



Contents lists available at ScienceDirect

# Expert Systems With Applications

journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)

## Dynamic portfolio optimization with inverse covariance clustering

Yuanrong Wang<sup>a</sup>, Tomaso Aste<sup>a,b,c,\*</sup><sup>a</sup> Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK<sup>b</sup> UCL Centre for Blockchain Technologies, University College London, Gower Street, London, WC1E 6BT, UK<sup>c</sup> Systemic Risk Center, London School of Economics and Political Sciences, London, WC2A 2AE, UK

### ARTICLE INFO

#### Keywords:

Dynamic portfolio optimization  
 Portfolio management  
 Financial market states  
 Market regimes  
 Temporal clustering  
 Information filtering networks  
 Covariance structure

### ABSTRACT

Market conditions change continuously. However, in portfolio investment strategies, it is hard to account for this intrinsic non-stationarity. In this paper, we propose to address this issue by using the Inverse Covariance Clustering (ICC) method to identify inherent market states and then integrate such states into a dynamic portfolio optimization process. Extensive experiments across three different markets, NASDAQ, FTSE and HS300, over a period of ten years, demonstrate the advantages of our proposed algorithm, termed Inverse Covariance Clustering-Portfolio Optimization (ICC-PO). The core of the ICC-PO methodology concerns the identification and clustering of market states from the analytics of past data and the forecasting of the future market state. It is therefore agnostic to the specific portfolio optimization method of choice. By applying the same portfolio optimization technique on a ICC temporal cluster, instead of the whole train period, we show that one can generate portfolios with substantially higher Sharpe Ratios, which are statistically more robust and resilient with great reductions in the maximum loss in extreme situations. This is shown to be consistent across markets, periods, optimization methods and selection of portfolio assets.

### 1. Introduction

In the field of asset management, the problem of portfolio allocation has gained unprecedented popularity over the past few years. Constructing a good portfolio combines the art and science of balancing between trade-offs and the aim to meet long-term financial goals. The simple core of any portfolio optimization is to assign optimal weights to each portfolio's component in order to minimize investment's risk and maximize the return. In 1952, Markowitz (1952) demonstrated that, by assuming risk to be quantifiable by the variance of the portfolio's returns, the optimal weights which minimize portfolio's variance at a given average portfolio's return can be computed with a simple and exact formula. However, the Markowitz's theoretical maximum is attained only in-sample, on the train dataset, whereas off-sample, on the test set where investment is made, performances of the Markowitz's portfolio can be largely sub-optimal.

Markowitz's modern portfolio theory is the foundation to modern quantitative asset management. There are however two main limitation in the Markowitz's assumptions. The first limitation concerns the use of the portfolio's variance as measure of risk. The variance (when defined) is indeed a measure for the width of the distribution but there are other properties that are better measure of risk (e.g. the value at risk) and might not be reducible to the variance when the underlying probability distribution is not a location-scale. The second limitation concerns the

ability to estimate the (future) means and covariance of the asset's returns in the portfolio.

After Markowitz's seminal work, many portfolio selection methodologies have been introduced to cure the first limitation concerning the reliance on variance for risk quantification and nowadays there are several well-established approaches that go well beyond the use of variance as sole risk measure and yield superior results (Dymova et al., 2021; Hult et al., 2012; Kaczmarek et al., 2020). Furthermore, with the enormous development of machine learning optimization techniques there are presently virtually no limitations in constructing optimal portfolios based on any kind of risk measure (Ban et al., 2018; Paiva et al., 2019; Samo & Vervuurt, 2016).

Addressing the second limitation is harder. Indeed, normally, one does not have information from the future that would allow to set the future properties of the asset's multivariate distribution. Several literature contributions have been focusing on estimating expected return and covariance. Specifically, for expected return estimation, various pricing models have been proposed over last decades in the traditional finance domain (see, for instance, Jensen et al., 1972; Merton, 1973) and factor models (Fama & French, 1992; Fama & MacBeth, 1973). However, those models suffers from high variance, which reduces the step-ahead prediction ability beyond a simple benchmark (Simin, 2008). Currently, modern signal processing and machine

\* Corresponding author at: Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK.

E-mail addresses: [yuanrong.wang@cs.ucl.ac.uk](mailto:yuanrong.wang@cs.ucl.ac.uk) (Y. Wang), [t.aste@ucl.ac.uk](mailto:t.aste@ucl.ac.uk) (T. Aste).

<https://doi.org/10.1016/j.eswa.2022.118739>

Received 4 April 2022; Received in revised form 20 June 2022; Accepted 29 August 2022

Available online 13 September 2022

0957-4174/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

learning community have also applied data-driven modelling to return and volatility prediction (Briola et al., 2020, 2021; Glosten et al., 1993; Zhong & Enke, 2019). However, it is well understood that these models tend to be prone to overfitting and large noise in the low signal-to-noise ratio financial environment. On the other hand for covariance estimation, research has been addressing the issues related with noise using covariance sparsification. Two algorithms e.g., Graphical Lasso (Friedman et al., 2008), Triangulated Maximally Filtered Graph (Masara et al., 2017) and etc. are discussed in Section 2.5. Nevertheless, the reliance on past observations and the assumption that they will significantly represent also the future, is hard to avoid. Nonetheless, markets are not stationary, it is common knowledge that they cyclically pass through bull and bear states and occasionally deepen into crisis periods. For each of these periods the market prices' returns have different statistical properties and they are not describable by means of a unique multivariate probability distribution. This is especially relevant for factors that matter most to the management of portfolio risk. Indeed, crisis periods have distribution with fatter tails and they tend to be more asymmetrical with the left tail having larger probability for large losses than the right tail for equivalent gains. Portfolio constructions must take into account these differences and devise different investment strategies for each market condition. This is indeed the ground basis for any dynamic asset allocation. However, such a wise allocation would imply the knowledge of the future market state and forecasting it from past observations is not an easy task (Ma et al., 2021; Soleymani & Paquet, 2021). In the literature there have been several contributions exploring dynamical portfolio optimization. Let us here mention a few such as Ang and Bekaert (2003), Guo and Ching (2021), Hansen (1992) and Ramchand and Susmel (1998), who leveraged on Hidden Markov Model (HMM). However, HMM suffers heavily from the curse of dimensionality, and this is highly problematic in this domain of application because the number of assets is usually large in portfolio management. Others researchers, such as Buckley et al. (2008), Delany (2007), Ketter et al. (2009), Kumar et al. (2018), Nayak et al. (2015) and Reynolds (2009), employ clustering methods, e.g., Gaussian Mixture, and K-Nearest Neighbours (KNN). However, these methods rely on strong assumptions and are not originally designed for time-series, e.g., Gaussian Mixture assumes Gaussian nature in all the base distributions, and KNN overlooks temporal consistency between single data point.

In this paper, we provide an algorithm termed Inverse Covariance Clustering-Portfolio Optimization (ICC-PO) to address the non-stationarity problem, by identifying the inherent market states and forecast the most likely future state. The Inverse Covariance Clustering (ICC) (Procacci & Aste, 2018) is a novel temporal clustering method for market states clustering. In this paper we propose to make use of this temporal clustering classification, constructing different optimal portfolios associated with two ICC market state clusters. By incorporating this method, we avoid many of the aforementioned caveats, as the sparse inverse covariance of the portfolio is robust and efficient, and the ICC impose temporal consistency during clustering. The clusters are constructed in the in-sample training set (the past) and then are used separately to train the portfolio optimizer of choice which is then tested on an off-sample period following the training set (the future). For the optimization we used two approached based on the classical Markowitz's approach but devised to have only positive weights (no short-sellings). They are the Sequential Least Square Quadratic Programming (SLS) and the Critical Line Algorithm (CLA). Let us note, that the ICC-PO approach allows the use of any optimization method of choice.

We tested the approach with three extensive experiments with daily data, from 2010 to 2020, from three different markets: NASDAQ, FTSE and HS300. For each market, we selected 100 largest market capitalization constituent stocks and quantified the off-sample performances of portfolios constructed from in-sample training data using separately the two ICC-market states. We demonstrate that the difference in returns

and risks (computed on the testing set) between the two optimal portfolios, constructed from the two ICC-market states (on the training set), is very large with Sharpe Ratios that more than double and with very large differences in the likelihoods of large negative returns that can have up to three time smaller quantiles (i.e. value at risk). We provide a simple criteria to forecast the best performing out of sample market state which we named 'State 0'. Our results also show that sparsification of the inverse covariance matrix through information filtering networks (Barfuss et al., 2016; Massara et al., 2017) is improving the results, this is a confirmation of a previous result (Procacci & Aste, 2021) extended however in this paper to a different dataset, different portfolio optimizers and different markets. The robustness of the method is tested by gathering statistics over 100 re-sampling of consecutive train-test sets randomly selected across the 10 years period 2010 to 2020. Furthermore, reliance on portfolio basket choices is tested by doing the same experiments with random selection of 100 stocks instead of the 100 most capitalized.

The main contribution of this paper consist in the demonstration that market observations at different times can be classified into different states. Such states have distinct statistical properties, and they continue to be separable in log-likelihood after the in-sample training period showing temporal persistence. Such persistence enables us to predict the best performing state with higher log-likelihood in the off-sample investment period. Moreover, in this paper we confirm the intuitive argument by Procacci and Aste (2021) that a model with larger log-likelihood must perform better for portfolio optimization with respect to one with lower log-likelihood'

The remaining of the paper is organized as follows: in Section 2, we review literature on market states clustering, dynamic portfolio optimization, and information filtering network applied in the experiments; in Section 3, we outline experimental methodologies and implementations. The final results are presented in Section 4 and discussed in Section 5.

## 2. Background literature

### 2.1. Mean-variance optimization

Despite the unquestionable merits and pioneering status of Markowitz's mean-variance optimization (MVO) approach, there are some major assumptions, and several bad applications, that reduce its efficacy for practical implementations (Kalayci et al., 2019). Firstly, MVO assumes that asset returns follow a finite-variance distribution and higher moments are monotonic with variance. Many financial theories simplify this assumption adopting a normal distribution, and consequently the models that utilize such theories do not account for extreme market situations. Moreover, the variance of a normal distribution as a risk measure does not distinguish between upside and downside moves in the market. Secondly, the MVO rely on the inversion of a covariance matrix and this operation makes the method highly sensitive to estimation error especially when the covariance is estimated on a relatively short time-period and when such a past period is not representative of the future. Indeed, historical financial market data is never a good representation of the true underlying distribution as the observations are often partial. Furthermore, most MVO implementations are assuming market stationarity, which is that the mean and variance are assumed constant in each asset, while the correlation is static between assets. MVO is designed to avoid unsystematic risks by optimizing diversification. However, the systematic risks from market movements are not addressed by the MVO methodology and this is usually the most significant factor for investment decisions. Lastly, a single-period investment will almost never work in reality. A constant re-allocation is vital to respond to the rapidly changing environment.

## 2.2. Fat-tailed and asymmetric return distributions

Normal distributions do not represent well the observed probability distribution of financial market's assets prices returns. Indeed, they instead have a larger number of small returns than what expected from normal statistics, but also a larger number of very large positive and negative returns of sizes that would be impossible with normal statistics (Küchler et al., 1999; Limpert & Stahel, 2011; Officer, 1972). They also have often asymmetric distributions with larger negative returns (losses) more likely than large positive ones (gains). Several alternative probability distributions have been used in the literature, namely, Student-t (Peiró, 1994; Platen & Rendek, 2008; Student, 1908), Laplace (Eltoft et al., 2006; Norton, 1984; Wilson, 1923) and Pareto-Levy (Coronel-Brizio & Hernández-Montoya, 2005; Mandelbrot, 1960) distributions. In addition, alternative approaches to account for asymmetry have been taken into account, with early works by Markowitz himself which in 1959 (Stuart & Markowitz, 1959) employed semi-covariance (the covariance from negative returns only) as a better risk measure to better describe the downside market moves. Furthermore, limited sample size is a critical contributing factor to estimation errors. Yet, simply extending the sample size introduces data from events happened far in the past which are likely to be less representative of present market conditions. Hence, methods ranging from shrinkage (Ledoit & Wolf, 2003), to LASSO regularization (Friedman et al., 2008; Tibshirani, 1996), and Monte Carlo based re-sampling (Carsey & Harden, 2013; Manly, 2020) have been used to reduce this issue.

## 2.3. Non-stationarity and dynamic portfolio allocation

Assumptions regarding market stationarity and portfolio re-allocation are often considered together, since multi-period investment is proven to be an effective solution to mitigate the effect of Market turmoil. Several contributions have shown that dynamic re-allocation brings improvements in the resilience to market volatility with respect to the original single-period portfolio diversification methods (Berger & Ofek, 1995; DeMiguel et al., 2009; Schmitt, 2016). Nonetheless, such methods still fail to address structured market movements. Indeed, accounting for such changes requires to forecast the future market state. Further studies on market states has been proposed to model and predict the intrinsic properties of these dynamics, and two main streams are discussed below. The first one uses Markov decision process to model the transition probability between different market regimes. Currently, Hidden Markov Model (HMM) has demonstrated great efficiency and validity (Guo & Ching, 2021; Hamilton, 1989). However, it often encounters problems mainly associated with the curse of dimensionality, as the dimensionality of hidden states is linear to the number of assets considered (Ang & Bekaert, 2003; Ramchand & Susmel, 1998). On the other stream, researchers believe that market comprises mixed multivariate distributions, and each state effectively corresponds to a distribution. Hence, temporal clustering methods such as Gaussian Mixture (Buckley et al., 2008; Ketter et al., 2009; Reynolds, 2009), K-Nearest Neighbours (KNN) (Delany, 2007; Kumar et al., 2018; Nayak et al., 2015) have been applied for this purpose. Then, portfolios can be re-adjusted according to the predicted state with a selected re-allocation period. Yet, these methods often based on strong assumptions and they are not originally designed for time-series, which results issues e.g., Gaussian Mixture assumes Gaussian nature in all the base distributions, and KNN overlooks temporal consistency between single data point. This is also to some extent the approach of the decision-theoretic Bayesian methods (Berry et al., 2010; Hee et al., 2016; Reyes & Ghosh, 2013), such as the Black-Litterman model (Black & Litterman, 1991, 1992), which includes in the optimization a Bayesian prior on the future state. The ICC-PO approach introduced in the present paper is in the same line of temporal clustering methods just mentioned. However, in our case the temporal clustering is the ICC method and we make no use of Gaussian Mixtures.

The traditional Markowitz model optimizes on a single-period only, and it relies heavily on the assumption of constant asset mean vectors and covariance matrix. Therefore, this static and long investment horizon is inadequate in a dynamic market place. Yet, the mean-variance criteria inspires the development in multi-period dynamic portfolio construction. The dynamic portfolio optimization field currently follows two main streams. A discrete-time model was proposed by Samuelson in 1969 (Samuelson, 1969) and developed since by Hakansson, Grauer and others (Grauer & Hakansson, 1993; Hakansson, 1971; Pliska, 1997). It separates an investment horizon into discrete periods, and the portfolio can be reallocated at the end of each period. In contrast, a continuous-time model was introduced by Merton (1969a) in the same year, and together with further studies described the continuous rebalancing of securities for a fixed planning horizon (Bajeux-Besnainou et al., 2001; Karatzas et al., 1987; Merton, 1971).

The two alternative assumptions that are often made in dynamic portfolio optimization problems are market completeness and investment horizon. A complete market is an approximation to the real market where friction, transaction costs and asset liquidity exist, and dynamic portfolio has to consider those real world factors (Björk, 1999; Cox & Huang, 1989; Jiao & Pham, 2011; Merton, 1969b). A more ideal scenario is instead the incomplete market where some conditions are waived so that research can only focus on dynamic asset selection process and ignore some practical issues (Bielecki & Jang, 2006; Brennan & Xia, 2000; Jiao et al., 2013; Karatzas, 1989). Similarly, infinite horizon is a naive assumption to finite horizon where investors will withdraw investment with an exit time. The earlier pioneers (Cox & Huang, 1989; Hakansson, 1971; Karatzas et al., 1987) in this field, such as Samuelson (1969) and Merton (1969a, 1971), began with the infinite horizon assumption, while later researchers in the 90s and the beginning of the millennium (Castañeda-Leyva & Hernández-Hernández, 2005; Jiao et al., 2013; Jiao & Pham, 2011; Zariphopoulou, 2001) led by He and Pearson (1991) and Karatzas et al. (1991) started to introduce the finite horizon into the problem.

## 2.4. Market states clustering

After the initial pitfall of Markov-Model-based methods (Ang & Bekaert, 2003; Hamilton, 1989; Ramchand & Susmel, 1998), mainly due to the curse of dimensionality, literature has started to look for alternative methods to cluster similar temporal data points into a same group based on certain comparison criteria (Khedmati & Azin, 2020). Such temporal clustering methods can mostly be divided into two approaches: subsequent clustering and point clustering. Subsequent clustering uses a sliding window to capture a period of data points and analyse for recurrent patterns (Aghabozorgi et al., 2015; Zolhavarieh et al., 2014). The four main methods of subsequent clustering are: (i) hierarchical (Bhattacharjee et al., 2019; Navarro et al., 1997; Zeitsch, 2019); (ii) partitioning (Dolnicar, 2002; Madhulatha, 2012); (iii) density-based (Campello et al., 2013; Liu & Cao, 2020; Tigani et al., 2019) and; (iv) pattern discovery (Ahn et al., 2005; Aitken et al., 1995; Li et al., 2002). These method have all shown applicability to financial data analysis and portfolio construction. An alternative approach is point clustering that, instead of measuring spatial similarity between two slices of time-series, it looks at each temporal point individually, and assigns this multivariate observation to an appropriate cluster based on distance metrics (Alkhatib et al., 2013; Cheng, 1995; Kanungo et al., 2002). Hence, in point clustering, the choice of distance is core. In macroeconomics, the market states are not the representation of solely upward or downward trends of the market, but also the relative dynamics of equity prices, which naturally makes correlations a convenient choice of collective dynamics. A stationary correlation structure was assumed as the common approach in the industry in the 90s (Black & Litterman, 1992; Duffie & Pan, 1997a), which was, however, later shown to be overly presumptive (Ang & Bekaert, 2002; Lin et al., 1994;



Musmeci et al., 2016). Consequently, research has been devoted to study time-varying correlations. Models, such as Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Bollerslev, 1990) and the Dynamic Conditional Correlation (DCC) (Engle, 2002) have been proposed for simulating and predicting this dynamical correlation. However, most of these models suffer from the curse of dimensionality and can only be applied to a limited number of assets, as numbers of parameters increases super-linearly with the number of variables.

In 2017, Hallac et al. proposed the Toeplitz Inverse Covariance Clustering (TICC) (Hallac et al., 2018) algorithm, originally devised for electric vehicles action sensor. It classifies states based on the likelihood measures of short subsequences of observations and corresponding sparse precision matrix. After clustering, the precision matrix of each state is estimated under a Toeplitz constraint. Inspired by TICC, Procacci and Aste in 2020 (Procacci & Aste, 2018) proposed a closed related methodology names Inverse Covariance Clustering (ICC). This approach provides a point clustering of observations also enforcing temporal consistency by penalizing switching between states. The ICC method also uses sparse precision matrices but sparsification is attained via information filtering networks (see next Subsection). One main advantage of ICC, compared to TICC, is its flexibility in the selection of similarity measures. It was also stated in their original paper that different clustering distances separate market states differently. For example, likelihood distance distinguishes better with pre- and post-crisis period, Euclidean distance discriminates well between bull and bear states, and Mahalanobis distance is a mixture of the above.

### 2.5. Information filtering networks

Many computational methods employ sparse approximation techniques to estimate the inverse covariance matrix. The sparsification is effective because the least significant components in a covariance matrix are often largely prone to small changes and can lead to instability. Sparsified models filters out these insignificant components, and thus improve the model resilience to noise. A widely used approach for inverse covariance sparsification is GLASSO (Friedman et al., 2008) that uses  $L_1$  norm regularization. An alternative approach that uses information filtering networks was shown to deliver better results with lower computational burden and larger interpretability (Barfuss et al., 2016). Information filtering network analysis of complex system data has advanced significantly in the past few years. Its aim is to model interactions in a complex system as a network structure of elements (vertices) and interactions (edges). The first and best know approach is the Minimum Spanning Tree (MST) that was firstly introduced by Boruvka in 1926 (Nešetřil et al., 2001) and it can be solved exactly (see Kruskal (1956) and Prim (1957) for two common approaches). The MST reduces the structure to a connected tree which retains the larger correlations. To better extract useful information, Aste and Matteo (2017) and Tumminello et al. (2005) introduced the use of planar graphs in the Planar Maximally Filtered Graph (PMFG) algorithm. Recent studies have extended the approach to chordal graphs of flexible sparsity (Massara & Aste, 2019; Massara et al., 2017). Research fields ranging from finance (Barfuss et al., 2016) to neural systems (Telesford et al., 2011) have applied this approach as a powerful tool to understand high dimensional dependency and construct a sparse representation. It was shown that, for chordal information filtering networks, such as the Triangulated Maximally Filtered Graph (TMFG) (Massara et al., 2017), one can obtain a sparse precision matrix that is positively definite and has the structure of the network paving the way for a proper  $L_0$ -norm topological regularization (Aste, 2020). This approach has been proved to be computationally more efficient and stable than GLASSO (Friedman et al., 2008), especially when few data points are available (Aste & Matteo, 2017; Barfuss et al., 2016).

### 2.6. Sparse inverse covariance for portfolio construction and market state prediction

The application of sparse inverse covariance for portfolio construction has been recently introduced in the literature (Lee & Seregina, 2020; Millington & Niranjana, 2017; Yuan et al., 2020). The general approach has been to make use of  $L_1$ -norm regularization via GLASSO (Friedman et al., 2008). In a recent paper the sparsification methodology via information filtering networks, was applied to the identification of inherent market states via ICC (Procacci & Aste, 2018). In a following paper, the sparsification with TMFG information filtering networks was applied to the portfolio construction problem revealing several advantages with respect to traditional mean-variance methods (Procacci & Aste, 2021).

## 3. Methodologies

In the present paper we combine ICC clustering with market state forecasting to be used for portfolio optimization. Let us list in this Section the main methods we use in our approach.

### 3.1. Inverse covariance temporal clustering for portfolio optimization (ICC-PO)

Let us consider a set of  $n$  assets with  $\mathbf{r}_t \in \mathbb{R}^{1 \times n}$  the vector of returns at time  $t$ . The corresponding vector of their expected values is  $\boldsymbol{\mu} = \mathbb{E}(\mathbf{r}_t) \in \mathbb{R}^{1 \times n}$  and their covariance matrix is  $\boldsymbol{\Sigma} = \mathbb{E}((\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})^\top) \in \mathbb{R}^{n \times n}$ . The ICC clustering method depends on the choice of a gain function,  $G_{i,k}$ , which is a measure which qualifies the gain when the time  $t$  returns,  $\mathbf{r}_t$ , are associated with cluster  $k$ . Indeed, the ICC approach gathers together in cluster  $k$  observations that have the largest gain in such a cluster with respect to any other cluster:  $G_{i,k} > G_{i,h}$  for all  $h \neq k$ . For instance, in Procacci and Aste (2018) it was used

$$G_{i,k}^{Eu} = -(\mathbf{r}_t - \hat{\boldsymbol{\mu}}_k)(\mathbf{r}_t - \hat{\boldsymbol{\mu}}_k)^\top \quad (1)$$

where  $\hat{\boldsymbol{\mu}}_k$  is the sample mean return computed from the observations in cluster  $k$ . This gain is minus the square of the euclidean distance between the observation and the centroid of cluster  $k$ . A distance associated with the likelihood for multivariate normal distributions is instead

$$G_{i,k}^{No} = \frac{1}{2} \ln |\hat{\boldsymbol{\Sigma}}_k^{-1}| - n \frac{d_{i,k}^2}{2}, \quad (2)$$

with

$$d_{i,k}^2 = (\mathbf{r}_t - \hat{\boldsymbol{\mu}}_k)^\top \hat{\boldsymbol{\Sigma}}_k^{-1} (\mathbf{r}_t - \hat{\boldsymbol{\mu}}_k) \quad (3)$$

the Mahalanobis distance where  $\hat{\boldsymbol{\Sigma}}_k$  is the sample covariance computed from the observations in cluster  $k$ . While for the multivariate Student-t one has

$$G_{i,k}^{St} = \frac{1}{2} \ln |\hat{\boldsymbol{\Sigma}}_k^{-1}| - \frac{\nu + n}{2} \ln \left( 1 + \frac{d_{i,k}^2}{\nu} \right) \quad (4)$$

where, in this case,  $\hat{\boldsymbol{\Sigma}}_k^{-1}$  is the scale matrix, which is  $1 - 2/\nu$  times the sample covariance when  $\nu > 2$  and the covariance is defined (note that also  $\hat{\boldsymbol{\Sigma}}_k$  in the expression for  $d_{i,k}^2$  is the scale matrix, in this case).

We extensively tested all these gain functions observing that  $G_{i,k}^{Eu}$  is particularly efficient in selecting clusters with prevalence of positive or negative returns but it is performing poorly in the portfolio optimization problem. The normal and Student-t likelihood related gains have similar performances, but  $G_{i,k}^{St}$  turns out to be in average superior and we adopted it for the experiments we present in this paper. We also tested an hybrid distance  $G_{i,k} = c_1 \ln |\hat{\boldsymbol{\Sigma}}_k^{-1}| - c_2 d_{i,k}^2$  with the two arbitrary constants,  $c_1$  and  $c_2$ , that allow to gauge between the effects of the determinant of the covariance (an entropic term) and the Mahalanobis distance term. The measure of the determinant of a covariance is an equivalent estimation of the differential entropy of the multivariate

system, while Mahalanobis distance measures distance between points and distributions.

ICC approach uses sparse inverse covariance that was shown to improve considerably results over the full covariance. As sparsification technique we used the sparse inverse constructed with TMFG information filtering graphs (Massara et al., 2017) using the local-global (LoGo) inversion procedure described in Barfuss et al. (2016), where the elements of the inverse are computed by inverting local sample covariance matrices from only four variables at the time and adding them up. The result is a sparse inverse covariance with  $3n - 6$  non-zero entries in the upper diagonal (instead of  $n(n - 1)/2$  in the full matrix). Such a matrix is positively defined, if the number of observations is larger than four, independently on the size of the whole matrix ( $n \times n$ ). Sparse portfolios are simply obtained by applying a portfolio optimization method (see next subsection) with a sparse inverse covariance instead of a full covariance as input.

A final key element of the ICC methodology is the temporal consistency of the cluster that is imposed by penalizing frequent switches between clusters. In this paper the penalizer parameter  $\gamma$  is estimated in the train set through a grid search so that the average cluster persistence is of a given length (30 days in this paper).

The assignment of the temporal instance  $t$  to a cluster number,  $k_t$ , is performed iteratively starting from an initial random cluster assignment. Specifically we evaluate the penalized gain

$$\tilde{G}_{t,k_t} = G_{t,k_t} - \gamma \delta_{k_{t-1},k_t}, \quad (5)$$

and assign observation  $t$  to the cluster with largest penalized gain. In the previous expression,  $\delta_{k_{t-1},k_t}$  is the Kronecker delta returning one if  $k_{t-1} = k_t$  and zero otherwise. After the assignment of the time- $t$  observation to a given cluster  $k_t$ , all cluster parameters (means and covariances) are recomputed with the new cluster assignments.

We then performed a mean variance portfolio optimization method independently for each ICC state. Obtaining optimal weights associated with each temporal cluster. To apply effectively such optimized weights to the portfolio problem we have to forecast the state that is most likely to be predominant in the future test set where the investment is performed. For this purpose we made use of the short term persistence of such states and we assigned as most likely future state the one that is predominant in the last part of the train set. In this paper we consider two clusters only.

### 3.2. Portfolio optimization methods

Our proposed methodology is made of three main stages. First, we use ICC for temporal clustering the train dataset into two market states. Second, we forecast which of the two states will be predominant in the future, test dataset, where the investment is made. Third, we perform portfolio optimization using train data from the forecasted predominant ICC state. Our approach is, to large extent, agnostic to the kind of optimization adopted. In this paper, for the experiments, we used two, mean-variance optimizations methods: 1. the Sequential Least Square Quadratic Programming approach and; 2. the Critical Line Algorithm method. Let us briefly recall the basic elements of these two portfolio optimization methods.

For the experiments in this paper Markowitz's optimal weights can be computed with the python package 'Numpy' for direct matrix multiplication. The exact solution is shown in Appendix E. In the literature, this solution is referred to as 'unconstrained' because, beside the normalization and average conditions, the weights have no other constraints. On the other hand, in some practical cases, one might want to add further conditions to the weights. For instance, many real world situations do not allow short selling, which hence makes necessary to impose only positive weights in the range  $w_i \in [0, 1]$ . This constrained optimization problem cannot be any longer analytically solved and numerical optimization methods must be adopted.

Two numerical optimization methods have been adopted in the experiment. The **sequential least square quadratic programming (SLS)** (Boggs & Tolle, 1996; Kraft, 1988; Nocedal & Wright, 2006) is considered to be one of the most efficient computational method to solve general nonlinear constrained optimization problems. Jackson et al. and Cesarone et al. demonstrate its effectiveness in finance (Cesarone et al., 2015; Jackson & Staunton, 1999). There is an easy-to-use package implemented in Python's SciPy.optimize library which we applied in our experiments. The **Critical Line Algorithm (CLA)** is an efficient alternative to the quadratic optimizer for mean-variance model, as it is specifically designed for inequality portfolio optimization. It was already originally introduced in the Markowitz Portfolio Selection paper (Markowitz, 1952), and its computational implementation has become increasingly popular (Markowitz et al., 2020; Singh et al., 2016). CLA also solves constrained problems with conditions in inequalities, but unlike SLS, it divides a constrained problem into series of unconstrained sub-problems. In our experiment, to compute CLA optimization for portfolio selection, we leveraged the implementation from the open-source **portfoliolab** Python library from Hudson and Thames (Bailey & de Prado, 2013). A key drawback of CLA is called the Curse of Markowitz, which is that a small change can lead to a very unstable inverse covariance matrix calculation. Our employment of sparse inverse covariance matrices via information filtering network produces more robust results that are more resilient to noise produced by small changes, and the overall model deliver better performances with respect to the model with full inverse covariances. Mathematical and algorithmic details of SLS and CLA are included in Appendix E for reference.

In summary, these two portfolio optimization methodologies output optimal portfolio weights  $\mathbf{W}$  from an input constituted of: (i) a set of observations  $\mathbf{r}_t$ ; (ii) a vector of mean returns  $\boldsymbol{\mu}$ ; (iii) a covariance  $\boldsymbol{\Sigma}$ . As we shall see shortly, in our implementation these inputs are provided in various combinations including selecting from ICC states and sparsifying.

## 4. Implementation

### 4.1. Data

We carried out several experiments using historical financial time-series data from three major capital markets: NASDAQ, FTSE and HS300. We selected 100 stocks from each of these three markets during the trading period between 01/01/2010 and 01/01/2020. For each stock, we calculated the daily log-return,  $r_t(t) = \log(P_t(t)) - \log(P_t(t-1))$ , using closing prices. For the 100 stocks, in the main paper, we selected the largest market capitalization constituents but in the appendix we repeat the experiments with random selection obtaining comparable results.

We have performed sensitivity analysis on the number of stocks in each case. The result suggests that the effectiveness of our proposed method is consistent when the number of stocks is larger than 25. The market state clustering algorithm, ICC, is a high dimensional multivariate method, which is usually applied on a dataset with high dimensionality. Therefore, the choice of 100 stocks in our experiment is well justified.

### 4.2. Experiments

The optimal portfolio weights are obtained from the data in the train set and performances are measured over the test set where portfolio weights are left unchanged. As performance indicators, we compute portfolio return, portfolio standard deviation (i.e. volatility) and Sharpe ratio over the investment horizon (test period). We report the annualized value of these quantities, estimated as the daily values multiplied by  $\sqrt{252}$ . For statistical robustness, for each market, we compute the above portfolio performance indicators over 100 randomly chosen

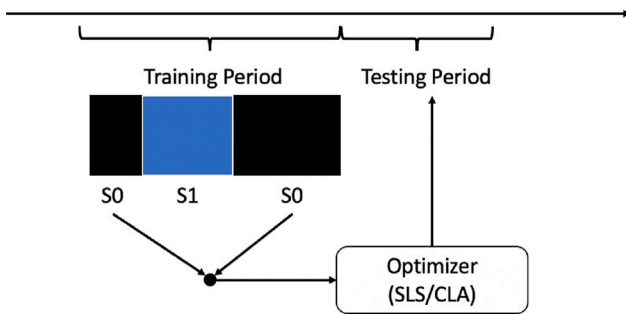


Fig. 1. The workflow schematics of ICC-PO, where S0 represents 'Sparse 0' state and S1 represents 'Sparse 1' state. S0 is defined by the most abundant state among the last 20 days of observations at the end of the train period. The optimal portfolio is obtained by training the optimizer solely on concatenated S0 data from sub periods, e.g., the two black-coloured sub periods labelled 'S0', and the performance is evaluated in the testing period immediate after the training period.

consecutive train-test periods within the ten years dataset. Results are reported for the mean performances and the 5%–95% quantile ranges over such re-sampling.

The test set length (investment horizon) was established at 30 days which is a reasonable value for practical applications, we however also report in appendix results for horizons of 10, 20 and 100 days finding consistent results. The train set length was established by performing experiments with train sets of  $L = 0.5, 1, 2, 3, 4$  years. Fig. 2 reports the annualized average Sharpe Ratio computed on the test set as function of the train set length. One can observe from the top figure that the lengths between one and two years yields to consistent good performances. We adopted the period of 2 years as optimal compromise between statistical robustness and best performances.

In the experiments outlined by Fig. 1, we first compute, on the training set, the ICC time clusters assuming two states and Student-t log-likelihood, Eq. (4) as gain function. The choice of two states is for the seek of simplicity, we tested also 3 states obtaining inferior but comparable results. We verified that Student-t likelihood is best performing among the tested gain functions, in Appendix we report results also for Normal log-likelihoods (Eq. (2)). The switching penalty parameter  $\gamma$  in ICC was set so that the average cluster size is around 30 days, i.e. consistent with the 30-day investment horizon. This selection of average cluster size and investment horizon is a somehow arbitrary choice based on the effective threshold of the portfolio performance measured by the Sharpe Ratio. We then labelled 'Sparse 0' the state that is most abundant among the last 20 days of observations at the end of the train period. Conversely, we labelled 'Sparse 1' the other. The term Sparse is used to indicate that this portfolio uses sparse inverse covariance. To set such a 'prevalence period' of 20 days we first performed a grid search over the combination of training duration  $L = 0.5, 1, 2, 3, 4$  years, and using prevalence periods of 10, 20, 30, 100,  $L/2, L$  days. This search confirmed that small values of prevalence periods, of 10, 20, 30 days, provide better results than larger prevalence periods. We therefore set a prevalence period of 20 days as it provides the most consistent results across the grid search and it is also consistent with the length of the test set.

The bottom plot in Fig. 2 reveals that Sparse 0 has consistent better performances over Full with, best results for training periods of one year. Let us notice that, having a ICC average cluster size of 30 days, it makes hard to cluster well a small training period of six months, and often unbalanced clusters where one cluster dominates the period are obtained. On the contrary, a large training duration (4 years) makes the model prone to unnecessary patterns and noise, and in turn reduces performance. Thus we chose 1 year as best compromise for the length of the training set. In Appendices A and C we see that similar results are obtained for the other two markets (FTSE, HS300).

We compute optimal portfolios using the two (SLS and CLA) optimization methods. We trained each optimization method both on the whole train dataset and, separately, on the two Sparse 0 and Sparse 1 states. We used the sample means for each of the respective sets and either the 'full' sample covariances (Pearson's estimate) or the 'sparse' sample covariances (TMFG-LoGo estimate Barfuss et al., 2016; Massara et al., 2017). Therefore, for each optimization method we have four optimized portfolios: two computed on the whole training set and with full or sparse inverse covariance (named 'Full' and 'Sparse' respectively); two computed on the two ICC market states and with sparse inverse covariance (named 'Sparse 0' and 'Sparse 1'). For benchmarking, these portfolios are also compared to a portfolio with equal weights,  $w_i = 1/n$  named 'Naive'. Overall, we have therefore  $4 \times 2$  plus 1 differently optimized portfolios that are recomputed 100 times over randomly sampled time-periods. Such optimized portfolio weights are applied, for each of the three markets, to 100 most capitalized stocks. In appendix we repeat the experiments for randomly selected stocks.

## 5. Results

### 5.1. Log-likelihood

We computed the daily Student-t log-likelihood, using Eq. (4), for each of the 30-day investment horizon. Fig. 3, reports for the averages of the differences for each day between the log-likelihood of Sparse 0 and full and also between Sparse 1 and full. The average is taken over the 100 random re-sampling.

Fig. 3, shows mostly positive gains for Sparse 0 indicating that, for most days across the investment horizon, it has larger log-likelihoods than Full. Sparse 1 gain instead reveals mostly negative results against full. This therefore indicates that while Sparse 0 is, in average, a better model to describe the multivariate nature of the log-returns in the test set with respect to Full; instead, Sparse 1 is in average worst. Since both Sparse 0 and 1 were sparsified using TMFG, the difference between them must therefore be a consequence of clustering. One might note that, even though some Sparse 1 log-likelihoods gains are in the positive domain, they anyway have smaller magnitudes than their Sparse 0 counterparts. This result clearly shows the effectiveness and importance of considering market states.

Let us note that the two ICC clusters gather together observations that maximize in-sample log-likelihood in the respective clustered periods. The fact that these models (i.e. in-sample means and covariance) from these clusters still correspond to different log-likelihood performances in the off-sample test set indicates that the states are still relevant off-sample. Further, the better off-sample performances of Sparse 0 state indicates predictability; i.e. if one state outperforms another during the training period, it will remain better performing throughout the test period. Furthermore, to illustrate the universality of the log-likelihood results, we have included similar Student-t log-likelihood (Appendix A) as well as Normal log-likelihood (Appendix C) plots for random 100 stocks selections for all three major indices (NASDAQ, FTSE, HS300), where similar patterns are observed.

### 5.2. Portfolio performance

We tested portfolio performances over 30-day investment horizon for 100 largest market capitalization constituent stocks of NASDAQ Composite computed with the four portfolio optimization methods, SLS and CLA and using as inputs Full, Sparse, Sparse 0 and Sparse 1. We also report the  $1/n$  Naive construction for benchmark. Tables 1 and 2 reports portfolio performances for the combination of portfolio constructions (column 'Solver') and inputs (column 'State'). Performances are quantified in terms of annualized portfolio return, annualized portfolio standard deviation (volatility) and the annualized Sharpe ratio over 30-day investment horizon. We report the 5% and 95% quantiles and the means computed from the 100 random resampling of

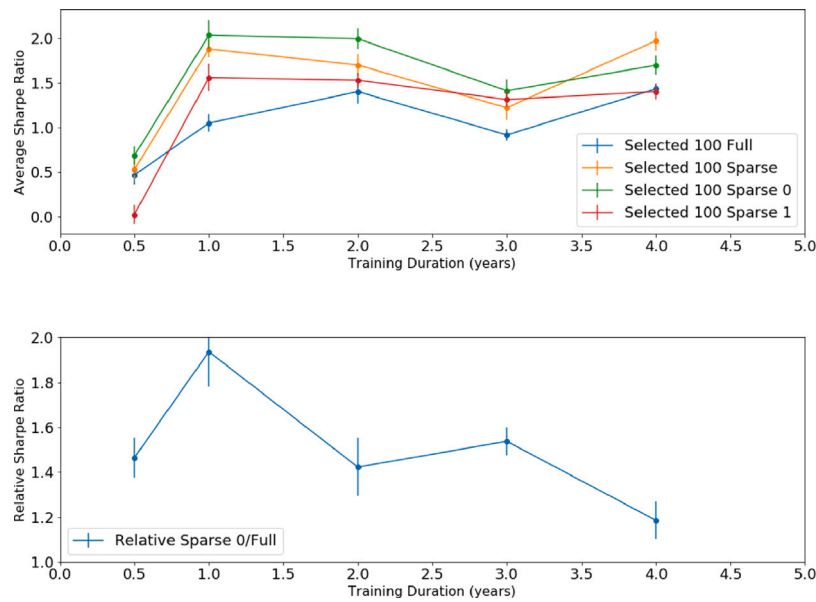


Fig. 2. Sharpe Ratio for portfolios with 100 largest market capitalization constituent stocks of NASDAQ Composite optimized using different training set durations. The top subplot reports the average Sharpe Ratios ( $SR$ ) with error bars reporting 1 standard deviation, for Full, Sparse, Sparse 0 and 1, statistics is on 100 training-testing periods chosen at random within the 10 years dataset. The bottom subplot report instead the relative Sharpe Ratios between Sparse 0 and Full,  $SR_{Sparse0}/SR_{Full}$ .

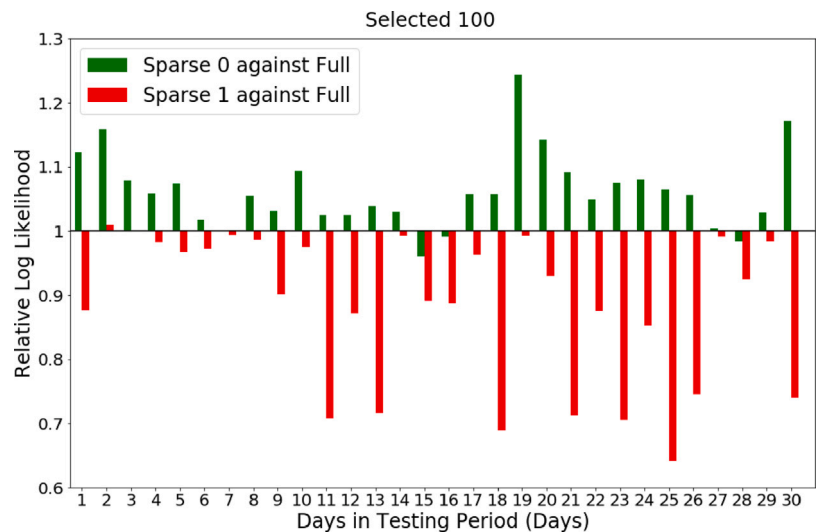


Fig. 3. Student-t log-likelihood for 100 largest market capitalization constituent stocks of NASDAQ Composite v.s. number of days in the test period after training. Each bar represents the average gain of the Sparse 0 (green) or 1 (red) with respect to the Full in each day. Averages are over 100 re-samplings.

consecutive training-investment periods chosen at random within the 10 years dataset. The maximum returns and Sharpe Ratios, as well as the minimum volatility, are highlighted in bold. Thus showing the best performer in each market-solver combination. In addition, we highlight the minimum 5th percentile returns to depict the state suffering the least loss, and the maximum 95th percentile volatility to showcase the most stable state in extreme market situations.

From the mean values reported in Tables 1 (Student-t log-likelihood), we observe that Sparse 0 outperforms Full, and this supremacy dominates for the two solvers. More specifically, Sparse 0 is on average 29.3%, 19.5% and 53.1% and 25.1% better in return, volatility and Sharpe Ratio and Sortino Ratio than Full across all two solvers. We observe instead that Sparse 1 is considerably worse than Full indicating therefore that the significant gain of State 0 come from filtering out the ‘disadvantageous’ Sparse 1 state rather than sparsification. This is indeed confirmed by the small observed gains of Sparse over Full. these results are confirmed by the analysis of the 5th

and 95th quantiles where we notice that Sparse 0 consistently achieves the least minimum extreme loss and the least maximum extreme risk. Specifically, Sparse 0 on average loses 66.0% less and is 13.6% less volatile than Full on 5th percentile return and 95th percentile volatility respectively. In other words, the integrated clustering portfolio optimization algorithm, ICC-PO, that we proposed can boost returns with less risk than the traditional benchmark, as well as provide extra resilience in extreme market situations.

To test the sensitivity of this method to the specific ICC clustering gain function, we performed the same analysis using normal log-likelihood gain function for ICC clustering. Results are reported in Table 2. Consistently with the previous results, we observe 28.6%, 18.8% and 42.0% and 44.0% improvements in return, volatility and Sharpe Ratio, with 73.1% and 10.2% gains in 5th percentile return and 95th percentile volatility. The comparison illustrates a 11.1% Sharpe Ratio improvement in Student-t log-likelihood and a 7.1% 5th percentile return advance in Normal log-likelihood. In other words,



**Table 1**

Portfolio performances obtained by using Student-t log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 30 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 30-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are 100 largest market capitalization constituent stocks of NASDAQ Composite. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state 1/n Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Solver	State	Return (%)	(5,95)th percentile	Volatility (%)	(5,95)th percentile	Sharpe	(5,95)th percentile	Sortino	(5,95)th percentile
	1/n Naive	14.46	(-36,55)	17.4	(14,28)	1.536	(-1.4,4.3)	2.379	(-1.3,4.6)
SLS	Full	22.71	(-28,96)	19.5	(14,28)	1.627	(-1.4,5.4)	2.413	(-1.2,6.1)
SLS	Sparse	21.81	(-23,74)	17.5	(14,26)	1.764	(-1.0,6.2)	2.506	(-1.4,6.8)
SLS	Sparse 0	<b>29.04</b>	<b>(-6,66)</b>	<b>16.0</b>	<b>(12,25)</b>	<b>2.478</b>	<b>(-0.3,6.9)</b>	<b>2.933</b>	<b>(-0.7,8.4)</b>
SLS	Sparse 1	5.35	(-49,57)	19.8	(14,34)	0.978	(-2.3,4.6)	1.299	(-2.7,4.7)
CLA	Full	21.97	(-69,97)	19.5	(14,31)	1.541	(-2.1,6.5)	2.378	(-2.4,7.3)
CLA	Sparse	22.27	(-32,85)	17.0	(12,27)	1.758	(-1.9,6.5)	2.521	(-1.4,6.3)
CLA	Sparse 0	<b>28.73</b>	<b>(-27,76)</b>	<b>15.8</b>	<b>(11,26)</b>	<b>2.372</b>	<b>(-1.5,7.6)</b>	<b>3.057</b>	<b>(-1.0,8.8)</b>
CLA	Sparse 1	12.48	(-57,86)	18.7	(12,32)	0.964	(-2.9,6.6)	1.533	(-2.6,6.3)

**Table 2**

Portfolio performances obtained by using Normal log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 30 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 30-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are 100 largest market capitalization constituent stocks of NASDAQ Composite. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state 1/n Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Solver	State	Return (%)	(5,95)th percentile	Volatility (%)	(5,95)th percentile	Sharpe	(5,95)th percentile	Sortino	(5,95)th percentile
	1/n Naive	14.46	(-36,55)	17.4	(14,28)	1.536	(-1.4,4.3)	2.268	(-1.1,4.6)
SLS	Full	22.98	(-28,96)	19.3	(14,28)	1.667	(-1.4,5.4)	2.240	(-0.6,5.3)
SLS	Sparse	21.96	(-23,74)	17.3	(14,26)	1.787	(-1.0,6.2)	2.346	(-0.4,6.3)
SLS	Sparse 0	<b>29.00</b>	<b>(-14,66)</b>	<b>15.9</b>	<b>(12,23)</b>	<b>2.260</b>	<b>(-0.8,4.6)</b>	<b>2.860</b>	<b>(-0.4,6.6)</b>
SLS	Sparse 1	6.84	(-43,63)	19.5	(14,30)	0.845	(-1.8,4.6)	1.536	(-1.6,3.7)
CLA	Full	20.63	(-76,97)	19.5	(14,31)	1.456	(-3.0,6.5)	1.932	(-2.4,6.3)
CLA	Sparse	21.15	(-53,85)	17.0	(12,27)	1.678	(-2.1,6.5)	1.958	(-2.4,6.3)
CLA	Sparse 0	<b>27.08</b>	<b>(-14,79)</b>	<b>15.6</b>	<b>(10,30)</b>	<b>2.175</b>	<b>(-0.7,6.6)</b>	<b>3.099</b>	<b>(-0.2,7.6)</b>
CLA	Sparse 1	11.54	(-69,77)	18.6	(14,36)	1.028	(-2.6,5.6)	1.136	(-2.9,5.3)

Student-t is a better model to the market and boosts portfolio performance. However, Normal log-likelihood generates a higher resilience to extreme loss. Since the average gain in return and volatility are similar in the two cases, the performance difference should mainly come from general upward-shifted ranges in the Student-t Sharpe Ratio. Moreover, we also observe a higher improvement in Normal Sortino Ratio than Student-t Sortino Ratio, which might be due to that the downside deviation captured by Sortino Ratio reduces.

Similar tables of optimization results using 10, 20, 30 and 100-day investment horizons, can be found in [Appendix B](#) for Student-t log-likelihood and [Appendix D](#) for Normal log-likelihood. These experiments were carried over randomly selected 100 stocks baskets (instead of the 100 most capitalized ones); the set of 100 random stocks was re-chosen for each of the 100 re-sampling. Most of the general patterns found earlier still hold regardless of the length of the testing period and underlying assets. The relative difference, namely, the gain between Sparse 0 and Full remains roughly the same. This consistency further confirms the generality of our ICC-PO model. In this case we report only the percentiles of the performance measures because being re-sampled on different constituents, mean values might be misleading.

## 6. Discussion

The results presented in Section 5.1 quantitatively demonstrate an effective gain in log-likelihood after applying temporal ICC clustering and computing the optimized sparse portfolio associated to the most persistent ICC cluster in the last 20 days of training (the Sparse 0 portfolio). We highlighted that the additional gain in the ICC-PO construction is mainly a consequence of the market states clustering and only partially consequence of sparsification. These results are extremely robust showing comparable patterns across experiments conducted for three major capital markets, using two different solving methodologies,

adopting four investment horizons and using both Student-t and normal log-likelihoods gain functions ([Appendices B and D](#)). The results in [Appendices B and D](#) obtained for 100 random stocks in the US, the UK and the Chinese markets show a broader variability but overall well aligned results. Our results also confirm the observation, by [Procacci and Aste \(2021\)](#), that models with larger likelihood better solve the portfolio optimization problem.

As for the analysis on the 100 NASDAQ's most capitalized stocks, also for the random selection and the three markets we observe that the Normal log-likelihood, Sparse 0 is 33.8% less than Full in the 5th percentile Return, whereas the Student-t log-likelihood is only 20.4%, which illustrates that the Normal statistically loses less money in extreme situations. Namely, it results that there are general advantages in using Student-t over Normal log-likelihood, yet, the latter performs a better at limiting risks. The edge in three main performance matrices depicts the Student-t's better market modelling property as suggested in the literature, especially for limited sample daily log-return. In contrast, the mere pitfall in risk measures may probably come from the fat-tail nature of the Student-t distribution.

It is difficult to assess the efficiency of ICC-PO by direct comparison to the literature, since our focused result is the relative difference between Sparse 0 from Full. The most informative measurements to the general performance used widely in the field of portfolio management are Sharpe Ratio (the risk adjusted return), Jensen's Alpha (the abnormal return over the theoretical expectation), Treynor Ratio (the risk adjusted excess return from a risk-free asset) and Roy Ratio (the risk adjusted excess return from the market index) ([Narsoo, 2017](#)). Literature identifies that Sharpe Ratio at values around 1 is commonly considered as the boundary between a good and bad investment strategy, while Sharpe Ratio at values around 2 represents an excellent standard, and 3 and above are more likely to be achieved in a High Frequency Trading (HFT) strategy ([Lo, 2002](#); [Sharpe, 1994](#)). During the 10-year we



investigated the annualized Sharpe Ratio for NASDAQ-100, FTSE-250 and HS300 have been respectively equal to 1.77, 0.42 and 1.07 (Fiance, 0000a, 0000b; Zhang, 2020). While our results for the various portfolio construction combinations generally lie in a reasonable range around these values, we note that the average Sharpe Ratio of the Sparse 0 based on the 100 largest market capitalization stocks from NASDAQ is 2.425, as well as 100 random stocks from NASDAQ is 2.132, from FTSE is 1.682 and from HS300 is 1.814 greatly exceeding the index's performances.

In addition, we also report Sortino Ratio (Sortino & Price, 1994), which is a variation of the Sharpe Ratio, as a risk-adjusted return indicator. It uses asset's downside deviation as opposed to the total standard deviation of a portfolio in the Sharpe Ratio. It is obvious that in all table, the Sortino Ratio of Sparse 0 is superior than others. Apart from Sharpe Ratio for general performance assessment, risk is often a critical consideration in portfolio investment due to the risk aversion nature of investors and the quadratic utility function assumption. Two widely used risk measures are value at risk (VaR) (Duffie & Pan, 1997b) and probable maximum loss (PML) (Ramsey & Goodwin, 2019), which are interpreted as the minimum and the maximum loss expected in a portfolio over a time period. As a proxy combination of VaR and PML, we reported the 5th percentile Return in the random re-sampling. The observed general 66.0%, 49.8%, 21.6% and 32.4% reductions in loss respectively for largest-market-capitalization NASDAQ, NASDAQ, FTSE and HS300 are a highly significant result indicating likely large improvements of both VaR and PML.

Lastly, as ICC-PO is computationally very efficient, it can be easily re-run for every allocation window making dynamic portfolio allocation easy.

## 7. Conclusions

Portfolio optimization lays at the core of quantitative investment. Automation in the dynamic allocation process is a challenging goal with a large community of academics and practitioners dedicated to this task which requires a precise and accurate modelling of the past market performance and a predictive inference of the future market state. However, it is never an easy task to predict the future, not to mention doing so constantly. Explanatory as they are, only certain signals possess the forecasting ability and normally only for a limited period of time. Hence, the results of our proposed algorithm ICC-PO are worth to be mentioned. Indeed, we improve the equal weight benchmark by over 50% in Sharpe Ratio, obtaining a statistically more robust and resilient investment performance especially in the extreme market situations with large reductions in losses.

In this paper we demonstrated that markets can be classified in different states with distinct statistical properties. By using two states, classified and clustered using log-likelihood as gain function and sparse inverse covariance estimation, we have shown that the two clustered states continue to be distinguishable in log-likelihood after the train (in-sample) period, with one having systematically larger log-likelihood than the one computed from the whole, unclustered, training sample. We have shown that the state with larger log-likelihood tends to be the one also with largest likelihood in the last period of training, indicating temporal persistence and providing a way for predictability of the best performing state in the off-sample investment period. Portfolios optimized with data from the best performing state's cluster give significantly better results than portfolios constructed from the full dataset or the other state. This also confirm the intuitive argument (see Procacci & Aste, 2021) that a model with larger likelihood must perform better for portfolio optimization purposes than a model with lower likelihood. These results were tested extensively across a period of ten years, across three different markets, with portfolios from two different optimizers, with clustering from two different log-likelihoods, and both by using a selected group of most capitalized stocks as well as by random picking

a stock basket, given the number of stocks is relatively large as the underlying ICC is most suitable in a high-dimensional space.

The choice of using two market states has been dictated by simplicity. Future work will investigate the effect of the number of ICC clusters on the results. Our results are based on a naive selection of stocks from major indices. Hence, with a carefully designed portfolio basket, as commonly done in industrial practices, we expect further improvement of the results. Also a wider application in asset classes is a straightforward extension of the method.

## CRedit authorship contribution statement

**Yuanrong Wang:** Writing – original draft, Editing, Performing experiments. **Tomaso Aste:** Writing – review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

TA acknowledges partial support from ESRC, UK (ES/K002309/1), EPSRC, UK (EP/P031730/1) and EC (H2020-ICT-2018-2 825215).

## Appendix A. Off sample log-likelihood and performances for Student-t log-likelihood construction

In this appendix we investigate the effect of the length of the train set on the Sharpe Ratio performance and off-sample (test set) log-likelihood using 100 randomly selected stocks drawn from NASDAQ, FTSE and HS300. They are in the similar format as Figs. 2 and 3 and demonstrate that identical patterns exist regardless underlying assets and capital markets. In Fig. 5, it is noticeable that the green bars in general sit above 0 and the red are below 0, which indicates the Sparse 0 has better off-sample log-likelihood than the Full, as illustrated in Fig. 3. In this appendix we compute Student-t likelihoods (see Fig. 4).

## Appendix B. Portfolio performances

In this appendix we extend the results in the main paper including 10, 20, 30 and 100-day investment horizons based on Student-t log-likelihood. Differently from the main text, portfolio are constructed with 100 random stocks drawn from NASDAQ, FTSE and HS300. They serve as complement and comparison for Table 1. It is noticeable that although shorter testing period yields numerically larger Sharpe Ratio due to a possible low Volatility and a overestimation of annualized Return on small sample, the relative difference, namely, the gain between the Sparse 0 and the Full remains roughly the same. This consistency further confirms the generality of our model. Besides, the patterns in three tables are generally consistent to the findings in Section 5.2 (see Tables 3–10).

## Appendix C. Normal log-likelihood: training duration and off-sample log-likelihood

This Appendix C section includes Sharpe ratio against training duration plots and off-sample Normal log-likelihood plots of 100 random stocks drawn from NASDAQ, FTSE and HS300. They are in the similar format as Figs. 2 and 3 and demonstrate that identical patterns exist regardless underlying assets and capital markets. In Fig. 7, it is noticeable that the green bars in general sit above 0 and the red are below 0, which indicates the Sparse 0 has better off-sample log-likelihood than the Full, as illustrated in Fig. 3 (see Fig. 6).

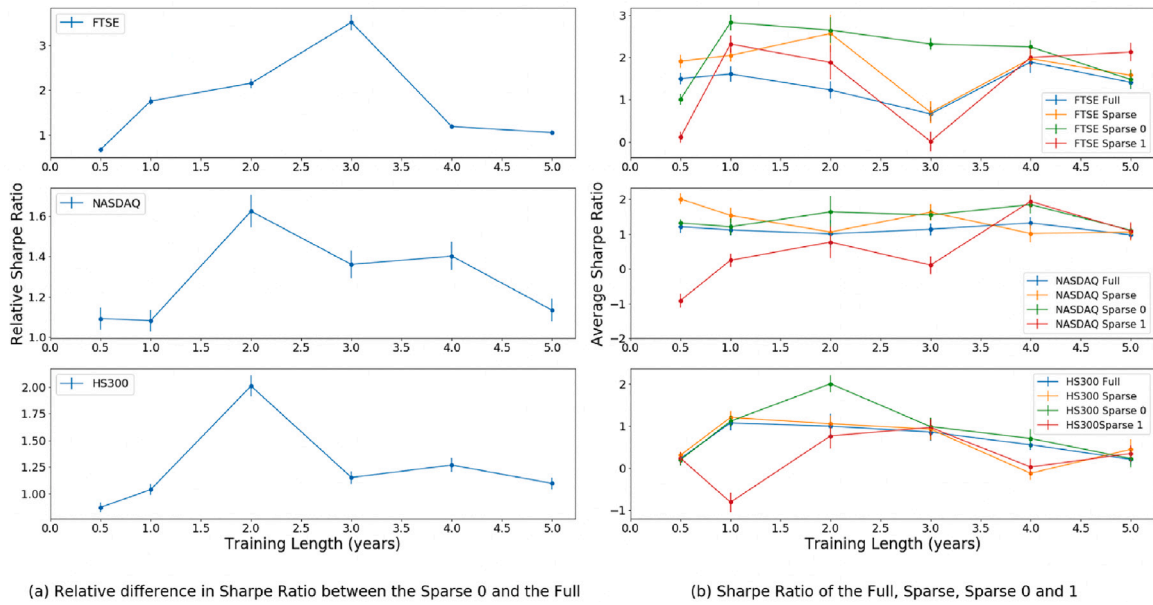


Fig. 4. Sharpe Ratio for portfolios with constituent stocks of three indices optimized using different training set durations by using Student-t log-likelihood for ICC clustering. The right subplot reports the average Sharpe Ratios ( $SR$ ) with 1 standard deviation for states, statistics is on 100 training–testing periods chosen at random within the 10 years dataset. The left subplot report instead the relative Sharpe Ratios between Sparse 0 and Full,  $SR_{Sparse0}/SR_{Full}$ .

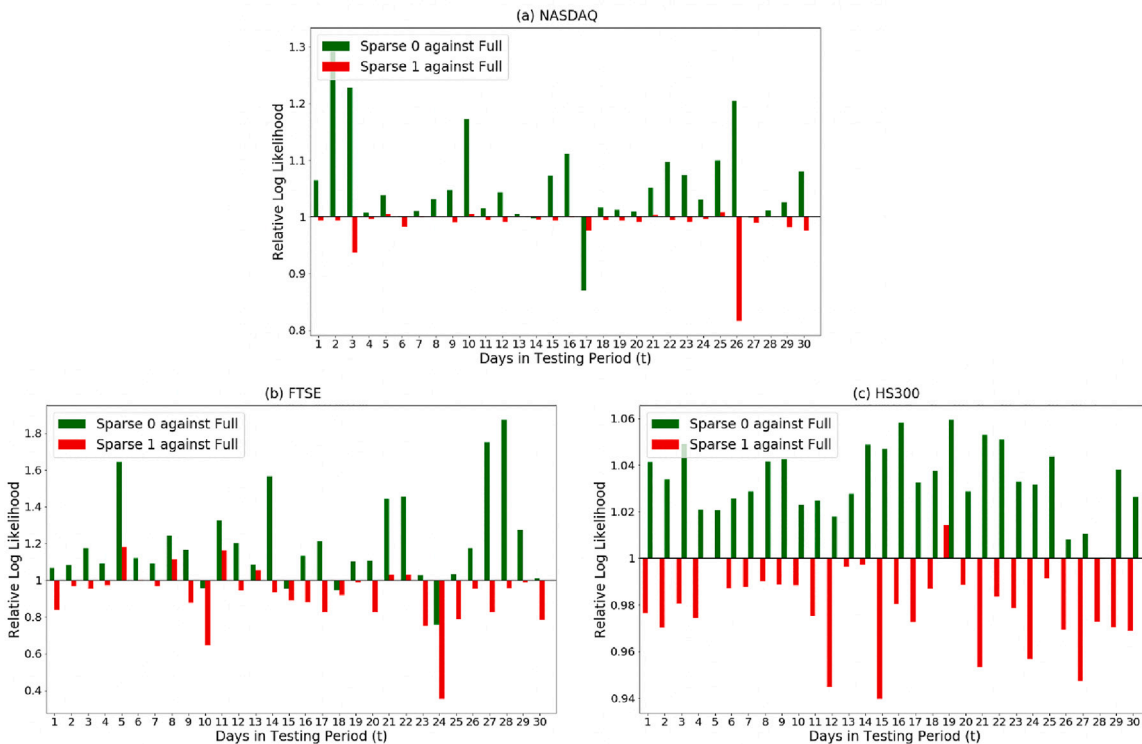


Fig. 5. Student-t log-likelihood for constituent stocks of (a) NASDAQ, (b) FTSE and (c) HS300 Composite v.s. number of days in the test period after training. Each bar represents the average gain of the Sparse 0 (green) or 1 (red) with respect to the Full in each day. Averages are over 100 re-samplings.

**Appendix D. Off sample log-likelihood and performances for Normal log-likelihood construction**

In this appendix we perform the same kind of investigations as in the previous appendix but ICC is computed using Normal log-likelihood. We notice similar patterns but the Student-t log-likelihood result are

more significant. However, the Normal log-likelihood performs better in risk matrices.

**Appendix E. Portfolio optimization**

In the original Markowitz’s mean variance optimization approach, the portfolio weights  $\mathbf{W} = (w_1, \dots, w_n) \in \mathbb{R}^{1 \times n}$  are chosen in

**Table 3**

Portfolio performances obtained by using Student-t log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 10 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 10-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-171,206)	(14.0,40.0)	(-8.7,12.5)	(-9.2,13.3)
NASDAQ	SLS	Full	(-210,138)	(13.0,69.0)	(-7.2,9.1)	(-7.9,9.0)
NASDAQ	SLS	Sparse	(-213,122)	(12.0,59.0)	(-7.6,9.1)	(-7.7,10.1)
NASDAQ	SLS	Sparse 0	(-157,125)	(9.0,52.0)	(-5.1,10.4)	(-7.1,11.0)
NASDAQ	SLS	Sparse 1	(-352,153)	(13.0,49.0)	(-7.1,7.6)	(-8.0,8.1)
NASDAQ	CLA	Full	(-198,190)	(13.0,57.0)	(-6.6,10.1)	(-7.1,10.5)
NASDAQ	CLA	Sparse	(-172,223)	(13.0,51.0)	(-8.4,10.3)	(-7.9,9.7)
NASDAQ	CLA	Sparse 0	(-181,185)	(10.0,49.0)	(-5.2,13.4)	(-6.0,13.9)
NASDAQ	CLA	Sparse 1	(-198,200)	(13.0,66.0)	(-8.3,7.4)	(-9.2,7.4)
FTSE		$1/n$ Naive	(-125,140)	(9.0,28.0)	(-9.0,17.0)	(-9.8,16.1)
FTSE	SLS	Full	(-91,147)	(7.0,27.0)	(-8.1,20.5)	(-8.2,19.5)
FTSE	SLS	Sparse	(-89,125)	(7.0,26.0)	(-8.8,15.5)	(-9.2,16.1)
FTSE	SLS	Sparse 0	(-64,150)	(8.0,22.0)	(-5.9,18.8)	(-7.2,19.2)
FTSE	SLS	Sparse 1	(-113,116)	(9.0,26.0)	(-8.9,11.6)	(-9.8,12.9)
FTSE	CLA	Full	(-147,137)	(8.0,22.0)	(-6.9,17.6)	(-7.6,17.1)
FTSE	CLA	Sparse	(-138,122)	(7.0,25.0)	(-9.0,18.0)	(-9.2,18.0)
FTSE	CLA	Sparse 0	(-119,129)	(7.0,23.0)	(-6.4,21.5)	(-7.4,22.9)
FTSE	CLA	Sparse 1	(-171,149)	(9.0,22.0)	(-10.5,16.3)	(-11.2,15.0)
HS300		$1/n$ Naive	(-228,198)	(10.0,60.0)	(-7.7,10.8)	(-8.2,10.1)
HS300	SLS	Full	(-250,216)	(12.0,42.0)	(-8.3,15.8)	(-9.5,16.1)
HS300	SLS	Sparse	(-283,252)	(11.0,44.0)	(-8.1,15.3)	(-9.4,16.1)
HS300	SLS	Sparse 0	(-181,284)	(11.0,46.0)	(-6.0,16.0)	(-7.2,17.3)
HS300	SLS	Sparse 1	(-317,192)	(13.0,58.0)	(-7.9,9.9)	(-7.9,8.3)
HS300	CLA	Full	(-250,216)	(12.0,42.0)	(-8.3,15.8)	(-8.2,14.1)
HS300	CLA	Sparse	(-283,252)	(11.0,44.0)	(-8.1,15.3)	(-7.2,14.1)
HS300	CLA	Sparse 0	(-194,284)	(11.0,41.0)	(-4.9,14.2)	(-4.2,15.5)
HS300	CLA	Sparse 1	(-277,223)	(13.0,53.0)	(-9.2,8.6)	(-9.0,8.9)

**Table 4**

Portfolio performances obtained by using Student-t log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 20 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 20-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-112,180)	(14.0,84.0)	(-4.7,5.9)	(-5.7,5.3)
NASDAQ	SLS	Full	(-126,131)	(14.0,85.0)	(-4.4,6.5)	(-5.5,6.5)
NASDAQ	SLS	Sparse	(-133,127)	(13.0,53.0)	(-5.1,6.5)	(-5.1,6.5)
NASDAQ	SLS	Sparse 0	(-124,120)	(12.0,83.0)	(-4.4,7.5)	(-4.7,7.9)
NASDAQ	SLS	Sparse 1	(-198,78)	(15.0,50.0)	(-6.1,4.5)	(-4.9,4.8)
NASDAQ	CLA	Full	(-147,127)	(13.0,84.0)	(-4.7,5.4)	(-4.9,6.9)
NASDAQ	CLA	Sparse	(-149,135)	(13.0,53.0)	(-4.6,6.4)	(-5.3,5.1)
NASDAQ	CLA	Sparse 0	(-101,144)	(12.0,46.0)	(-2.8,7.9)	(-3.9,8.8)
NASDAQ	CLA	Sparse 1	(-187,88)	(15.0,62.0)	(-5.4,4.1)	(-5.7,4.9)
FTSE		$1/n$ Naive	(-77,111)	(11.0,26.0)	(-4.7,8.1)	(-4.9,7.9)
FTSE	SLS	Full	(-58,94)	(9.0,26.0)	(-4.5,11.4)	(-5.7,11.3)
FTSE	SLS	Sparse	(-59,95)	(10.0,25.0)	(-4.9,10.6)	(-5.5,10.7)
FTSE	SLS	Sparse 0	(-39,96)	(9.0,17.0)	(-3.4,11.4)	(-4.3,12.9)
FTSE	SLS	Sparse 1	(-72,75)	(9.0,22.0)	(-5.4,7.5)	(-4.8,6.9)
FTSE	CLA	Full	(-82,84)	(10.0,25.0)	(-5.6,10.3)	(-4.1,11.9)
FTSE	CLA	Sparse	(-62,81)	(10.0,20.0)	(-6.2,10.1)	(-4.8,11.9)
FTSE	CLA	Sparse 0	(-59,79)	(9.0,22.0)	(-4.3,11.7)	(-4.7,12.7)
FTSE	CLA	Sparse 1	(-100,80)	(10.0,23.0)	(-6.0,9.0)	(-5.7,8.6)
HS300		$1/n$ Naive	(-102,236)	(11.0,44.0)	(-4.4,9.3)	(-5.7,8.9)
HS300	SLS	Full	(-133,246)	(15.0,42.0)	(-5.1,10.7)	(-5.7,10.2)
HS300	SLS	Sparse	(-125,234)	(14.0,42.0)	(-5.0,10.3)	(-5.6,10.5)
HS300	SLS	Sparse 0	(-101,218)	(13.0,40.0)	(-2.8,11.3)	(-2.7,13.9)
HS300	SLS	Sparse 1	(-142,202)	(13.0,46.0)	(-5.2,7.9)	(-5.7,7.8)
HS300	CLA	Full	(-133,246)	(15.0,42.0)	(-5.1,10.7)	(-5.7,11.7)
HS300	CLA	Sparse	(-125,234)	(14.0,42.0)	(-5.0,10.3)	(-5.7,11.4)
HS300	CLA	Sparse 0	(-62,247)	(12.0,41.0)	(-2.8,10.2)	(-3.6,12.1)
HS300	CLA	Sparse 1	(-131,187)	(14.0,48.0)	(-5.0,8.2)	(-5.6,9.9)

**Table 5**

Portfolio performances obtained by using Student-t log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 30 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 30-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-110,62)	(13.0,35.0)	(-4.5,5.7)	(-4.9,6.5)
NASDAQ	SLS	Sparse	(-133,115)	(14.0,73.0)	(-4.7,4.7)	(-4.5,4.9)
NASDAQ	SLS	Sparse 0	(-99,112)	(14.0,71.0)	(-2.9,6.4)	(-3.3,6.6)
NASDAQ	SLS	Sparse 1	(-128,72)	(14.0,63.0)	(-4.4,4.4)	(-4.1,4.7)
NASDAQ	CLA	Full	(-87,121)	(16.0,69.0)	(-3.0,5.8)	(-3.2,5.6)
NASDAQ	CLA	Sparse	(-86,130)	(15.0,72.0)	(-2.6,6.2)	(-3.5,7.7)
NASDAQ	CLA	Sparse 0	(-40,134)	(14.0,74.0)	(-2.5,6.5)	(-3.1,7.9)
NASDAQ	CLA	Sparse 1	(-101,86)	(15.0,73.0)	(-3.3,4.0)	(-2.5,4.7)
FTSE		$1/n$ Naive	(-69,90)	(11.0,28.0)	(-3.0,6.6)	(-3.5,6.8)
FTSE	SLS	Full	(-63,80)	(10.0,26.0)	(-4.7,8.0)	(-4.5,7.7)
FTSE	SLS	Sparse	(-56,73)	(10.0,22.0)	(-4.8,8.3)	(-4.5,8.1)
FTSE	SLS	Sparse 0	(-52,87)	(9.0,20.0)	(-3.5,7.3)	(-3.7,8.7)
FTSE	SLS	Sparse 1	(-68,62)	(11.0,22.0)	(-4.5,6.3)	(-4.9,5.1)
FTSE	CLA	Full	(-56,79)	(10.0,24.0)	(-4.8,8.0)	(-4.4,9.7)
FTSE	CLA	Sparse	(-53,73)	(10.0,20.0)	(-4.6,9.1)	(-4.5,8.3)
FTSE	CLA	Sparse 0	(-47,72)	(9.0,20.0)	(-4.2,9.0)	(-4.3,10.1)
FTSE	CLA	Sparse 1	(-81,64)	(11.0,24.0)	(-5.8,7.1)	(-4.8,8.9)
HS300		$1/n$ Naive	(-90,172)	(11.0,38.0)	(-3.3,7.1)	(-4.4,6.9)
HS300	SLS	Full	(-127,173)	(16.0,40.0)	(-4.0,6.3)	(-5.1,7.2)
HS300	SLS	Sparse	(-98,160)	(15.0,36.0)	(-4.1,7.2)	(-4.5,7.7)
HS300	SLS	Sparse 0	(-65,172)	(13.0,43.0)	(-2.9,7.3)	(-3.2,7.7)
HS300	SLS	Sparse 1	(-118,142)	(15.0,45.0)	(-3.8,5.7)	(-4.6,6.9)
HS300	CLA	Full	(-127,173)	(16.0,40.0)	(-4.0,6.2)	(-5.1,6.3)
HS300	CLA	Sparse	(-98,160)	(15.0,36.0)	(-4.1,7.2)	(-5.0,6.7)
HS300	CLA	Sparse 0	(-64,173)	(13.0,38.0)	(-2.7,7.4)	(-4.5,6.9)
HS300	CLA	Sparse 1	(-98,142)	(15.0,36.0)	(-4.2,5.4)	(-4.9,6.0)

**Table 6**

Portfolio performances obtained by using Student-t log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 100 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 100-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-167,172)	(11.0,48.0)	(-7.6,8.4)	(-8.6,7.3)
NASDAQ	SLS	Full	(-210,138)	(13.0,69.0)	(-7.2,9.1)	(-6.6,8.2)
NASDAQ	SLS	Sparse	(-213,122)	(12.0,59.0)	(-7.6,9.1)	(-7.6,10.2)
NASDAQ	SLS	Sparse 0	(-189,160)	(11.0,41.0)	(-5.4,8.4)	(-6.6,10.0)
NASDAQ	SLS	Sparse 1	(-290,114)	(13.0,66.0)	(-8.5,5.6)	(-7.9,6.3)
NASDAQ		Full	(-198,190)	(13.0,57.0)	(-6.6,10.1)	(-7.6,10.4)
NASDAQ	CLA	Sparse	(-172,223)	(13.0,51.0)	(-8.4,10.3)	(-7.6,10.4)
NASDAQ	CLA	Sparse 0	(-165,187)	(11.0,46.0)	(-6.1,14.2)	(-6.9,15.4)
NASDAQ	CLA	Sparse 1	(-257,166)	(13.0,62.0)	(-6.8,6.6)	(-7.5,8.4)
FTSE		$1/n$ Naive	(-186,140)	(8.0,22.0)	(-10.3,23.0)	(-9.6,20.4)
FTSE	SLS	Full	(-91,147)	(7.0,27.0)	(-8.1,20.5)	(-7.6,19.0)
FTSE	SLS	Sparse	(-89,125)	(7.0,26.0)	(-8.8,15.5)	(-10.6,16.4)
FTSE	SLS	Sparse 0	(-75,161)	(7.0,21.0)	(-7.4,29.2)	(-8.6,28.4)
FTSE	SLS	Sparse 1	(-110,123)	(9.0,27.0)	(-9.5,13.6)	(-9.6,14.4)
FTSE	CLA	Full	(-147,137)	(8.0,22.0)	(-6.9,17.6)	(-7.6,16.5)
FTSE	CLA	Sparse	(-138,122)	(7.0,25.0)	(-9.0,18.0)	(-9.5,17.2)
FTSE	CLA	Sparse 0	(-80,145)	(8.0,23.0)	(-5.7,24.2)	(-7.2,23.5)
FTSE	CLA	Sparse 1	(-194,138)	(9.0,25.0)	(-12.0,16.3)	(-11.5,17.3)
HS300		$1/n$ Naive	(-228,198)	(10.0,60.0)	(-7.7,10.7)	(-7.8,11.4)
HS300	SLS	Full	(-250,216)	(12.0,42.0)	(-8.3,15.8)	(-6.9,15.5)
HS300	SLS	Sparse	(-283,252)	(11.0,44.0)	(-8.1,15.3)	(-8.6,15.3)
HS300	SLS	Sparse 0	(-237,249)	(11.0,50.0)	(-5.6,16.5)	(-4.6,18.1)
HS300	SLS	Sparse 1	(-237,193)	(14.0,45.0)	(-8.4,9.3)	(-7.7,10.8)
HS300	CLA	Full	(-250,216)	(12.0,42.0)	(-8.3,15.8)	(-8.1,15.5)
HS300	CLA	Sparse	(-283,252)	(11.0,44.0)	(-8.1,15.3)	(-8.6,15.3)
HS300	CLA	Sparse 0	(-186,298)	(10.0,44.0)	(-7.6,13.9)	(-7.6,15.3)
HS300	CLA	Sparse 1	(-302,186)	(13.0,56.0)	(-7.5,9.5)	(-7.6,19.2)



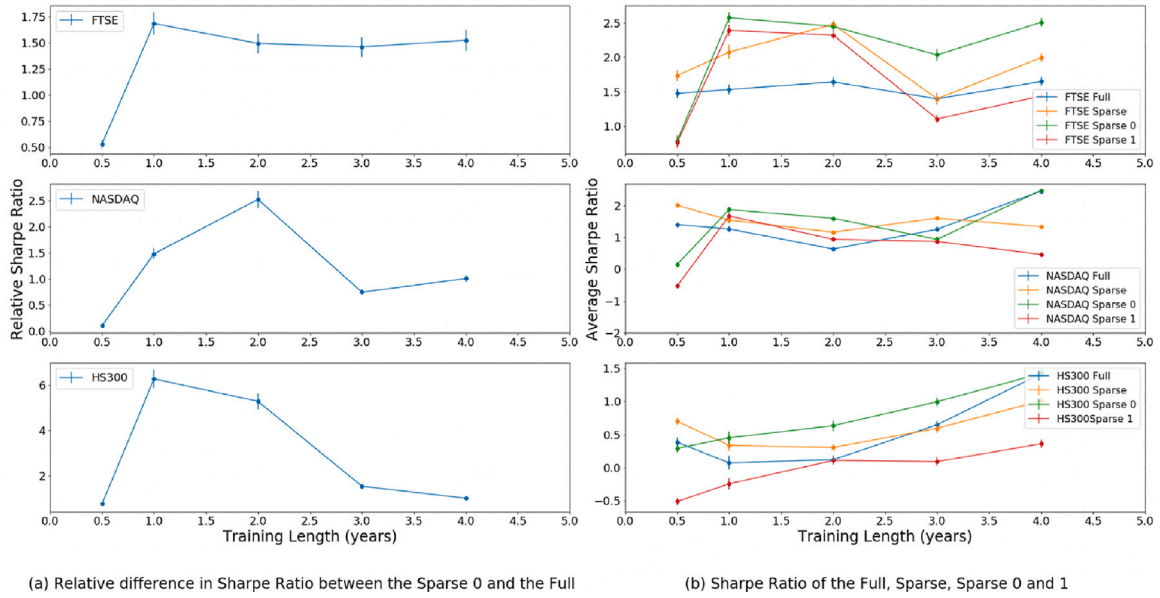


Fig. 6. Sharpe Ratio for portfolios with constituent stocks of three indices optimized using different training set durations by using Normal log-likelihood for ICC clustering. The right subplot reports the average Sharpe Ratios (SR) with 1 standard deviation for states, statistics is on 100 training-testing periods chosen at random within the 10 years dataset. The left subplot report instead the relative Sharpe Ratios between Sparse 0 and Full,  $SR_{Sparse0}/SR_{Full}$ .

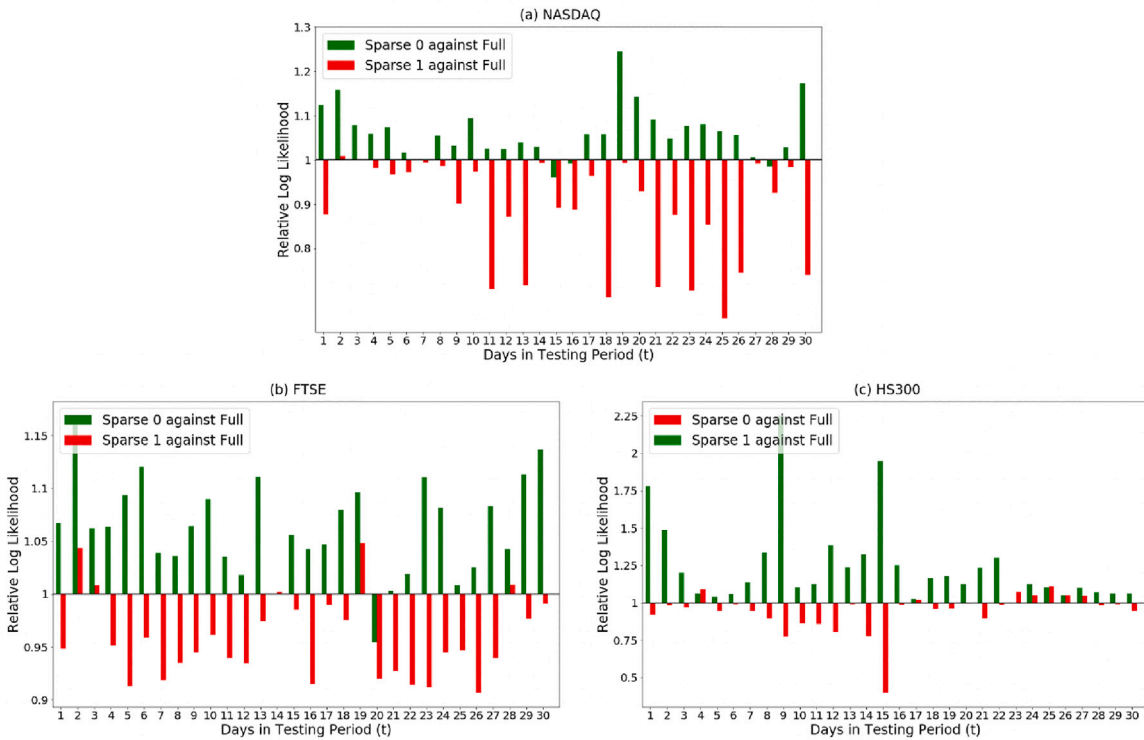


Fig. 7. Normal log-likelihood for constituent stocks of (a) NASDAQ, (b) FTSE and (c) HS300 Composite v.s. number of days in the test period after training. Each bar represents the average gain of the Sparse 0 (green) or 1 (red) with respect to the Full in each day. Averages are over 100 re-samplings.

order to minimize portfolio's variance  $\sigma_p^2 = \mathbf{W}\Sigma\mathbf{W}^T$  for a given value, of the portfolio's expected return  $\mu\mathbf{W}^T = \bar{r}_p$ . Specifically,

$$\begin{aligned} \mathbf{W}^* &= \min_{\mathbf{W}} \mathbf{W}\Sigma\mathbf{W}^T \\ \text{s.t. } & \mathbb{1}\mathbf{W}^T = 1, \\ & \text{and } \mu\mathbf{W}^T = \bar{r}_p, \end{aligned} \tag{6}$$

The exact solution can be obtained analytically by setting to zero the derivatives with respect to  $\mathbf{W}$ , using the Lagrange multiplier technique

to account for the constraints. Namely the minimum of the following Lagrangian is computed

$$L(\mathbf{W}, \lambda) = \mathbf{W}\Sigma\mathbf{W}^T + \lambda_1 \mu\mathbf{W}^T + \lambda_2 \mathbb{1}\mathbf{W}^T, \tag{7}$$

and the solution is

$$\mathbf{W}^* = \Sigma^{-1}(\lambda_1 \mu + \lambda_2 \mathbb{1})^T, \tag{8}$$

where  $\lambda_1$  and  $\lambda_2$  are the Lagrange multipliers.

**Table 7**

Portfolio performances obtained by using Normal log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 100 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 10-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-171,206)	(14.0,40.0)	(-8.7,12.5)	(-7.6,11.9)
NASDAQ	SLS	Full	(-192,190)	(13.0,57.0)	(-5.4,10.1)	(-5.7,10.3)
NASDAQ	SLS	Sparse	(-160,223)	(13.0,51.0)	(-7.4,10.3)	(-7.3,10.2)
NASDAQ	SLS	Sparse 0	(-144,174)	(11.0,49.0)	(-5.0,12.6)	(-4.7,13.6)
NASDAQ	SLS	Sparse 1	(-181,218)	(14.0,59.0)	(-8.7,7.8)	(-7.7,6.9)
NASDAQ	CLA	Full	(-192,198)	(13.0,57.0)	(-5.4,10.1)	(-4.7,10.1)
NASDAQ	CLA	Sparse	(-160,223)	(13.0,51.0)	(-7.4,10.3)	(-6.7,12.2)
NASDAQ	CLA	Sparse 0	(-169,171)	(12.0,36.0)	(-5.0,14.2)	(-4.2,15.2)
NASDAQ	CLA	Sparse 1	(-256,144)	(14.0,67.0)	(-6.5,6.5)	(-5.4,6.9)
FTSE		$1/n$ Naive	(-161,117)	(7.0,34.0)	(-9.2,15.4)	(-9.7,15.1)
FTSE	SLS	Full	(-163,116)	(8.0,28.0)	(-8.7,14.1)	(-9.7,15.6)
FTSE	SLS	Sparse	(-148,108)	(8.0,33.0)	(-8.9,14.2)	(-9.5,15.5)
FTSE	SLS	Sparse 0	(-111,138)	(7.0,22.0)	(-7.1,18.8)	(-8.0,18.0)
FTSE	SLS	Sparse 1	(-199,118)	(9.0,40.0)	(-12.1,13.0)	(-10.9,12.3)
FTSE	CLA	Full	(-163,116)	(8.0,28.0)	(-8.7,14.1)	(-6.9,14.5)
FTSE	CLA	Sparse	(-148,108)	(8.0,33.0)	(-8.9,14.2)	(-9.7,13.9)
FTSE	CLA	Sparse 0	(-111,146)	(7.0,21.0)	(-8.3,18.0)	(-6.3,20.1)
FTSE	CLA	Sparse 1	(-176,123)	(8.0,36.0)	(-11.0,13.7)	(-11.5,14.5)
HS300		$1/n$ Naive	(-228,198)	(10.0,60.0)	(-7.7,10.8)	(-8.2,11.9)
HS300	SLS	Full	(-250,216)	(12.0,42.0)	(-8.3,15.8)	(-8.7,16.5)
HS300	SLS	Sparse	(-283,252)	(11.0,44.0)	(-8.1,15.3)	(-8.1,16.5)
HS300	SLS	Sparse 0	(-165,276)	(11.0,42.0)	(-5.6,15.2)	(-6.3,16.5)
HS300	SLS	Sparse 1	(-289,176)	(13.0,44.0)	(-8.3,7.9)	(-8.7,9.5)
HS300	CLA	Full	(-250,216)	(12.0,42.0)	(-8.3,15.8)	(-8.5,14.5)
HS300	CLA	Sparse	(-283,252)	(11.0,44.0)	(-8.1,15.3)	(-9.1,16.2)
HS300	CLA	Sparse 0	(-131,250)	(11.0,45.0)	(-5.3,18.8)	(-5.7,19.8)
HS300	CLA	Sparse 1	(-331,218)	(12.0,58.0)	(-8.9,9.9)	(-7.9,10.0)

**Table 8**

Portfolio performances obtained by using Normal log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 20 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 20-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-129,147)	(14.0,36.0)	(-4.9,7.8)	(-4.5,8.1)
NASDAQ	SLS	Full	(-147,124)	(13.0,84.0)	(-4.7,5.4)	(-4.9,6.3)
NASDAQ	SLS	Sparse	(-149,135)	(13.0,53.0)	(-4.6,6.4)	(-5.1,7.4)
NASDAQ	SLS	Sparse 0	(-98,133)	(12.0,64.0)	(-3.7,8.4)	(-4.2,9.8)
NASDAQ	SLS	Sparse 1	(-147,101)	(14.0,71.0)	(-5.6,5.3)	(-5.9,6.8)
NASDAQ	CLA	Full	(-147,127)	(13.0,84.0)	(-4.7,5.4)	(-3.9,5.8)
NASDAQ	CLA	Sparse	(-149,135)	(13.0,53.0)	(-4.6,6.4)	(-5.2,6.8)
NASDAQ	CLA	Sparse 0	(-95,127)	(11.0,51.0)	(-3.1,7.8)	(-3.9,8.9)
NASDAQ	CLA	Sparse 1	(-149,111)	(14.0,69.0)	(-4.8,5.5)	(-4.2,5.8)
FTSE		$1/n$ Naive	(-83,104)	(10.0,32.0)	(-6.9,9.1)	(-5.4,8.8)
FTSE	SLS	Full	(-79,100)	(9.0,30.0)	(-5.5,9.5)	(-5.2,9.8)
FTSE	SLS	Sparse	(-63,84)	(9.0,27.0)	(-5.6,9.6)	(-5.4,9.8)
FTSE	SLS	Sparse 0	(-49,82)	(9.0,27.0)	(-5.7,11.7)	(-5.9,12.8)
FTSE	SLS	Sparse 1	(-92,94)	(11.0,34.0)	(-5.6,8.0)	(-5.9,9.8)
FTSE	CLA	Full	(-79,100)	(9.0,30.0)	(-5.5,9.5)	(-5.9,9.9)
FTSE	CLA	Sparse	(-63,84)	(9.0,27.0)	(-5.6,9.6)	(-4.6,10.0)
FTSE	CLA	Sparse 0	(-68,82)	(9.0,23.0)	(-5.2,10.7)	(-5.8,11.8)
FTSE	CLA	Sparse 1	(-110,101)	(9.0,29.0)	(-7.3,8.3)	(-8.1,8.9)
HS300		$1/n$ Naive	(-102,236)	(11.0,44.0)	(-4.4,9.3)	(-3.8,9.9)
HS300	SLS	Full	(-133,246)	(15.0,42.0)	(-5.1,10.7)	(-5.1,10.2)
HS300	SLS	Sparse	(-125,234)	(14.0,42.0)	(-5.0,10.3)	(-4.4,10.4)
HS300	SLS	Sparse 0	(-92,231)	(13.0,45.0)	(-3.8,11.4)	(-4.0,12.8)
HS300	SLS	Sparse 1	(-143,204)	(14.0,43.0)	(-5.6,7.4)	(-5.9,8.1)
HS300	CLA	Full	(-133,246)	(15.0,42.0)	(-5.1,10.7)	(-5.9,9.9)
HS300	CLA	Sparse	(-125,234)	(14.0,42.0)	(-5.0,10.3)	(-5.3,9.8)
HS300	CLA	Sparse 0	(-94,223)	(12.0,46.0)	(-3.0,10.0)	(-4.0,9.8)
HS300	CLA	Sparse 1	(-146,209)	(12.0,39.0)	(-5.5,8.0)	(-4.9,7.8)

**Table 9**

Portfolio performances obtained by using Normal log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 30 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 30-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-112,137)	(15.0,41.0)	(-3.3,7.0)	(-3.0,7.9)
NASDAQ	SLS	Full	(-105,113)	(16.0,71.0)	(-3.0,5.0)	(-2.4,5.8)
NASDAQ	SLS	Sparse	(-135,120)	(15.0,73.0)	(-3.2,5.5)	(-3.0,6.0)
NASDAQ	SLS	Sparse 0	(-52,116)	(12.0,78.0)	(-2.6,5.6)	(-2.3,6.6)
NASDAQ	SLS	Sparse 1	(-169,86)	(16.0,74.0)	(-3.9,3.3)	(-3.0,4.0)
NASDAQ	CLA	Full	(-105,113)	(16.0,71.0)	(-3.0,5.0)	(-3.1,5.7)
NASDAQ	CLA	Sparse	(-135,120)	(15.0,73.0)	(-3.2,5.5)	(-3.3,6.2)
NASDAQ	CLA	Sparse 0	(-61,116)	(14.0,71.0)	(-2.5,6.2)	(-3.0,7.5)
NASDAQ	CLA	Sparse 1	(-146,78)	(15.0,79.0)	(-4.1,3.7)	(-4.3,4.0)
FTSE		$1/n$ Naive	(-46,68)	(11.0,31.0)	(-3.0,5.7)	(-3.1,7.1)
FTSE	SLS	Full	(-45,75)	(11.0,26.0)	(-2.9,6.8)	(-2.5,8.0)
FTSE	SLS	Sparse	(-48,69)	(11.0,24.0)	(-3.3,7.7)	(-3.9,7.3)
FTSE	SLS	Sparse 0	(-32,68)	(11.0,21.0)	(-2.6,8.2)	(-2.3,9.2)
FTSE	SLS	Sparse 1	(-69,67)	(11.0,29.0)	(-3.8,6.6)	(-3.0,6.4)
FTSE	CLA	Full	(-45,75)	(11.0,26.0)	(-3.0,6.8)	(-2.4,7.3)
FTSE	CLA	Sparse	(-48,69)	(11.0,24.0)	(-3.3,7.7)	(-2.3,8.0)
FTSE	CLA	Sparse 0	(-43,67)	(11.0,22.0)	(-3.0,7.6)	(-4.3,10.0)
FTSE	CLA	Sparse 1	(-56,58)	(12.0,29.0)	(-3.2,5.3)	(-4.3,6.0)
HS300		$1/n$ Naive	(-91,165)	(11.0,51.0)	(-3.3,6.0)	(-3.1,7.4)
HS300	SLS	Full	(-127,168)	(16.0,42.0)	(-4.0,6.3)	(-3.8,7.6)
HS300	SLS	Sparse	(-94,163)	(15.0,38.0)	(-4.1,7.1)	(-3.5,7.9)
HS300	SLS	Sparse 0	(-78,182)	(12.0,38.0)	(-3.0,6.5)	(-2.4,7.8)
HS300	SLS	Sparse 1	(-121,135)	(14.0,55.0)	(-3.9,5.5)	(-4.5,6.1)
HS300	CLA	Full	(-127,168)	(16.0,42.0)	(-4.0,6.3)	(-4.2,8.0)
HS300	CLA	Sparse	(-94,163)	(15.0,38.0)	(-4.1,7.1)	(-4.3,8.0)
HS300	CLA	Sparse 0	(-54,165)	(12.0,51.0)	(-2.3,7.8)	(-2.3,9.3)
HS300	CLA	Sparse 1	(-110,138)	(14.0,43.0)	(-4.2,6.0)	(-4.3,6.5)

**Table 10**

Portfolio performances obtained by using Normal log-likelihood for ICC clustering. We report annualized return, annualized volatility, annualized Sharpe Ratio and annualized Sortino Ratio computed on 100 days investment period after the 1 year training set. The values are averages and 5th and 95th percentiles computed over 100-day investment horizon from obtained from 100 re-sampling of consecutive training-investment periods chosen at random within the 10 years dataset. The underlying assets are constituent stocks of NASDAQ, FTSE and HS300. Highlight in bold are return, volatility and Sharpe Ratio indicating the optimal state in each market solver combination, while highlights in 5th return and 95th volatility showcase the extreme behaviours (excluding the state Market). The state  $1/n$  Naive is the equally weighted un-optimized portfolio and it is reported as benchmark.

Market	Solver	State	Return (5, 95) percentile	Volatility (5, 95) percentile	Sharpe (5, 95) percentile	Sortino (5, 95) percentile
NASDAQ		$1/n$ Naive	(-25,42)	(15.0,33.0)	(-1.3,2.6)	(-1.0,3.5)
NASDAQ	SLS	Full	(-33,52)	(17.0,53.0)	(-1.4,2.8)	(-1.5,3.9)
NASDAQ	SLS	Sparse	(-24,56)	(16.0,36.0)	(-1.2,3.1)	(-1.0,4.4)
NASDAQ	SLS	Sparse 0	(-17,38)	(14.0,37.0)	(-0.8,2.3)	(-0.9,4.2)
NASDAQ	SLS	Sparse 1	(-41,51)	(17.0,51.0)	(-1.9,2.1)	(-1.0,1.9)
NASDAQ	CLA	Full	(-33,52)	(17.0,53.0)	(-1.4,2.8)	(-1.5,3.9)
NASDAQ	CLA	Sparse	(-24,56)	(16.0,36.0)	(-1.2,3.1)	(-1.0,4.1)
NASDAQ	CLA	Sparse 0	(-24,52)	(15.0,36.0)	(-1.2,2.4)	(-1.1,2.8)
NASDAQ	CLA	Sparse 1	(-34,34)	(16.0,49.0)	(-1.7,1.7)	(-0.9,2.0)
FTSE		$1/n$ Naive	(-37,49)	(11.0,30.0)	(-2.3,5.3)	(-3.3,6.4)
FTSE	SLS	Full	(-34,55)	(11.0,26.0)	(-2.2,5.0)	(-2.4,5.6)
FTSE	SLS	Sparse	(-38,55)	(11.0,26.0)	(-2.1,5.0)	(-1.8,5.6)
FTSE	SLS	Sparse 0	(-27,44)	(10.0,18.0)	(-1.9,6.3)	(-1.0,7.9)
FTSE	SLS	Sparse 1	(-39,59)	(12.0,30.0)	(-2.3,5.1)	(-2.5,5.2)
FTSE	CLA	Full	(-34,55)	(11.0,26.0)	(-2.2,5.0)	(-2.3,6.3)
FTSE	CLA	Sparse	(-38,55)	(11.0,26.0)	(-2.1,5.0)	(-2.3,5.5)
FTSE	CLA	Sparse 0	(-28,57)	(11.0,18.0)	(-1.9,7.2)	(-2.8,8.6)
FTSE	CLA	Sparse 1	(-45,44)	(11.0,34.0)	(-2.6,4.5)	(-0.9,5.6)
HS300		$1/n$ Naive	(-47,112)	(15.0,51.0)	(-1.9,5.0)	(-2.8,6.9)
HS300	SLS	Full	(-68,77)	(17.0,52.0)	(-2.0,3.8)	(-2.3,4.8)
HS300	SLS	Sparse	(-50,96)	(16.0,43.0)	(-1.8,4.9)	(-2.4,5.6)
HS300	SLS	Sparse 0	(-60,109)	(15.0,46.0)	(-1.8,5.5)	(-1.0,7.0)
HS300	SLS	Sparse 1	(-74,114)	(17.0,48.0)	(-2.0,4.1)	(-2.3,4.6)
HS300	CLA	Full	(-68,77)	(17.0,52.0)	(-2.0,3.8)	(-2.2,3.9)
HS300	CLA	Sparse	(-50,96)	(16.0,43.0)	(-1.8,4.9)	(-2.3,5.3)
HS300	CLA	Sparse 0	(-54,86)	(15.0,51.0)	(-1.4,5.3)	(-2.1,5.8)
HS300	CLA	Sparse 1	(-71,91)	(17.0,46.0)	(-2.0,4.3)	(-2.5,4.9)

The **sequential least square quadratic programming (SLS)** (Boggs & Tolle, 1996; Kraft, 1988; Nocedal & Wright, 2006) is considered to be one of the most efficient computational method to solve general nonlinear constrained optimization problems. Jackson et al. and Cesarone et al. demonstrate its effectiveness in finance (Cesarone et al., 2015; Jackson & Staunton, 1999). SLS solves the optimization problem iteratively with a gradient descent strategy starting with an initial setting  $\mathbf{W}^0$ , and updating  $\mathbf{W}^{k+1}$  from  $\mathbf{W}^k$  by:

$$\mathbf{W}^{k+1} = \mathbf{W}^k + \alpha^k \mathbf{d}^k \tag{9}$$

where  $\mathbf{d}^k$  is the search direction at the  $k$ th step and  $\alpha^k$  is the associated step size. In each iteration, the descent search direction,  $\mathbf{d}$ , is determined by the solution of a sub-problem. Given the loss function

$$f(\mathbf{W}) = \mathbf{W}\Sigma\mathbf{W}^\top \tag{10}$$

that we want to minimize under a set of non-linear constraints  $g_j(\mathbf{W}) = 0$  for  $j \in [1, m_e]$  and  $g_j(\mathbf{W}) \geq 0$  for  $j \in [m_e + 1, m]$ , at each iteration, the problem of finding the optimal descent direction can be addressed by solving the standard quadratic programming sub-problem (Wilson, 1963):

$$\begin{aligned} \mathbf{d}^{k+1} = \min_{\mathbf{d}} \quad & \frac{1}{2} \mathbf{d}^\top \nabla^2 L(\mathbf{W}^k, \lambda) \mathbf{d} + \nabla f(\mathbf{W}^k) \mathbf{d}^\top \\ \text{s.t.} \quad & \nabla g_j(\mathbf{W}^k) \mathbf{d}^\top + g_j(\mathbf{W}^k) = 0, \quad j = 1, \dots, m_e \\ & \nabla g_j(\mathbf{W}^k) \mathbf{d}^\top + g_j(\mathbf{W}^k) \geq 0, \quad j = m_e + 1, \dots, m \end{aligned} \tag{11}$$

where  $L(\mathbf{W}, \lambda)$  is the associated Lagrangian

$$L(\mathbf{W}, \lambda) = f(\mathbf{W}) - \sum_{j=1}^m \lambda_j g_j(\mathbf{W}). \tag{12}$$

A step size  $\alpha = 1$  is optimal near a local optimum, but when far from the optimum, the step size will need to be modified to guarantee a global convergence. Han (1997), Hock and Schittkowski (1980), Powell (1978) and Rockafellar (1973) have introduced the use of penalty functions in the nonlinear programming to control the step size.

The **Critical Line Algorithm (CLA)** is an efficient alternative to the quadratic optimizer for mean-variance model, as it is specifically designed for inequality portfolio optimization. It was already originally introduced in the Markowitz Portfolio Selection paper (Markowitz, 1952), and its computational implementation has become increasingly popular (Markowitz et al., 2020; Singh et al., 2016). CLA also solves constrained problems with conditions in inequalities, but unlike SLS, it divides a constrained problem into series of unconstrained sub-problems by invoking the concept of turning point. A turning point is a constrained minimum variance portfolio whose vicinity contains other constrained minimum variance portfolios of different free assets.

Similar to quadratic programming, an initial solution is required on the constrained minimum variance frontier. To construct the initial solution, assets are ranked with respect to their expected returns. Then, one increases the weight of the first asset of the highest expected returns,  $w_1$ , from a defined lower bound  $l_1 = 0$  to an upper bound  $u_1$  if  $w_1 \leq 1$ . Subsequently, the following assets have their weights increased until  $\sum_i w_i = 1$ . Typically, the weights of the first and the last few assets are set to the upper and lower bound which are called bounded assets, while only one in the middle has its weight between bounds and referred as the free asset. The free weight is expressed as:

$$w_f = 1 - \sum_{i \in \mathbf{U}} w_i - \sum_{i \in \mathbf{L}} w_i \tag{13}$$

where  $\mathbf{U}$  and  $\mathbf{L}$  represents two sets of upper and lower bounded weights. Then in the following iterations, by decreasing the Lagrange multiplier for the constraint on expected portfolio return,  $\lambda$  to move to the next lower turning point, two cases need to be considered to compute  $\mathbf{W}$ . A formally free asset moves to its bound, or vice versa, a bounded asset wants to become free. In both situations, the maximum threshold  $\lambda_{inside}$  and  $\lambda_{outside}$  for the former and the later will be found. Subsequently, the larger one characterizes the new turning point, and

the asset is moved accordingly, and weights are re-assigned. As the free and bounded assets do not interchange between turning points, the constrained solution between two turning points is in fact the solution of unconstrained optimization on only the free assets. Therefore, the constrained problem reduces to solving the unconstrained problem on the free assets. When no new threshold can be found, the lowest turning point is said to be reached and the algorithm is terminated on the optimized  $\mathbf{W}$ .

**References**

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering - a decade review. *Information Systems*, 53, 16–38.

Ahn, H.-J., Cai, J., & Cheung, Y.-L. (2005). Price clustering on the limit-order book: Evidence from the stock exchange of Hong Kong. *Journal of Financial Markets*, 8, 421–451.

Aitken, M., Brown, P., Buckland, C., Izan, H., & Walter, T. S. (1995). Price clustering on the Australian stock exchange. *Pacific-Basin Finance Journal*, 4, 297–314.

Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M. (2013). Stock price prediction using K-nearest neighbor (kNN) algorithm.

Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15, 1137–1187.

Ang, A., & Bekaert, G. (2003). *How do regimes affect asset allocation?* Working paper 10080, National Bureau of Economic Research, <http://dx.doi.org/10.3386/w10080>, URL: <http://www.nber.org/papers/w10080>.

Aste, T. (2020). Topological regularization with information filtering networks. arXiv preprint arXiv:2005.04692.

Aste, T., & Matteo, T. (2017). Sparse causality network retrieval from short time series. *Complex*, 2017, 4518429:1–4518429:13.

Bailey, D., & de Prado, M. (2013). An open-source implementation of the critical-line algorithm for portfolio optimization. *Algorithms*, 6(1).

Bajoux-Besnainou, I., Jordan, J. V., & Portait, R. (2001). An asset allocation puzzle: Comment. *American Economic Review*, 91(4), 1170–1179. <http://dx.doi.org/10.1257/aer.91.4.1170>, URL: <https://www.aeaweb.org/articles?id=10.1257/aer.91.4.1170>.

Ban, G.-Y., Karoui, N. E., & Lim, A. E. B. (2018). Machine learning and portfolio optimization. *Management Science*, 64, 1136–1154.

Barfuss, W., Massara, G. P., Di Matteo, T., & Aste, T. (2016). Parsimonious modeling with information filtering networks. *Physical Review E*, 94(6), <http://dx.doi.org/10.1103/physreve.94.062306>.

Berger, P. G., & Ofek, E. (1995). Diversification's effect on firm value. *Journal of Financial Economics*, 37, 39–65.

Berry, S., Carlin, B., Lee, J.-K. J., & Müller, P. (2010). Bayesian adaptive methods for clinical trials.

Bhattacharjee, B., Shafi, M., & Acharjee, A. (2019). Network mining based elucidation of the dynamics of cross-market clustering and connectedness in Asian region: An MST and hierarchical clustering approach. *Journal of King Saud University - Computer and Information Sciences*, 31, 218–228.

Bielecki, T., & Jang, I. (2006). Portfolio optimization with a defaultable security. *Asia-Pacific Financial Markets*, 13, 113–127.

Björk, T. (1999). Arbitrage theory in continuous time.

Black, F., & Litterman, R. B. (1991). Asset allocation. *The Journal of Fixed Income*, 1(2), 7–18. <http://dx.doi.org/10.3905/jfi.1991.408013>, URL: <https://jfi.pm-research.com/content/1/2/7>. arXiv:<https://jfi.pm-research.com/content/1/2/7.full.pdf>.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5), 28–43.

Boggs, P., & Tolle, J. (1996). Sequential quadratic programming. *Acta Numerica*, 4(1).

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, 72(3), 498–505, URL: <http://www.jstor.org/stable/2109358>.

Brennan, M., & Xia, Y. (2000). Dynamic asset allocation under inflation. *The Journal of Finance*, 57, 1201–1238.

Briola, A., Turiel, J. D., & Aste, T. (2020). Deep learning modeling of the limit order book: A comparative perspective. *ERN: Other Econometrics: Econometric & Statistical Methods - Special Topics (Topic)*.

Briola, A., Turiel, J. D., Marccaccioli, R., & Aste, T. (2021). Deep reinforcement learning for active high frequency trading. arXiv, abs/2101.07107.

Buckley, I., Saunders, D., & Seco, L. (2008). Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research*, 185, 1434–1461.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *PAKDD*.

Carsey, T., & Harden, J. J. (2013). Monte Carlo simulation and resampling methods for social science.

Castañeda-Leyva, N., & Hernández-Hernández, D. (2005). Optimal consumption-investment problems in incomplete markets with stochastic coefficients. *SIAM Journal on Control and Optimization*, 44, 1322–1344.



- Cesarone, F., Scozzari, A., & Tardella, F. (2015). Portfolio selection problems in practice: a comparison between linear and quadratic optimization models. *Computational Management Science*, 12(3).
- Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 790–799.
- Coronel-Brizio, H. F., & Hernández-Montoya, A. (2005). On fitting the Pareto-Levy distribution to stock market index data: selecting a suitable cutoff value. *Physica A. Statistical Mechanics and its Applications*, 354, 437–449.
- Cox, J., & Huang, C.-F. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49, 33–83.
- Delany, S. J. (2007). K-nearest neighbour classifiers.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22, 1915–1953.
- Dolnicar, S. (2002). A review of unquestioned standards in using cluster analysis for data-driven market segmentation.
- Duffie, D., & Pan, J. (1997a). An overview of value at risk.
- Duffie, D., & Pan, J. (1997b). An overview of value at risk. *The Journal of Derivatives*, 4(3), 7–49. <http://dx.doi.org/10.3905/jod.1997.407971>, URL: <https://jod.pm-research.com/content/4/3/7>. arXiv:<https://jod.pm-research.com/content/4/3/7.full.pdf>.
- Dymova, L., Kaczmarek, K., & Sevastjanov, P. V. (2021). A new approach to the bicriteria multi-period fuzzy portfolio selection. *Knowledge-Based Systems*, 234, Article 107582.
- Eltoft, T., Kim, T., & Lee, T.-W. (2006). On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13, 300–303.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3), 339–350, URL: <http://www.jstor.org/stable/1392121>.
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47, 427–465.
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81, 607–636.
- Fiance, Y. (0000). Shelton capital management nasdaq-100 index fund direct shares. URL: <https://finance.yahoo.com/quote/NASDX/risk/>.
- Fiance, Y. (0000). iShares Core FTSE 100 UCITS ETF GBP (Dist) (ISF.L). URL: <https://finance.yahoo.com/quote/ISF.L?p=ISF.L&.tsrc=fin-srch>.
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9, 3, 432–441.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779–1801.
- Grauer, R. R., & Hakansson, N. H. (1993). On the use of mean-variance and quadratic approximations in implementing dynamic investment strategies: A comparison of returns and investment policies. *Management Science*, 39(7), 856–871, URL: <https://EconPapers.repec.org/RePEc:inm:ormnsc:v:39:y:1993:i:7:p:856-871>.
- Guo, S., & Ching, W.-K. (2021). High-order Markov-switching portfolio selection with capital gain tax. *Expert Systems with Applications*, 165, Article 113915.
- Hakansson, N. H. (1971). Capital growth and the mean-variance approach to portfolio selection. *Journal of Financial and Quantitative Analysis*, 6(1), 517–557, URL: <https://EconPapers.repec.org/RePEc:cup:jfinq:v:6:y:1971:i:01:p:517-557.02>.
- Hallac, D., Vare, S., Boyd, S., & Leskovec, J. (2018). Toeplitz inverse covariance-based clustering of multivariate time series data. arXiv:1706.03161.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384, URL: <http://www.jstor.org/stable/1912559>.
- Han, S. (1997). A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications*, 22, 248–256.
- Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: Testing the markov switching model of GNP. *Journal of Applied Econometrics*, 7, S61–S82, URL: <http://www.jstor.org/stable/2284984>.
- He, H., & Pearson, N. D. (1991). Consumption and portfolio policies with incomplete markets and short-sale constraints: The infinite dimensional case. *Journal of Economic Theory*, 54, 259–304.
- Hee, S., Hamborg, T., Day, S., Madan, J., Miller, F., Posch, M., Zohar, S., & Stallard, N. (2016). Decision-theoretic designs for small trials and pilot studies: A review. *Statistical Methods in Medical Research*, 25, 1022–1038.
- Hock, W., & Schittkowski, K. (1980). Test examples for nonlinear programming codes. *Journal of Optimization Theory and Applications*, 30, 127–129.
- Hult, H., Lindskog, F., Hammarlid, O., & Rehn, C. J. (2012). Risk and portfolio analysis.
- Jackson, M., & Staunton, M. (1999). Quadratic programming applications in finance using excel. *Journal of the Operational Research Society*, 50(12).
- Jensen, M. C., Black, F., & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests. *Capital Markets: Asset Pricing & Valuation*.
- Jiao, Y., Kharroubi, I., & Pham, H. (2013). Optimal investment under multiple defaults risk: a BSDE-decomposition approach. *Annals of Applied Probability*, 23, 455–491.
- Jiao, Y., & Pham, H. (2011). Optimal investment with counterparty risk: a default-density model approach. *Finance and Stochastics*, 15, 725–753.
- Kaczmarek, K., Dymova, L., & Sevastjanov, P. V. (2020). A simple view on the interval and fuzzy portfolio selection problems. *Entropy*, 22.
- Kalayci, C. B., Ertenlice, O., & Akbay, M. A. (2019). A comprehensive review of deterministic models and applications for mean-variance portfolio optimization. *Expert Systems with Applications*, 125, 345–368.
- Kanungo, T., Mount, D., Netanyahu, N., Piatko, C., Silverman, R., & Wu, A. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 881–892.
- Karatzas, I. (1989). Optimization problems in the theory of continuous trading. *SIAM Journal on Control and Optimization*, 27, 1221–1259.
- Karatzas, I., Lehoczky, J., & Shreve, S. (1987). Optimal portfolio and consumption decisions for a “small investor” on a finite horizon. *SIAM Journal on Control and Optimization*, 25, 1557–1586.
- Karatzas, I., Lehoczky, J., Shreve, S., & Xu, G. (1991). Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal on Control and Optimization*, 29, 702–730.
- Ketter, W., Collins, J., Gini, M. L., Gupta, A., & Schrater, P. (2009). Detecting and forecasting economic regimes in multi-agent automated exchanges. *Econometrics eJournal*.
- Khedmati, M., & Azin, P. (2020). An online portfolio selection algorithm using clustering approaches and considering transaction costs. *Expert Systems with Applications*, 159, Article 113546.
- Kraft, D. (1988). *A software package for sequential quadratic programming*. Tech. rep. DFVLR-FB, 88.
- Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem.
- Küchler, U., Neumann, K., Sørensen, M., & Streller, A. (1999). Stock returns and hyperbolic distributions. *Mathematical and Computer Modelling*, 29, 1–15.
- Kumar, I., Dogra, K., Utreja, C., & Yadav, P. (2018). A comparative study of supervised machine learning algorithms for stock market trend prediction. In *2018 second international conference on inventive communication and computational technologies (IICCT)* (pp. 1003–1007).
- Ledoit, O., & Wolf, M. (2003). Honey, I shrunk the sample covariance matrix. *Capital Markets: Asset Pricing & Valuation*.
- Lee, T.-H., & Seregina, E. (2020). Optimal portfolio using factor graphical lasso. *Econometrics*, arXiv.
- Li, F., Sheng, H., & Zhang, D. (2002). Event pattern discovery from the stock market bulletin. In *Discovery science*.
- Limpert, E., & Stahel, W. A. (2011). Problems with using the normal distribution – and ways to improve quality and efficiency of data analysis. *PLoS One*, 6.
- Lin, W.-L., Engle, R., & Ito, T. (1994). Do bulls and bears move across borders? International transmission of stock returns and volatility. *Review of Financial Studies*, 7, 507–538.
- Liu, C., & Cao, Y. (2020). Task re-pricing model based on density-based spatial clustering of applications. *Applied Soft Computing*, 96, Article 106608.
- Lo, A. W. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52. <http://dx.doi.org/10.2469/faj.v58.n4.2453>, arXiv:<https://doi.org/10.2469/faj.v58.n4.2453>.
- Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, 165, Article 113973.
- Madhulatha, T. (2012). An overview on clustering methods. arXiv, abs/1205.1117.
- Mandelbrot, B. (1960). The pareto-Levy law and the distribution of income\*. *International Economic Review*, 1, 79.
- Manly, B. (2020). Randomization, bootstrap and Monte Carlo methods in biology.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1).
- Markowitz, H. M., Starer, D., Fram, H., & Gerber, S. (2020). Avoiding the downside: A practical review of the critical line algorithm for mean-semivariance portfolio optimization. In J. B. Guerard, & W. T. Ziemba (Eds.), *World scientific book chapters, Handbook of applied investment research* (pp. 369–415). World Scientific Publishing Co. Pte. Ltd., Chapter 17.
- Massara, G. P., & Aste, T. (2019). Learning clique forests. arXiv, abs/1905.02266.
- Massara, G. P., Matteo, T., & Aste, T. (2017). Network filtering for big data: Triangulated maximally filtered graph. arXiv, abs/1505.02445.
- Merton, R. C. (1969a). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3), 247–257, URL: <http://www.jstor.org/stable/1926560>.
- Merton, R. C. (1969b). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51, 247–257.
- Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4), 374–413.
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41, 867–887.
- Millington, T., & Niranjan, M. (2017). Robust portfolio risk minimization using the graphical lasso. In *ICONIP*.
- Musmeci, N., Aste, T., & Matteo, T. D. (2016). What does past correlation structure tell us about the future? An answer from network filtering. *Portfolio Management*, arXiv.
- Narsoo, J. (2017). Performance analysis of portfolio optimisation strategies: Evidence from the exchange market. *International Journal of Economics and Finance*, 9, 124–132.

- Navarro, J., Frenk, C., & White, S. (1997). A universal density profile from hierarchical clustering. *Astrophysical Journal*, 490, 493–508.
- Nayak, R., Mishra, D., & Rath, A. (2015). A naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. *Applied Soft Computing*, 35, 670–680.
- Nešetřil, J., Milková, E., & Nešetřilová, H. (2001). Otakar boruvka on minimum spanning tree problem translation of both the 1926 papers, comments, history. *Discrete Mathematics*, 233(1–3), 3–36.
- Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer-Verlag.
- Norton, R. M. (1984). The double exponential distribution: Using calculus to find a maximum likelihood estimator. *The American Statistician*, 38, 135–136.
- Officer, R. (1972). The distribution of stock returns. *Journal of the American Statistical Association*, 67, 807–812.
- Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, 115, 635–655.
- Peiró, A. (1994). The distribution of stock returns: international evidence. *Applied Financial Economics*, 4(6), 431–439.
- Platen, E., & Rendek, R. (2008). Empirical evidence on student-t log-returns of diversified world stock indices. *Journal of Statistical Theory and Practice*, 2(2), 233–251.
- Pliska, S. R. (1997). *Introduction to mathematical finance discrete time models*. Malden: Mass Blackwell, URL: <http://openlibrary.org/books/OL661705M>. Includes bibliographical references (p. [254]-256) and index.
- Powell, M. (1978). A fast algorithm for nonlinearly constrained optimization calculations.
- Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal*, 36, 1389–1401.
- Procacci, P. F., & Aste, T. (2018). Forecasting market states. *Quantitative Finance*, 19, 1491–1498.
- Procacci, P. F., & Aste, T. (2021). Portfolio optimization with sparse multivariate modelling. arXiv:2103.15232.
- Ramchand, L., & Susmel, R. (1998). Cross correlations across major international markets. *Journal of Empirical Finance*, 5(4), 397–416.
- Ramsey, A. F., & Goodwin, B. K. (2019). Value-at-risk and models of dependence in the U.S. federal crop insurance program. *Journal of Risk and Financial Management*, 12(2), URL: <https://www.mdpi.com/1911-8074/12/2/65>.
- Reyes, E. M., & Ghosh, S. (2013). Bayesian average error-based approach to sample size calculations for hypothesis testing. *Journal of Biopharmaceutical Statistics*, 23, 569–588.
- Reynolds, D. (2009). Gaussian mixture models. In *Encyclopedia of biometrics*.
- Rockafellar, R. (1973). The multiplier method of hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications*, 12, 555–562.
- Samo, Y.-L. K., & Vervuurt, A. (2016). Stochastic portfolio theory: A machine learning perspective. *Advanced Risk & Portfolio Management® Research Paper Series*.
- Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3), 239–246, URL: <http://www.jstor.org/stable/1926559>.
- Schmitt, J. (2016). Portfolio selection efficient diversification of investments.
- Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, 21(1), 49–58. <http://dx.doi.org/10.3905/jpm.1994.409501>, URL: <https://jpm.pm-research.com/content/21/1/49>. arXiv:<https://jpm.pm-research.com/content/21/1/49.full.pdf>.
- Simin, T. T. (2008). The poor predictive performance of asset pricing models. *Journal of Financial and Quantitative Analysis*, 43, 355–380.
- Singh, R., Barford, L., & Harris, F. (2016). Accelerating the critical line algorithm for portfolio optimization using gpus. *Advances in Intelligent Systems*, 448.
- Soleymani, F., & Paquet, E. (2021). Deep graph convolutional reinforcement learning for financial portfolio management - DeepPocket. *Expert Systems with Applications*, 182, Article 115127.
- Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework.
- Stuart, A., & Markowitz, H. (1959). Portfolio selection: Efficient diversification of investments. *A Quarterly Journal of Operations Research*, 10, 253.
- Student (1908). On the probable error of the mean. *Biometrika*, 6, 1–25.
- Telesford, Q. K., Simpson, S., Burdette, J., Hayasaka, S., & Laurienti, P. (2011). The brain as a complex system: Using network science as a tool for understanding the brain. *Brain Connectivity*, 1(4), 295–308.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B-Methodological*, 58, 267–288.
- Tigani, S., Chaibi, H., & Saadane, R. (2019). Gaussian mixture and kernel density-based hybrid model for volatility behavior extraction from public financial data. *Data*, 4, 19.
- Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30), 10421–10426. <http://dx.doi.org/10.1073/pnas.0500298102>, URL: <https://www.pnas.org/content/102/30/10421>. arXiv:<https://www.pnas.org/content/102/30/10421.full.pdf>.
- Wilson, E. B. (1923). First and second laws of error. *Journal of the American Statistical Association*, 18, 841–851.
- Wilson, R. (1963). *A simplicial algorithm for concave programming* (PhD dissertation), Harvard University.
- Yuan, X., Yu, W., Yin, Z., & Wang, G. (2020). Improved large dynamic covariance matrix estimation with graphical lasso and its application in portfolio selection. *IEEE Access*, 8, 189179–189188.
- Zariphopoulou, T. (2001). A solution approach to valuation with unhedgeable risks. *Finance and Stochastics*, 5, 61–82.
- Zeitsch, P. J. (2019). A jump model for credit default swaps with hierarchical clustering. *Physica A. Statistical Mechanics and its Applications*, 524, 737–775.
- Zhang, T. (2020). Stock picking strategy based on exploration of chip distribution indicators. In *2020 international conference on computing and data science (CDS)* (pp. 276–282). <http://dx.doi.org/10.1109/CDS49703.2020.00061>.
- Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, 5, 1–20.
- Zolhavarieh, S., Aghabozorgi, S., & Teh, Y. W. (2014). A review of subsequence time series clustering. *The Scientific World Journal*, 2014.