Portfolio optimization with sparse multivariate modeling

Pier Francesco Procacci1 · Tomaso Aste1,2

Revised: 21 July 2022 / Accepted: 26 July 2022 / Published online: 25 August 2022
© The Author(s), under exclusive licence to Springer Nature Limited 2022

Abstract
Portfolio optimization approaches inevitably rely on multivariate modeling of markets and the economy. In this paper, we address three sources of error related to the modeling of these complex systems: 1. oversimplifying hypothesis; 2. uncertainties resulting from parameters’ sampling error; 3. intrinsic non-stationarity of these systems. For what concerns point 1. we propose a $L_0$-norm sparse elliptical modeling and show that sparsification is effective. We quantify the effects of points 2. and 3. by studying the models’ likelihood in- and out-of-sample for parameters estimated over different train windows. We show that models with larger off-sample likelihoods lead to better performing portfolios only for shorter train sets. For larger train sets, we found that portfolio performances deteriorate and detach from the models’ likelihood, highlighting the role of non-stationarity. Investigating the out-of-sample likelihood of individual observations we show that the system changes significantly through time. Larger estimation windows lead to stable likelihood in the long run, but at the cost of lower likelihood in the short term: the “optimal” fit in finance needs to be defined in terms of the holding period. Lastly, we show that sparse models outperform full-models and conventional GARCH extensions by delivering higher out of sample likelihood, lower realized volatility and improved stability, avoiding typical pitfalls of conventional portfolio optimization approaches.

Keywords Portfolio construction · Market states · Mean-variance · Information filtering · TMFG · Sparse covariance · Correlation structure

Introduction
Quantitative approaches to asset management have accumulated unprecedented popularity over the last few decades. Of all the algorithms and strategies developed, portfolio selection models are among those that have received wider attention. The essence of portfolio investing is to find the best way of assigning weights to a given portfolio of assets to maximize future portfolio returns while minimizing the investment risk. The exploration of this filed starts with Markowitz’s mean–variance optimization process Markowitz (1952).

The theory of mean–variance-based portfolio selection is still today a cornerstone of modern asset management. It rests on the presumption that rational investors choose among risky assets purely on the basis of expected return and risk, with risk measured as portfolio variance. The theoretical foundation of this framework is sound if either: investors exhibit quadratic utility, in which case they ignore non-normality in the data (Gollier 2001), or all the higher moments of the portfolio distribution can be expressed as a function of mean and variance and hence all optimal solutions satisfy the mean–variance criterion. Also, the “optimality” of the mean–variance portfolios is based on the assumption that investors live in a one-period world, while in reality they have an investment horizon that lasts longer than one period. Markets indeed constantly change over time and investors are subject to inflows/outflows forcing them to adjust their allocation and take corrective actions.

Form a general, high level, perspective all portfolio optimization approaches are based on a multivariate model of the variables in the market and the economy. The optimization strategies are devised to maximize profits and minimize risks based on such models. In modeling these complex systems, there are, however, several sources of inaccuracies and errors with the three main ones being: 1. oversimplifying hypothesis (such as the use of normal distributions); 2. uncertainties...
resulting from the estimation of the parameters from datasets of limited sizes; 3. intrinsic non-stationarity of these systems, which makes in-sample estimations, based on past observations, inadequate for the estimation of off-sample, future properties. Most likely all three of these factors—and others—contribute to undermining the predictive power of any attempt of modeling markets.

While a large deal of literature has been devoted to relaxing some of the most unrealistic model assumptions (point 1.), the current main pitfall of portfolio optimization is attributed to error maximization (point 2.). This effect has long been established in literature (Michaud 1989; Nawrocki 1996). Essentially, inputs into any risk-return optimization are measured with uncertainty, and the optimization procedure tends to pick those assets which appear to have the most attractive features—but these are outlying cases where estimation error is likely to be the highest, hence maximizing the impact of estimation error on portfolios’ weights. The estimation error is also amplified by market evolution which makes the training error maximization (point 2.). This effect has long been established in literature (Michaud 1989; Nawrocki 1996). Essentially, inputs into any risk-return optimization are measured with uncertainty, and the optimization procedure tends to pick those assets which appear to have the most attractive features—but these are outlying cases where estimation error is likely to be the highest, hence maximizing the impact of estimation error on portfolios’ weights. The estimation error is also amplified by market evolution which makes the training error maximization (point 2.). This effect has long been established in literature (Michaud 1989; Nawrocki 1996).

In this paper, we propose a $L_0$-norm topologically regularized sparse elliptical modeling (Aste 2020) and show that it over-performs traditional models both in likelihood and in portfolio performances. We quantify the effects of estimation error and non-stationarity on portfolio performances (point 2. and 3.) by assessing the goodness of models’ statistical likelihood for estimates over train sets of different lengths. Specifically, we study how the realized portfolio variance reacts to different out-of-sample likelihoods of the input parameters and, particularly, to sparse models. Lastly, we analyze how sparse precision matrices impact the magnitude and the stability of portfolio weights. We refer to the plain mean–variance optimization process being typically point estimates—we can compare the performances of sparse models to conventional conditional correlation models.

The remainder of this paper is organized as follows: In “Literature review” section, we briefly review the theory around portfolio construction, highlighting the pitfalls on assumptions and estimation error and the main solutions proposed in literature; in “Methodology” section we outline our methodology and experiments design, and in “Results”, we present the results. Appendices A and B are devoted to recalling some useful aspects of Elliptical distributions and Appendices C and D report details of the PC-GARACH estimation and comparison.

Literature review

Modern portfolio theory

Considering a portfolio of $n$ assets with weights $W = (w_1, ..., w_n)$, returns $R = (r_1, ..., r_n)$ and portfolio returns $r_p = WR^T$, the standard mean–variance optimization problem consists in minimizing the portfolios’ variance $\sigma_p^2$ for fixed levels of expected returns $\mathbb{E}[r_p] = \bar{r}_p$:

$$\min_W \sigma_p^2 = W\Sigma W^T$$

s.t. $\mathbb{E}[r_p] = \bar{r}_p$, (2)

and $W1 = 1$,

where $\Sigma \in \mathbb{R}^{nxn}$ is the assets’ covariance matrix and $1 \in \mathbb{R}^{nx1}$ is a basis column vector with all elements equal 1. Solving for $W$ for different values of $\bar{r}_p$, one can obtain the optimal weights (i.e., the weights that minimize the portfolio variance) corresponding to different portfolio expected returns $\bar{r}_p$ yielding the so-called efficient frontier—i.e., the set of optimal weights which provide the lowest variance for each level of expected return.

As discussed in Introduction, this optimization is only concerned with the first two moments of the distribution of portfolios’ returns and it does not deal with multiperiod investment decisions. These pitfalls have largely been discussed in literature. Kritzman (2000) provides a clear review of what are the assumptions under which repeatedly investing in one-period-efficient portfolios will also result in multiperiod efficiency. Also, many models have been proposed to deal explicitly with multiperiod optimality (see, for example, Mei et al. (2016), Li and Ng (2000) or Arditti and Levy (1976)). With respect to non-normality of returns, a rich literature is available on both alternative parametrization of data (Bamberg and Dorfleitner 2001) and optimization frameworks that consider other distribution moments (Lasance and Vrins 2021; Hogan and Warren 1974; Bawa 1978) or other measures of risk/return (Yao et al. 2021; Quaranta and Zaffaroni 2008; Harlow 1991).

From the optimization problem in Eq. 2, it is also clear that the optimization does not treat the error and uncertainty around the parameters $\Sigma$ and $\mu$. The difference between the estimated and true distribution parameters is called estimation error. It arises from both the sampling procedure or availability of data and non-stationarity. The error coming from sampling, also referred to as sampling error, is due to parameters used in the portfolio optimization process being typically point estimates—we can only expect these estimates to equal the true distribution parameters if our sample is infinitely large and stationary.
Indeed, assuming stationary data, sampling error could be fixed by increasing the number of observations in the estimation sample with convergence rate towards the true distribution in the inverse of the square-root of the sample size. This would come handy in our times of increasing data availability. However, a second source of estimation error comes from non-stationarity. A time series is said to be non-stationary if its distribution parameters (or the distribution itself) change over time—in this case, extending the length of observations might reduce the contribution of sampling error to estimation error, but at the same time, it could increase that of non-stationarity (Broadie 1993).

Many techniques have been proposed in literature to deal with this phenomenon, both relying on heuristic methods and decision-theoretic foundations (Scherer 2006). Heuristic approaches mainly propose to constrain the optimization problem in order to impose feasible optimal weights.

Michaud and Michaud (1998) address explicitly the sampling error proposing a Monte Carlo-based procedure called resampling. In order to model the randomness of the input mean vector and covariance matrix, portfolio resampling consists in repeatedly drawing from the return distribution given by the point estimates and creating $n$ artificial new samples. For each sample, an efficient frontier is estimated and the final, resampled-efficient frontier is given by the average weight across all of the resampled portfolios.

From a decision-theoretic perspective, Bayesian techniques have recently played a primary role in literature. The rationale behind Bayesian statistics for portfolio construction is to include non-sample information to tackle the effect of parameter uncertainty on optimal portfolio choice. Instead of a point estimate, Bayesian approaches produce a density function for the parameters involved, by combining sample information (likelihood) with prior belief, potentially coming from non-sample information. A special case of this general approach is the seminal work of Black and Litterman (1992). In their pioneering work, the authors assume assets’ returns to be normally distributed with mean equal to the “equilibrium returns” (that is, the mean returns that would return the market portfolio if used in a mean–variance optimization) and combine this “sample” information with investors’ views on the assets. In this way, in absence of an informative prior from investors, the model would return the market or “equilibrium” portfolio. In presence of investors’ priors, instead, the allocation would diverge from the equilibrium portfolio accounting for investors’ views and proportionally to their confidence level. Other than being highly appealing from a practitioner’s perspective, the model proposed in Black and Litterman (1992) highlights the flexibility of the Bayesian framework, with many sources of information that could potentially be used in combination or to update the in-sample information. This is a very active area of research with recent notable examples including (Scherer et al. 2012) and (De Franco et al. 2019).

More recently, entropy is receiving increasing attention as alternative measure of uncertainty in information theory, econometrics, and finance [16]. Starting from the pioneering work of Philippatos and Wilson (1972), entropy-based portfolio allocation models are increasingly popular in the financial literature. Entropy in place of variance as measure of uncertainty and diversification for the portfolio selection problem has proven to provide greater diversification and stability, avoiding classical corner solutions of the mean–variance approach Pola (2016); Batra and Taneja (2022). Further entropy is a nonparametric function designed to accommodate non-normality and asymmetry and no covariance estimation is required as the joint entropy dependence structure can be captured in the objective function Mercurio et al. (2020). Lastly, entropy provides a flexible framework also in mixing multiple sources of information into the joint probability definition Meucci (2010).

**Conditional correlation models**

Modeling volatility in financial time series has been the object of much attention ever since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model in the seminal paper of Engle (1982). Subsequently, numerous variants and extensions of ARCH models have been proposed investigating and leveraging different effects observed in financial time series - for a survey of ARCH-type models, see Bollerslev et al. (1992) and Bera and Higgins (1993), among others.

While modeling volatility of the returns has been the main centre of attention, understanding the comovements of financial returns is of great practical importance. A large body of literature has therefore developed studying the evolution and temporal dependence of correlations, with the main approaches being Multivariate extensions of the GARCH model. Similarly to the univariate case, many different model specifications have been proposed trying to balance flexibility and number of parameters. For as survey, please refer to Boudt et al. (2019) and Bauwens et al. (2006).

More recently, copula Sklar (1959)-based models are increasingly emerging as useful tools to deal with non standard multivariate distributions remedying to various shortcomings of the GARCH structures, with the copula approach being effective in describing the non-linear, asymmetric, and possible tail dependence between markets. Copula-GARCH models combine the use of GARCH models and a copula function to allow flexibility on the choice of marginal distributions and dependence structures and particularly the vine-copula method has been gaining attention recently in that a multi-dimensional density can be decomposed into a product of conditional bivariate copulas and marginal

---

**Conditional correlation models**

Modeling volatility in financial time series has been the object of much attention ever since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model in the seminal paper of Engle (1982). Subsequently, numerous variants and extensions of ARCH models have been proposed investigating and leveraging different effects observed in financial time series - for a survey of ARCH-type models, see Bollerslev et al. (1992) and Bera and Higgins (1993), among others.

While modeling volatility of the returns has been the main centre of attention, understanding the comovements of financial returns is of great practical importance. A large body of literature has therefore developed studying the evolution and temporal dependence of correlations, with the main approaches being Multivariate extensions of the GARCH model. Similarly to the univariate case, many different model specifications have been proposed trying to balance flexibility and number of parameters. For as survey, please refer to Boudt et al. (2019) and Bauwens et al. (2006).

More recently, copula Sklar (1959)-based models are increasingly emerging as useful tools to deal with non standard multivariate distributions remedying to various shortcomings of the GARCH structures, with the copula approach being effective in describing the non-linear, asymmetric, and possible tail dependence between markets. Copula-GARCH models combine the use of GARCH models and a copula function to allow flexibility on the choice of marginal distributions and dependence structures and particularly the vine-copula method has been gaining attention recently in that a multi-dimensional density can be decomposed into a product of conditional bivariate copulas and marginal.
densities. Jondeau and Rockinger (2006). Vine structure is an approach to effectively solve the problem of the dynamic correlation structure between multiple variables, and it provides an effective solution to the matter of variable correlation with complex dependency patterns. The vine-copula method has been gaining attention recently in that a multi-dimensional density can be decomposed into a product of conditional bivariate copulas and marginal densities. He et al. (2021); Luca et al. (2020). Several authors show that compared to the traditional methods the vine structures are better in capturing the dependence between variables and in risk management applications Brechmann et al. (2013); Zhang et al. (2014).

Given the wide adoption among financial academics and practitioners, we will consider the Orthogonal GARCH (O-GARCH) as baseline method to compare our results ("O-GARCH Comparison" section). Considering a dataset of \( T \) returns \( \times n \) assets the observations are assumed to be generated by an orthogonal transformation of \( n \) (or a smaller number of) univariate GARCH processes. The matrix of transformation is the orthogonal matrix (or a subsection) of eigenvectors of the covariance matrix of the returns. In the generalized version, this matrix must only be invertible.

In the orthogonal GARCH model of Alexander and Chibumba (1997), the \( n \times n \) time-varying variance matrix \( H_t \) is generated by \( n \) univariate GARCH models

\[
 r_t = G z_t
\]  

(3)

where \( G \) is a non-singular \( n \times n \) matrix. In the generalized specification of the O-GARCH model of van der Weide (2002), the uncorrelated factors \( z_t \) are standardized to have unit unconditional variances \( \mathbb{E}[z_t^2] = 1 \). The principal components (i.e., unobservable factors) are estimated from the data through \( G \) and the factors \( z_t \) are assumed to follow a GARCH process with the \( n \times n \) diagonal matrix of conditional variances \( Z_t \) defined as:

\[
 H_t = (I - A - B) + A \odot (z_{t-1} z_{t-1}^T) + B H_{T-1}
\]  

(4)

where \( A \) and \( B \) are diagonal \( n \times n \) parameter matrices and \( \odot \) denotes the Haramard (i.e., element-wise) product. Therefore, the conditional covariance matrix of \( r_t \) can be expressed as:

\[
 H_t = G H_{T}^{1/2} G'
\]  

(5)

The linear mapping \( G \) is constructed via singular value decomposition of the returns covariance matrix \( \mathbb{E}[r_T r_T'] = \Sigma \):

\[
 G = U \Lambda^{1/2} V
\]  

(6)

where the \( U \) is the matrix of the eigenvectors of \( \Sigma \) and the diagonal matrix \( \Lambda \) holds its eigenvalues.

It is worth emphasizing that while these models have the potential to offer great flexibility, they are inevitably exposed to the curse of dimensionality in that as the number of assets and parameters increases, the estimation of these models becomes quickly unfeasible Caporin and McAleer (2012).

**Precision matrix estimation with topological regularization**

The literature concerning parameter estimation has greatly developed in recent years. While the increasing availability of data has represented the fuel for data-greedy machine learning models, it has also exacerbated problems related to the curse of dimensionality and overfitting. In essence, the goal is to extract the largest amount of information from data with a model that avoids overfitting, generalizing well with new data. Furthermore, model interpretability (or explainability) has become increasingly important, and for this purpose, simpler and sparser models with a smaller number of parameters are preferred. To this extent, understanding the dependency structure among variables has proven to be essential to capture the collective behaviour of the systems and many techniques have been introduced in literature. Examples are dimensionality reduction methods (Van Der Maaten et al. 2009) or shrinkage techniques (Ledoit and Wolf 2020; Friedman et al. 2008).

One possible approach to represent the set interactions in a complex system is to model them as a network structure where the vertices are the system’s elements and edges between vertices indicate the interactions between the corresponding elements (Lauritzen 1996). Information filtering networks (Tumminello et al. 2005) aim at retrieving the relevant subnetwork of interactions among the elements of the system. In the pioneering work of Mantegna (1999), it was proposed to investigate financial systems by the extraction of a minimal set of relevant interactions associated with the strongest correlations belonging to the minimum spanning tree (MST). The MST structure is, however, a drastic filtering tool and is likely to discard valuable information. Tumminello et al. (2005) and Aste and Di Matteo (2006) expanded the concept by introducing the general idea of information filtering networks. In particular, they show that graphs of different complexities can be constructed by iteratively linking the most strongly connected nodes under the constraint of generating planar or hyperbolic graphs. There is now a large body of literature proving network filtering to be a powerful tool to associate a sparse network to a high-dimensional dependency measure with applications ranging from financial markets (Barfuss et al. 2016) to biological systems (Song et al. 2012) and econophysics (Mantegna and Stanley 2000). Recent developments extend the information filtering network methodology to chordal graphs (Barfuss
et al. 2016; Massara et al. 2015; Massara and Aste 2019). This allows to directly associate the information filtering network with a positive definite sparse inverse covariance matrix that is a $\ell_0$-norm topological regularization of the full covariance estimate (Aste 2020).

Robust portfolio selection methods that are less sensitive to the estimation error of the covariance matrix have received substantial investigations in literature. Several authors propose to impose norm constraints on the portfolio weights (Kremer et al. 2020; Xing et al. 2014; Brodie et al. 2009) or to shrink the covariance matrix estimator by assuming structural models on the covariance matrix rooted on a reference factor model or on Bayesian estimation (Han 2020; Ledoit and Wolf 2004, 2003). In our analysis, we use the TMFG-LoGo network filtering approach (Massara et al. 2015; Barfuss et al. 2016). TMFG-LoGo approach has proven to be more efficient and better performing, particularly when few data are available (Barfuss et al. 2016; Aste and Di Matteo 2017), with respect to $L_1$-norm sparsification techniques such as GLASSO (Friedman et al. 2008). Moreover, the TMFG procedure is computationally highly efficient and this is well suited in our resampling experiment.

Despite the rich literature on portfolio construction techniques, to the best of our knowledge, the effects of non-stationarity on financial markets have not been explicitly investigated in literature. In this paper, we aim at discussing the link between traditional measures of estimation goodness, portfolio performances and non-stationarity, showing that regularisation and sparsification of the covariance matrix are effective remedies in dealing with the non-stationary, entropic evolution of financial markets.

**Methodology**

The estimation error is quantified by measuring how well the functional form of the multivariate probability distribution, $f_{\theta}(X)$ defined via the parameters $\theta$ estimated in-sample, describes the actual data out-of-sample. The statistical measure that describes how likely are observations to belong to the estimated probability function is the likelihood. The likelihood principle is a cornerstone of statistical analysis in that maximum likelihood estimators are guaranteed to be asymptotically efficient under mild conditions (Wald 1949, Cramér 1946, Daniels 1965).

In this section, we introduce our methodology and discuss our results for the multivariate normal case. In Appendix A, we discuss further the generality of this approach and show that it in this section, we introduce our methodology and discuss our extends to other distributions of the elliptical family including, in particular, the multivariate Student-t.

The logarithm of the likelihood for the normal case is proportional to:

$$\ln \mathcal{L}(\theta; X_t) = \ln |J| - (X_t - \mu)^T J (X_t - \mu) + k,$$

where $X_t = (x_{t,1}, x_{t,2}, ..., x_{t,n})$ is the $n$-dimensional multivariate returns observation vector at time $t$; $\theta$ is the model parameter set, which includes $\mu$ the vector of means and $J$ the generalized precision matrix and; $k$ is a constant which is independent from $\mu$, $J$ or $X_t$ (see Appendix A). In the multivariate normal case $J = \Sigma^{-1}$ is the inverse of the covariance. Our results generalize to other elliptical distributions with defined covariance where $J$ is proportional to the inverse covariance, which we assume is defined and invertible. Results for the Student-t are explicitly reported in appendix A. Our goal is to study the log-likelihood in Eq. 7 using different estimation windows and comparing how the precision matrices, estimated through maximum likelihood and TMFG-LoGo, perform.

We designed a resampling experiment in which we select 100 stocks at random among the 342 and a random trading day spanned in our dataset. Starting from the randomly selected trading day and going back in time, we define five train sets of different sizes by including an increasing number of observations. We start at 101 observations, then 150, 250, 500, 1000 and finally 1500 observations. We then use a fixed-length test set of 500 observations following the randomly selected trading day. We keep the test set length fixed to avoid biases and selected 500 observations so that, for all estimation windows, the main crisis event (i.e., Global Financial Crisis in 2008) can be randomly included in or out of sample. Figure 1 shows a sketched example of our train/test split with different estimation windows.

We use the train set to estimate the mean vector $\mu$, the maximum-likelihood covariance matrix $\Sigma$ and the sparse TMFG LoGo covariance matrix $\Sigma_{TMFG}$. These parameters are then used to compute the log-likelihood in Eq. 7 for both in-sample and out-of-sample observations. We then investigate how the different estimates used in a portfolio optimization procedure affect the optimal weights and portfolio characteristics. To this extent, we considered the standard, unconstrained Markowitz optimization problem described in "Modern portfolio theory" section. This is done to avoid any bias coming from constraints in our analysis and to keep the framework as plain as possible. We focus therefore our analysis on the minimum variance portfolio, that is the efficient portfolio that minimizes the expected variance. To obtain the solution for the minimum variance portfolio, the portfolio optimization problem in Eq. 2 rewrites:

$$\min_W \sigma^2 = W \Sigma W^T$$

s.t. $W 1 = 1$,

which gives the optimal, minimum variance weights:
where $c = \frac{1}{\sqrt{\Sigma^{-1}}}$. \hspace{1cm} (9)

For each asset $i = 1, \ldots, n$, we calculated the corresponding daily returns $x_{t,i} = P_{t,i}/P_{t(-1),i} - 1$, where $P_{t,i}$ is the closing price of stock $i$ at time $t$, for a total of 4026 daily multivariate observations. Figure 2 reports the means distribution and the correlation matrix of the stocks in the sample. Particularly for the purposes of this paper, Fig. 2b reports the pairwise correlations heatmap for the stocks in the sample. The chart shows that the correlations are positive in the vast majority of the cases, with most correlations ranging in $[0, 0.5]$ and only a few cases of higher and slightly negative correlation. While this is a standard and common setting for any large cap focused investment universe, it is worth emphasizing that the TMFG filtered correlation matrices are independent on the sign of the correlations and built only based on the magnitude of the correlations. Therefore, the filtering effects discussed in in the next Section are general and apply to any investment universe.

**Dataset**

We considered a standard dataset of daily closing prices of US stocks entering among the constituents of the S &P 500 index between 02/01/1997 and 31/12/2015. After screening for those continuously traded and those not displaying abnormal returns, we reached a final dataset of 342 stocks.
Results

Likelihood comparison

Figure 3 reports the average log-likelihood, estimated using the maximum-likelihood and the TMFG covariances, for the train data (Fig. 3a) and for the test data (Fig. 3b) computed across 100 resamplings. The larger the log-likelihood is, the better the parameters $\theta$ are at describing the data, for the assumed model. Figure 3a shows that, as expected and by definition, the maximum-likelihood estimate of the covariance matrix provides a higher in-sample likelihood compared to the TMFG covariance, although the latter tracks quite closely. Also, the likelihood is strictly decreasing with the number of observations included in the estimation window. Indeed, as the number of observations decreases relative to the parameters, the model overfits the sample yielding larger in-sample likelihoods. Filtering the covariance matrix and reducing the number of parameters limits the overfitting potential of the model. This is shown by the lower levels of likelihood attained by TMFG when fewer observations are used which, therefore, results in a larger gap in likelihood relative to the maximum likelihood.

Perhaps more interestingly, Fig. 3b reports the likelihoods obtained out-of-sample using the two different in-sample estimates of the covariance matrix. The first observation is that TMFG-LoGo provides a substantially larger log-likelihood, especially for short estimation windows. This result is exacerbated by the fact that when 101 observations are considered, the number of stocks is very close to the number of observations in our samples. While the resulting covariance is still full-rank (number of observations > number of variables), it leads to unstable estimation in the maximum-likelihood covariance (i.e., the so-called the curse of dimensionality”), whereas TMFG-LoGo is still well defined. Note that there is a break y axis of the figure to allow a better inspection of the results. The figure shows that for longer estimation windows, the out-of-sample log-likelihood computed with the maximum-likelihood covariance tends to converge to the TMFG likelihood which, however, a) always provides the best out-of-sample likelihood in our experiment and b) provides quite stable likelihood values also for shorter estimation windows. We conclude that the TMFG-LoGo algorithm does a good job at filtering the correlation structure providing higher out-of-sample likelihood and stable results with shorter estimation windows, confirming the results with stationary time series previously reported in (Barfuss et al. 2016).
Impact of precision matrix estimate on optimal portfolios

We now address empirically the question of what is the impact of different parameter estimates on portfolios weights and performances, when these parameters are used as inputs in the portfolio optimization problem in Eq. 2. Having focused our attention on the minimum variance portfolio on the efficient frontier, we report in Fig. 4 the realized standard deviation of portfolios obtained using the maximum-likelihood and TMFG covariances, estimated over different estimation windows and across 100 resamplings.

The chart shows that, overall, the out-of-sample portfolio variance decreases as the likelihood increases up until when 750 observations are used. This is coherent with the likelihood results that reported increasing likelihoods for the same estimation windows. In particular, when fewer observations are used in-sample, the TMFG-LoGo covariance matrix provides portfolios with significantly lower realized variance than the maxim-likelihood portfolios. Also, little changes are observed in the TMFG portfolios realized variance when number of observations in sample ranges from 101 to 750, signalling that the TMFG-LoGo extract the relevant dependency links also when only few observations are available. The gap in performance between TMFG and maximum-likelihood portfolios tends to reduce as the number of observations in the estimation window increases, with the TMFG-LoGo portfolios always displaying lower volatility. However, when more than 750 daily observations are included, while the out of sample likelihood remains flat or slightly increases, the portfolios’ variance tends to increase.

To further investigate this pattern, we report in Fig. 5 the volatilities for all 100 resamplings and considering steps of 25 observations in the estimation windows. The figure confirms that the TMFG-LoGo covariances delivered overall less volatile portfolios across resamplings and estimation windows. Secondly, the figure shows that the portfolios obtain the lowest out-of-sample variance when approximately 2 to 3 years of daily observations (450–700 observations) are included in the train set. This pattern is clear for the maximum likelihood portfolios, with means, quintiles and outliers drifting upwards when more than 750 observations are included. The TMFG filtered covariance regularizes and smooths this effect as well, but still when

Fig. 4 Realized standard deviation. Increasing the estimation window and for higher values of likelihood (Fig. 3), the realized standard deviation of portfolios decreases. Y axis break to fit the scale for 101 days estimation window

Fig. 5 Portfolio realized volatility across resamplings for different estimation windows. The box-plot shows the distribution of the variances obtained for 100 resampled portfolios and the the blue line overlaid shows the average variance (i.e., the mean of the variances’ distribution)
more than 750 observations are included, the resulting portfolios exhibit a slightly higher variance. This is consistent with the literature showing that longer estimation windows provide worse forecasts in financial time series due to the regime-changing nature of financial markets (Procacci and Aste 2019). It is also worth emphasizing that the different features of TMFG-LoGo and the Maximum-Likelihood portfolios are due solely to the different estimates of the covariance matrix as these are the only inputs used in the Minimum Variance optimization Eq. 9.

Finally, we address the impact of sparsity on optimal portfolio weights. Figure 6 reports the number of long (Fig. 6a) and short (Fig. 6b) positions (i.e., positive and negative weights assigned to the stocks in portfolio) on average across the 100 resamplings. The first observation is that the number of long positions tends to increase as the estimation window increases and coherently the short positions diminish accordingly. Using TMFG-LoGo precision matrices anticipates this behavior, in that TMFG portfolios always display a greater number of long positions also for short estimation windows. Recalling that the Minimum-Variance optimisation (Eq. 2) is constrained to sum to 1, the intuition behind this phenomenon is that the fewer the observations used in the estimation of the covariance, the higher is the tendency of the Minimum-variance portfolios to exhibit extreme negative and positive weights. In other words, the weights still sum up to 1, but with a combination of large long and short bets. Over the long term, estimates are more stable and possible outliers in assets’ variances and correlations are polished, leading to more stable portfolios.

This intuition is confirmed by looking at the distribution of weights across resamplings in Fig. 7. This chart (note the different scales) shows that using the TMFG-LoGo covariance matrix significantly improves the stability of the optimal solutions, reducing outliers and avoiding “corner,” i.e., extreme solutions which are a typical pitfall of the unconstrained Markowitz optimization. This results shows that the correlation coefficient among assets plays an important role in that for high correlation levels, the optimization procedure would prefer one stock in place of another for slightly more appealing variance features. Having filtered the correlation structure in the TMFG-LoGo procedure, we obtained a portfolio that is much more general (hence the anticipated larger number of long positions) and less sensitive to single assets features given the filtered correlation among stocks. Lastly, considering the standard unconstrained optimization problem in Eq. 2, both the maximum likelihood and the TMFG matrices produce portfolios that are in the vast majority of cases investing in all assets. In other words, even considering a sparse precision matrix like in the TMFG-LoGo case, we very rarely found weights equal to zero assigned to some assets.

### Non stationarity

From the results discussed in previous section and shown in Figs. 3 and 4, we found that the portfolio performances improve coherently with the likelihood up until approximately 3 years of observations are used in the train set to estimate the models parameters. However, when more
Observations are included in the train set, the likelihood of the parameters detaches from the portfolio performances and we speculate that this is due to the role of non-stationarity. To further investigate this phenomenon, Fig. 8 reports the likelihood corresponding to each out-of-sample observation in our experiment.

Figure 8a shows the average likelihood across 100 resamplings for each out-of-sample observation. We note that when shorter estimation windows are used to estimate the models' parameters (i.e., 125 days) the likelihood is higher in the days immediately following the estimation window, but tends to rapidly decrease as the observations depart from the training window. Larger estimation windows (i.e., 750 or 1,500 days) instead, lead to a more stable likelihood in the long run, but at the cost of a lower likelihood for the observations closer to the estimation set. Figure 8b shows the observation-wise box-plot of the likelihood computed across the resamplings when 750 observation are used in the train set. The box plot reports the 25–75% quantile interval (dark blue) and the max-min interval (“whiskers” light blue) having excluded the “outliers” that are below the whiskers (Langford 2006). Other than decreasing means, the figure shows that as the observations depart from the train set, the amount of observations posting a significantly lower likelihood increases, together with the downside volatility. In other words, it is more likely to
have observations that are far from the model estimated in sample, supporting the conclusions drawn from Fig. 8a.

As we discussed in (Procacci and Aste 2019), market states tend to be persistent in daily observations. Shorter estimation windows, therefore, tend to better describe the system belonging to the same “state” which is likely to be persistent for adjacent observations. Notice that by considering the aggregate behaviors across 100 resamplings, we want to avoid specific market conditions and state shifts, but rather focus on the general behaviour. The evolution of the financial system is obviously very dynamic and the goodness of parameters is certainly dependent on both systemic and idiosyncratic events.

These results also provide further insights on the findings discussed in Figs. 3 and 4 in that our conclusions are dependent on the number of out-of-sample observations that in our case coincides with the portfolio holding period—i.e., 500 days in our experiments. Shorter estimation windows provide better fit in the short term, while larger estimation windows provide robustness in the long run. The optimal balance between these two effects depends on the holding period and in our experiments it is achieved with approximately 3 years of observations in the estimation window. Short holding periods do not require robustness in the long run (i.e., shorter estimation windows would deliver better results). As the holding period increases, the long-term robustness becomes more relevant than the short-term fit and larger estimations windows have to be preferred.

**O-GARCH comparison**

Having assessed the impact of sparsity on portfolio weights’ stability, in this section we compare the features of portfolios obtained with TMFG-LoGo covariances with a baseline standard approach. As discussed in “Literature review” section, time varying conditional models represent the most widely used models in dealing with covariance structures. As such, we selected the general O-GARCH model as baseline model.

Following the approach of van der Weide (2002) recalled in "Conditional correlation models" section, for each resampling we estimated the maximum likelihood covariance eigenvectors and the corresponding principal components (PC). The PCs are assumed to follow a GARCH process and the corresponding parameters are estimated for each principal component time series. We tested for different model specifications, allowing for both the ARCH and GARCH parameters to range from 1 to 3 lags and selected the best model based on the AIC and BIC criteria. Tables 3 and 4 in Appendix C report the average AIC and BIC statistics for every model specifications across different estimation windows. For all estimation windows, both the AIC and BIC criteria support the GARCH(1,1) specification. Furthermore, in Appendix C we report the median, 5th and 95th percentiles of all the AIC and BIC statistics obtained, showing that the GARCH(1,1) specification is at any point the preferred specification.
We carried out the same stability experiments described in "Impact of precision matrix estimate on optimal portfolios" section using the O-GARCH(1,1) covariance in the portfolio optimization. Figure 7c presents the distribution of weights across resamplings for the optimal minimum-variance weights obtained using the OGARCH covariance and compared to the TMFG and maximum-likelihood covariances. The figure shows that the O-GARCH covariance presents instability problems similar to the full covariance (Fig. 7a). Similarly, Fig. 11 in the Appendix extends Fig. 6 by comparing the number of long (Fig. 11a) and short (Fig. 11b) positions when also the optimal minimum variance weights obtained using the OGARCH forecasted covariance matrix are considered. The O-GARCH positions closely track the full covariance behaviour, thus the same conclusions drawn in "Conditional correlation models" section apply.

In essence, all the stability problems that apply to the maximum-likelihood covariance still apply to the O-GARCH forecasted covariance matrix when used for portfolio construction.

**Backtest**

To further compare the features of sparse TMFG portfolios to the conventional O-GARCH, we backtest a simple trading strategy and compared performance results. The strategy follows a rolling estimation scheme with daily rebalance: we pick a random trading day, we use the previous q days to estimate the parameters and we compute the optimal minimum variance portfolio weights (Eq. 9). We then roll the estimation window on a daily basis and rebalance the portfolio accordingly, assuming therefore to buy at the close price and hold until close the following day. This process is reiterated through 500 observations following the randomly picked starting trading day. In accordance with the testing framework considered throughout the paper, for each estimation window we run 100 resampling where we pick a different set of 100 stocks and a different start trading day. We do no include transactions costs in the performance presented, as we separately investigate the turnover of both the strategies.

In terms of parameters re-estimation, on a daily basis we consider the previous q observations and re-estimate the covariance matrix. For the O-GARCH, we forecast the one day ahead conditional volatility of the principal components and reconstruct the corresponding covariance matrix. The model is fit only once, with the first estimation window, and we then keep the ARCH and GARCH parameters fixed. As the new observations come in, we estimate the new covariance matrix, the corresponding principal components and use it forecast the one day ahead covariance matrix.

Table 1 presents the out-of-sample annualized standard deviation of the strategy with minimum variance portfolios constructed based on TMFG precision matrices compared to the OGARCH-based portfolios in Table 2. For each estimation window, we report the median, 5th and 95th percentiles across 100 resamplings.

As shown in the tables, TMFG-based portfolios delivered more stable performances with lower realized variance in the vast majority of cases.

More important than performance in the context of this paper is the stability of portfolios weights through rebalances. As previously mentioned, the performance metrics

| Train obs | 5th  | Median | 95th  |
|-----------|------|--------|-------|
| 101       | 0.078| 0.097  | 0.197 |
| 125       | 0.079| 0.106  | 0.198 |
| 250       | 0.080| 0.100  | 0.196 |
| 500       | 0.084| 0.112  | 0.209 |
| 750       | 0.085| 0.097  | 0.214 |
| 1000      | 0.088| 0.114  | 0.212 |
| 1500      | 0.092| 0.119  | 0.223 |

Table 1 TMFG portfolios performance metrics

| Train obs | 5th  | Median | 95th  |
|-----------|------|--------|-------|
| 101       | 0.104| 0.168  | 0.415 |
| 125       | 0.101| 0.178  | 0.384 |
| 250       | 0.097| 0.137  | 0.293 |
| 500       | 0.098| 0.151  | 0.316 |
| 750       | 0.099| 0.117  | 0.325 |
| 1000      | 0.103| 0.164  | 0.314 |
| 1500      | 0.105| 0.161  | 0.334 |

Table 2 OGARCH portfolios performance metrics

Annualized standard deviation of a daily rebalancing minimum volatility strategy. Optimal weights computed using using TMFG precision matrices. Median, 5th and 95th percentiles across 100 resamplings.
Fig. 9 Average daily turnover across 100 resamplings for different estimation windows. Comparison of minimum variance portfolios constructed based on TMFG (blue line) and O-GARCH (orange line) precision matrices. (Colour figure online)
presented in Tables 1 and 2 do not take into account transaction costs to allow a separate investigation on the impact of the precision matrix on portfolios stability. We turned our attention to daily turnover, computes as:

\[ T = \sum_{i=1}^{100} |W_{i,t} - W_{i,t-1}| \]  

where \( i \) is the \( i \)-th stock among the 100 randomly sampled.

Figure 9 reports the average daily turnover of the two strategy across our 100 resamplings. Across all the estimation windows, TMFG portfolios reported significantly lower turnover, in most of cases two to three times less than O-GARCH portfolio. Plotting all daily turnover data across all resamplings in Fig. 12 further shows that turnover on TMFG portfolios is actually less than 5% for most of the days across our resamplings, with a distribution that is heavily positively skewed. O-GARCH portfolios, on the other hand, displays a much less favorable turnover distribution, with extremes that reach 300% turnover. These results come with no surprise given the much higher portfolio stability already observed in Fig. 7c.

Conclusions

Portfolio construction is a cornerstone of financial theory and practice. However, it is still today a controversial topic for both academics and practitioners. Any portfolio optimization strategy relies on assumptions and modeling of the future market structure. However, inferring such structure from past observations is a very challenging task, plagued by uncertainty around parameters estimation and relying on some non fully satisfied assumptions.

We identify three main sources of inaccuracies and errors: 1. model oversimplification; 2. limited size of the estimation set; 3. non-stationarity. We address oversimplification by introducing a modeling that uses a \( L_0 \)-norm regularized elliptical multivariate distribution, demonstrating that it over-performs traditional models both in likelihood and in portfolio variance performances. We test the effect of sample size by training the models on windows of different sizes and find that performances initially increase with sample size but then eventually decrease for windows above 750 days. We attribute the initial improvement in performance to sampling error, which is reduced when more observations are included, and we interpret the decay in performance when more the 750 observations are included as an instance of non-stationarity. We further investigate this phenomenon by studying the likelihood corresponding to individual observations out-of-sample and show that shorter estimation windows deliver higher out-of-sample likelihood in the days immediately following the train window, but it tends to rapidly decrease afterwards. As more observations are included in the training set, the out-of-sample likelihood gains stability, with larger values in the long term, but at the cost of lower likelihood in the short term. We conclude that the financial system changes significantly through time and the “optimal” fit in finance needs to be defined in terms of the holding period. Lastly, we compare the performances and stability of sparse TMFG portfolios to a conventional conditional correlation model, showing that sparse portfolios lead to 4 times less turnover than conventional covariance structures. Our conclusion is that sparsity represent a valuable tool for portfolio construction which provides significant, practical advantages and in our view should be implemented in addition (and not alternatively) to more conventional correlation forecasting techniques.

Our main contribution to the literature on portfolio construction is the demonstration of the relationship between the goodness of the model, measured as out-of-sample likelihood, and the realized portfolio volatility. We show that higher likelihood obtained with filtered TMFG-LoGo precision matrices correspond to lower portfolio volatility out-of-sample. The relationship between larger likelihood and lower realized volatility is also verified in the maximum-likelihood estimate of the covariance matrix when computed over train sets of different lengths. Further, we show that sparse, filtered covariance matrices can significantly reduce estimation errors coming from both sampling error and non-stationarity, other than significantly improving many of the instability problems related to mean–variance optimal weights.

Finally, all of the analysis and conclusions drawn in this paper are based on different estimates of the covariance matrix. While forecasting future returns remains of primary importance in trading and wealth management, we showed that the correlation structure, sometimes overlooked in the asset allocation literature, plays a key role in portfolio
construction and a good deal of performances depends upon it.

**Appendix A. Elliptical distributions**

Consider an \( n \)-dimensional vector of multivariate returns \( X = (x_1, x_2, \ldots, x_n) \). If \( X \) is elliptical distributed, then its probability density function is defined as:

\[
f_X(X) = c_n |J|^{1/2} g_n \left[ (X - \mu) J (X - \mu)^T \right], \tag{A.1}
\]

where \( \mu \in \mathbb{R}^n \) is the vector of location (mean) parameters and \( c_n \) is a normalization constant. The matrix, \( J = \Omega^{-1} \in \mathbb{R}^{n \times n} \) is the generalized precision matrix, a positively defined matrix which is the inverse of the dispersion matrix \( \Omega \). When the covariance is defined (as we assume in this paper) then \( \Sigma = (-\psi'(0))^{-1} \), that is, \( \Omega \) is proportional to the covariance matrix and the proportionality factor is the inverse of the first derivative of the characteristic generator evaluated at 0. The function, \( g_n(\cdot) \), is called density generator.

Also, let us stress that \( (X - \mu) J (X - \mu)^T \) - i.e., the generalized, square Mahalanobis distance - is a quadratic term and hence a nonnegative quantity provided that the matrix \( \Omega \) is positive definite. To ease the notation, for the remaining of the paper we shall refer to the generalized Mahalanobis distance as \( d^2 \):

\[
d^2 = (X - \mu) J (X - \mu)^T. \tag{A.2}
\]

For different density generators \( g_n(\cdot) \) we obtain different distributions of the elliptical family. It is easy to see, for example, that the normal distribution is obtained by using:

\[
g(u) = e^{-u/2}, \tag{A.3}
\]

and \( \Omega = \Sigma \).

Similarly the Student-t distribution is obtained by using:

\[
g_n(u) = \left( 1 + \frac{u}{\nu} \right)^{-\frac{\nu + 1}{2}}, \tag{A.4}
\]

where \( \nu \) is the degrees of freedom, and \( \Omega = \frac{\nu - 2}{\nu} \Sigma \).

The validity of the mean–variance framework for elliptical distributions has long been established in literature (Owen and Rabinovitch 1983). This proposition is derived easily from two properties of the elliptical distributions. First, for every elliptical distribution with defined mean and variance, the distribution is completely specified by them (Owen and Rabinovitch (1983) or Chamberlain (1983)), with all the higher moments being either zero or proportional to the first or second moment. Second, any linear combination of multivariate elliptically distributed variables is also an elliptically distributed variable. In the case of normal distribution and Student-t distribution they also have the same density generator function. Further details on these properties are provided in Appendix B.

It follows that, if asset returns have a multivariate elliptical distribution \( X \sim \mathcal{E}_n(\mu, \Omega, g_n) \), then the portfolio expected return and dispersion are given by, respectively, \( \mathbb{E}[r_p] = W \mu \) and \( \sigma_p^2 = W \Omega W^T \), matching the optimization framework outlined in "Modern portfolio theory" section.

With respect to our likelihood analysis, considering distributions with probability density function of the form specified in Eq. A.1, the corresponding likelihood function is of the form:

\[
L_{\Theta|X} = |J|^{1/2} g_n(d^2). \tag{A.5}
\]
where $ED$ denotes the general Elliptical Distributions and we omitted the constant of integration.

To stress the general validity of our analysis for other elliptical distributions, we repeated the experiments discussed in "Methodology" section considering the Student-t generator.

Assuming a Student-t distribution of the log returns, the log-likelihood (Eq. A.5) is:

$$\ln L_{\text{Student}} = \frac{\ln |J|}{2} - \frac{n + v}{2} \ln \left(1 + \frac{d^2}{v-2}\right)$$  \hspace{1cm} \text{(A.6)}

where $n$ is the sample size and $v$ is the degree of freedom. Figure 10a reports the likelihood comparison for the same resamplings as in Fig. 3 but using a Student-t log-likelihood as in Eq. A.6. Here we used $n = 500$ observations (i.e., the out-of-sample size) and $v = 3$. We verified that this findings are robust across different degrees of freedom in the range $v = [2.1, 4]$.

### Appendix B. Properties of elliptical distributions

In this section we recall some useful properties of Elliptical Distribution which we referred to in our discussion and particularly in Section A.

**Property 1** (Distribution Definition) Consider an $n$-dimensional random vector $X = (X_1, ..., X_n)$. $X$ has a multivariate elliptical distribution with location parameter $\mu$ and dispersion parameter $\Omega$, written as $X \sim \mathcal{E}(\mu, \Omega)$ if its characteristic function $\phi$ can be expressed as:

$$\phi_X(w) = \mathbb{E}(e^{iwX}) = e^{iw\mu} \psi \left( \frac{1}{2} w \Omega w^T \right). \hspace{1cm} \text{(B.1)}$$

for some location parameter $\mu \in \mathbb{R}^n$, positive-definite dispersion matrix $\Omega \in \mathbb{R}^{nxn}$ and for some function $\psi(\cdot) : [0, \infty) \to \mathbb{R}$ such that $\psi\left(\sum_{i=1}^{n} w_i^2\right)$ is a characteristic function, which is called characteristic generator. If $X \sim \mathcal{E}(\mu, \Omega)$ and if its density $f_X(x)$ exists, it is of the form defined in Eq. A.1.

**Property 2** (Density Generator) The function $g(\cdot)$ defined in Section Appendix A is guaranteed to be density generator if the following condition holds:

$$\int_0^\infty x^{n/2-1} g_n(x)dx < \infty. \hspace{1cm} \text{(B.2)}$$

**Property 3** (Affine Equivariance) If $X = (X_1, ..., X_n)$ is an $n$-dimensional elliptical random variable with location parameter $\mu$ and dispersion parameter $\Omega$, then for any vector $a \in \mathbb{R}^{1xn}$ and any matrix $B \in \mathbb{R}^{mxn}$ the following affine equivariance holds:

$$Y = a + BX \sim \mathcal{E}_\Gamma(a + B\mu, B\Omega B). \hspace{1cm} \text{(B.3)}$$

In other words, any linear combination of multivariate elliptical distributions is another elliptical distribution.

### Table 3 Average AIC information criterion for different GARCH specifications across 100 resamplings for different estimation windows

| Train Obs | (1, 1) | (1, 2) | (2, 1) | (2, 2) | (2, 3) | (3, 2) | (3, 3) |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| 101       | -664   | -661   | -660   | -660   | -659   | -659   | -657   |
| 125       | -806   | -803   | -803   | -803   | -800   | -801   | -799   |
| 250       | -1495  | -1492  | -1491  | -1491  | -1489  | -1490  | -1488  |
| 500       | -2914  | -2908  | -2911  | -2911  | -2908  | -2909  | -2907  |
| 750       | -4316  | -4305  | -4313  | -4313  | -4308  | -4311  | -4310  |
| 1000      | -5643  | -5628  | -5641  | -5641  | -5635  | -5639  | -5639  |
| 1500      | -8062  | -8040  | -8061  | -8061  | -8056  | -8059  | -8060  |

### Table 4 Average BIC information criterion for different GARCH specifications across 100 resamplings for different estimation windows

| Train obs | (1, 1) | (1, 2) | (2, 1) | (2, 2) | (2, 3) | (3, 2) | (3, 3) |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| 101       | -657   | -651   | -648   | -648   | -643   | -644   | -639   |
| 125       | -798   | -792   | -789   | -789   | -784   | -784   | -780   |
| 250       | -1485  | -1478  | -1474  | -1474  | -1469  | -1469  | -1464  |
| 500       | -2902  | -2892  | -2890  | -2890  | -2883  | -2884  | -2878  |
| 750       | -4302  | -4287  | -4290  | -4290  | -4281  | -4284  | -4278  |
| 1000      | -5629  | -5609  | -5617  | -5617  | -5606  | -5610  | -5605  |
| 1500      | -8046  | -8019  | -8035  | -8035  | -8024  | -8028  | -8023  |
In the special cases of normal, Student-$t$ and Cauchy distributions, the induced density generators are $m$-dimensional version of the original generator of $X$.

For the proof of Properties 1, 2 and 3, we refer to Fang et al. (1990).

This implies that any portfolio $Y = \beta_1 X_1 + \ldots + \beta_n X_n$ of elliptically distributed variables is distributed accordingly with a (univariate) elliptical distribution, which is a location-scale distribution. Furthermore, for any univariate elliptical distribution all moments can be obtained from the first and second moments (if defined). In particular, for centered variables with zero mean ($\mu_Y = 0$), the resulting distribution of $Y$ is symmetrical around zero and it has all odd moments equal to zero and all even moments given by:

$$\mu_{2m} = c_m\mu_2^m,$$

with
$c_m = \frac{(2m)!}{(2^{m!})(\psi^{(1)}(0))^m}$.

Where $\psi^{(m)}(0)$ indicated the $m^{th}$ derivative of $\psi(\omega)$ computed at $\omega = 0$.

As an example, in the normal (0,1) case, $c_1 = 1$, $c_2 = 0$ for all $m = 1, 2, ..., $ the kurtosis is $\mu_4(\phi) = \frac{4!}{24^4} = 3$, and $\mu_2(\phi) = \frac{(2m)!}{(2^{m!})}$. For the proof we refer to Berkane and Bentler (1986), which derived this property by successive differentiation of $\phi(\cdot)$, and to Maruyama and Seo (2003), which attained the same result by expressing the elliptical distribution in terms of a random vector with uniform distribution on the unit sphere.

Therefore, the mean–variance optimization is of general applicability and relevance for any portfolio generated from multivariate elliptically distributed variables.

**Appendix C. Orthogonal GARCH estimation and further results**

In this section, we report details on the O-GARCH estimation discussed in Section 6. In particular, Tables 3 and 4 report the average AIC and BIC statistics across the 100 resamplings we considered in the experiment. As discussed in "O-GARCH comparison" section, both the AIC and BIC criteria support a GARCH(1,1) specification.

Table 5 below reports Median, 5th and 95th percentiles all the statistics obtained. The table shows that the GARCH(1,1) specification delivered the lowest AIC and BIC statistics for each of the main percentiles considered. In other words, the GARCH(1,1) specification selected in our experiment is the preferred specification according to the AIC and BIC criteria in all cases, and not only in mean across resamplings.

Having selected the GARCH(1,1) specification, for each resampling we estimated the model parameters (Table 6 reports the estimated parameters and corresponding p-Values), forecasted one steps ahead principal components and then reconstruct the covariance matrix as in Eq. 5.

Having selected the O-GARCH(1,1) specification we performed the same experiments discussed in "Impact of precision matrix estimate on optimal portfolios" section to compare the optimal minimum variance portfolio features when the O-GARCH covariance is used in the optimization. Figure 11 reports the number of active long and short positions obtained with the O-GARCH covariance matrix. As discussed in "O-GARCH comparison" section, the O-GARCH positions closely track the full covariance behaviour leading to the same conclusions and instability problems.

**Appendix D. Turnover comparison**

In "Backtest" section we compared the performances and stability of daily rebalanced portfolios, constructed based on OGARCH and TMFG precision matrices. Particularly relevant for the purpose of this paper is the stability of portfolios to new observations that, from a practitioner standpoint, results in lower turnover and transaction costs. We reported the average turnover in Fig. 9, while Fig. 12 reports a histogram of all daily turnovers computed across 100 resamplings and comparing TMFG to OGARCH for each estimation window. The histograms show that, other than delivering a significantly lower turnover on average as noted in "Backtest" section, TMFG portfolios’ turnover is heavily positively skewed, with a pick in the distribution around
zero, supporting and strengthening our claim of superior stability of TMFG portfolios.

Acknowledgements The authors acknowledge partial support from ESRC (ES/K002309/1), EPSRC (EP/P031730/1) and EC (H2020-ICT-2018-2 825215).

Fig. 12 Average daily turnover across 100 resamplings for different estimation window lengths

Declarations

Conflict of interest The author declare that they have no conflict of interests.
References

Alexander, C., and A. Chibumba. 1997. Multivariate orthogonal factor garch. Mimeo: University of Sussex.

Arditti, F., and H. Levy. 1976. Portfolio efficiency analysis in three moments: The multiperiod case. Journal of Finance 30: 797–809.

Aste, T., and T. Di Matteo. 2006. Dynamical networks from correlations. Physica A: Statistical Mechanics and its Applications 370: 156–161. https://doi.org/10.1016/j.physa.2006.04.019

Bamberg, G., and G. Dorfleitner. 2001. Multivariate orthogonal factor garch models: a survey. Journal of Applied Econometrics 21 (1): 79–109.

Baw, V.S. 1978. Safety-first, stochastic dominance, and optimal portfolio choice. Journal of Financial and Quantitative Analysis 13 (2): 255–271.

Bera, A.K., and M.L. Higgins. 1993. Arch models: Properties, estimation and testing. Journal of Economic Surveys 7 (4): 305–366.

Berkane, M., and P. Bentler. 1986. Moments of elliptically distributed random variates. Statistics & Probability Letters 4 (6): 333–335.

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

Bollerslev, T., R. Chou, and K.F. Kroner. 1992. Arch modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52 (1–2): 5–59.

Boudt, K., A. Galanos, S. Paysier, and E. Zivot. 2019. Chapter 7 - multivariate garch models: a survey. Journal of Applied Econometrics 26 (4): 736–751.

Boukhat, K., A. Galanos, S. Paysier, and E. Zivot. 2019. Chapter 7 - multivariate garch models: a survey. Journal of Applied Econometrics 26 (4): 736–751.

Brechmann, E.C., K. Hendrich, and C. Czado. 2013. Conditional copula simulation for systemic risk stress testing. Insurance: Mathematics and Economics 53 (3): 722–732.

Broadie, M. 1993. Computing efficient frontiers using estimated parameters. Annals of Operations Research 45: 215–229.

Brodie, J., I. Dubachev, C. De Mol, D. Giannone, and I. Loris. 2009. Sparse and stable markowitz portfolios. Proceedings of the National Academy of Sciences 106 (30): 12267–12272.

Caporin, M., and M. McAleer. 2012. Do we really need both bekk and dcs? a tale of two multivariate garch models. Journal of Economic Surveys 26 (4): 736–751.

Chamberlain, G. 1983. A characterization of the distributions that imply mean-variance utility functions. Journal of Economic Theory 29 (1): 185–201.

Cramér, H. 1946. Mathematical methods of statistics. Princeton landmarks in mathematics and physics. Princeton: Princeton University Press.

Daniels, H.E. 1965. The asymptotic efficiency of a maximum likelihood estimator. Mathematika 9 (1): 149–161.

De Franco, C., J. Nicolle, and H. Pham. 2019. Bayesian learning for the Markowitz portfolio selection problem. International Journal of Theoretical and Applied Finance (IJTAF) 22 (07): 1–40.
Maruyama, Y., and T. Seo. 2003. Estimation of moment parameter in elliptical distributions. *Journal of the Japan Statistical Society* 33: 215–229.

Massara, G. P., Aste, T., 2019. Learning clique forests. arXiv preprint https://arxiv.org/abs/1905.02266.

Massara, G. P., di Matteo, T., Aste, T., 2015. Network filtering for big data: Triangulated maximally filtered graph. http://arxiv.org/abs/1505.02445.

Mei, X., V. DeMiguel, and F.J. Nogales. 2016. Multiperiod portfolio optimization with multiple risky assets and general transaction costs. *Journal of Banking & Finance* 69: 108–120.

Mercurio, P.J., Y. Wu, and H. Xie. 2020. An entropy-based approach to portfolio optimization. *Entropy* 22 (3): 332.

Meucci, A., 2010. Fully flexible views: Theory and practice. https://ideas.repec.org/p/arn/papers/1012.2848.html

Michaud, R., 1989. The markowitz optimization enigma: Is optimized optimal? Working Paper, University of Augsburg (45), 31–42.

Michaud, R., and R. Michaud. 1998. *Efficient asset management: A practical guide to stock portfolio optimization and asset allocation*. Boston: Harvard Business School Press.

Nawrocki, D. 1996. Portfolio analysis with a large universe of assets. *Applied Economics* 28: 1191–1198.

Owen, J., and R. Rabinovitch. 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance* 38 (3): 745–752.

Philippatos, G.C., and C.J. Wilson. 1972. Entropy, market risk, and the selection of efficient portfolios. *Applied Economics* 4 (3): 209–220.

Pola, G. 2016. On entropy and portfolio diversification. *Journal of Asset Management* 17 (4): 218–228.

Procacci, P.F., and T. Aste. 2019. Forecasting market states. *Quantitative Finance* 19 (9): 1491–1498.

Quaranta, A.G., and A. Zaffaroni. 2008. Robust optimization of conditional value at risk and portfolio selection. *Journal of Banking & Finance* 32 (10): 2046–2056.

Scherer, Bernd. 2006. Portfolio construction and risk budgeting. New York: Risk Books.

Scherer, B., K. Winston, and C. O’Cinneide. 2012. *Bayesian methods in investing*. Oxford: Oxford University Press.

Sklar, M. J. 1959. Fonctions de repartition a n dimensions et leurs marges. *Publications de l’Institut Statistique del’Université de Paris* 8: 229–231.

Song, W.-M., T. Di Matteo, and T. Aste. 2012. Building complex networks with platonic solids. *Physical Review E* 85: 046115.

Tumminello, M., T. Aste, T. Di Matteo, and R.N. Mantegna. 2005. A tool for filtering information in complex systems. *Proceedings of the National Academy of Science* 102: 10421–10426.

Van Der Maaten, L., E. Postma, and J. Van den Herik. 2009. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research* 10: 66–71.

van der Weide, R. 2002. Go-garch: A multivariate generalized orthogonal garch model. *Journal of Applied Econometrics* 17 (5): 549–564.

Wald, A. 1949. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20 (4): 595–601.

Xing, X., J. Hu, and Y. Yang. 2014. Robust minimum variance portfolio with 1-infinity constraints. *Journal of Banking & Finance* 46: 107–117.

Yao, H., J. Huang, Y. Li, and J. Humphrey. 2021. A general approach to smooth and convex portfolio optimization using lower partial moments. *Journal of Banking & Finance* 129: 106167.

Zhang, B., Y. Wei, J. Yu, X. Lai, and Z. Peng. 2014. Forecasting var and es of stock index portfolio: A vine copula method. *Physica A: Statistical Mechanics and its Applications* 416: 112–124.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Pier Francesco Procacci graduated in Finance and Economics from Sapienza, University of Rome in 2011. He received a PhD in Computer Science from UCL in 2022. From 2016 to 2017, Pier worked at Banca Finnat Euramerica as Assistant Portfolio Manager. In 2017 he joined Eurlizon as Data Scientist. From 2018 to 2020 he joined the Quant Equitydepartment of Citigroup as Senior Associate and in 2021 he has been Vice President of Quant Equity. His current researchactivities are focused on non-stationary complex systems and end their time evolution.

Tomaso Aste obtained a degree in Physics from the University of Genoa in 1990; in 1994 he was then awarded a PhD inMaterial Sciences from the Politecnico di Milano. After two Post-Docs at the University of Strasbourg (’94–’98) and atthe University of Genoa (’98–’01), from 2002 to 2006 he joined the Department of Applied Mathematics, at The Australian National University (Canberra, Australia), from 2006 to 2009 he has been Associate Professor in the same University. From 2009 to 2012 he moved to UK taking a position as Reader at The School of Physics at The University of Kent inCanterbury UK. Then since 2012 he moved to UCL, Computer Science where he is Head of the Financial Computing & Analytics Group and Director of the MSc programme in Financial Risk Management. His current researchinterests concern the investigation of complex systems and complex materi-als by combining network theory and statisticalmethods to decode the properties of these systems and to extract information concerning their organization.