
[90] B. Ans, J. Hérault, and C. Jutten, "Architectures neuromimétiques adaptatives: Détection de primitives," in *Proc. Cognitiva*, vol. 2, Paris, France, 1985, pp. 593–597.
[91] A. Back and A. Weigend, "A first application of independent component analysis to extracting structure from stock returns," *Int. J. Neural Syst.*, vol. 8, no. 4, pp. 473–484, 1997.
[92] E. Chin, A. S. Weigend, and H. Zimmermann, "Computing portfolio risk using Gaussian mixtures and independent component analysis," in *Proc. IEEE/IAFE Conf. Comput. Intell. Financial Eng. (CIFEr)*, New York, NY, USA, Apr. 1999, pp. 74–117.
[93] J. Belda, L. Vergara, G. Safont, and A. Salazar, "Computing the partial correlation of ICA models for non-Gaussian graph signal processing," *Entropy*, vol. 21, no. 1, p. 22, Dec. 2018.
[94] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Austral. New Zealand J. Statist.*, vol. 46, no. 4, pp. 657–664, 2004.
[95] J. Belda, L. Vergara, A. Salazar, and G. Safont, "Estimating the Laplacian matrix of Gaussian mixtures for signal processing on graphs," *Signal Process.*, vol. 148, pp. 241–249, Jul. 2018.
[96] M. Ames, G. Bagnarosa, G. W. Peters, P. Shevchenko, and T. Matsui, "Forecasting covariance for optimal carry trade portfolio allocations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5910–5914.
[97] M. Ames, G. Bagnarosa, G. Peters, and P. Shevchenko, "Understanding the interplay between covariance forecasting factor models and risk-based portfolio allocations in currency carry trades," *J. Forecasting*, vol. 37, no. 8, pp. 805–831, 2018.
[98] M. Ames, G. Bagnarosa, and G. Peters, "An online appendix to: 'Violations of uncovered interest rate parity and international exchange rate dependences,'" SSRN 2638103, Aug. 2015, pp. 1–13. [Online]. Available: https://ssrn.com/abstract=2638103

COLE VAN JAARSVELDT received the B.Sc. degree in physics and the B.Sc. degree (Hons.) in mathematics from Nelson Mandela University, Gqeberha, South Africa, the Postgraduate Diploma degree in actuarial science and the M.Sc. degree in mathematical finance from the University of Cape Town, Cape Town, South Africa, and the master's degree (Hons.) from the African Institute of Financial Markets and Risk Management (AIFMRM) under Prof. David Taylor. He is currently pursuing the Ph.D. degree in actuarial mathematics and statistics with Heriot-Watt University, Edinburgh, U.K., with Prof. Gareth W. Peters and Prof. Mike Chantler as co-supervisors.

GARETH W. PETERS (Senior Member, IEEE) received the M.Sc. (via research) degree from Cambridge University, Cambridge, U.K., and the Ph.D. degree in statistics by publication from the University of New South Wales (UNSW), Australia, in 2009. He is currently the Janet and Ian Duncan Endowed Chair Professor in Actuarial Science and a Chair Professor of Statistics for Risk and Insurance with the Department of Applied Probability and Statistics, University of California at Santa Barbara. He has published more than 150 peer-reviewed articles on risk and insurance modeling, two research textbooks on operational risk and insurance as well as being the editor and the contributor to three edited textbooks on spatial statistics and Monte Carlo methods.

MATTHEW AMES received the M.Sc. degree (Hons.) in advanced computer science from The University of Manchester, in 2012, and the Ph.D. degree in statistics from University College London (UCL), London, U.K., in 2017. He was a Researcher with the Rakuten Institute of Technology and a Non-Fungible Token Quantitative Analyst. He is currently the CTO and the Co-Head of Research at the startup technology company ResilientML which brings resilient, robust, and interpretable machine learning solutions to users by combining cutting-edge academic research with industry expertise. He is an experienced Researcher. He received the Peter Jones Memorial Prize for the M.Sc. degree.

MIKE CHANTLER received the Ph.D. degree in image processing from Heriot-Watt University, Edinburgh, U.K. He has been a Professor of computer science with Heriot-Watt University for the past 40 years where he manages the Texture Laboratory and the Strategic Futures Laboratory which specializes in developing and deploying AI tools that provide strategic insight and promote evidence-based, auditable, and decision-making. After graduating, he was a Software Engineer in national defense projects and did consultancy in Europe before taking up a post at Heriot-Watt University. He has published over 100 peer-reviewed articles, has chaired leading international conferences and workshops, and has attracted multi-million pound EU and national funding.