Combining robust covariance estimates and new dependency measures : an innovative approach to portfolio allocation

Thibault Soler

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

UNIVERSITÉ PARIS I PANTHÉON-SORBONNE
ÉCOLE DOCTORALE ED-465
CENTRE D'ÉCONOMIE DE LA SORBONNE

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS I PANTHÉON-SORBONNE

Spécialité : Sciences Économiques

présentée et soutenue par

Thibault Soler

Vendredi 26 Février 2021

Combining Robust Covariance Estimates and New Dependency Measures: An Innovative Approach to Portfolio Allocation

Thèse dirigée par Christophe Chorro et Philippe de Peretti

JURY :

Monica Billio	Professeur, Università Ca' Foscari Venezia	Rapporteur
Christophe Chorro	Maître de Conférences HDR, Université Paris I	Directeur de Thèse
Serge Darolles	Professeur, Université Paris-Dauphine	Examinateur
Jean-David Fermanian	Professeur, ENSAE	Examinateur
Olivier Guéant	Professeur, Université Paris I	Président du jury
Emmanuelle Jay	Docteur, La Banque Postale AM	Encadrant Fideas Capital
Philippe de Peretti	Maître de Conférences HDR, Université Paris I	Directeur de Thèse
Giulia Rotundo	Professeur, Sapienza Università di Roma	Rapporteur

Remerciements

Cette thèse a été menée au sein du laboratoire Centre d'économie de la Sorbonne (CES) et de Fideas Capital.

Tout d'abord, je tiens à remercier Christophe Chorro, Emmanuelle Jay et Philippe de Peretti qui m'ont fait l'honneur de m'encadrer durant cette thèse. Ce fut un plaisir de travailler avec eux et de bénéficier de leur expertise. J'ai énormément progressé à leur côté et je leur en serai toujours reconnaissant. Je tiens à exprimer plus particulièrement toute ma gratitude envers Christophe Chorro et Emmanuelle Jay pour leur gentillesse, leur disponibilité et le temps qu'ils m'ont accordé depuis le début de ma thèse. Je tiens aussi à remercier Jean-Philippe Ovarlez pour son apport non négligeable à ce travail, et de m'avoir initié à certaines problématiques de la théorie du signal.

Je remercie les membres de mon jury de thèse, pour l'intérêt qu'ils ont montré à mon travail. C'est un grand honneur pour moi que Monica Billio et Giulia Rotundo aient accepté de rapporter ma thèse. Je suis également très honoré que Serge Darolles, Jean-David Fermanian et Olivier Guéant aient participé à mon jury de thèse.

Je tiens à remercier également toute l'équipe de Fideas Capital, et tout particulièrement Pierre Filippi et Alexis Merville pour m'avoir fait confiance dès mon stage de master et m'avoir offert la possibilité de réaliser une thèse au sein de Fideas Capital. Je les remercie aussi pour leur disponibilité et leur constante interaction avec l'équipe de recherche.

Mes derniers remerciements vont bien évidemment à mes parents, ma soeur, mon frère, François, Catherine et Bernard. Ils ont tous été importants pour moi durant ces années de thèse. Enfin, mes pensées les plus affectueuses vont à celle qui est à mes côtés depuis plusieurs années et qui m'a toujours soutenu durant cette thèse.

Pour finir, je dédie ce travail à mes grands-parents qui m'ont toujours soutenu et sans qui je ne serais pas l'homme que je suis devenu aujourd'hui.

Thibault Soler

"Caminante no hay camino, se hace camino al andar"
Antonio Machado, Proverbios y cantares (XXIX, 1917)

Contents

Abstract

This thesis tackles portfolio allocation problems in studying robust covariance matrix estimators and the dynamic dependence between financial assets to improve the overall performance of risk-based allocation strategies. Today, it is well-established that the solution proposed by Markowitz leads to poor out-of-sample performance, due to estimation errors on the input variables, especially for the expected return. Despite several extensions to the mean-variance strategy over the last decades, most practitioners prefer simpler and more robust models such as the Minimum Variance portfolio (MinVar), the Equal Risk Contribution portfolio (ERC), and the Most Diversified portfolio (MDP), where expected returns are put aside and the covariance matrix estimation is the sole focus. However, two main issues remain: first, the Sample Covariance Matrix estimator (SCM) is inaccurate under non-Gaussian assumptions (asymmetry and heavy tails) and small sample sizes; second, the covariance matrix does not capture the dependency structure among financial assets (spillover and feedback effects), leading to incomplete risk assessments of investment universes.

In the first part of this thesis, we focus on the covariance matrix estimation and we develop for it a robust and de-noised estimator adapted to more realistic assumptions on financial asset returns. This estimator based on the Tyler-M estimator and the Random Matrix Theory (RMT) is adapted to non-Gaussian distributions (elliptical distribution) and we show that the assets should be preferably classified in homogeneous groups before applying the proposed methodology.

The second part of this thesis is dedicated to assessing the dynamic dependence between financial assets using the Generalized Partial Directed Coherence measure (GPDC) to take into account both the direction and the strength of causal relationships among financial assets. However, we show that a naive estimation of the Vector Autoregressive model (VAR) leads to poor results for the GPDC measure. To accurately capture diffusion patterns, we propose a parsimonious estimation (mBTS-TD) of the VAR model (no estimation of non-significant coefficients) by combining two subset selection methods, the modified Backward-in-Time Selection method (mBTS) and the Top-Down strategy (TD).

Finally, in the last part, we derive from the local directed weighted clustering coefficient an indicator adapted to the number of connections in the network in order to remove the most unstable assets (systemic and influenced) before allocating portfolios. Moreover, an empirical study is carried out that demonstrates that combining all of the results of the different chapters significantly improves risk-based allocation strategies.

Keywords: Portfolio allocation, Multivariate time series, Covariance matrix, Random matrix theory, Factor model, Elliptical distributions, Vector AutoRegressive model, Subset selection methods, Causality measures, Frequency causality measures, Financial networks, Clustering coefficient

Notations

$\mathbb{Z}$	Set of integer numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{C}$	Set of complex numbers
i	$\sqrt{-1}$
z^*	The conjugate of z
$\|.\|$	The absolute value or modulus for complex numbers
$\mathbf{v}$	Column vectors (bold lowercase letters)
v_j	The element j of $\mathbf{v}$
$\mathcal{L}(.)$	Toeplitz operator (1.4.1)
$\mathbf{0}_m$	The $m \times 1$ vector of zeros
$\mathbf{1}_m$	The $m \times 1$ vector of ones
$\|\mathbf{v}\|_1$	The L_1-norm $\|\mathbf{v}\|_1 = \sum_j \|v_j\|$
$\|\mathbf{v}\|_2$	The L_2-norm $\|\mathbf{v}\|_2 = \sqrt{\sum_j \|v_j\|^2}$
$\boldsymbol{A}$	Matrices (bold capital letters)
a_{jk}	The element j,k of $\boldsymbol{A}$
$\widehat{\mathbf{A}}$	Any estimate of matrix $\mathbf{A}$
$\boldsymbol{I}_m$	The $m \times m$ identity matrix
$\boldsymbol{A}'$	The transpose of $\boldsymbol{A}$
$\boldsymbol{A}^H$	The Hermitian (conjugate transpose) of $\boldsymbol{A}$
$\det(\boldsymbol{A})$	The determinant of $\boldsymbol{A}$
$Tr(\mathbf{A})$	The trace of $\boldsymbol{A}$
$\|\boldsymbol{A}\|_1$	The L_1-norm $\|\mathbf{A}\|_1 = \max_k \sum_j \|a_{jk}\|$
$\|\mathbf{A}\|$	The spectral norm (or L_2-norm $\|.\|_2$) $\|\mathbf{A}\| = \sqrt{\lambda_{max}(\mathbf{A}^H\mathbf{A})}$ where λ_{max} is the maximum eigenvalue of $\mathbf{A}^H\mathbf{A}$
$\|\mathbf{A}\|_F$	The Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\sum_{j,k} \|a_{jk}\|^2}$
$\mathcal{T}(.)$	Toeplitz rectification operator (1.4.1)
vec	The column stacking operator
$\otimes$	The Kronecker product
$\mathbb{1}_{(.)}$	The indicator function (1 if the condition is fulfilled, 0 otherwise)
$\mathbb{E}[.]$	Expected value
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution
$\chi^2(k)$	Chi-squared distribution
$\mathbf{G}$	Network of nodes and edges
$\mathbf{V}$	Set of nodes
$\mathbf{E}$	Set of edges (links)
$\mathbf{Z}$	The $m \times m$ adjacency matrix
$\mathbf{Z^u}$	Unweighted network where $z^u_{jk} = 1$ if $(j,k) \in \mathbf{E}$ and 0 otherwise
$\mathbf{Z^w}$	Weighted network where $z^w_{jk} \in \mathbb{R}^+$ if $(j,k) \in \mathbf{E}$ and 0 otherwise

Abbreviations

AHC	Ascending Hierarchical Clustering
AIC	Akaike Information Criterion
AP	Affinity Propagation
BIC	Bayesian Information Criterion
bp	basis points (one basis point is equal to 0.01%)
BU	Bottom-Up strategy
CAPM	Capital Asset Pricing Model
CH	Caliński-Harabasz criterion
DC	Directed Coherence measure
$\mathcal{DR}$	Diversity Ratio
DTF	Direct Transfer Function measure
ERC	Equal Risk Contribution portfolio
EW	Equally Weighted portfolio
GC	Granger's non-causality tests
GMVP	Global Minimum Variance portfolio
GPDC	Generalized Partial Directed Coherence measure
Lasso	Least Absolute Shrinkage and Selection Operator
LS	Least Squares estimator
LW	Ledoit & Wolf shrinkage estimator
mBTS	modified Backward-in-Time Selection method
MDP	Most Diversified Portfolio
MIME	Mutual Information on Mixed Embedding
MinVar	Minimum Variance portfolio
MLE	Maximum likelihood estimator
MP	Marčenko-Pastur
OLS	Ordinary Least Squares estimator
PDC	Partial Directed Coherence measure
PMIME	Partial Mutual Information on Mixed Embedding
RIE	Rotational Invariant Estimator
RMT	Random Matrix Theory
SCM	Sample Covariance Matrix
TD	Top-Down strategy
TE	Transfer Entropy
TT	Testing Procedure (t-test)
VAR	Vector AutoRegressive model
VarMax	Maximum Variety portfolio (equivalent to the Most Diversified Portfolio)
$\mathcal{VR}$	Variety Ratio (equivalent to $\mathcal{DR}$)

Chapter 1

Introduction

1.1 Motivations

The main challenge in asset management has always been how to allocate financial assets in order to maximize an investment's net worth while efficiently managing the portfolio risk. In 1952, Markowitz [1, 2] proposed a solution to the asset selection problem under the mean-variance framework, supposing that investors only care about the expected return (the mean) and the risk level (the variance) of their portfolio. The optimal weights among assets are thus obtained by maximizing the expected return for a given risk level, or minimizing the risk for a given expected return. All such portfolios form the efficient frontier that gives the best expected return for a given risk level or vice versa. For an investor, it represents the trade-off between risk and expected return when allocating a portfolio. Moreover, the efficient frontier illustrates the benefits of diversification, because a well-diversified portfolio can reduce the risk while preserving the same level of expected return, or even increase the portfolio return without increasing risk. Despite its powerful theoretical framework, the mean-variance strategy suffers from major pitfalls in practice. First, the input parameters are not known *a priori* and must be estimated, leading to estimation errors, especially for the expected return. For instance, Chopra and Ziemba in [3] showed that estimation errors in the means are about ten times greater than those in the variances, and twenty times greater than those in the covariances. Second, the mean-variance solution is very sensitive to the input parameters, again mainly for the expected returns [4, 5]. If these drawbacks are ignored, the traditional mean-variance portfolio provides a highly concentrated portfolio with extreme weights, unstable composition over time and poor out-of-sample performances [6, 7, 8].

To overcome these limitations, several extensions have emerged in the literature over the last fifty years. These extensions can be divided in two classes. The first class includes the Bayesian approach, to estimate unknown parameters reducing the estimation errors such as predictive distribution of returns [9, 10], the Bayes-Stein approach based on shrinkage estimators [11, 12, 13, 14], or asset pricing models providing informative prior distributions of future returns [15, 16]. The second class includes more heterogeneous approaches such as robust portfolio allocation rules using bounded parameters or confident intervals [17, 18, 19], moment restrictions based on factor models [20], covariance matrix estimation [21, 22, 23] or specific optimization constraints [24, 25, 26]. Although these extensions reduce the portfolio sensitivity to the parameter estimates, they also increase the computational complexity while not guaranteeing

out-of-sample performances [27, 28]. Furthermore, most practitioners prefer simpler and more robust models where expected returns are put aside and the covariance matrix estimation is the sole focus.

The most well-known alternative allocation strategies are the naive Equally Weighted portfolio (EW) [28] and risk-based allocation strategies such as the Minimum Variance portfolio (MinVar) [3], the Equal Risk Contribution portfolio (ERC) [29] and the Most Diversified Portfolio[1] (MDP) [30].

The naive EW portfolio is the simplest way to allocate portfolios, since assets are allocated with the same weights without any parameter estimation nor complex optimization. Under the mean-variance framework, this portfolio is optimal only if all assets have the same expected returns, variances and covariances. However, if the risk levels are very heterogeneous, this strategy leads to poor risk diversification since even if the assets have the same weight in the portfolio, their contribution to the risk of the portfolio is higher for a risky asset than for a low-risk asset. Nevertheless, despite its shortcomings, the EW is widely used in practice by investors as documented in [31, 32, 33] and can even outperforms various extensions of mean-variance strategies [28].

As suggested in [3], the easiest way to put aside the expected return of the mean-variance strategy is to assume that all assets have the same expected return. Under this assumption, the optimal portfolio is the MinVar portfolio. This strategy only minimizes the variance of the final portfolio and the solution is unique. In [34, 35, 36, 37], the authors showed that MinVar portfolios improve returns with lower volatilities when compared to the natural cap-weighted[2] strategy. Moreover, since the financial crisis of 2007-2008, MinVar portfolios have been widely used by investors while providing higher performances than traditional factor strategies (e.g. dividend, growth, momentum, value, etc.), reinforcing the low volatility anomaly concept[3] [38, 39]. Nevertheless, in practice, the MinVar strategy leads to highly concentrated portfolios, especially in highly volatile and hence highly correlated markets requiring individual constraints for reducing idiosyncratic risk.

The risk parity strategy was first used by the asset management company Bridgewater in the 1990s. The original strategy allocated assets in proportion to their inverse volatility without considering covariances. In [40, 41], the author introduced the concept of risk budget considering both variances and covariances, thus extending the original risk parity. This strategy, now more commonly known as the ERC portfolio thanks to [29] is halfway between the EW and MinVar portfolios, allocating assets according to their contribution to the risk of the portfolio. It preserves the benefits of investing in all assets such as the EW portfolio while improving the risk diversification when risks are heterogeneous in the investment universe. In addition, it is now well established that risk diversification can improve portfolio returns [42, 43, 44]. Note that for high-dimensional universes, the ERC algorithm is time-consuming and does not always converge [45].

[1] In the following chapters, the Most Diversified Portfolio (MDP) will be referred to as the Maximum Variety portfolio (VarMax).

[2] The capitalization-weighted strategy allocates assets using market capitalizations (e.g. S&P 500, CAC 40, etc.)

[3] The low-volatility anomaly is the observation that low-volatility stocks can provide higher returns than high-volatility stocks, challenging assumptions about risk and return.

Since these works, different approaches to risk parity and/or ERC extensions have been proposed based on the market exposure, value-at-risk, expected shortfall, systematic risk, etc. [46, 47, 48, 49, 50, 51]

An alternative allocation strategy has been proposed by Choueifaty and Coignard in [30], based directly on portfolio diversification to reduce common risk exposures, providing an efficient alternative to cap-weighted portfolios [30, 37, 52, 53]. The Most Diversified Portfolio (MDP) maximizes the ratio of the weighted arithmetic mean of asset volatilities over the portfolio volatility. For long-only portfolios[4], the purpose of maximizing the Diversity Ratio ($\mathcal{DR}$) is to buy the most independent risks within a universe to reduce the portfolio volatility. Indeed, if we consider two independent assets with the same volatilities, the $\mathcal{DR}$ is equal to $\sqrt{2}$, thus reducing the portfolio's volatility by $\sqrt{2}$ and by $\sqrt{m}$ for m independent assets. Note that if average asset returns increase proportionally with volatility, then the MDP portfolio is the tangent portfolio on the efficient frontier [35](it has the highest Sharpe ratio [54]). Moreover, the MDP portfolio fulfills some interesting invariance properties [52]: duplication invariance, i.e. if an asset is duplicated in the universe, then MDP will be unchanged giving half the weight to each duplicated asset; leverage invariance, i.e the weighting remains unchanged whatever the policy of the underlying company in terms of leverage; positive linear combination invariance (po-li-co invariance), i.e. it stays unchanged if a positive linear combination of the assets of the universe is added as a new asset. Unlike MDP, the EW portfolio does not verify invariance properties, the MinVar portfolio only satisfies the duplication invariance, and the ERC porfolio only the leverage invariance. Nevertheless, as with the MinVar portfolio, the MDP strategy suffers from a high portfolio concentration requiring individual constraints in practice [37]. In addition, these strategies have a natural order of portfolio volatility, where the MinVar portfolio is unsurprisingly the least volatile, the MDP portfolio is the second least, followed by the ERC portfolio, and finally the EW portfolio [29].

Despite the fact that risk-based allocation strategies focus solely on the covariance matrix estimation, this step plays a central role and should not be neglected to ensure more stable portfolios and better out-of-sample performances [55, 56, 57, 58, 23, 59, 60, 61, 62, 63]. Estimating a parameter such as the covariance matrix using a given statistical method is a multifaceted task. The value of the parameter is expected to be as close as possible to its theoretical value. In financial modeling, the widely used estimator of the covariance matrix is the Sample Covariance Matrix (SCM). The SCM is actually the optimal estimator in the case of multivariate Gaussian samples and coincides in this case with the maximum likelihood estimator (MLE). Let $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_T) \in \mathbb{R}^{m \times T}$ be the matrix of observations where $\forall t \in [1, T]$, $\mathbf{r}_t$ is a m-vector of independent Gaussian with zero-mean and covariance matrix $\mathbf{\Sigma}$. Then, when $T \to \infty$ for a fixed m, the law of large numbers ensures that:

$$\left\| \frac{1}{T} \mathbf{R}\,\mathbf{R}' - \mathbf{\Sigma} \right\| \xrightarrow{a.s.} 0 \tag{1.1}$$

However, this biased estimator has two major drawbacks: first, the estimation becomes inaccurate when the number of observations T is not too large relative to the m variables (small sample sizes); second, the lack of robustness for non-Gaussian

[4]Long-only portolfio: all quantities invested in assets are necessarily greater than or equal to 0.

distributions (asymmetry and heavy tails). These issues are now well-identified in many areas such as signal processing and finance, which confirm the very poor performances of the SCM estimator. Furthermore, it is well-known that asset returns are non-Gaussian [64, 65, 66, 67, 68] usually exhibiting stylized facts such as asymmetry, fat tails, and tail dependence, leading to large estimation errors. The assumption made on this distribution plays a fundamental role in the estimation accuracy and should therefore question the choice of the estimator.

Another key aspect to consider in portfolio allocation problems is the dynamic dependence between financial assets, including spillover and feedback effects. Indeed, according to their market importance, economic fragility, business activity and/or geographical position, assets will not react identically if a market shock occurs. Ignoring the systemic or influenced nature of an asset leads to incomplete risk assessments on the investment universe. Therefore, it is essential to assess the causal relationships between assets, since the covariance matrix used as a parameter in the risk-based allocation strategies only quantifies the risk through volatilities and behavior similarities. Over the last two decades and since the seminal work of Mantegna [69], the use of network theory to represent the market dependency structure (financial network) has played an important role in the literature on portfolio allocation problems [70, 71, 72, 73, 74, 75, 76]. Such approach provides useful insights on the portfolio selection process to understand complex interactions. Nevertheless, common approaches to recover the network topology such as the sample correlation matrix or Granger non-causality tests [77] lead to partial information, never providing indications of both the direction and strength of causal relationships. Recovering an accurate network topology (directed weighted network) requires the use of dependency measures that assess both the direction and strength of causal relationships.

1.2 Literature review

1.2.1 Covariance matrix estimators

Estimating covariance matrices is a classical problem in multivariate statistics. As already stated, the covariance matrix estimation is conditioned by both the number of variables m relative to the number of observations T and the underlying multivariate distribution of the variables. When the variables are Gaussian and $T >> m$, the SCM estimator $\widehat{\mathbf{\Sigma}}_{scm}$ (1.1) converges almost surely to the theoretical covariance matrix $\mathbf{\Sigma}$. Nevertheless, in financial modeling, these assumptions are not fulfilled. In practice, few historical observations are used (1 or 2 years of daily returns), in order to provide only the most recent information in the optimization process, and the asset returns are non-Gaussian (asymmetry, fat tails and tail dependence). Several approaches have therefore been proposed to address these issues.

According to several works [78, 79, 80], it has been shown that when $m \geq 3$, there exists a better estimator than $\widehat{\mathbf{\Sigma}}_{scm}$ in terms of mean square error [80]. This estimator uses external information to bias $\widehat{\mathbf{\Sigma}}_{scm}$ and leads to the linear shrinkage first introduced by Haff [81]. This estimator is simply a weighted average of the SCM and the identity matrix $\mathbf{I}_m$

$$\widehat{\mathbf{\Sigma}}_{LS}(\xi) = \xi\, \widehat{\mathbf{\Sigma}}_{scm} + (1-\xi)\, \mathbf{I}_m$$

where $\xi \in [0,1]$ is the shrinkage parameter. If $\xi = 1$, this estimator coincides with the SCM estimate. This estimator shows that incorporating external information improves the estimate in terms of mean square error. The benefits of the linear shrinkage when compared to the SCM have been precisely quantified in [82]. Note that this estimator is also very useful for rank-deficient matrices ($m > T$).

Following this approach, Ledoit & Wolf (LW) [58] introduced a shrinkage estimator particularly adapted to financial asset returns and based on the single factor model of Sharpe [83], where the factor is a market index. The LW estimate is a linear shrinkage of the SCM and the covariance matrix containing the market information. This model can be written as follows:

$$r_{j,t} = \alpha_j + \beta_j\, F_t + \epsilon_{j,t},\ \forall j \in [1,m] \text{ and } \forall t \in [1,T] \tag{1.2}$$

where $r_{j,t}$ is the return of asset j at time t , α_j is the active return of the asset j, β_j is the asset sensitivity to the market index return, F_t is the market index return at time t, and $\epsilon_{j,t}$ is the idiosyncratic return for asset j at t assumed to follow independent Gaussian with zero-mean and variance σ^2. This latter term is assumed to be uncorrelated to the market index. Under these assumptions, the covariance matrix writes:

$$\mathbf{\Sigma}_r = \sigma_F^2\, \boldsymbol{\beta}\, \boldsymbol{\beta}' + \mathbf{\Omega}_\epsilon$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)'$, σ_F^2 is the variance of the market returns and $\mathbf{\Omega}_\epsilon$ the covariance matrix of the idiosyncratic error.

An estimator for $\mathbf{\Sigma}_r$ is given by:

$$\widehat{\mathbf{\Sigma}}_r = \hat{\sigma}_F^2\, \widehat{\boldsymbol{\beta}}\, \widehat{\boldsymbol{\beta}}' + \widehat{\mathbf{\Omega}}_\epsilon$$

where each $\hat{\beta}_j$ is estimated individually using the Ordinary Least Squares (OLS) estimator based on equation (1.2) and $\widehat{\mathbf{\Omega}}_\epsilon$ is a diagonal matrix composed of the OLS residual variances. Finally, $\hat{\sigma}_F^2$ is the sample variance of the market returns.

The LW estimator is therefore equal to:

$$\widehat{\mathbf{\Sigma}}_{LW}(\xi) = \xi\, \widehat{\mathbf{\Sigma}}_{scm} + (1-\xi)\, \widehat{\mathbf{\Sigma}}_r \tag{1.3}$$

where ξ is computed by minimizing the quadratic loss function $\left\| \xi\, \widehat{\mathbf{\Sigma}}_{scm} + (1-\xi)\, \widehat{\mathbf{\Sigma}}_r - \mathbf{\Sigma} \right\|_F^2$ as in [58].

Nonetheless, it is now well established that several factors can better capture common risks [84, 85, 86, 87, 88, 89], challenging the single market factor assumption of the Capital

Asset Pricing Model (CAPM) [83]. Since these works, multi-factor models have emerged based either on observable or statistical factors. Using this approach, the covariance matrix can be estimated by imposing some factor structure (systematic part of the risk) to reduce the number of parameters to be estimated [22, 90]. Let's assume that the investment universe contains m assets whose returns at each time $t = 1, \cdots, T$ are stored in the m-vector $\mathbf{r}_t$. Suppose also that $\mathbf{r}_t$ admits a K factors structure, where the $K < m$ common factors (known or unknown), and that the additive noise follows a multivariate centered Gaussian distribution. The model for $\mathbf{r}_t$ is as follows:

$$\mathbf{r}_t = \mathbf{B}_t \, \mathbf{f}_t + \boldsymbol{\epsilon}_t \tag{1.4}$$

where $\mathbf{r}_t$ is the m-vector of returns at time t, $\mathbf{B}_t$ is the $m \times K$-matrix of asset sensitivities to each factor at time t, $\mathbf{f}_t$ is the K-vector of factor values at t supposed to be common to all assets, and $\boldsymbol{\epsilon}_t$ is a m-vector of Gaussian white noises with variances σ^2 uncorrelated to the factors. Given equation (1.4) the covariance matrix is given at time t by:

$$\boldsymbol{\Sigma}_t = \mathbf{B}_t \, \boldsymbol{\Sigma}_t^f \, \mathbf{B}_t' + \boldsymbol{\Omega}_t$$

where $\boldsymbol{\Sigma}_t^f = \mathbb{E}[\mathbf{f}_t \, \mathbf{f}_t']$ is the factor-related term and $\boldsymbol{\Omega}_t$ is the covariance matrix of $\boldsymbol{\epsilon}_t$. In the case where the factors are known *a priori*, we deal only with the estimation of $\mathbf{B}_t$ and $\boldsymbol{\epsilon}_t$. However, if the factors are unobservable and determined from the asset universe, which is very frequent in practice, statistical methods must be used. Determining, the number of factors is a tough task in all the model order selection problems. Subspace methods such as the Random Matrix Theory (RMT) aim to identify the K highest eigenvalues of $\boldsymbol{\Sigma}_t$ supposed to represent the K-factors, especially when the power of the factors is greater than the noise power.

The Random Matrix Theory (RMT) was originally introduced in nuclear physics by Eugène Wigner [91] to model the spectrum of heavy atoms and has now many applications in several domains. In this framework, the first result on the behavior of $\boldsymbol{\Sigma}$ comes from the paper by Marčenko and Pastur in 1967 [92]. This result allows one to quantify the error committed by the analysis of the eigenvalue spectral density. The authors found an equation for the eigenvalue density when $T, m \to \infty$, and $m/T \to c \in]0, \infty[$. More precisely, they establish in this case that if the variables are independent and identically distributed according to the centered Gaussian distribution with variance σ^2, then the eigenvalue spectral density of the SCM tends towards the Marčenko-Pastur's law (MP):

$$f_c(\lambda) = \frac{1}{2\pi\sigma^2 c\lambda} \sqrt{(\lambda^+ - \lambda)(\lambda - \lambda^-)} \, \mathbb{1}_{]\lambda^-, \lambda^+[}(\lambda) \tag{1.5}$$

where $\lambda^\pm = \sigma^2(1 \pm \sqrt{c})^2$. Note when we consider the normalized-SCM $\widehat{\boldsymbol{\Sigma}}_{nscm}$, σ^2 is equal to 1. This result shows the influence of c on the covariance matrix estimation accuracy. The empirical eigenvalues are noisy estimators of the true eigenvalues independently of T, because the eigenvalue spectral density diverges more and more from the Dirac delta function at 1 ("perfect estimation") as c becomes large (Fig. 1.1).

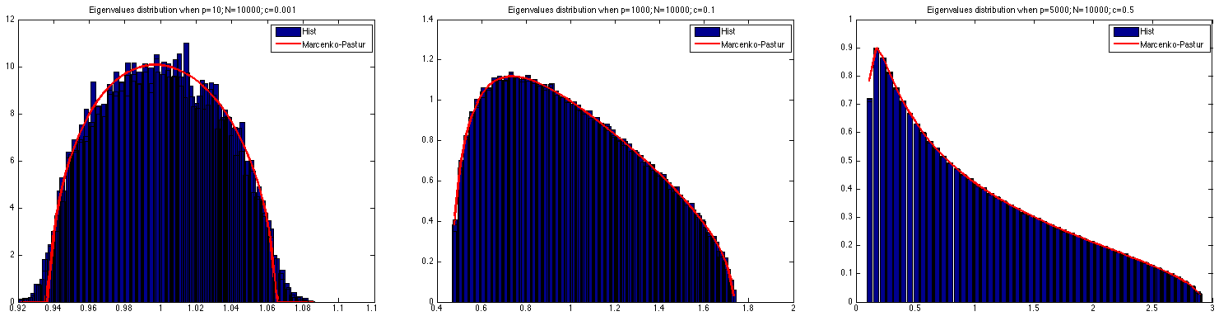


Figure 1.1: Histogram of the SCM eigenvalues computed from samples of reduced centered multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}_m, \mathbf{I}_m)$. Left side: the case of $m/T = 0.01$; Middle: the case of $m/T = 0.1$; Right: the case of $m/T = 0.5$. In all cases, $m = 1000$, and T varies in order to obtain the value of $c = m/T$. The theoretical distribution of Marčenko-Pastur's is shown in red.

According to the RMT, it is therefore shown that any eigenvalue above the Marčenko-Pastur's threshold $\lambda^+ = \sigma^2(1 + \sqrt{c})^2$ represents a dimension of the signal space and can therefore be detected as a target (or factor) of interest for the current study. The number of eigenvalues beyond this threshold determines the order of the model, and thus the number of factors carrying information in the universe. Moreover, empirical studies have confirmed the validity of MP law on financial data [55, 56, 57, 59]. These studies show that RMT helps identifying a solution to filter noise by considering that all eigenvalues within the MP theoretical distribution are only noise, and that those above the upper bound λ^+ distribution are signals (or factors) [55, 56]. In [56], Laloux et al proposed a de-noising procedure of $\widehat{\mathbf{\Sigma}}_{nscm}$ called "Eigenvalue clipping" and defined as follows:

$$\widehat{\mathbf{\Sigma}}_{clip} = \sum_{k=1}^{m} \lambda_k^{clip} \mathbf{u}_k \mathbf{u}_k' \tag{1.6}$$

with $\mathbf{u}_k$ the eigenvector associated to the eigenvalue λ_k of $\widehat{\mathbf{\Sigma}}_{nscm}$, and λ_k^{clip} defined as follows:

$$\lambda_k^{clip} = \begin{cases} \lambda_k, \text{ if } \lambda_k \geq (1 + \sqrt{c})^2 \\ \widetilde{\lambda}, \text{ otherwise} \end{cases}$$

where $\widetilde{\lambda}$ is chosen such that $Tr(\widehat{\mathbf{\Sigma}}_{clip}) = Tr(\widehat{\mathbf{\Sigma}}_{nscm})$.

This method simply sets the noisy eigenvalues as a constant value such that the trace of the matrix is preserved, thus creating a real difference between signal and noise significantly reducing the values close to λ^+. Although this method provides competitive out-of-sample results [61], in practice it suffers from shortcomings. It appears that the value of c is significantly different from the "effective" value [56] due to either small temporal autocorrelation in the time series [93, 94, 95] and/or the underlying hypothesis $\mathbf{\Sigma} = \mathbf{I}_m$ [96, 97], leading to detection in most cases of only the first component (market factor) which is not completely satisfactory.

In recent works [63, 98, 99], the authors proposed an optimal and fully observable estimator of the "true" covariance matrix, valid for large matrices. Using the theoretical quantification of the relationship between sample and population eigenvectors provided

in [100], they obtained an optimal rotational invariant estimator (RIE) for general covariance matrices by computing the overlap between the true and sample eigenvectors. This estimator also provides a tractable implementation when compared to [101, 102]. For large m, the optimal RIE of $\widehat{\mathbf{\Sigma}}_{nscm}$ is given by:

$$\widehat{\mathbf{\Sigma}}_{RIE} = \sum_{k=1}^{m} \lambda_k^{RIE} \, \mathbf{u}_k \, \mathbf{u}_k' \tag{1.7}$$

with $\mathbf{u}_k$ the eigenvector associated to the eigenvalue λ_k of $\widehat{\mathbf{\Sigma}}_{nscm}$, and λ_k^{RIE} defined as follows:

$$\lambda_k^{RIE} = \frac{\lambda_k}{|1 - c + c \, z_k \, s(z_k)|^2}$$

where $z_k = \lambda_k - i \, T^{-1/2}$ and $s(z)$ denotes the discrete form of the limiting Stieltjes transform

$$s(z) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{z - \lambda_j}$$

where $Tr(\widehat{\mathbf{\Sigma}}_{RIE}) = Tr(\widehat{\mathbf{\Sigma}}_{nscm})$ is ensured by multiplying each λ_k^{RIE} by η with $\eta = \sum_{k=1}^{m} \lambda_k / \sum_{k=1}^{m} \lambda_k^{RIE}$. The authors also observed that the optimal RIE systematically underestimates the small eigenvalues and thereby proposed a regularization procedure for small eigenvalues too close to 0 which attempts to correct the estimation error (see chapter 9 in [99]). This estimator applied to portfolio allocation strategies gives promising results providing lower volatility portfolios than those obtained when using the SCM, Ledoit & Wolf (LW) and Eigenvalue clipping methods.

The methods defined above address the estimation accuracy, only focusing on the sample size issue (m/T). Nevertheless, the non-Gaussian distribution of asset returns must also be considered. The field of robust estimation [103, 104, 105, 106] intends to deal with this problem. The theory of robustness was first studied in the 1960s by Huber and Tukey [105, 106, 103, 104, 107, 108]. Huber introduces M-estimators which are robust estimators generalizing the concept of MLE. A M-estimator of $\mathbf{\Sigma}$ is defined as the solution of the following equation:

$$\widehat{\mathbf{\Sigma}}_M = \frac{1}{T} \sum_{t=1}^{T} u\left(\frac{1}{m} \, \mathbf{r}_t' \, \widehat{\mathbf{\Sigma}}_M^{-1} \, \mathbf{r}_t \right) \mathbf{r}_t \, \mathbf{r}_t' \tag{1.8}$$

The existence and uniqueness of the solution of (1.8) has been shown by Maronna in [107], provided that the function $u(.)$ satisfies a set of general assumptions such as for example, being non-negative, non-increasing, and continuous on $\mathbb{R}^+$ (the whole set of assumptions can be found in [107]). Moreover, equation (1.8) admits a unique solution and is a consistent estimate of $\mathbf{\Sigma}$ that can be obtained by a classical iterative procedure. The Huber's M-estimator is defined for the following $u(.)$ function:

$$u(s) = \frac{1}{\beta} \min_{s \in \mathbb{R}^+} \left(1, \frac{a}{s}\right) = \frac{1}{\beta} \left(\mathbb{1}_{s \leq a} + \frac{a}{s} \mathbb{1}_{s > a} \right) \tag{1.9}$$

Replacing (1.9) in (1.8) gives the complete expression of the Huber's M-estimator of $\mathbf{\Sigma}$. Parameters a and β have to be set [103]. They control the percentage of attenuated

data. As an example, the greater is a and the closest to $\widehat{\mathbf{\Sigma}}_{scm}$ is the Huber's estimator. Figure 1.2 shows the shape of Huber's $u(.)$ function.

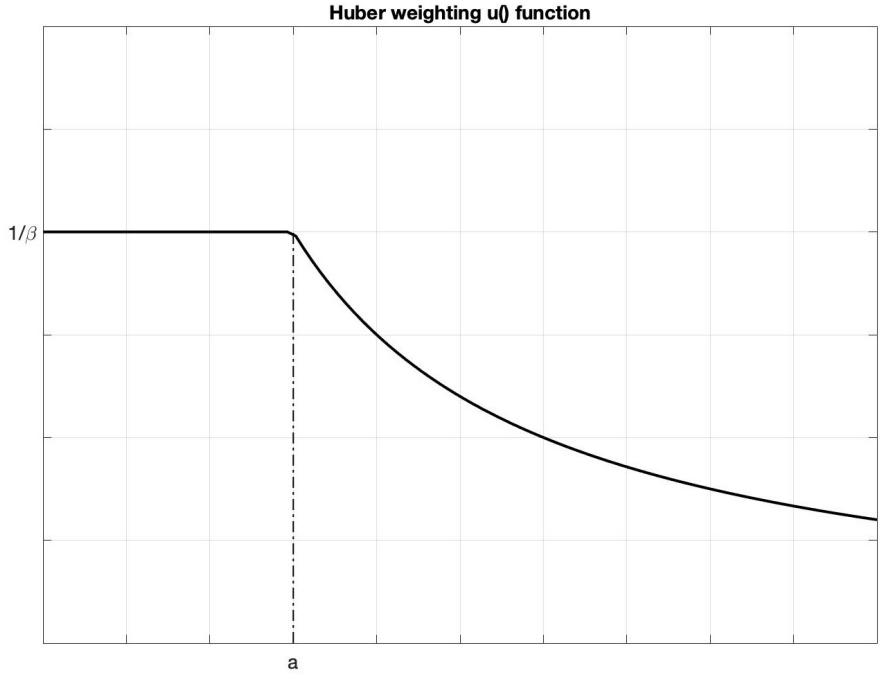


Figure 1.2: Huber's $u(.)$ function.

In [108], Tyler defines a special case of the above estimator (1.8) since instead of considering any $u(.)$ function, he defines $u(s) = s^{-1}$ satisfying Maronna's general assumptions [107] and providing more flexibility as it is a distribution-free estimator. Also known as the Fixed-Point (FP) estimator, the Tyler-M estimator when estimated under non-Gaussian assumptions (elliptical distributions with zero-mean [109, 110, 111, 112, 113]) is shown to be the "most robust" covariance matrix estimator in the sense of minimizing the maximum asymptotic variance [108, 114]. This estimator is defined as the unique solution, up to a scale factor, of the following equation:

$$\widehat{\mathbf{\Sigma}}_{tyl} = \frac{m}{T} \sum_{t=1}^{T} \frac{\mathbf{r}_t \, \mathbf{r}_t'}{\mathbf{r}_t' \, \widehat{\mathbf{\Sigma}}_{tyl}^{-1} \, \mathbf{r}_t} \tag{1.10}$$

This estimator can be normalized, for example, such that $Tr(\widehat{\mathbf{\Sigma}}_{tyl}) = m$. The solution can be found using the following recursive algorithm [115]:

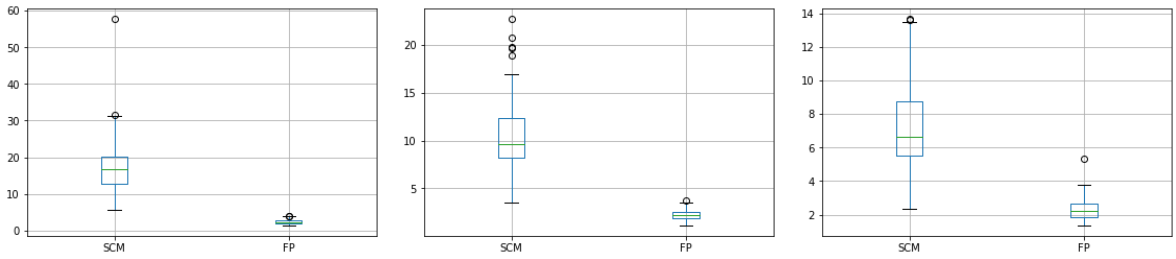
- Initialize $\widehat{\mathbf{\Sigma}}_{tyl,(1)}$ with any full-rank matrix (e.g. $m \times m$ identity matrix),
- Iterate over $k \geq 1$:

$$\widetilde{\mathbf{\Sigma}}_{tyl}^{(k+1)} = \frac{m}{T} \sum_{t=1}^{T} \frac{\mathbf{r}_t \, \mathbf{r}_t'}{\mathbf{r}_t' \, (\widehat{\mathbf{\Sigma}}_{tyl}^{(k)})^{-1} \, \mathbf{r}_t'}$$

$$\widehat{\mathbf{\Sigma}}_{tyl}^{(k+1)} = \frac{m}{Tr(\widetilde{\mathbf{\Sigma}}_{tyl}^{(k+1)})} \widetilde{\mathbf{\Sigma}}_{tyl}^{(k+1)}$$

  until the relative Frobenius norm between two consecutive values becomes lower than a fixed threshold $\epsilon \in \mathbb{R}^+$, i.e. until: $\dfrac{\|\widehat{\mathbf{\Sigma}}_{tyl}^{(k+1)} - \widehat{\mathbf{\Sigma}}_{tyl}^{(k)}\|_F}{\|\widehat{\mathbf{\Sigma}}_{tyl}^{(k)}\|_F} < \epsilon$

To illustrate the efficiency of the Tyler-M estimator compared to the SCM under non-Gaussian assumptions, we ran the following test: we simulate $T = 256$ observations of a size $m = 40$ sampled from a highly correlated Student's T distribution having shapes parameter $\nu = [5, 7, 10]$, and a Toeplitz-structured covariance matrix whose coefficient $\rho = 0.95$ (each element j, k of the Toeplitz matrix is defined by $\rho^{|j-k|}, j, k = 1, \cdots, m$). In Fig. 1.3, it appears clearly that the Tyler-M estimator is more robust than the SCM under Student's T distribution, having a significantly lower Frobenius norm between the theoretical and the estimated covariance matrix.



(a) (b) (c)

Figure 1.3: Boxplot of Frobenius norm between the true covariance matrix and the covariance estimation using the sample covariance matrix estimator (SCM) or the Tyler-M estimator (FP). The estimates are based on samples from a correlated multivariate Student's T distribution with sample size $T = 256$, $m = 40$, Toeplitz-structured covariance matrix with $\rho = 0.95$ and three degrees of freedom $\nu = 5$ (a), $\nu = 7$ (b), and $\nu = 10$ (c).

Note that classical robust covariance estimators require $T >> m$; otherwise, they do not perform well or are not defined. In this way, recent works [116, 117, 118, 115, 119, 62] based on the RMT have also considered robust estimation when $T \simeq m$. In [116, 117, 115], the authors proposed a hybrid robust shrinkage covariance matrix estimator on Tyler's M-estimator [108] and LW's shrinkage [82]. Moreover in [62], the authors show that applying an adapted estimation methodology (Shrinkage Tyler-M estimator) leads to achieving superior performance over many other competing methods under the MinVar framework.

1.2.2 Dependency measures and indicators in financial networks

In recent years and since the seminal work of Mantegna [69], networks have become a popular tool for representing financial markets due to their ability to describe the interactions or relationships between assets in a simple model. A network $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ is a set of objects with $\mathbf{V}$ the set of nodes (assets) and $\mathbf{E}$ the set of edges (links) between nodes. The edge (j, k) connects a pair of nodes j and k. The mathematical representation of a network is the $m \times m$ adjacency matrix $\mathbf{Z} = (z_{jk})$. Four types of network can be considered to represent the topology of financial markets. Undirected networks (symmetric matrix) which only describe the relationship between two assets, can be either weighted ($z^w_{jk} \in \mathbb{R}^+$ if $(j, k) \in \mathbf{E}$ and 0 otherwise) or unweighted ($z^u_{jk} = 1$ if $(j, k) \in \mathbf{E}$ and 0 otherwise), i.e. whether the strength of the relationships is quantified or not. Directed networks (non-symmetric matrix) where the relationship is characterized by both its existence and direction, can also be weighted or not.

The network approach provides useful information for understanding multiple interactions and describing complex systems such as financial markets [120, 121, 122, 123]. In particular, portfolio allocation problems ideally involve modeling the relationships among financial assets to capture diffusion schemes (spillover and feedback effects) to efficiently identify and manage portfolio risk. Indeed, if a market event occurs, the assets will react differently depending on their multiple connections, and neglecting these complex relationships leads to a lack of information on the nature of assets (central, peripheral, systemic or influenced) [124, 70, 71, 72, 125, 73, 74, 75, 76]. However, in such an approach, the network topology relies on the choice of dependency measures to accurately assess and quantify the system's complex interactions.

The sample correlation matrix is the common measure to recover the dependencies between assets through undirected networks [69, 120, 121, 122, 123, 124, 71, 72, 75, 76]. In [71], using correlation-based networks, Pozzi et al describe the highly connected assets as "central" and the poorly connected ones as "peripheral", combining several centrality/peripherality measures. They find that investing in peripheral assets leads to a more diversified portfolio and increases the return/volatility ratio relative to central assets. In the continuation of this work [72], they theoretically prove a negative relationship between the mean-variance optimal weights and the centrality of assets, showing that central assets tend to have "value" characteristics, i.e. having rather large market capitalizations, being undervalued and financially risky. Recent alternative works [75, 76] propose to use, for risk-based allocation strategies, an interconnectedness matrix derived from the clustering coefficient[5] [126, 127] instead of the classical covariance matrix. Such studies have confirmed the usefulness of correlation-based networks to improve the portfolio selection processes. But, while this approach is easy to implement and give interpretable results, several drawbacks exist. First, the sample correlation matrix is likely to provide a complete network, i.e. that all nodes are connected to each other. In this case, reduction dimension tools are required to keep only the most prominent relationships such as statistical tests or filtering methods (Minimum Spanning Trees [69] and Planar Maximally Filtered Graphs [128]). Second, the correlation coefficient does not distinguish between direct and indirect correlations, i.e. the behavior similarity between two assets may come from a third asset. To this end, the partial coefficient correlation can be used to correct this shortcoming as in [129, 130, 131]. Finally, the main limitation of correlation approaches remains that it cannot reflect cause-and-effect relationships and capture diffusion patterns, resulting in a significant loss of information on the network topology. Accordingly, causality measures are needed to distinguish whether any two assets interact directly.

Since the seminal work of Sims [132] in macroeconomics, Granger's non-causality tests [77] are probably the most popular of the available causality measures [133, 134, 135]. The concept of non-causality defined by Granger [77] is based on the idea that, if a time series $x_k(t)$ causes another time series $x_j(t)$, then the past of $x_k(t)$ will significantly decrease the forecast error of $x_j(t)$. Such tests have become standard tools to assess the direction of relationships and recover the network topology (directed network) in neuroscience [136, 137, 138, 139, 140] and in finance [70, 125, 73, 74]. Since these tests take into account causal relationships and can therefore capture diffusion patterns between assets, several works have studied their behavior before, during, and after

[5]The clustering coefficient measures how a node is embedded in the network (1.21).

market shocks (e.g. 2008 financial crisis, sovereign debt crisis, etc.). In [70], Billio et al show that hedge funds, banks, brokers/dealers and insurance companies were highly interdependent during the period 2000-2010 with a high level of systemic risk. They also show that banks are the most systemic assets relative to the other three industries, since they are more likely to transmit shocks. In the same way, the work in [125, 74] aim to identify in the networks the nodes' state change to anticipate phase transitions on financial markets.

Granger's non-causality tests first rely on the estimation of a Vector AutoRegressive (VAR) model to capture temporal relationships between time series. Let $\mathbf{x}(t) = (x_1(t), \ldots, x_m(t))'$ be a zero-mean m-dimensional stationary process admitting the following VAR(p) representation:

$$\mathbf{x}(t) = \mathbf{A}_1\mathbf{x}(t-1) + \cdots + \mathbf{A}_p\mathbf{x}(t-p) + \boldsymbol{\epsilon}(t), \quad t \in \mathbb{Z} \tag{1.11}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $(m \times m)$ coefficient matrices that describe the temporal relationships between the m time series, p is the model order, and $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \ldots, \epsilon_m(t))'$ is a $(m \times 1)$ vector of white noises with $\mathrm{E}[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}'(s)] = 0$ for $t \neq s$ and $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$. The matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ helps to test the VAR estimation accuracy or to determine the contemporaneous or instantaneous effects between the time series. Moreover, the VAR(p) is said to be stable, i.e. the VAR(p) is stationary with time-invariant means, variances, and autocovariances, if the roots $z^1, \ldots, z^p$ of the equation $det(\boldsymbol{I}_m - \boldsymbol{A}_1 z^1 - \ldots - \boldsymbol{A}_p z^p) = 0$ have modulus strictly greater than one.

A classical estimator of VAR coefficients[6] is the Least Squares estimator (LS), either in a multivariate (LS) or univariate (OLS) environment (equation by equation) [141]. The LS estimation for the VAR(p) model defined in (1.11) is as follows:

$$\widehat{\boldsymbol{\beta}} = ((\boldsymbol{X}\boldsymbol{X}')^{-1}\boldsymbol{X} \otimes \boldsymbol{I}_m)\boldsymbol{y} \tag{1.12}$$

where

- $\boldsymbol{y} = vec(\mathbf{x}(1), \ldots, \mathbf{x}(T))$ is the $(mT \times 1)$ vector of observations,
- $\boldsymbol{X}$ is the $(mp \times T)$ matrix with $\boldsymbol{X}(t) = \mathrm{vec}(\mathbf{x}(t-1), \ldots, \mathbf{x}(t-p))$,
- $\widehat{\boldsymbol{\beta}}$ is the $(m^2p \times 1)$ vector of the estimated VAR coefficients $\mathrm{vec}(\widehat{\boldsymbol{A}}_1, \ldots, \widehat{\boldsymbol{A}}_p)$

For the univariate case, the OLS estimator is given by:

$$\hat{\boldsymbol{b}}_j = (\boldsymbol{X}\boldsymbol{X}')^{-1}\boldsymbol{X}\boldsymbol{y}_j \tag{1.13}$$

where

- $\boldsymbol{y}_j = (x_j(1), \ldots, x_j(T))'$ is the $(T \times 1)$ vector of observations,
- $\boldsymbol{X}$ is the $(mp \times T)$ matrix defined in (1.12),
- $\hat{\boldsymbol{b}}_j$ is the $(mp \times 1)$ vector of the estimated coefficients.

[6] Maximum Likelihood and Yule-Walker estimators [141] are also widely used to estimate VAR coefficients

In addition to estimating VAR coefficients, the model order p is in most cases unknown and must also be estimated. This step is crucial to accurately capture the full dynamics of the system. The lag order p is often chosen to minimize an information criterion that takes the following general form:

$$\mathrm{IC}(p) = \ln\left(\det\left(\widehat{\boldsymbol{\Sigma}}_\epsilon(p)\right)\right) + g_T \frac{pm^2}{T}$$

where $\widehat{\boldsymbol{\Sigma}}_\epsilon(p)$ is either the unbiased or the maximum likelihood estimates of $\boldsymbol{\Sigma}_\epsilon(p)$ and g_T is the penalty factor which is a function of the sample size T. The most popular information criteria are the Akaike information criterion ($g_T^{AIC} = 2$) [142, 143], the Hannan-Quinn information criterion ($g_T^{HQ} = \ln\ln T$) [144], and the Bayesian information criterion ($g_T^{BIC} = \ln T$) [145]. The main difference between those three information criteria arises from the penalty factor strength. For a fixed $T \geq 16$, the AIC criterion tends to the largest order, BIC the smallest one and HQ is in between [141]. Moreover, the use of the unbiased estimator of $\boldsymbol{\Sigma}_\epsilon(p)$, which increases the penalty factor, may play an important role in model order selection.

Once the model order p and the VAR coefficients are estimated, Granger's non-causality can be assessed with a Wald multiple restrictions test [141] that tests whether the coefficients are jointly significant, e.g. $a_{kj}(1) = \ldots = a_{kj}(p) = 0$. The general null hypothesis is given by $\mathrm{H}_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{c}$, where $\boldsymbol{C}$ is a $(q \times m^2 p)$ matrix called the restriction matrix of the VAR coefficients (1 for tested coefficients and 0 otherwise), q denotes the number of restrictions, $\boldsymbol{\beta}$ is a $(m^2 p \times 1)$ vector with $\boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$, and $\boldsymbol{c}$ is a $(q \times 1)$ vector with $\boldsymbol{c} = \mathbf{0}_q$ for Granger non-causality. The Wald statistic is therefore

$$\Gamma = (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})' \left[\boldsymbol{C}\left((\boldsymbol{X}\boldsymbol{X}')^{-1} \otimes \widetilde{\boldsymbol{\Sigma}}_\epsilon\right)\boldsymbol{C}'\right]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$$

where $\widehat{\boldsymbol{\beta}}$ is the $m^2 p$-vector of the estimated VAR coefficients $\boldsymbol{\beta}$, $\boldsymbol{X}$ is the $(mp \times T)$ matrix defined in (1.12) and $\widetilde{\boldsymbol{\Sigma}}_\epsilon$ is the unbiased estimator of $\boldsymbol{\Sigma}_\epsilon(p)$ given by

$$\widetilde{\boldsymbol{\Sigma}}_\epsilon(p) = \frac{\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}'(t)}{T - pm - 1}$$

Under the null hypothesis H_0: $\Gamma \sim \chi^2(q)$. H_0 is not rejected (non-causality) for a given probability α if $\Gamma \leq \chi^2_\alpha(q)$, where $\chi^2_\alpha(q)$ is the α quantile of the distribution. The adjacency matrix $\mathbf{Z}^{GC} = (z_{jk}^{GC})$ $\forall j, k \in [1, \ldots, m]$ is therefore defined as follows:

$$z_{jk}^{GC} = \begin{cases} 1, & \text{if } \Gamma_{k\to j} > \chi^2_\alpha(q) \\ \\ 0, & \text{otherwise} \end{cases} \tag{1.14}$$

Though Granger's non-causality tests assess the existence and direction of interactions whereas the sample correlation matrix does not, they also suffer from some limitations. First the VAR estimation may fail for many well-known reasons: information criteria often do not provide optimal lags [141, 146, 147, 148]; VAR dimensionality which requires $m^2 p$ parameters to be estimated, leading to estimation errors for either high-dimensional system (large m) or for small sample size (small T) [149, 150, 151, 140]; contemporaneous or instantaneous effects between the time series creating spurious temporal relationships

(structural VAR model) [132]. Moreover, if a causality is detected (rejection of non-causality hypothesis), Granger's non-causality tests do not give any information about the causality strength, providing only directed unweighted networks, which inevitably leads to a loss of key information to evaluate interactions.

Another possible approach to recover the network topology is to extend Granger's non-causality tests to nonlinear dependency measures such as Transfer Entropy (TE) [152]. The TE quantifies the amount of time-delayed information between two dynamical systems. Given an information set, the transfer entropy from the time series $x_k(t)$ to the time series $x_j(t)$ is the amount of Shannon uncertainty [153] reduction in the future values of $x_j(t)$ when including the knowledge of the past value of $x_k(t)$. In [154], Barnett et al study the differences between Granger's non-causality tests (autoregressive model) and TE (information theory) measures and they point out that under Gaussian assumptions, these two causality measures are entirely equivalent, up to a factor of 2. However, the TE measure does not require any specific assumptions or models about the underlying system. Although this measure appears attractive in theory with promising results in many cases, TE has mostly been applied in bivariate cases, since become problematic for high-dimensional systems due to the "curse of dimensionality" even when appropriate estimators are used [135, 140]. To overcome this issue, dimension reduction was implemented in TE for both the bivariate (mutual information on mixed embedding (MIME) [155]) and the multivariate case (partial mutual information on mixed embedding (PMIME) [156, 73]).

Once the network is built, the identification of specific nodes is a key issue in network analysis and many indicators of interest exist for studying the network topology. In the financial literature related to network theory, two classes of indicators can be distinguished: first, indicators that assess the centrality/peripherality of a node such as degree, betweenness centrality, eccentricity, closeness and eigenvector centrality [157]; second, indicators that quantify to what extend a node is embedded into the network such as the clustering coefficient [126, 127] or community structure [158].

Centrality measures characterize the importance and influence of a given node within the network. Such measures are widely used in finance to classify each asset in terms of its relative position in the network [70, 159, 160, 125, 73] and in portfolio allocation problems to select the most central/peripheral assets before allocating portfolios as in [124, 71, 72]. In the sequel, we briefly introduce the most common centralities such as degree, betweenness centrality, eccentricity, closeness and eigenvector centrality [157].

For an unweighted network $\mathbf{Z}^u$, the degree centrality for a node j denoted by d_j measures its total number of edges. A high value indicates that the asset j is heavily central in the network. It is defined as follows:

$$d_j = \left(\mathbf{Z}^{u\prime} + \mathbf{Z}^u\right)_j \mathbf{1}_m \tag{1.15}$$

For a directed network, the degree includes the in-degree (number of edges pointing toward a given node) and out-degree (number of edges starting from j) given by:

$$d_j^{in} = \mathbf{Z}_j^{u\prime}\mathbf{1}_m \tag{1.16}$$

$$d_j^{out} = \mathbf{Z}_j^{u}\mathbf{1}_m \tag{1.17}$$

In a weighted network $\mathbf{Z}^w$, the node strength can also be measured. For a node j, the strength denoted by s_j is the weighted sum of all edges

$$s_j = \left(\mathbf{Z}^{w\prime} + \mathbf{Z}^{w}\right)_j \mathbf{1}_m \tag{1.18}$$

For a directed weighted network, the strength also includes the in-strength (weighted sum of edges pointing towards a given node) and out-strength (weighted sum of edges starting from j) given by:

$$s_j^{in} = \mathbf{Z}_j^{w\prime}\mathbf{1}_m \tag{1.19}$$

$$d_j^{out} = \mathbf{Z}_j^{w}\mathbf{1}_m \tag{1.20}$$

A node with a high degree/strength value is considered to be central in the network.

The betweenness centrality measures how a node relates to the other nodes and characterizes its place in the information diffusion in the network. The betweenness centrality of the node j is the number of shortest paths that pass through j divided by the total number of shortest paths existing in the network. It is defined as follows:

$$BC(j) = \sum_{j \neq k \neq n \in V} \frac{\theta(k, n|j)}{\theta(k, n)}$$

where $\theta(k, n|j)$ is the the number of shortest paths from k to n that contains j and $\theta(k, n)$ is the number of shortest paths from k to n. A node that appears in many shortest paths will have a high betweenness centrality value and will be considered as central in the network.

The closeness centrality measures how close a node is to all other nodes in the network. The closeness of a node j is the inverse of the total length of the shortest paths between it and all others. It is as follows:

$$C(j) = \sum_{k \in V} \frac{1}{d(j, k)}$$

where $d(j, k)$ is the distance between nodes j and k. If the sum of the distances is small, then the closeness centrality value is high and a node can be considered as central in the network.

The eccentricity centrality of a node j is the greatest distance between this node and the other nodes. A node with a low eccentricity centrality value can be considered as central in the network. It is defined as follows:

$$E(j) = \max_{k \in V}(d(j, k))$$

Finally, the eigenvectors centrality measures the importance of a node by also considering the importance of its neighbors. The eigenvector centrality of the j-th node is the j-th value of the first eigenvector, i.e. the eigenvector associated with the largest eigenvalue λ_1. The eigenvector centrality $\mathbf{u}$ is defined as follows:

$$\mathbf{Z}\,\mathbf{u} = \lambda_1 \mathbf{u}$$

A node with a high eigenvector value centrality can be considered as central in the network. Note that the eigenvector centrality value will be even more important if the node is connected to other nodes that have high eigenvector centrality value.

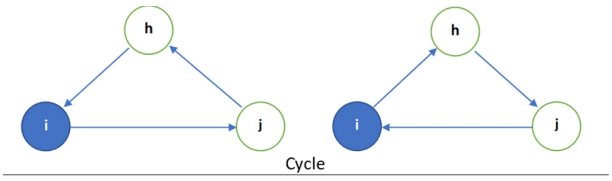
The above indicators only indicate the relative position of a node with respect to its neighbors (node-level measures) without taking into account how information is spreading across the network (spillover and feedback effects). According to this idea, the clustering coefficient was introduced in [161] and extended to directed networks in [126]. This coefficient measures how a node is embedded into the network by quantifying its number of triangles out of all of its possible triangles. In [127], this tool was extended to directed and weighted networks (the local directed weighted clustering coefficient[7]). The local directed weighted clustering coefficient takes into account the strength of a node in the denominator, which ensures that the measure is always between 0 and 1 whatever the network. A high value indicates that the asset j is heavily embedded in the network. For an asset j, it is defined as follows:

$$h_j = \frac{\frac{1}{2}\left[\left(\mathbf{Z}^w + \mathbf{Z}^{w\prime}\right)\left(\mathbf{Z}^u + \mathbf{Z}^{u\prime}\right)^2\right]_{jj}}{s_j\left(d_j - 1\right) - 2s_j^{\leftrightarrow}} \tag{1.21}$$

where $\mathbf{Z}^u$ is the unweighted version of $\mathbf{Z}^w$ ($z^u_{jk} = 1$ if $z^w_{jk} \neq 0$, and 0 otherwise), d_j and s_j are respectively the total degree (1.15) and the total strength (1.18) of the asset j and $s_j^{\leftrightarrow} = \dfrac{\left(\mathbf{Z}^w\,\mathbf{Z}^u + \mathbf{Z}^u\,\mathbf{Z}^w\right)_{jj}}{2}$ is the strength of bilateral edges between j and k.

Moreover, the local directed clustering coefficient can be divided into four types of triangle patterns (cycle, middleman, out, and in) [126] giving rise to a completely different interpretation. The four type of triangle patterns are defined as follows:

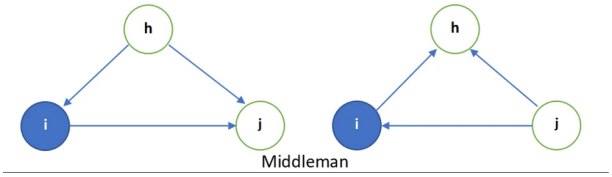
- cycle: a triangle in which every edge has the same direction



$$h_j^{cycle} = \frac{\frac{1}{2}\left[\mathbf{Z}^w\,\mathbf{Z}^{u\,2} + \mathbf{Z}^{w\prime}\left(\mathbf{Z}^{u\prime}\right)^2\right]_{jj}}{\frac{1}{2}\left(s_j^{in}\,d_j^{out} + s_j^{out}\,d_j^{in}\right) - s_j^{\leftrightarrow}} \tag{1.22}$$

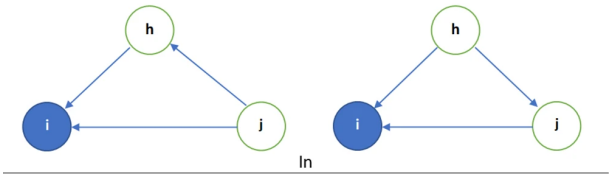
[7]For directed unweighted networks, the local directed weighted clustering coefficient [127] is equivalent to the Fagiolo coefficient [126].

- middleman: a triangle such that j has two edges of different directions and with an edge between k and l without forming a cycle



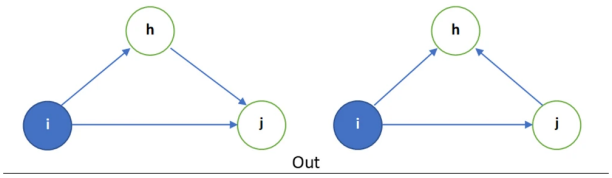
$$h_j^{middle} = \frac{\frac{1}{2}\left[\mathbf{Z}^{w\prime}\,\mathbf{Z}^{u}\mathbf{Z}^{u\prime} + \mathbf{Z}^{w}\,\mathbf{Z}^{u\prime}\mathbf{Z}^{u}\right]_{jj}}{\frac{1}{2}\left(s_j^{in}\,d_j^{out} + s_j^{out}\,d_j^{in}\right) - s_j^{\leftrightarrow}} \tag{1.23}$$

- in: a triangle in which there are two edges pointing toward j



$$h_j^{in} = \frac{\frac{1}{2}\left[\mathbf{Z}^{w\prime}\left(\mathbf{Z}^{u} + \mathbf{Z}^{u\prime}\right)\mathbf{Z}^{u}\right]_{jj}}{s_j^{in}\left(d_j^{in} - 1\right)} \tag{1.24}$$

- out: a triangle in which there are two edges starting from j



$$h_j^{out} = \frac{\frac{1}{2}\left[\mathbf{Z}^{w}\left(\mathbf{Z}^{u} + \mathbf{Z}^{u\prime}\right)\mathbf{Z}^{u\prime}\right]_{jj}}{s_j^{out}\left(d_j^{out} - 1\right)} \tag{1.25}$$

where d_j^{in}, d_j^{out}, s_j^{in} and s_j^{out} are respectively the in-degree (1.16), the out-degree (1.17), the in-strength (1.19) and the out-strength (1.20).

The cycle and the middleman coefficients represent the diffusion of information between assets (indirect diffusion), while the out coefficient can be directly assimilated to systemic assets (influential assets) and the in coefficient to influenced assets. Note that the local directed clustering coefficient h_j is a weighted average of the four coefficients. The weights are the denominator of each triangle pattern divided by the denominator of h_j.

1.2.3 Main limitations

Risk-based allocation strategies require the covariance matrix estimation, where the number of observation T is not too large relative to the number of pre-selected assets m and the distribution of asset returns is not Gaussian. In such a case, the SCM estimator is no longer optimal and leads to noisy estimates. Several approaches have been addressed to reduce the noise in the SCM dealing with sample size issues (m/T) such as the LW estimator [58], covariance matrix estimation based on factor models [22, 90], the Eigenvalue clipping estimator [55, 56] and the optimal RIE [98, 63, 99]. All of these methods have already shown good out-of-sample portfolio performances and offer some advantages, such as:

- LW estimator: particularly suitable for asset returns as it uses market information in the shrinkage process,
- Factor models: capturing common risk factors and using them to reduce dimensionality issue ($K < m$),
- Eigenvalue clipping method: identifying the signal part (or factor) thanks to the upper bound of the MP law [92] to de-noise the covariance matrix,
- optimal RIE: using the overlap between the true and sample eigenvectors to correct the estimate

However, since they operate on the SCM, they are not suitable for non-Gaussian distributions. To deal with non-Gaussian distributions, the M-estimators [105, 106, 103, 104, 107] and especially Tyler-M estimator [108] offer an interesting alternative to the classical SCM, thanks to their robustness properties among elliptical distributions. But, such estimators are not free from sample size issues and they generally require $m << T$, which also leads to combining the methods presented above to de-noise the estimate [116, 117, 118, 115, 119, 62]. Moreover, estimating a covariance matrix assumes that the observations are sampled from a single multivariate distribution (Gaussian or not), which seems too restrictive in financial modeling.

Modeling the structure and dynamics of financial markets through financial networks requires recovering the network topology as accurately as possible. In order to obtain exhaustive information on asset relationships and to accurately capture diffusion patterns, the measures must take into account the existence, direction, and strength of the interactions; otherwise, we can rightly suppose that we are missing information about the dependency structure. Nevertheless, the most common measures used to recover the network topology are not fully satisfactory. Correlation-based networks require dimension reduction tools [69, 128], else they are not suitable in practice and the hierarchical structure cannot be specified. They also lead to undirected weighted/unweighted networks and cannot capture spreading information across assets (spillover and feedback effects). Indeed, if an asset fails, the contagion pattern will not be identifiable. To this end, Granger's non-causality tests can help for capturing information spread across assets (directed networks). However, they are based on the VAR estimation which may fail for many well-known reasons (incorrect model order p [141, 146, 147, 148], high-dimensional systems [149, 150, 151, 140] and correlated white noises [132]) and therefore, if careful attention is not paid to the estimation, the temporal relationships will not be captured

with precision. Moreover, although they capture diffusion patterns, they do not quantify the strength of the relationships, which is problematic to establishing the interaction intensity and thus the systematic risk of assets. To this purpose, nonlinear causality measures such as transfer entropy (TE) may be a possible alternative, but it remains a tough task to obtain accurate estimations in high-dimensional systems and small sample sizes [135, 140], requiring dimension reduction methods [156, 155, 73].

Finally, once the financial network is built, usually centrality measures or the clustering coefficient are used to select a fixed number of assets (the most peripherical or least systemic) independently of the connectivity level in the network [124, 71, 72]. The indicator used must therefore adapt to the number of connections in the network and removes only the most embedded assets for each period. Depending on market cycles, assets are more or less connected to each other and according to such cycles it does not make sense to always eliminate the same number of assets [70, 125, 73, 74]. If the network is very disconnected, few or no assets should be removed; conversely, if the network is very connected a much larger number should be removed.

1.3 Objectives

In this thesis, we address portfolio allocation issues through distinct but complementary approaches to improve the overall performance of risk-based allocation strategies.

The first objective is to develop a robust and de-noised estimator of the covariance matrix adapted to more realistic assumptions on financial asset returns. This estimator must obviously be adapted to non-Gaussian distributions that play a fundamental role in the estimation accuracy, but also must consider that asset returns might be non-homogeneously distributed among themselves or over time.

The second objective is to assess the dynamic dependence between financial assets to recover the network topology and identify the most embedded assets (systemic risk). The dependency measures used must be suitable for high-dimensional systems, but must also quantify the causal strength among the relationships to obtain exhaustive information and thus to accurately capture diffusion patterns.

The third objective is to derive a network indicator adapted to the number of connections in the network and remove only the most unstable assets (systemic and influenced) for each period, to reduce systemic risk within the initial investment universe before allocating portfolios.

Finally, the last objective of this thesis is to combine these different approaches to propose a complete portfolio allocation methodology, where systemic risk has been reduced from the initial investment universe and then the covariance matrix of the remaining assets is estimated with a clean and robust estimate.

1.4 Main contributions

1.4.1 Chapter 2: Robust covariance matrix estimation

In order to reflect the very specific behavior of asset returns, we model them using a multi-factor model with a multivariate correlated non-Gaussian additive noise, assuming that the number of unknown K-factors drives the undiversified part (common risks) of asset returns and that the additive noise is part of elliptical distributions and is correlated according to a Toeplitz structure.

The class of elliptical distributions introduced by Kelker [109] and well-studied in [110, 111, 112] generalizes the multivariate Gaussian distribution. This class includes many well-known distributions such as the multivariate Gaussian distribution, the multivariate t-distribution, the multivariate symmetric α-stable distribution, and the multivariate symmetric generalized hyperbolic distribution. Notice that the generalized hyperbolic distribution also contains the hyperbolic, the normal-inverse Gaussian, and the generalized Laplace distributions. This class of distributions reflects the key stylized facts of asset returns, and the number of multivariate distributions covered makes it possible to assume that the assets have different distributions among themselves or over time. Such distributions are widely used in signal processing applications [162, 97, 163, 164, 165], but also for modeling asset returns [166, 167, 168].

Let $\mathbf{r} \in \mathbb{R}^m$ be a random vector with location parameter $\boldsymbol{\mu} \in \mathbb{R}^m$ and scatter matrix $\mathbf{C} \in \mathbb{R}^{m \times m}$. $\mathbf{r}$ follows an elliptical distribution if and only if there exists a random variable $\tau \in \mathbb{R}^+$, a matrix $\boldsymbol{\Delta} \in \mathbb{R}^{m \times k}$ with $\mathbf{C} = \boldsymbol{\Delta}\boldsymbol{\Delta}'$, a random vector $\mathbf{x} \in \mathbb{R}^k$ independent of τ and uniformly distributed in the k dimensional sphere, $\mathbb{S}^{k-1}$, such that

$$\mathbf{r} = \boldsymbol{\mu} + \tau \boldsymbol{\Delta} \mathbf{x} \tag{1.26}$$

The elliptical distribution is denoted $\mathbf{r} \sim \mathrm{EC}_m(\boldsymbol{\mu}, \mathbf{C}, \tau)$. From this definition, all distributions are simply obtained by setting the distribution of τ. For example, the multivariate Gaussian distribution (1.27) or the multivariate t-distribution (1.28) can be defined as follows:

$$\mathbf{r} = \boldsymbol{\mu} + \sqrt{\chi^2_k} \boldsymbol{\Delta} \mathbf{x} \tag{1.27}$$

$$\mathbf{r} = \boldsymbol{\mu} + \sqrt{\nu \chi^2_k / \chi^2_\nu} \boldsymbol{\Delta} \mathbf{x} \tag{1.28}$$

where χ^2_k is the Chi-squared distribution and ν the degrees of freedom. Note that for $\nu \to \infty$ the multivariate Gaussian distribution is obtained from (1.28).

Moreover, here $\mathbf{C}$ is assumed to be a nonnegative definite Toeplitz matrix and this assumption made on $\mathbf{C}$ is required to use theoretical results found in [163, 164, 165]. A Toeplitz matrix is a matrix with similar elements on the same diagonal. The symmetric Toeplitz matrix also provides a flexible framework to generate a positive-definite covariance matrix. Their properties have been widely studied and is generally used to describe stationary processes [169].

Thus, the proposed model for the asset returns at time t is given by:

$$\mathbf{r}_t = \mathbf{B}_t \, \mathbf{f}_t + \sqrt{\tau_t} \, \mathbf{C}^{1/2} \, \mathbf{x}_t \tag{1.29}$$

where

- $\mathbf{r}_t$ is the m-vector of returns at time t,
- $\mathbf{B}_t$ is the $m \times K$-matrix of coefficients that define the assets sensitivities to each factor at time t,
- $\mathbf{f}_t$ is the K-vector of random factor values at t, supposed to be common to all of the assets,
- $\mathbf{x}_t$ is a m-vector of independent Gaussian whites noise with unit variance and non-correlated with the factors, i.e. $\mathbb{E}[\mathbf{x}_t \, \mathbf{f}_t'] = \mathbf{0}_{m \times K}$,
- $\mathbf{C}$ is called the $m \times m$ scatter matrix that is supposed to be Toeplitz structured and time invariant over the period of observation,
- τ_t is a family of i.i.d positive random variables with expectation τ that is independent of the noise and the factors and drives the variance of the noise. These random variables are time-dependent and generate the elliptical distribution of the noise.

Given equation (1.29), the covariance matrix writes at time t:

$$\mathbf{\Sigma}_t = \mathbf{B}_t \, \mathbf{\Sigma}_t^f \, \mathbf{B}_t' + \tau \, \mathbf{C}$$

where $\mathbf{\Sigma}_t^f = \mathbb{E}[\mathbf{f}_t \, \mathbf{f}_t']$ is the factor-related term and $\tau \, \mathbf{C}$ is the covariance matrix of the additive noise.

In this framework, we apply the robust Tyler-M estimator (1.10) and the RMT results (1.6), in order to filter the noise part of the observations and thus to estimate only the covariance matrix from the subspace generated by the K eigenvectors linked to the K eigenvalues (K factors).

The Tyler-M estimator applied to model (1.29) turns out to be the "most robust" covariance matrix estimator [108, 114] for the true scatter matrix $\mathbf{C}$, and also independent of the τ_t distribution. Moreover, the *Consistency Theorem* found in [163, 164, 165] shows that if $\mathbf{C}$ admits a Toeplitz structure, then applying a Toeplitz rectification operator $\mathcal{T}(.)$ on $\widehat{\mathbf{C}}_{tyl}$ provides an estimator of $\mathbf{C}$ that almost surely converges in spectral norm under the RMT regime, i.e. when $T, m \to \infty$, such as $m/T \to c \in]0, \infty[$. This theorem ensures the convergence of the covariance matrix of observation towards the covariance matrix of additive noise, independently of the variance of the noise.

A Toeplitz operator $\mathcal{L}(.)$ is a linear operator that transforms any matrix $\mathbf{A}$ into a matrix with a Toeplitz structure. There exist many Toeplitz operators since the operator transforms the matrix in such a way to have a matrix with constant diagonals. Here, as in [163, 164, 165], we use the Toeplitz rectification operator which replaces any element $a_{j,k}$ from $\mathbf{A}$ of size (m, m) with the sum of the elements of its diagonal i divided by m. For any vector $\mathbf{x}$ of size m, $\mathcal{L} : \mathbf{x} \mapsto \mathcal{L}(\mathbf{x})$ is defined as the associated symmetric square matrix of size m obtained through the Toeplitz operator: $([\mathcal{L}(\mathbf{x})]_{j,k}) = x_{|j-k|+1}$. For any

square matrix $\mathbf{A} = [a_{j,k}]$ of size m, $\mathcal{T}(\mathbf{A})$ represents the matrix $\mathcal{L}(\check{\mathbf{a}})$ where $\check{\mathbf{a}}$ fulfills $\check{a}_j = (\sum_{k=j}^{m} a_{k,k-j+1})/m$.

Once the covariance matrix $\widetilde{\mathbf{C}}_{tyl} = \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}\right)$ is estimated, the observations can be whitened and we can thus apply the RMT results under the same conditions stipulated by the theory, i.e. multivariate non-correlated Gaussian distribution with unit variance. Then, it is possible to identify the largest K eigenvalues of the estimated covariance matrix thanks to the upper bound of the Marčenko-Pastur law [92], and thus obtain a clean and robust estimator of the covariance matrix of observations (whitening process).

However, the whitening process proposed above is made under the implicit assumption that the asset returns are drawn from a unique multivariate law and are therefore homogeneous in law. Indeed, the Toeplitz structure characterizes a stationary process that assumes that all asset returns come from the same process and are "spatially" stationary[8]. The temporal stationarity of returns is generally observable, but not the spatial stationarity between assets. In the case where the assets are heterogeneous in distribution, the Toeplitz structure assumption on the covariance matrix $\mathbf{C}$ is difficult to verify. Under the Toeplitz structure assumption, if the observation order of the assets changes, the covariances between two assets should not change, whereas in practice this assumption is not verified. Considering this phenomenon, we now assume that the asset returns might be non-homogeneously distributed extending the results presented in [168] to no longer be dependent on the observation order. We therefore propose to split the m assets into $p < m$ groups, each composed of $\{m_q\}_{q=1}^{p}$ assets (with $\sum_{q=1}^{p} m_q = m$), and formed to be composed of assets having similar distributions. Under this new assumption, the model (1.29) applies for each group q as follows:

$$\mathbf{r}_t^{(q)} = \mathbf{B}_t^{(q)}\,\mathbf{f}_t + \sqrt{\tau_t^{(q)}}\,\mathbf{C}_{(q)}^{1/2}\,\mathbf{x}_t,$$

where the complete scatter matrix $\mathbf{C}$ is therefore block-constructed, and block-Toeplitz.

Under the assumption of non-homogeneous assets returns, we propose to form groups of assets before applying the whitening process. The groups are built using the Ascending Hierarchical Clustering (AHC) method that requires the number of groups to be fixed a priori or determined using a predefined criterion (Caliński-Harabasz (CH) criterion [170]), and the Affinity Propagation (AP) method [171] that self-determines the number of groups.

Empirical tests are carried out on two different asset universes (European and US equity universes) both allocated with the Maximum Variety[9] or the Minimum Variance strategy. The results are compared to those obtained with several classical estimators such as Ledoit & Wolf (1.3), Eigenvalue clipping method (1.6), and RIE (1.7). These tests extend our preliminary results and show that the way the assets are grouped might improve the portfolio performance even more.

[8] By "spatially" stationary, we refer to the stationarity of the dependency structure of assets.
[9] The Maximum Variety portfolio (VarMax) refers to the Most Diversified Portfolio [30].

1.4.2 Chapter 3: Frequency causality measures and parsimonious VAR estimation

VAR representation (1.11) allows either in the time domain or in the frequency domain to define the interactions between time series. Their use in many fields comes from their straightforward theoretical framework for understanding the dynamical structure of systems by capturing complex temporal relationships among time series. However, in the time domain, Granger's tests provide partial information about interactions since they only assess the existence and direction of causal relationships without any information about the causal strength. In neuroscience, several measures of connectivity in the frequency domain have been developed to deal with this point. Two types of connectivity measures should be distinguished: first, coupling measures such as coherence and partial coherence measures [172], respectively related to the cross-correlation or partial cross-correlation; second, frequency causality measures able to quantify the strength of causal relationships, extending the concept of Granger causality such as the Directed Coherence (DC) [173], the Direct Transfer Function measure (DTF) [174, 175], the Partial Directed Coherence measure (PDC) [176, 177], and the Generalized Partial Directed Coherence measure (GPDC) [178]. However, such measures do not provide the same information [172, 179, 180]. The PDC and GPDC are measures of direct causality while DC and DTF measure both direct and indirect causality. The difference between direct and indirect causality are illustrated in Fig. 1.4.

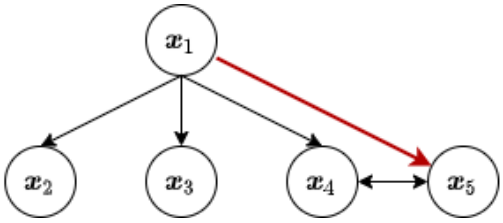


Figure 1.4: Causal structure of S in chapter 3: the time series $\boldsymbol{x}_1$ causes $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4$, while $\boldsymbol{x}_4$ and $\boldsymbol{x}_5$ are causing each other. The black lines represent the direct causalities and the red ones represent the indirect causalities.

1.4.2.1 Frequency causality measures

The above causality measures first require the estimation of a VAR(p) model (1.11) and then the estimated coefficients are shifted into the frequency domain via the Fourier transform. The discrete Fourier transform of the coefficients $a_{jk}(1), \ldots, a_{jk}(p)$ at a given frequency f is defined by:

$$\tilde{a}_{jk}(f) = \begin{cases} 1 - \sum_{l=1}^{p} a_{jk}(l) e^{-2i\pi fl}, & \text{if } j = k \\ \\ - \sum_{l=1}^{p} a_{jk}(l) e^{-2i\pi fl}, & \text{otherwise} \end{cases} \tag{1.30}$$

where f are the discrete frequencies lying in $\left[-\frac{1}{2}; \frac{1}{2}\right]$. Note that for a discrete time series sampled at frequency f_e, its Fourier Transform will reveal information for frequencies

lying in $\left[-\frac{f_e}{2}; \frac{f_e}{2}\right]$. In our case $f_e = 1$, we can therefore choose the interval $\left[-\frac{1}{2}; \frac{1}{2}\right]$ with a step of $\frac{1}{F-1}$, where F is the number of frequencies.

From the $\tilde{\boldsymbol{A}}(f)$ matrix (1.30), the transfer matrix of the system at frequency f is given by:

$$\boldsymbol{H}(f) = \tilde{\boldsymbol{A}}^{-1}(f) \tag{1.31}$$

Furthermore, from the transfer matrix (1.31), the coherence and partial coherence measures can be defined using the cross-spectral power density matrix [181] given by:

$$\boldsymbol{S}(f) = \boldsymbol{H}(f)\boldsymbol{\Sigma}_{\epsilon}\boldsymbol{H}(f)^{H}$$

where the superscript ^{H} is the Hermitian transpose and $\boldsymbol{\Sigma_\epsilon}$ is the covariance matrix of white noises (1.11).

Directed Coherence and Direct transfer function

The DC, introduced by Saito and Harashima [173] for bivariate cases, has been extended for multivariate cases by Baccalá et al. [182]. The DC is based on the transfer matrix $\boldsymbol{H}$ defined in (1.31). For two time series $x_j(t)$ and $x_k(t)$ the DC from k to j at each frequency f is defined as follows:

$$\psi_{jk}(f) = \frac{\sigma_{kk}h_{jk}(f)}{\sqrt{\sum_{n=1}^{m}\sigma_{nn}^2 h_{jn}(f)h_{jn}^*(f)}}$$

where σ_{kk}^2 is the k-th element of the diagonal of $\boldsymbol{\Sigma_\epsilon}$ and $h_{jk}(f)$ is the element j,k of $\boldsymbol{H}(f)$. Moreover, the DC/DTF is represented as a power spectral density, i.e. $|\psi_{jk}(f)|^2$ and has the following normalization properties:

$$0 \leqslant |\psi_{jk}(f)|^2 \leqslant 1 \tag{1.32}$$

$$\sum_{n=1}^{m} |\psi_{jn}(f)|^2 = 1, \forall j = 1, \ldots, m. \tag{1.33}$$

The DTF [174, 175] is a version of the DC where $\sigma_{kk}^2 = 1\ \forall k \in [1, m]$. The DC $\psi_{jk}(f)$ and the DTF at frequency f give causal influence between k and j, but even if the two time series are not directly causal, an indirect causality may also emerge from a third time series. In Fig. 1.5, we provide an example of the DC measure on the causal structure defined in in Fig. 1.4.

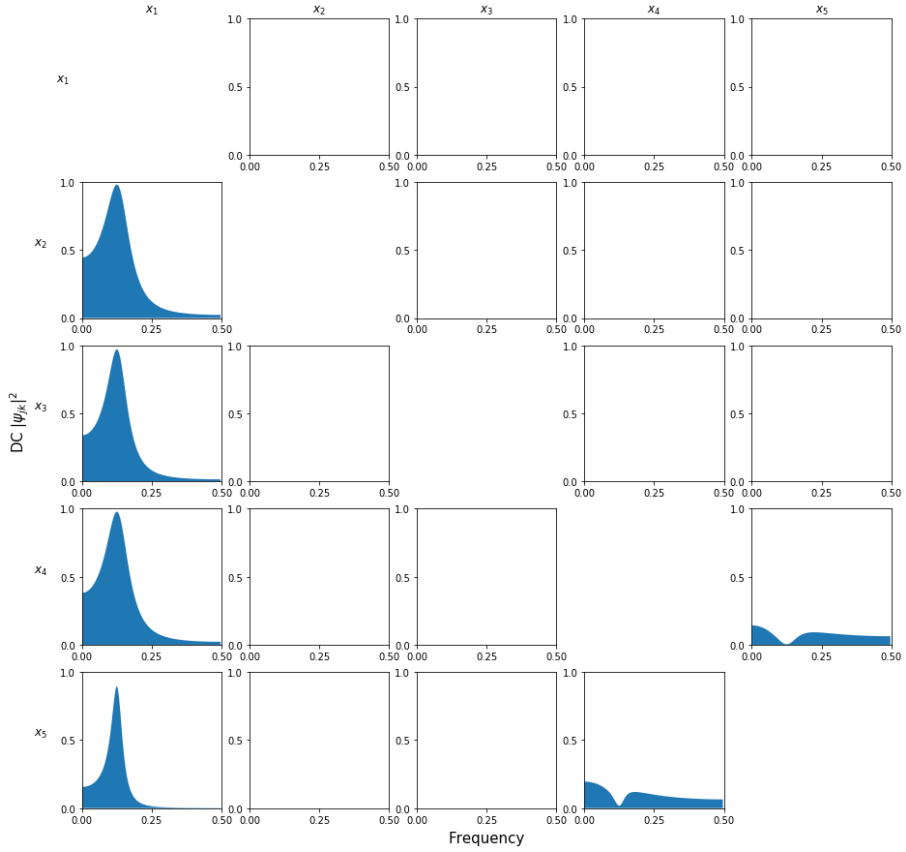


Figure 1.5: Theoretical DC of causal structure represented in Fig. 1.4 (column k causes row j). The DC is computed using the true coefficients and the identity matrix for the residual correlation matrix ($\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \boldsymbol{I}_5$). Interpreting the DC: $\boldsymbol{x}_1$ causes $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4$. In contrast, $\boldsymbol{x}_1$ causes $\boldsymbol{x}_5$ indirectly via $\boldsymbol{x}_4$.

Generalized Partial Directed Coherence measure (GPDC)

The GPDC [178] is an extended version of the PDC introduced by Baccalá and Sameshima [176, 177]. It is a generalization of the multivariate case of the DC, based on the partial coherence that describes the mutual interaction between two time series when the effects of all others have been subtracted. Unlike DC/DTF measure, it quantifies only direct connections. For two time series $x_j(t)$ and $x_k(t)$ the GPDC is defined from k to j at each frequency f as follows:

$$\omega_{jk}(f) = \frac{\frac{1}{\sigma_{jj}}\tilde{a}_{jk}(f)}{\sqrt{\sum_{n=1}^{m}\frac{1}{\sigma_{nn}^2}\tilde{a}_{nk}(f)\tilde{a}_{nk}^*(f)}}$$

As for the DC/DTF, the PDC is a version of the GPDC where $\sigma_{kk}^2 = 1\ \forall k \in [1, m]$. The two normalization properties (1.32) and (1.33) defined for the DC/DTF hold for the GPDC/PDC. However, the GPDC/PDC is normalized by the total outflows, then (1.33) becomes:

$$\sum_{n=1}^{m}|\omega_{nk}(f)|^2 = 1, \forall k = 1, \ldots, m. \tag{1.34}$$

The GPDC is a measure of direct causality representing the relative strength of an interaction with respect to a given signal source (outflow to all the outflows) and is therefore more suitable than DC/DTF for capturing spreading information between time series. Indeed, the GPDC provides the multivariate relationships from a partial perspective. Moreover, the GPDC resolves the shortcomings of the PDC [183]: i) the GPDC is not affected when multiple signals are emitted from a given source; ii) GPDC is a scale invariant measure; iii) GPDC allows to draw conclusions about the absolute causal strength. In Fig. 1.6, we provide an example of the GPDC measure on the causal structure defined in in Fig. 1.4.

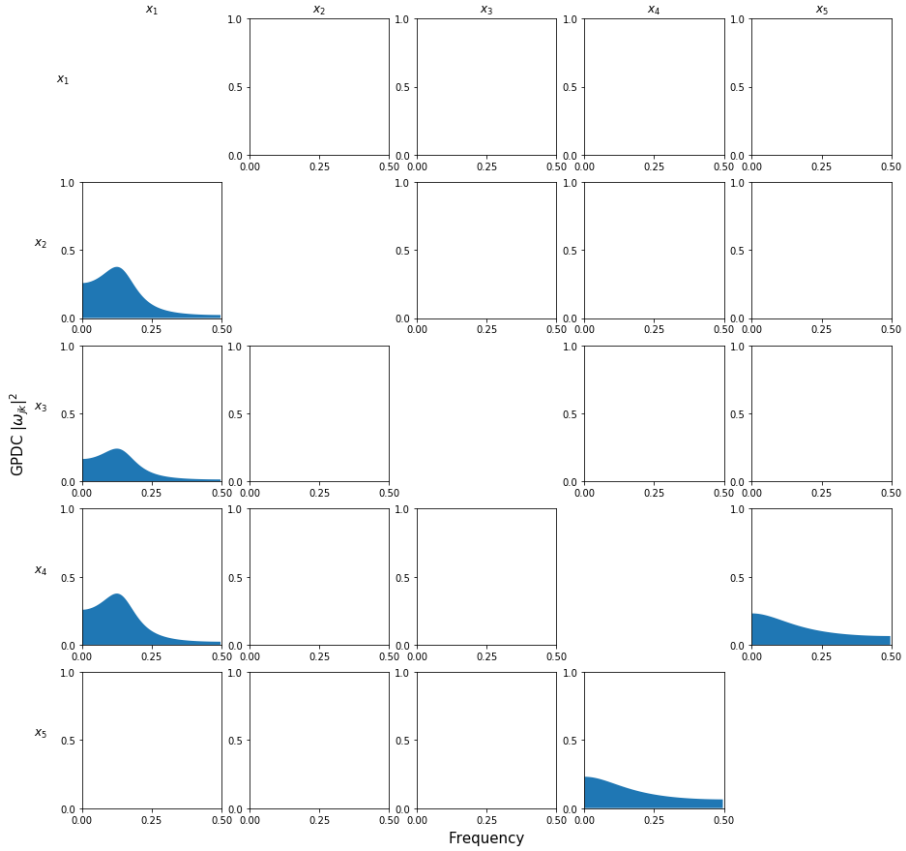


Figure 1.6: Theoretical GPDC of causal structure represented in Fig. 1.4 (column k causes row j). Interpreting the GPDC: $\boldsymbol{x}_1$ causes $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4$. In contrast, $\boldsymbol{x}_1$ causes $\boldsymbol{x}_5$ indirectly via $\boldsymbol{x}_4$, but as GPDC quantifies only direct interactions, the causality values for all frequencies are equal to zero. If $j = k$, the GPDC represents the part that is not explained by other signals. Since it is quite difficult to interpret, the diagonal is not reported here.

Although GPDC is a powerful measure for detecting and quantifying causal influences in multivariate systems with respect to Granger non-causality tests, it requires a precise VAR estimation. Indeed, since the VAR coefficients are directly used to compute the GPDC, the problem is intrinsically linked to the VAR estimation and it is obvious that a flawed estimation of the VAR model will lead to both spurious causalities and inaccurate causal strengths (cascading errors). A classical VAR model (unrestricted VAR model) assumes that a time series depends on all lagged variables in the system. This is a very strong assumption, in fact unreasonably so when modeling multivariate systems which admit more parsimonious structures with only a few non-zero coefficients. This

assumption is unrealistic, as in a multivariate system it is unusual for all time series to be mutually dependent at each lag. The classical VAR estimation that focuses only on determining the "optimal" lag order p leads to non-significant coefficient estimates and strong biases on the GPDC measure. It is therefore essential to have a precise estimation of the VAR coefficients, especially for those that are non-significant. Otherwise spurious causalities will appear, biasing the true ones due to both the compensatory effect in VAR estimation and the GPDC normalization property (1.34). The standard VAR estimation is thus not well-suited to parsimonious models, inducing cascading errors in the GPDC, which can be tenfold for high-dimensional system and/or small sample sizes [184]. To this end, restricted VAR estimations (subset VAR models) can help to estimate only the significant VAR coefficients to reduce the cascading errors in both causal and non-causal parts.

1.4.2.2 Subset VAR models

Determining the best possible VAR estimate ideally involves testing all possible subset VAR models and selecting the optimal one for a given criterion. In practice, this procedure is hardly feasible because even for small m and p, the number of possible susbet VAR models is huge (2^{m^2p} possibilities). To deal with this "curse of dimensionality", alternative procedures have been developed.

In the literature on VAR models, three procedures can be considered to remove non-significant coefficients: first, procedures based on an information criterion to add or delete coefficients (Bottom-Up strategy, Top-Down strategy [141] and modified Backward-in-Time Selection [151]); second, procedures using hypothesis testing, such as the t-test, likelihood ratio and Wald test [141, 149, 150]; and finally, procedures based on shrinkage methods such as Ridge Regression [185], Lasso Regression [186], and Elastic-Net [187].

Information-based subset selection methods

This type of subset selection methods will either remove (Top-Down strategy) or add (Bottom-Up strategy) coefficients by comparing the restricted models to each other according to an information criterion such as AIC, HQ or BIC.

The Top-Down strategy (TD) [141] starts from the full VAR(p) model defined in (1.11). The significance of the VAR coefficients is tested separately in the m equations. The goal is to eliminate the coefficients for each equation by evaluating the information criterion. The order of the tested terms is arbitrary, but for example in [141], the largest lag p is tested first for all variables from $x_m(t-p)$ to $x_1(t-p)$, then the lag $p-1$ with the same order of variables, and the process is iterated until $p=1$. An alternative approach to remove the variables is to first delete the one that leads to the largest improvement in the information criterion. The advantage to the TD strategy is that all coefficients in each equation are tested, but it is very sensitive to the full VAR estimation which determines the model order. Indeed, if the model order is under-fitted and coefficients were significant, then this strategy will not be able to compensate for the estimation errors made. This strategy is precisely defined in chapter 3.

Unlike the TD strategy, the Bottom-Up strategy (BU) [141] starts from the null VAR model (empty coefficients matrix) and adds the variables progressively. For the j-th

equation, only the first variable $(x_1(t-1), \ldots, x_1(t-p))$ is considered, and the optimal lag p_1^* for that variable is selected by testing $p \in \{1, \ldots, p_{max}\}$ where p_{max} is fixed *a priori*. As the TD strategy, the optimal lag p_1^* is determined with an information criterion and all lagged terms up to p_1^* are included. Then, the same procedure is applied by adding one by one the remaining $m-1$ variables and thus determining for each variable its optimal lag. Note that for one or more variables it is possible that the information criterion may not be improved in which case the variables do not enter into the equation. The main advantage of the BU strategy is that by starting from the null model and adding one by one the variables, it makes it possible to work with high-dimensional systems. However, this strategy has two drawbacks. First, the p_{max} is fixed *a priori* and can create estimation biases. Indeed, if it is fixed too small, the model will be under-fitted; and on the other hand, if it is too large, then all lagged variables up to this p_{max} can be included. Second, by including all lagged variables up to the optimal lag, this strategy does not take into account possible non-significant lags. To overcome this last drawback, the TD strategy can be applied either at the level of each variable inclusion or at the level of the final equation.

A more recent method has been developed by Vlachos and Kugiumtzis [151] called the modified Backward-in-Time Selection (mBTS). This method is also a BU strategy and it is based on Dynamic Regression models [188]. However, the mBTS method adds progressively lagged variables, starting from the first lag for all variables and moving backward in time. Contrary to the BU strategy which includes the variables one by one by testing the lags, it tests all the variables at each lag which provides a better temporal order of the variables. Moreover, it includes in the equations only significant lags for each variable correcting the main drawback of the classical BU strategy. Nonetheless, it still has the inconvenience of p_{max} fixed *a priori*. This strategy is also precisely defined in chapter 3.

Such strategies are directly related to the choice of information criteria. The AIC criterion tends to select more variables while the BIC criterion tends to choose the most parsimonious model [149, 150].

Testing Procedure (TT)

Alternative procedures to information criteria are based on hypothesis testing. In this approach, two types of tests have to be distinguished: individual tests (t-test) or multivariate tests, i.e. that jointly test the variables (e.g. likelihood ratio and Wald tests) [141, 149, 150]. In VAR models, the coefficients are directly related to the temporal relationships. Thus, if the VAR has several lags, the use of multivariate tests can either lead to removing all lags from a variable, or on the contrary, to keeping all of them, which suffers from the same drawback as the classical BU strategy. In this section, we therefore focus only on the individual t-test.

The significant coefficients can be chosen with the individual t-ratio, i.e. by excluding all of the smallest absolute values of t-ratios until all absolute t-ratios are greater to a threshold η. This procedure called Testing Procedure (TT) [149, 150] is a similar approach to the TD strategy, but the coefficients are deleted using the individual t-ratios. This strategy is much faster than the TD strategy because it immediately identifies which variable is deleted in the next step, whereas in the TD strategy each coefficient has to be retested. However, the threshold η has to be fixed *a priori*. In

general, the value is fixed to 2 which corresponds roughly to the 5% significant level, or alternatively using the quantile of the t-distribution $\mathcal{T}_{\nu}^{\alpha/2}$ for a given probability α and ν degrees of freedom. Moreover, Brüggemann and Lütkepohl [149] show that if the critical value η is chosen as a function of the sample size T, the number of initial variables, the information criterion, and the reduction step, then the TT and the TD strategy (elimination by the greatest improvement in the information criterion) are equivalent.

For the j-th equation obtained from the full VAR model (1.11) estimated in a multivariate environment (LS), the TT procedure is applied as follows:

1. Compute the t-ratios associated to $\hat{\boldsymbol{b}}_j$ the $(q \times 1)$ vector of the estimated coefficients for the j-th equation, where q is the number of lagged variables (for the first step $q = mp$). For the coefficient $(\hat{b}_j)_n$, the t-ratio is defined as follows:

$$\varphi_{n,j} = \frac{(\hat{b}_j)_n}{\left(\dfrac{\hat{\boldsymbol{\epsilon}}_j' \hat{\boldsymbol{\epsilon}}_j}{T - q - 1} (\boldsymbol{X}\boldsymbol{X}')_{nn}^{-1}\right)^{1/2}} \tag{1.35}$$

   where $n \in \{1, \ldots, q\}$, $\boldsymbol{\epsilon}_j$ is the $(T \times 1)$ vector of residuals $(\boldsymbol{y}_j - \hat{\boldsymbol{b}}_j' \boldsymbol{X})$ with $\boldsymbol{y}_j = (x_j(1), \ldots, x_j(T))'$ the $(T \times 1)$ vector of observations defined in (1.13), $\boldsymbol{X}$ the $(q \times T)$ matrix defined in (1.12) and $(\boldsymbol{X}\boldsymbol{X}')_{nn}^{-1}$ the n-th element of the diagonal of $(\boldsymbol{X}\boldsymbol{X}')^{-1}$.

2. Delete the coefficient n with the lowest absolute t-ratio, if and only if $|\varphi_{n,j}| < \mathcal{T}_{T-q}^{\alpha/2}$.

3. Re-estimate the coefficients for the j-th equation by removing the n-th row from the $\boldsymbol{X}$ matrix and recompute the t-ratios again from (1.35) with the new residuals obtained and by decreasing the number of lagged variables q by one.

4. Repeat steps 2 and 3 until $|\varphi_{n,j}| \geq \mathcal{T}_{T-q}^{\alpha/2}$ $\forall n \in \{1, \ldots, q\}$.

The significance level chosen plays an important role, because if it is too severe, true causalities can be removed and conversely if it is too low, spurious causalities will remain, biasing the GPDC in both cases.

Shrinkage methods

The Ridge Regression [185] or the Lasso method (Least Absolute Shrinkage and Selection Operator) [186, 189] are Least Squares (LS) based methods with respectively a L_2-norm constraint or L_1-norm constraint on the VAR coefficients. The Ridge Regression estimator (1.36) and the Lasso estimator (1.37) for the VAR(p) model in (1.11) are defined as follows:

$$\widehat{\boldsymbol{A}} = \underset{\mathbf{A}}{\operatorname{argmin}} \, ||\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}||_2^2 + \xi ||\boldsymbol{A}||_2^2 \tag{1.36}$$

$$\widehat{\boldsymbol{A}} = \underset{\mathbf{A}}{\operatorname{argmin}} \, ||\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}||_2^2 + \xi ||\boldsymbol{A}||_1 \tag{1.37}$$

where

- $\boldsymbol{Y} = (\mathbf{x}(1), \ldots, \mathbf{x}(T))$ is a $(m \times T)$ matrix of observations,
- $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$ is the $(m \times mp)$ matrix of VAR coefficients,
- $\boldsymbol{X}$ is the $(mp \times T)$ matrix defined in (1.12),
- $\xi \in \mathbb{R}^+$ is the tuning parameter.

If $\xi = 0$, these estimators coincide with the LS estimate. If $\xi > 0$, the least significant coefficients in $\widehat{\boldsymbol{A}}$ are shrunk to zero. However, for the Ridge Regression, the coefficients are never completely shrunk to zero. Indeed, the L_2 penalty term only limits the value of coefficients whereas the L_1 penalty term imposes sparsity among the coefficients. The Lasso method has the advantage of estimating coefficients and selecting variables simultaneously. In typical cases, a cross-validation procedure may be used, as Tibshirani [186] suggested, to choose both the lag p and the tuning parameter ξ. In this thesis we only select ξ by considering that the lag p is predetermined by the VAR(p) model order estimation as in [189]. Finally, an extension of these two methods is the Elastic-Net method [187] that combines both of them. By adding the L_2 penalty term, this method allows one to correct some limitations of the Lasso method such as the limitation on the number of selected variables when $m \times p > T$, but also tends to select several variables from a group. Nonetheless, it also adds complexity in the estimation compared to the Lasso or Ridge Regression, with more parameters to calibrate.

The challenge here is obviously to use the most efficient subset method in terms of VAR coefficient estimation accuracy, but also to work with high-dimensional systems for assessing the structure and dynamics of financial markets. According to this idea, we propose to combine the mBTS method with the TD strategy (mBTS-TD). We first use the modified Backward-in-Time Selection (mBTS) to estimate the VAR coefficients as it includes one by one only the terms that improve the prediction of the equation starting from the null VAR model and thus makes it possible to work with high-dimensional systems like $K = 20$ in [151]. Moreover, as already shown in [184], the mBTS method dramatically improves GPDC accuracy. However, a maximum lag p_{max} must be fixed *a priori*; if it is too small this leads to an under-fitted model where the internal dynamics of the system are not completely captured. On the other hand, if p_{max} is too large, undesirable lagged variables may appear in the model, revealing spurious causalities. Thus, to be less dependent on the choice of p_{max}, we use also the TD strategy. Despite

the fact that the TD strategy is very sensitive to the initial estimation of the VAR, in our case, it operates already on a parsimonious model and allows us to test the significance of the variables in the opposite direction to produce if necessary a more parsimonious model when p_{max} is set at a high value.

Using Monte Carlo simulations, our extended subset selection method (mBTS-TD) is compared with classical subset selection methods (TD, Lasso, TT and mBTS) to quantify the GPDC accuracy. All subset selection methods improve the accuracy compared to standard VAR estimation and are therefore better suited. Nonetheless, the mBTS-TD method stands out clearly from the other four methods by drastically reducing the cascading errors in both causal and non-causal terms. What is more, we also show that mBTS-TD outperforms mBTS and TT whatever value of p_{max} is chosen. Furthermore, a recent work has shown [140] that linear measures (Granger non-causality and GPDC) can provide competitive performance with respect to nonlinear measures (TE [152] and PMIME [156]) also on nonlinear systems, especially when subset VAR methods are applied first.

Finally, we make use of the GPDC measure, estimated with the mBTS-TD method for modeling financial market dependency structures. This approach provides not only a precise network topology taking into account both the direction and the strength of the relationship between assets via the GPDC, but also solves the dimensionality puzzle via the mBTS-TD estimation that produces a parsimonious causal structure. In addition, via the local directed weighted clustering coefficient the most systemic assets are excluded improving the EW portfolio performances. To the best of our knowledge, we are the first to apply GPDC to financial networks, even though the PDC or the GPDC have already been applied in the field of neuroscience [190, 191, 192].

1.4.3 Chapter 4: GPDC financial networks and asset selection

In this last chapter, in order to obtain a complete methodology for asset selection problems, we propose a dynamic indicator to identify in the network the assets presenting major risks due to their influence (systemic) or, on the contrary, those too influenced to recover quickly from a market shock. Ignoring such assets in the selection process inevitably leads to underestimating the portfolio's risks, especially for world or multi-asset strategies. Thus, identifying and removing unstable assets from the universe before allocating portfolios improves risk diversification in the sense that the remaining assets are less interconnected. Such pre-selection can also reduce asset allocation errors from being complementary to the use of the covariance matrix which does not quantify diffusion patterns. As an example, a diversified strategy will seek the least correlated assets, but correlations do not reflect interconnectedness and an asset with low correlation may turn out to be potentially highly systemic since it carries a new risk premium (e.g. US subprime crisis, Italy/Spain sovereign debt crisis, etc.) or is strongly influenced and overreacting to every market shock. Even though centrality measures or clustering coefficients are useful tools in network theory to identify such assets [124, 70, 71, 125, 72, 73, 74], they are not able to eliminate a variable number of assets depending on the structure of the network. In portfolio allocation selection, these measures are used to select/remove a fixed number of assets (the most peripheral or the least systemic) and then allocate the portfolio regardless of the connectivity level

within the network [124, 71, 72]. But, depending on economic/political events or market cycles, assets are more or less connected to each other and according to such cycles it does not make sense to always eliminate the same number of assets [70, 125, 73, 74]. If the network is very disconnected, few or no assets should be removed; conversely, if the network is very connected a much larger number should be removed.

From this, we propose to derive from the local directed weighted clustering coefficient [127] a dynamic indicator essentially based on the out (1.25) and in (1.24) information spreading patterns, neglecting the middleman and cycle ones (indirect diffusion). The out triangle pattern identifies causal effects ($h_j^{out} \neq 0$) and thus can be directly assimilated to systemic assets, while the in identifies the assets that are caused ($h_j^{in} \neq 0$) and therefore influenced or that overreact if a market shock occurs. This asset selection procedure has several advantages. First, it focuses essentially on unstable assets (systemic or influenced) in order to reduce systemic risk within the universe. Second, by construction it adapts itself to the number of connections in the network and therefore removes only the most embedded assets for each period. Indeed, if the network is very disconnected, no asset will be removed. Hence, it is not necessary to set either a number of excluded assets or an exclusion threshold on the local directed weighted clustering coefficient.

The proposed dynamic asset selection process is applied on the GPDC financial network (chapter 3) and examined on two equity universes. The first universe is composed of national indices belonging to the MSCI ACWI (All Country World Index) and the second is composed of GICS [193] sector indices within four geographical areas (MSCI Emerging Markets, MSCI Europe, MSCI Japan, MSCI United States). These two universes allow us to focus on different characteristics such as time delay between areas (feedback effects) and global/regional macroeconomic effects for national universe and economic activity issues within or between geographical areas for the sector universe.

The methodology is applied on the EW, ERC, MinVar and VarMax[10] portfolios, using the SCM estimate first for risk-based allocation strategies. Regarding the EW, ERC and VarMax portfolios the dynamic pre-selection significantly improves portfolio performances with respect to a Granger-based network or applied on the whole universe. As expected, for both universes the most significant improvement is for the VarMax strategy. The asset selection process succeeds in identifying the least correlated assets that are either the least performing (influenced) or the riskiest (systemic) without significantly reducing portfolio diversification. As for the MinVar portfolio, the results are more contrasted. The methodology fails to improve return/volatility ratio on the country universe, increasing the idiosyncratic risk even more and thus making it more complicated to manage drastic market changes as in 2008, 2015 and 2018. Finally, when we associate the dynamic asset selection with the whitening procedure (chapter 2), the results are even more improved compared to the use of the SCM (5 times out of 6), the only failure is still on the MinVar portfolio applied to the sector universe. This empirical study highlights that by combining all of the results of the different chapters, we can significantly improve several classical allocation strategies.

[10]The VarMax portfolio refers to the Most Diversified Portfolio [30].

1.5 Outline of the thesis

The structure of the thesis is therefore as follows:

- In chapter 2, we propose a cleaned and robust covariance matrix estimation (whitening procedure). The asset returns are modeled as a multi-factor model embedded in correlated elliptical and symmetric noise, extending the classical Gaussian assumptions. We also consider that the asset returns might be non-homogeneously distributed, extending the results presented in [168]. The whitening procedure combines the robust Tyler-M estimator with the RMT results that are adapted to correlated and non-Gaussian assumptions as suggested in [96, 163, 164, 165]. Nonetheless, if the assets come from different distributions, other approaches should be considered. Thus, before applying the whitening procedure, we first classify the assets into homogeneous groups using two classification methods: the Ascending Hierarchical Clustering and the Affinity Propagation method [171]. Finally, the final covariance estimate is obtained only using the de-noised part of the observations.

- In chapter 3, we focus on frequency causality measures. In particular, we use the Generalized Partial Directed Coherence measure (GPDC) [178] to assess both the direction and the strength of causal relationships among financial assets. Nevertheless, this measure is based on VAR models and it is obvious that a flawed estimation of the VAR model will translate into inaccurate measure (cascading errors). We quantify these errors and then propose a parsimonious estimation (mBTS-TD) of the VAR model (no estimation of non-significant coefficients) by combining two subset selection methods (modified Backward-in-Time Selection method (mBTS) [188, 151] and Top-Down strategy (TD) [141]. Finally, we make use of the GPDC measure, estimated with the proposed mBTS-TD method, to recover financial network topology.

- Finally, in chapter 4, we carry out an empirical study based on four portfolio allocation strategies (EW, ERC, MinVar, VarMax), applying the results found in chapters 2 and 3. We first recover the financial network topology with the GPDC measure estimated with the mBTS-TD method. Then, we propose a dynamic pre-selection method based on the in and out triangle patterns of the local directed weighted clustering coefficient [127]. This pre-selection method allows us to remove essentially the most unstable assets (systemic and influenced) to reduce systemic risk and thus obtain a well-diversified universe. What is more, this procedure adapts to the number of connections in the network and therefore removes only the most embedded assets. Hence, it is not necessary to either set a number of excluded assets or to set an exclusion threshold on the local directed weighted clustering coefficient as in chapter 3. Finally, the covariance matrix of the remaining assets is estimated with whitening procedure.

Chapter 2

Improving portfolios global performance using a cleaned and robust covariance matrix estimate

This chapter is based on one poster session and two published articles in collaboration with Emmanuelle Jay, Eugénie Terreaux, Jean-Philippe Ovarlez, Frédéric Pascal, Philippe De Peretti and Christophe Chorro.

Improving portfolios global performance using a cleaned and robust covariance matrix estimate,
E. Jay, T. Soler, E. Terreaux, J. P. Ovarlez, F. Pascal, P. De Peretti, C. Chorro,
Conference on Dynamics of Socio Economic Systems (DySES), October 2018, Paris, France

Improving portfolios global performance using a cleaned and robust covariance matrix estimate,
E. Jay, T. Soler, E. Terreaux, J. P. Ovarlez, F. Pascal, P. De Peretti, C. Chorro,
Soft Computing, 24, 8643-8654, March 2020

Robust Covariance Matrix Estimation and Portfolio Allocation: the case of non-homogeneous asset,
E. Jay, T. Soler, J. P. Ovarlez, P. De Peretti, C. Chorro,
2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020, Barcelona, Spain

Abstract

This paper presents how the use of a cleaned and robust covariance matrix estimate can improve significantly the overall performance of Maximum Variety and Minimum Variance portfolios. We assume that the asset returns are modelled through a multi-factor model where the error term is a multivariate and correlated elliptical symmetric noise extending the classical Gaussian assumptions. The factors are supposed to be unobservable and we focus on a recent method of model order selection, based on the robust Tyler M-estimator and the Random Matrix Theory (RMT) to identify the most informative subspace and then to obtain a cleaned (or de-noised) covariance matrix estimate to be used in the Maximum Variety and Minimum Variance portfolio allocation processes. We focus on the fact that the assets should preferably be classified in homogeneous groups before applying the proposed methodology which is to whiten the data before estimating the covariance matrix. We apply our methodology on real market data and show the improvements it brings if compared with other techniques especially for non-homogeneous asset returns.

2.1 Introduction

Modern portfolio theory introduced by Markowitz [1] lays the foundation for optimal portfolio construction with the so-called mean-variance strategy. This optimization problem maximizes the expected return for a given risk level in order to obtain the optimal weights. Nevertheless, its practical implementation relies on the knowledge of the empirical expected return, a quantity classically known to be very hard to estimate. To overcome these drawbacks, allocation methods focusing solely on the covariance matrix estimation have been developed, such as the Global Minimum Variance Portfolio (GMVP) or the Equal Risk Contribution Portfolio [37], [29].

An alternative method has been proposed in [30, 52], based on portfolio diversification and having only the covariance matrix as an input parameter. This method seeks the most diversified portfolio by maximizing the variety (or diversification) ratio to reduce common risk exposures.

In financial modeling, a widely used estimator of the covariance matrix is the Sample Covariance Matrix (SCM), optimal under Gaussian assumptions. Nevertheless, it is well-known that asset returns usually exhibit departures from the optimal framework as asymmetry, fat tails, tail dependence, thus leading to large estimation errors. To deal with this point, covariance matrix estimation has been extended under non-Gaussian distributions [108, 107]. These robust estimators are generally adapted when $T > m$, where T is the sample size and m is the number of assets. Indeed, for singular covariance matrix estimate ($T < m$) regularization approaches are required and some authors have therefore proposed an hybrid robust shrinkage covariance matrix estimators [117, 115, 116] based on Tyler's robust M-estimator [108] and Ledoit-Wolf's shrinkage approach [82].

Recent works [118, 117, 115, 62] based on Random Matrix Theory (RMT) have therefore considered robust estimation when $T < m$. In [62], the covariance estimation approach is based on the Shrinkage-Tyler M-estimator and the authors show that applying an adapted estimation methodology leads to achieving superior performance over many other competing methods under the GMVP framework. Another way to lower the estimation errors of the covariance matrix is to distinguish the signal part from the noisy part using filters. It is now well documented in financial literature that the introduction of multiple sources of risks is a key factor to challenge the Capital Asset Pricing Model (CAPM) single market factor assumption [83]. Multi-factor models have therefore emerged based either on statistical factors or on observable factors [88, 89, 84, 86], and are designed to capture common risk factors (systematic risks). In this setup, the covariance matrix estimate of the assets depends solely of the systematic part of the risk, as in [88]. Statistical multi-factor models are also very interesting tools. Instead of choosing the factors among many others and from empirical studies, the factors are determined from the assets universe, using statistical methods. Whereas the principal component analysis may fail in distinguishing informative factors from the noisy ones, RMT helps identifying a solution to filter noise as in [55, 56, 57, 59] by correcting the eigenvalues of the covariance matrix, thanks to the upper bound of the Marčenko-Pastur distribution [194]. This method called "Eigenvalue clipping" provides competitive out-of-sample results [61], even though in most cases only the first

component (market factor) is detected which is not completely satisfactory. Other recent works [98, 63, 100] deal with the class of Rotational Invariant Estimators (RIE) that use all of the information on both eigenvectors and eigenvalues of the covariance matrix. The methodology proposed in [98] leads to portfolios having a lower volatility that those obtained when using SCM, Ledoit & Wolf (LW) and Eigenvalue clipping methods.

In this paper, we extend the results presented in [168] by considering that the assets returns might be non-homogeneously distributed. Indeed, as in [168], we assume that the asset returns are still modelled through a multi-factor model where the error term is a multivariate and correlated elliptical symmetric noise. However, in our approach the whitening procedure is now applied by group of homogeneous assets and the final covariance estimate obtained only using the de-noised part of the observations as suggested in [96, 163, 164, 165]. This paper also focuses on assets classification to determine the sub-groups of homogeneous assets and compares two different methods: the Ascending Hierarchical Clustering (AHC) method that requires the number of groups to be fixed a priori or determined using a predefined criterion (we choose here the Caliński-Harabasz (CH) criterion [170]), and the Affinity Propagation (AP) method [171] that self-determines the number of groups. Empirical tests are carried out on two different assets universes: a set of European assets and a set of American assets, both allocated with the Maximum Variety or the Minimum Variance process. These tests extend our preliminary results and show that the way the assets are grouped might improve again the portfolio performance.

This article is organized as follows: section 2.2 introduces the selected methods of portfolio allocation for this paper: the Maximum Variety (or VarMax) portfolio and the Minimum Variance (or MinVar) portfolio. Section 2.3 presents the classical model and the related assumptions. Section 2.4 describes the covariance matrix estimation methodology for the case of non-homogeneous asset returns. Section 2.5 provides empirical illustrations ascertaining the efficiency of the proposed method compared to the conventional ones. Section 2.6 concludes and discusses our results.

2.2 Portfolio allocation

Portfolio allocation is a widely studied problem. Depending on the investment objective, the resulting portfolio allocation differs. In this section two allocation methods are described: the Maximum Variety process and the Global Minimum Variance one. Both of them depend on a single parameter that is the covariance matrix of the asset returns. In practice, the Minimum Variance portfolio is known to lead to low-diversified but performing portfolios over recent years reinforcing the low volatility anomaly concept, whereas the Maximum Variety process leads to well-diversified (by construction) but less performing portfolios.

Here, we focus only on "Long only" portfolios, i.e. all the quantities invested in assets are necessarily greater than or equal to 0. This choice is motivated for two reasons. First, when short selling the assets, borrowing costs have to be taken into account to compute portfolio performances. What is more, these costs are not uniform among assets being dependent on their liquidity. Second, when building a long/short portfolio the size of the two legs must be defined as a constraint in the optimization process, otherwise the result obtained will not be realistic with possible strong leverage effects.

2.2.1 Maximum Variety (VarMax) Portfolio

We consider m financial assets used to build an investment portfolio perfectly characterized by the allocation vector $\mathbf{w} = [w_1, \ldots, w_m]'$ where w_j represents the proportion invested in asset j. In particular, we have $0 \leq w_j \leq 1 \ \forall j \in [1, m]$ and $\sum_{j=1}^{m} w_j = 1$. In [30], the authors provide a strong mathematical definition of portfolio diversification introducing the Variety Ratio ($\mathcal{VR}$) associated with $\mathbf{w}$ that is none other than the ratio of the weighted arithmetic mean of volatilities over the portfolio volatility:

$$\mathcal{VR}(\mathbf{w}, \boldsymbol{\Sigma}) = \frac{\mathbf{w}' \boldsymbol{\sigma}}{(\mathbf{w}' \boldsymbol{\Sigma} \mathbf{w})^{1/2}}, \tag{2.1}$$

where $\boldsymbol{\Sigma}$ is the variance covariance matrix of the m assets and $\boldsymbol{\sigma} = [\sqrt{\boldsymbol{\Sigma}_{11}}, \ldots, \sqrt{\boldsymbol{\Sigma}_{mm}},]'$ the m-vector of corresponding volatilities. Thus, the Maximum Variety (or VarMax) strategy, denoted by $\mathbf{w}^*_{vr}$, is obtained as the solution of the following optimization problem under convex constraints on weights

$$\mathbf{w}^*_{vr} = \underset{\mathbf{w}}{\operatorname{argmax}} \, \mathcal{VR}(\mathbf{w}, \boldsymbol{\Sigma}). \tag{2.2}$$

The VarMax Portfolio verifies some interesting properties, as described in [52]:

- VarMax is invariant by duplication: if an asset is duplicated in the universe, then VarMax will be unchanged giving half the weight to each duplicated asset,
- VarMax stays unchanged if a positive linear combination of the assets of the universe is added as a new asset,
- any asset of the universe not held in VarMax is more correlated to the portfolio than to any asset of the portfolio. Furthermore, the more diversified a long-only portfolio is, the greater its correlation with VarMax.

VarMax portfolios are often considered as interesting diversifying investments with respect to the other investments. The above last property would therefore suggest that the other portfolios might then be weakly diversified portfolios.

2.2.2 Minimum Variance (MinVar) Portfolio

The Global Minimum Variance Portfolio (or GMVP) is obtained by computing the portfolio whose m-vector of weights $\mathbf{w}_{gmvp}$ minimizes the variance of the final portfolio. It can be formulated as a quadratic optimization problem including the linear constraint that the sum of the weights is equal to 1:

$$\min_{\mathbf{w}} \ \sigma^2(\mathbf{w}, \boldsymbol{\Sigma}) = \min_{\mathbf{w}} \ \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w}, \quad \text{s.t.} \ \ \mathbf{w}' \mathbf{1}_m = 1 \tag{2.3}$$

with $\mathbf{1}_m$ being a m-vector of ones.

The solution to (2.3), when there is no other constraint on the weight values, is then: $\mathbf{w}_{gmvp} = \dfrac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_m}{\mathbf{1}_m'\boldsymbol{\Sigma}^{-1}\mathbf{1}_m}$, and the corresponding portfolio variance writes $\sigma^2(\mathbf{w}_{gmvp}, \boldsymbol{\Sigma}) = \dfrac{1}{\mathbf{1}_m'\boldsymbol{\Sigma}^{-1}\mathbf{1}_m}$.

As for the VarMax portfolio, the covariance matrix needs to be estimated. If we denote $\widehat{\boldsymbol{\Sigma}}$ an estimate of $\boldsymbol{\Sigma}$, then we have:

$$\widehat{\mathbf{w}}_{gmvp} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_m}{\mathbf{1}_m'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_m}.$$

In [195], the authors derive an optimal optimization strategy in order to minimize the realized portfolio variance, under an assumption of spiked structures[1] of both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$. In our case, the weights have to be positive, so that the optimal minimum variance portfolio weights cannot be obtained in a closed form expression, we therefore use the optimization process. We will nevertheless compare several competing methods of covariance matrix estimation in order to get the GMVP.

To get solutions for (2.2) and (2.3), the unknown covariance matrix $\boldsymbol{\Sigma}$ has to be determined or estimated. This is a challenging problem in portfolio allocation due to the strong sensitivity of the optimisation process to outliers and estimation errors. Apart from the classical SCM or the Minimum Covariance Determinant (MCD, [196]) that is a method robust to outliers, reside subspace methods that aim at separating the signal space from the noise space, using the eigen-decomposition of the SCM. The noise and signal subspaces are usually identified according to the eigenvalues magnitudes: the eigenvectors related to the lowest ones represent the noise whereas those related to the highest eigenvalues identify the signal. But the open question remains how to choose the separating threshold? In this paper we propose a robust and original technique that applies the Random Matrix Theory (RMT) results on the eigen-decomposition of a robust M-estimator leading to a denoised and robust covariance matrix estimate.

2.3 Model and assumptions

Let us assume that the investment universe contains m assets whose returns at each time $t = 1, \cdots, T$ are stored in the m-vector $\mathbf{r}_t$. We suppose also that $\mathbf{r}_t$ admits a K factors structure, where the $K < m$ common factors are unknown, and that the additive noise is a multivariate Elliptical Symmetric noise (1.26) [109, 113]. The assumed model for $\mathbf{r}_t$ writes as follows:

$$\mathbf{r}_t = \mathbf{B}_t\,\mathbf{f}_t + \sqrt{\tau_t}\,\mathbf{C}^{1/2}\,\mathbf{x}_t, \tag{2.4}$$

where

- $\mathbf{r}_t$ is the m-vector of returns at time t,
- $\mathbf{B}_t$ is the $m \times K$-matrix of coefficients that define the assets sensitivities to each factor at time t,

[1] A spiked structure denotes a covariance model where some eigenvalues are located out of the "bulk", like outliers.

- $\mathbf{f}_t$ is the K-vector of random factor values at t, supposed to be common to all the assets,
- $\mathbf{x}_t$ is a m-vector of independent Gaussian white noise with unit variance and non-correlated with the factors, i.e. $\mathbb{E}[\mathbf{x}_t\,\mathbf{f}_t'] = \mathbf{0}_{m\times K}$,
- $\mathbf{C}$ is called the $m \times m$ scatter matrix that is supposed to be Toeplitz[2] structured [169] and time invariant over the period of observation,
- τ_t is a family of i.i.d positive random variables with expectation τ that is independent of the noise and the factors and drives the variance of the noise. These random variables are time-dependent and generate the Elliptical distribution [110] of the noise.

The Toeplitz assumption made on $\mathbf{C}$ is a required assumption for the proposed methodology described in section 2.4.1. This hypothesis imposes a particular structure for the covariance matrix of the additive noise, and is generally used to describe stationary processes [169]. In the case of model (2.4) this hypothesis is plausible as it states that the additional white noise admits a Toeplitz-structured covariance matrix. In the case of financial time series where we only observe one sample at each time, the stationarity of the dependence structure of the assets is a statistical hypothesis really difficult to test in practice. This motivates the extension we propose in this paper, described in section 2.4.4, to splitting the assets universe into groups composed of assets having similar distributions, and being most probingly sampled from a stationary process representing a unique distribution for each group.

Given equation (2.4) the covariance matrix writes for a fixed period of time t:

$$\mathbf{\Sigma}_t = \mathbf{B}_t\,\mathbf{\Sigma}_t^f\,\mathbf{B}_t' + \tau\,\mathbf{C}, \tag{2.5}$$

that is a $m\times m$-matrix composed of two terms: the factor-related term with $\mathbf{\Sigma}_t^f = \mathbb{E}[\mathbf{f}_t\,\mathbf{f}_t']$ being of rank K, and the noise-related term being of rank m. Subspace methods aim at identifying the K highest eigenvalues of $\mathbf{\Sigma}_t$ supposed to represent the K- factors especially when the power of the factors is higher than the noise power.

Determining K, the number of factors is a tough task in all the model order selection problems, like e.g. when estimating the number of emitting sources in any received signal or when trying to unmix sources in hyperspectral images [197]. In financial applications, the K factors serve in building portfolios and also to identify the main sources of risks within the investment universe under study [198, 199, 200, 201], and is therefore of main importance in such cases.

In the next section we give a detailed description of our methodology that combines the robust Tyler M-estimator of the covariance matrix and the RMT results adapted to the above non Gaussian and multivariate model.

[2]A Toeplitz matrix is a diagonal-constant matrix (defined in 1.4.1).

2.4 Proposed Methodology

2.4.1 General framework

The Tyler M-estimator [108] of the covariance matrix for the m-vector $\mathbf{r}_t$ is defined as being the solution of the following fixed-point equation:

$$\mathbf{X} = \frac{m}{T} \sum_{t=1}^{T} \frac{\mathbf{r}_t \, \mathbf{r}_t'}{\mathbf{r}_t' \, \mathbf{X}^{-1} \, \mathbf{r}_t}, \tag{2.6}$$

where the trace of the resulting matrix is equal to m, and T is the number of observations for $\mathbf{r}_t$. Applied to model (2.4) and under non-Gaussian assumptions, the resulting Tyler-M estimate (that we denote $\widehat{\mathbf{C}}_{tyl}$) is shown to be the most robust covariance matrix estimator [108, 114] for the true scatter matrix $\mathbf{C}$. $\widehat{\mathbf{C}}_{tyl}$ is also independent of the distribution of the variable τ.

Under the white noise assumption, extracting information from the observed signal using RMT is quite straightforward and has been proposed in many applications, like in source detection [202], in radar detection [203], or signal subspace estimation using an adapted MUSIC (MUltiple SIgnal Classification) detection algorithm [204]. Nevertheless, when the noise is correlated, RMT results do not apply directly as the variance of the Marčenko-Pastur threshold has to be estimated, and only numerical methods can help in finding the resulting threshold [96, 97]. In some cases, secondary data that do not contain any sources can be used to estimate the covariance matrix and then whiten the observed data.

However, recent works [163, 164, 165] brought a solution that consists of applying a biased Toeplitz operator on $\widehat{\mathbf{C}}_{tyl}$ (defined in 1.4.1), let us say $\widetilde{\mathbf{C}}_{tyl} = \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}\right)$, which was proven to spectrally converge towards the theoretical scatter matrix $\mathbf{C}$. This result refers to the *Consistency Theorem* in [163, 164, 165], and asserts that whenever the sources are present in the observations, the resulting scatter matrix estimate is a consistent estimation of its theoretical value.

Consistency Theorem [163, 164, 165]
Under the RMT regime assumption, i.e. that $T, m \to \infty$, such as $m/T \to c \in]0, \infty[$, we have the following spectral convergence:

$$\left\| \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}\right) - \mathbf{C} \right\| \xrightarrow{a.s.} 0. \tag{2.7}$$

The first step of our methodology consists therefore in estimating $\widetilde{\mathbf{C}}_{tyl}$ from T observations of $\mathbf{r}_t$ in order to whiten the observations leading to the T whitened observations $\mathbf{r}_{w,t} = \widetilde{\mathbf{C}}_{tyl}^{-1/2} \, \mathbf{r}_t$.

Given the whitened observations $\{\mathbf{r}_{w,t}\}$ and their Tyler's covariance matrix $\widehat{\mathbf{\Sigma}}_{tyl}$, it has been shown in [165] that the eigenvalues distribution of $\widehat{\mathbf{\Sigma}}_{tyl}$ fit the predicted bounded distribution of Marčenko-Pastur [194] (1.5). As $m, T \to \infty$ such that $m/T \to c \in]0, \infty[$,

if $\mathbf{R}_w = \{\mathbf{r}_{w,t}\}_{t\in[1,T]}$ does not contain any factor, then:

$$\left\| \hat{\mathbf{\Sigma}}_{tyl} - \frac{1}{T}\mathbf{X}\,\mathbf{X}' \right\| \xrightarrow{a.s.} 0$$

However, if one or several sources are contained in the observations, being powerful enough to be detected, then there will be as many eigenvalues as there are sources standing outside the upper bound of the Marčenko-Pastur distribution, given in that case by $\bar{\lambda} = \sigma^2\left(1+\sqrt{c}\right)^2$ where $c = m/T$ and $\sigma^2 = 1$ (due to the preceding whitening process σ^2 is equal to one). Once the K largest eigenvalues larger than $\bar{\lambda}$ are detected, we process as for the Eigenvalue clipping in [56] to set the values of the remaining $m-K$ lowest eigenvalues to a unique value equal to $\left(Tr\left(\widehat{\mathbf{\Sigma}}_{tyl}\right) - \sum_{k=K+1}^{m}\lambda_k\right)/(m-K)$. Using also the corresponding eigenvectors, we then build back the de-noised assets covariance matrix to be used in (2.2) and (2.3) or in any other objective function. The whitening procedure is detailed more precisely in the next subsection.

2.4.2 Detailed whitening procedure

Given $\mathbf{R}$ the $m\times T$ matrix of observations, the de-noised covariance matrix estimate $\widehat{\mathbf{\Sigma}}_w$ is obtained through the following procedure steps:

1. Set $\widehat{\mathbf{C}}_{tyl}$ as the Tyler-M estimate of $\mathbf{R}$, solution of (2.6),
2. Set $\widetilde{\mathbf{C}}_{tyl} = \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}\right)$, the Toeplitz rectification matrix built from $\widehat{\mathbf{C}}_{tyl}$ for the Toeplitz operator $\mathcal{T}(.)$ (defined in 1.4.1),
3. Set $\mathbf{R}_w = \widetilde{\mathbf{C}}_{tyl}^{-1/2}\mathbf{R}$, the $m\times T$ matrix of the whitened observations,
4. Set $\widehat{\mathbf{\Sigma}}_{tyl}$ as the Tyler-M estimate of $\mathbf{R}_w$, solution of (2.6),
5. Set $\widehat{\mathbf{\Sigma}}_{tyl}^{clip} = \mathbf{U}\,\mathbf{\Lambda}^{clip}\,\mathbf{U}'$ where $\mathbf{U}$ is the $m\times m$ eigenvectors matrix of $\widehat{\mathbf{\Sigma}}_{tyl}$ and $\mathbf{\Lambda}^{clip}$ is the $m\times m$ diagonal matrix of the eigenvalues $(\lambda_k^{clip})_{k\in[1,m]}$ corrected using the Eigenvalue clipping method [56] (1.6),
6. Finally, $\widehat{\mathbf{\Sigma}}_w = \left(\widetilde{\mathbf{C}}_{tyl}^{1/2}\right)\widehat{\mathbf{\Sigma}}_{tyl}^{clip}\left(\widetilde{\mathbf{C}}_{tyl}^{1/2}\right)'$.

2.4.3 Simulation example

To illustrate the efficiency of the whitening process, we ran the following test: we simulate $T = 1000$ observations of a $m = 100$ here, sampled from a highly correlated K-distributed process [113] having a shape parameter $\nu = 0.5$, and a Toeplitz-structured covariance matrix whose coefficient $\rho = 0.8$ (each element j,k of the Toeplitz matrix is defined by $\rho^{|j-k|}, j,k = 1,\cdots,m$). We then embed $K = 3$ sources of information in the non-Gaussian and correlated noise, and we compare the eigenvalues distribution of the observations with the Marčenko-Pastur upper bound when the eigenvalues are computed from i) the SCM (on the left), ii) the Tyler M-estimate matrix (in the middle), and iii) the Tyler M-estimate matrix of the whitened observations. It appears clearly that the $K = 3$ factors can be identified quite easily only in the case where the observations are firstly whitened.

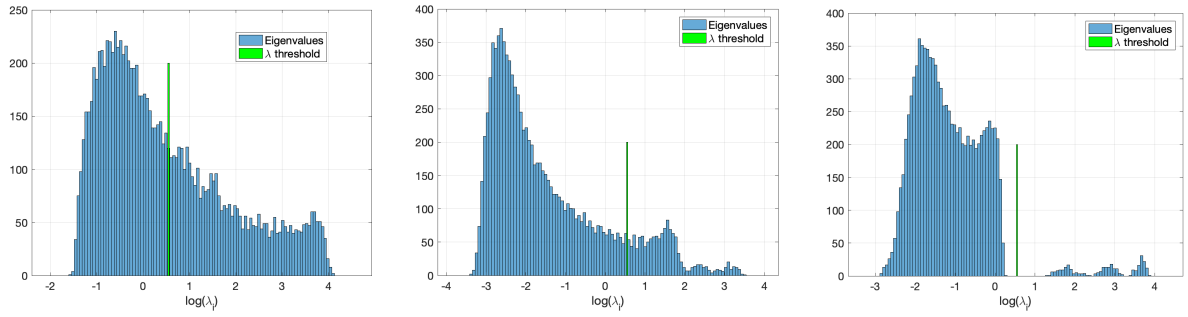


Figure 2.1: Distributions of the logarithm of the eigenvalues of three covariance matrix estimates. Left side: Eigenvalues (log) of the SCM of the observations; Middle: Eigenvalues (log) of the Tyler covariance matrix of the observations; Right side: Eigenvalues (log) of the Tyler covariance matrix of the whitened observations. Observations contain $K = 3$ sources embedded in a multivariate K-distributed noise with shape parameter $\nu = 0.5$, and a Toeplitz coefficient $\rho = 0.8$. $m = 100$, $T = 1000$ ($c = 0.1$), and the (log) Marčenko-Pastur upper bound is here: $\log(\bar{\lambda}) = \log(1.7325)$.

2.4.4 The case of non-homogeneous assets returns

The whitening process proposed above is made under the implicit assumption that the assets returns are drawn from a unique multivariate law and are therefore homogeneous in law. As described hereafter this assumption is unrealistic for financial time series of returns. We therefore propose to split the m assets into $p < m$ groups, each composed of $\{m_q\}_{q=1}^{p}$ assets (with $\sum_{q=1}^{p} m_q = m$), and formed to be composed of assets having similar distributions. We set a fixed number of groups, and group the assets regarding their returns distributions. Under this new assumption, model (2.4) applies for each group q as follows:

$$\mathbf{r}_t^{(q)} = \mathbf{B}_t^{(q)} \mathbf{f}_t + \sqrt{\tau_t^{(q)}}\, \mathbf{C}_{(q)}^{1/2} \mathbf{x}_t, \tag{2.8}$$

Then, the full model (2.4) rewrites:

$$\begin{bmatrix} \mathbf{r}_t^{(1)} \\ \vdots \\ \mathbf{r}_t^{(p)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_t^{(1)} \\ \vdots \\ \mathbf{B}_t^{(p)} \end{bmatrix} \mathbf{f}_t + \begin{pmatrix} \tau_t^{(1)}\mathbf{C}_{(1)} & \mathbf{0}_{1,2} & \cdots & \mathbf{0}_{1,p} \\ \mathbf{0}_{2,1} & \tau_t^{(2)}\mathbf{C}_{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{p-1,p} \\ \mathbf{0}_{p,1} & \cdots & \mathbf{0}_{p,p-1} & \tau_t^{(p)}\mathbf{C}_{(p)} \end{pmatrix}^{1/2} \mathbf{x}_t,$$

where $\mathbf{0}_{j,k}$ denotes the null matrix of size $m_j \times m_k$, $j, k = 1, \cdots, p$, corresponding to the additional hypothesis that the groups are uncorrelated each others. The complete scatter matrix $\mathbf{C}$ is therefore block-constructed, and block-Toeplitz.

2.4.4.1 Assets classification

Under the assumption of non-homogeneous assets returns, we propose to form groups of assets before applying the whitening process. There are many classification methods that differ according to their mode of learning (supervised, unsupervised, hierarchical, etc.). No matter the method chosen, a clustering algorithm will always be used in such a way as to obtain a partition of assets in p groups. Usually, the main question is the choice of p

that is in general unknown. To overcome the choice of p, we use the Affinity Propagation algorithm (AP) [171] that does not require to specify the number of clusters. We also use the classical Ascending Hierarchical Classification with Caliński-Harabasz [170] criterion to set the number of groups.

2.4.4.1.1 Affinity Propagation algorithm (AP)

The Affinity Propagation algorithm (AP) [171] is an iterative partitioning method similar to the K-means, but instead of regrouping individuals around central values, AP algorithm regroups them around exemplar values and all individuals can be considered as potential exemplars. The algorithm is based on a similarity matrix $\mathbf{S}$, where $s_{j,k} = -\|\mathbf{v}_j - \mathbf{v}_k\|^2$ for $j \neq k$, and with $\mathbf{v}_j$ and $\mathbf{v}_k$ the input variables vectors of the asset j and k. The number of groups is influenced by the main diagonal of $\mathbf{S}$ ($s_{j,j} \, \forall j \in [1, m]$) also called "preferences" parameters. In order to moderate the number of groups p, the parameters are set to a common value using the median of pairwise similarities as in [171].

In the AP algorithm, data points exchange information by passing messages until a set of exemplars are obtained. Two messages are used in the AP algorithm. The first one, called "responsibility" $y_{j,k}$ quantities how the element k is suitable to serve as an exemplar for the element j taking into account other potential exemplars. The "responsibility" $y_{j,k}$ is updated as follows:

$$y_{j,k} = s_{j,k} - \max_{k' \neq k} \{a_{j,k'} + s_{j,k'}\},$$

The second one, called "availability" $a_{j,k}$ quantifies how appropriate it would be for the element j to choose the element k as its exemplar, taking into account the consideration from other elements that element k should be an exemplar. The "availability" $y_{j,k}$ is updated as follows:

$$a_{j,k} = \min\{0, y(k,k) + \sum_{j' \notin \{j,k\}} \max\{0, y_{j',k}\}\},$$

where the self-availability $a_{k,k}$ reflects accumulated evidence that element k is an exemplar, based on the positive responsibilities of k towards other elements. The self-availability $a_{k,k}$ is defined as follows:

$$a_{k,k} = \sum_{j' \neq k} \max\{0, y_{j',k}\},$$

Finally, after updating passing messages, exemplars can be identified as follows:

$$z_{j,k} = y_{j,k} + a_{j,k}.$$

where if $k = j$, then the element j is selected as an exemplar, otherwise k is the exemplar of the element j.

2.4.4.1.2 Ascending Hierarchical Classification (AHC)

The classical Ascending Hierarchical Classification (AHC) is an iterative and unsupervised method. The algorithm is based on the distances between the variables $(\mathbf{v}_j)_{j\in[1,m]}$ used to represent individuals to be grouped and seeks at each step to build the groups by aggregation. AHC ensures to get homogeneous groups for which the intra-group variances are smaller than the inter-group variances. We use AHC with the Euclidean distance and the Ward measure [205] to form the p groups. The number of groups p is determined arbitrary or with Caliński-Harabasz (CH) criterion [170].

The algorithm can be described as follows:

- The initial classes are the m singletons (each individual represents a class). All distances between the m individuals ($m(m-1)/2$ distances) are computed and the two nearest individuals are aggregated into a new element. A first partition of $m-1$ classes is thus obtained.
- The distances are computed between the new element and the individuals. The other distances are unchanged. We again aggregate the two nearest elements and we obtain a new partition with $m-2$ classes. The process is repeated until there is only one element that includes all the individuals and is the last one partition.

2.4.4.2 Detailed whitening procedure by group

Given $\mathbf{R}$ the $m\times T$-matrix of observations, and $\mathbf{R}^{(q)}$ the $m_q\times T$-matrix of observations for group (q), the de-noised covariance matrix estimate $\widehat{\mathbf{\Sigma}}_w$ is obtained through the following procedure steps:

1. Compute the p groups using the methods described in 2.4.4.1 with $(\mathbf{v}_j)_{j\in[1,m]}$ composed of the mean μ_j, the standard deviation σ_j and of several quantiles computed from $\tilde{\mathbf{r}}_j = (\mathbf{r}_j - \mu_j\,\mathbf{1}_T)/\sigma_j$ the "standardized" returns, where $\mathbf{1}_T$ is the T-vector of ones,
2. Set $\widehat{\mathbf{C}}_{tyl}^{(q)}$ the Tyler-M estimate of $\mathbf{R}^{(q)}$, solution of (2.6),
3. Set $\widetilde{\mathbf{C}}_{tyl}^{(q)} = \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}^{(q)}\right)$, the Toeplitz rectification matrix built from $\widehat{\mathbf{C}}_{tyl}^{(q)}$ for the Toeplitz operator $\mathcal{T}(.)$ (defined in 1.4.1),
4. Set $\mathbf{R}_w^{(q)} = \left(\widetilde{\mathbf{C}}_{tyl}^{(q)}\right)^{-1/2}\mathbf{R}^{(q)}$, the $m_q\times T$ matrix of the whitened observations of group q,
5. Set $\widehat{\mathbf{\Sigma}}_{tyl}$ as the Tyler-M estimate of $\mathbf{R}_w$, solution of (2.6), where $\mathbf{R}_w = [\mathbf{R}_w^{(1)'}\ldots\mathbf{R}_w^{(p)'}]'$ of size $m\times T$,
6. Set $\widehat{\mathbf{\Sigma}}_{tyl}^{clip} = \mathbf{U}\,\mathbf{\Lambda}^{clip}\,\mathbf{U}'$ where $\mathbf{U}$ is the $m\times m$ eigenvectors matrix and $\mathbf{\Lambda}^{clip}$ is the $m\times m$ diagonal matrix of the eigenvalues $(\lambda_k^{clip})_{k\in[1,m]}$ corrected using the Eigenvalue clipping method [56] (1.6),
7. Finally, $\widehat{\mathbf{\Sigma}}_w = \left(\widetilde{\mathbf{C}}_{tyl}^{1/2}\right)\widehat{\mathbf{\Sigma}}_{tyl}^{clip}\left(\widetilde{\mathbf{C}}_{tyl}^{1/2}\right)'$.

2.5 Application

In this section we apply our methodology to the Maximum Variety and Minimum Variance portfolios. Two investment universes are tested: the first one consists of European equity indices ($m = 43$) and the second one to US equity indices ($m = 30$)[3]. These indices represent industry sub-sectors (e.g. transportation or materials), factor-based indices (e.g. momentum or growth), and also countries (e.g. Sweden or France) for the European universe. Using a blend of equities instead of individual stock allows capturing collective risks (systematic) rather than idiosyncratic ones and reinforce portfolio diversification without having to impose constraints to reduce a stock-specific and liquidity risk. Our daily track record spans from July 2000, the 27th to May 2019, the 20th. We use closing prices, i.e. the last traded price during stock exchange trading hours.

To build the portfolios, the weights are computed as follows: we estimate every four weeks the covariance matrix of assets using the last year of daily returns (T = 260 weekdays) and we optimize the objective function of Maximum Variety (2.1) or Minimum Variance (2.3) to obtain the vector of weights. Finally, the weights remain constant between two rebalancing periods of four weeks. When applicable, assets are classified either by the AP algorithm ("RMT-Tyler (AP)") or by AHC where the number of groups is set to $p = 6$ ("RMT-Tyler (AHC-6)") or set according to the CH criterion ("RMT-Tyler (AHC-CH)"). The quantiles used for the clustering algorithms are q_θ and $q_{1-\theta}$ with $\theta \in \{1\%, 2.5\%, 5\%, 10\%, 15\%, 25\%, 50\%\}$.

We compare the results with those obtained using the whitening process applied on the whole universe ("RMT-Tyler (all)"), the "SCM" and also with three other competing methods: the first one, denoted as "RMT-SCM" uses the Eigenvalue clipping of [56] (1.6), the second one, that we denote as "LW", is the method that uses the Ledoit & Wolf shrinkage of [58] (1.3), and finally the method using the Rotational Invariant Estimator of [98, 63] (1.7), denoted as "RIE". These methods are described in 1.2.1. We also add for comparison the equally weighted portfolio and the respective benchmark for each universe (MSCI® Europe Index or S&P® 500 Index)

We report several portfolios statistics computed over the whole period in order to quantify the benefits of the proposed methodology: the annualized return, the annualized volatility, the ratio between the annualized return and the annualized volatility, the value of the maximum drawdown (that is the return between the highest and the lowest portfolio levels observed during the whole period), and the average of the Variety Ratios computed at each rebalancing date. The higher is the return/volatility ratio, the lower is the maximum drawdown and the higher is the variety ratio, and better performing is the portfolio.

Moreover, all portfolio performances are "net of transaction fees", considering 0.07% of fees (or 7 basis points denoted as "bp") applied to any weight change from one time to the next one. Measuring the total weights changes is referred as the turnover of the portfolio. We assume that the turnover between two consecutive periods t and $t+1$ is measured by $\sum_{j=1}^{m} |w_{j,t+1} - w_{j,t}|$. If, for example, the turnover is equal to 0.15 for changing

[3]Data are available upon request.

weights from t to $t+1$, then the portfolio performance computed between t and $t+1$ will be decreased by 0.15×7 bp $= 0.0105\%$. Turnover is an important number in portfolio allocation. If you ever find an apparently well performing strategy that indicates you to change the overall portfolio at each time, then the cost of changing the overall portfolio will surely be equivalent or larger than would be the performance of the strategy itself. Here, the proposed technique leads to increase the cumulated turnover, but reasonably enough to let the improvement be a significant improvement that do not cost all the benefits of the technique. Limiting the turnover is often added as an additional non linear constraint to any optimization process like (2.2) or (2.3).

2.5.1 EU Variety Maximum (VarMax) portfolios results

Figure 2.2 shows the evolution of the VarMax portfolios wealth on the Europe (EU) universe, starting at 100 at the beginning of the first period. The "SCM", "RMT-SCM", "LW", "RIE", "RMT-Tyler (all)", "RMT-Tyler (AHC-6)", "RMT-Tyler (AHC)" and "RMT-Tyler (AP)" VarMax portfolios are respectively in red, pink, cyan, blue, purple, orange, brown and green. The naive equi-weighted portfolio is reported as the dotted grey line, and the price of the benchmark, also rebased at 100 at the beginning of the period, is the black line.

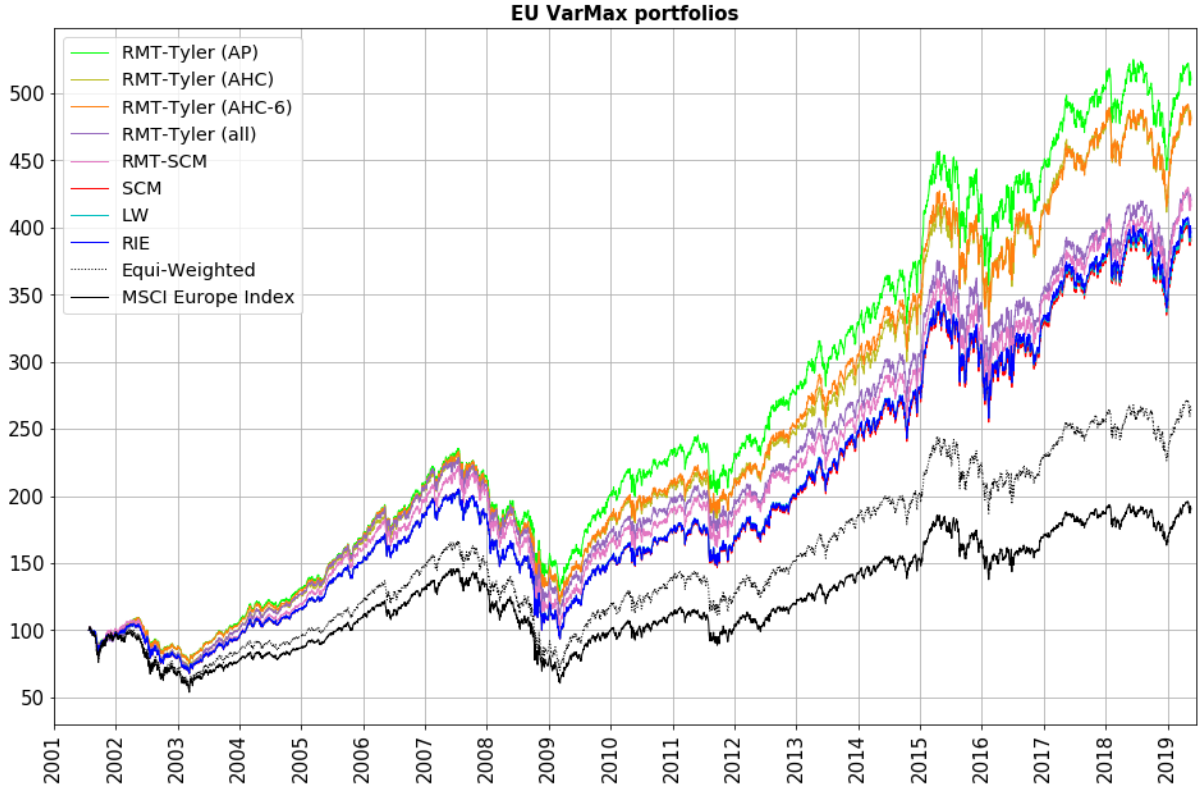


Figure 2.2: VarMax portfolios wealth from July 2001 to May 2019 on the EU universe. The proposed "RMT-Tyler (AP)" (green line) leads to improved performances vs the "RMT-Tyler (AHC)" (brown), the "RMT-Tyler (AHC-6)" (orange), the "RMT-SCM" (pink), the "LW" (cyan), the "RIE" (blue) and the "SCM" (red). All whitening process applied by group shown in Table 2.1, provide higher annualized returns, lower annualized volatilities, lower maximum drawdowns and higher Diversification Ratios. But it results in a twice higher turnover: we then have taken into account 7bp (or 0.07%) of transactions fees to compare the portfolios wealth.

The proposed "RMT-Tyler"-based techniques clearly outperform the conventional ones. Moreover, whitening homogeneous groups of data instead of the whole data set improves even more the results especially for the "RMT-Tyler (AP)". Regarding the other methods, "RMT-SCM" is the only one that outperforms significantly "SCM", but shows weaker performances than our proposed method does; "LW" and "RIE" are quite similar to "SCM".

Finally, we report on Table 2.1 some statistics on the overall portfolios performance: we compare, for the whole period, the annualized return, the annualized volatility, the ratio between the return and the volatility, the maximum drawdown and the average value of the diversification ratio, for the portfolios and the benchmark. All the indicators related to the proposed technique show a significant improvement with respect to the other methods: a higher annualized return, a lower volatility (so a higher return/volatility ratio), a lower maximum drawdown and a higher diversification ratio.

EU VarMax Portfolios	Ann. Ret.	Ann. Vol.	Ratio Ret/Vol	Max Drawdown	$\mathcal{VR}$ (avg)
RMT-Tyler (AP)	9.87%	12.14%	0.81	45.37%	1.46
RMT-Tyler (AHC-6)	9.65%	12.03%	0.80	46.84%	1.57
RMT-Tyler (AHC)	9.58%	12.45%	0.77	48.16%	1.51
RMT-Tyler (all)	8.90%	13.16%	0.68	51.18%	1.44
RMT-SCM	8.94%	13.79%	0.65	54.15%	1.27
RIE	8.65%	13.65%	0.63	54.44%	1.38
LW	8.59%	13.57%	0.63	54.28%	1.40
SCM	8.56%	13.68%	0.63	54.45%	1.38
Equi-Weighted	*6.60%*	*15.37%*	*0.43*	*57.82%*	*1.19*
MSCI Europe Index	*4.71%*	*14.87%*	*0.32*	*58.54%*	

Table 2.1: Performance numbers for the EU VarMax portfolios with 0.07% of fees from July 2001 to May 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

2.5.2 EU Minimum Variance (MinVar) portfolios results

Results obtained for the EU MinVar portfolios also show some improvements but less important than for the VarMax portfolios. Figure 2.3 shows that whitening by group process improves the performance whereas whitening the whole assets ("RMT-Tyler (all)") do not bring improvement with respect to all the other approaches, even if the variety ratio is higher. "RMT-SCM", "LW" and "RIE" provide lower or similar performances if compared to "SCM".

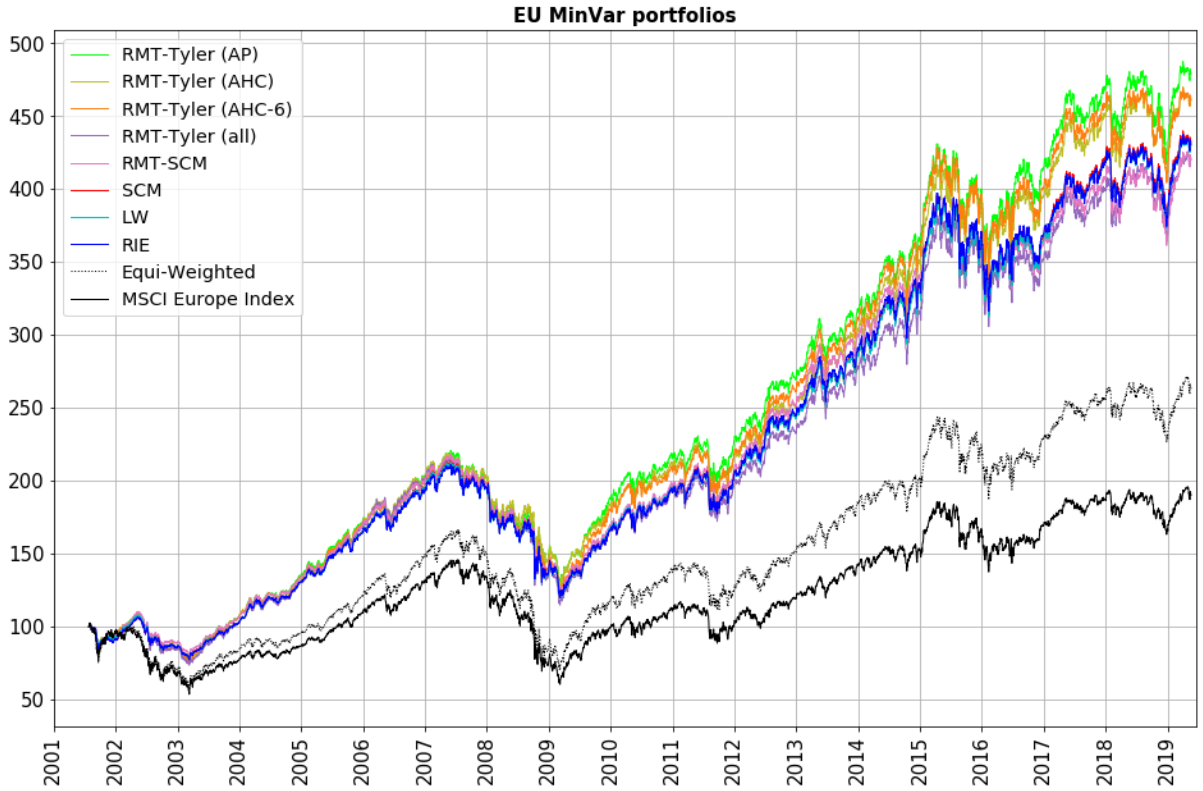


Figure 2.3: MinVar portfolios wealth from July 2001 to May 2019 on the EU universe. The proposed "RMT-Tyler (AP)" (green line) leads to improved performances vs the "RMT-Tyler (AHC)" (brown), the "RMT-Tyler (AHC-6)" (orange), the "RMT-SCM" (pink), the "LW" (cyan), the "RIE" (blue) and the "SCM" (red), as shown in Table 2.2. MinVar portfolios are known to result in poorly diversified portfolios and to invest in the lowest volatile assets. But surprisingly, the low-volatility anomaly applies in such cases.

As for the VarMax portfolios, Table 2.2 reports the EU MinVar portfolios statistics. Again, the indicators related to the proposed technique show an improvement if compared to the classical techniques.

EU MinVar Portfolios	Ann. Ret.	Ann. Vol.	Ratio Ret/Vol	Max Drawdown	$\mathcal{VR}$ (avg)
RMT-Tyler (AP)	9.39%	11.06%	0.85	41.76%	1.39
RMT-Tyler (AHC-6)	9.35%	11.08%	0.84	41.07%	1.52
LW	8.75%	10.75%	0.81	43.69%	1.21
RIE	8.76%	10.78%	0.81	43.24%	1.19
RMT-Tyler (AHC)	9.20%	11.32%	0.81	41.91%	1.44
SCM	8.74%	10.92%	0.80	43.78%	1.19
RMT-SCM	8.62%	10.80%	0.80	43.95%	1.14
RMT-Tyler (all)	8.72%	11.58%	0.75	46.50%	1.36
Equi-Weighted	*6.60%*	*15.37%*	*0.43*	*57.82%*	*1.19*
MSCI Europe Index	*4.71%*	*14.87%*	*0.32*	*58.54%*	

Table 2.2: Performance numbers for the EU MinVar portfolios with 0.07% of fees from July 2001 to May 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

Moreover, we know that minimizing the portfolio variance leads to choosing the assets having the lowest volatilities. Then, using a robust approach does flatten the volatility differences between assets and then the ex-post portfolio volatility, computed classically, will be higher than the ex-post portfolio volatility computed using the robust matrix. Nevertheless, our process leads to higher performance that the classical SCM exhibiting a higher diversification ratio, and also a lower maximum drawdown.

To illustrate this purpose, Figure 2.4 plots the standard deviations of the invested assets versus the resulting weights obtained for MinVar/SCM weights (on the top graph) the VarMax/SCM (on the bottom graph). The same conclusion arises for the whitening process applied by group. It shows explicitly which assets are preferred and when, according to their volatility level.

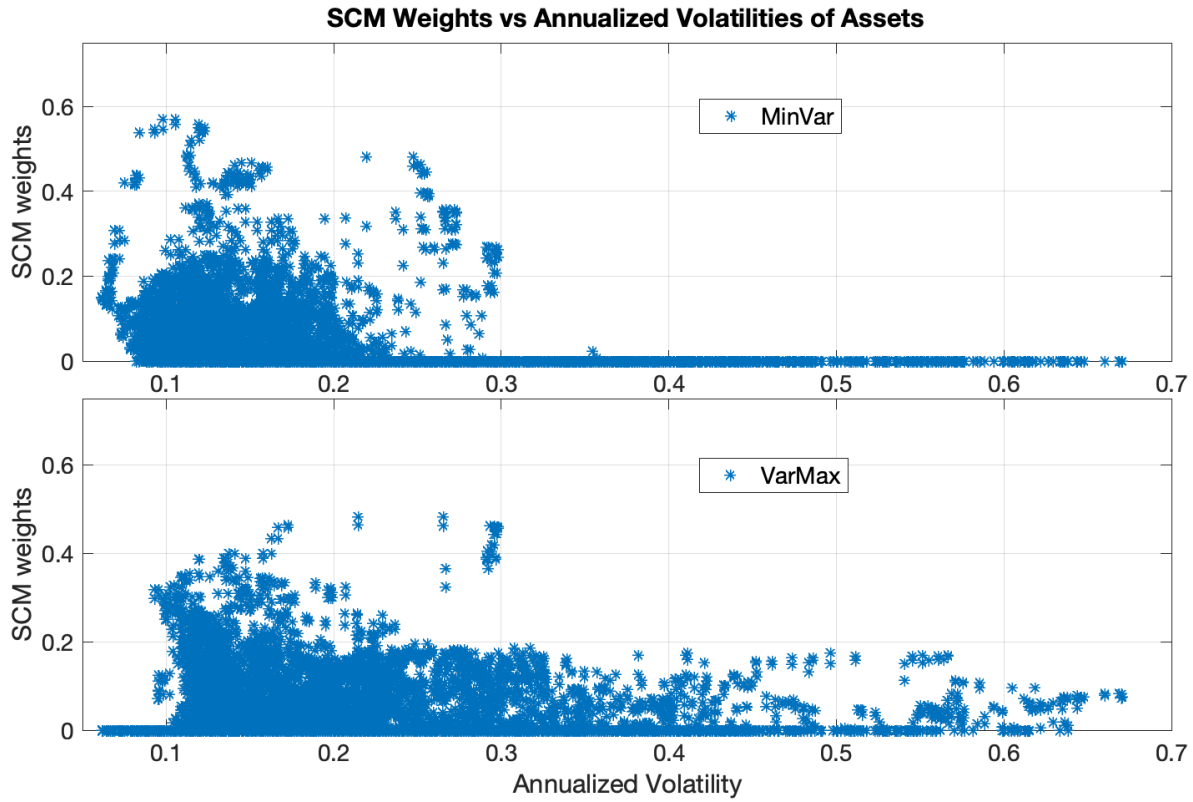


Figure 2.4: VarMax and MinVar SCM weights versus the assets volatilities. As expected, MinVar weights are mostly non-zeros for the assets having the lowest volatilities. VarMax weights are more indifferent to the volatility levels.

On a similar way, Figure 2.5 shows that VarMax assigns non-zeros weights to the less correlated assets if compared to the non-zeros MinVar weights.

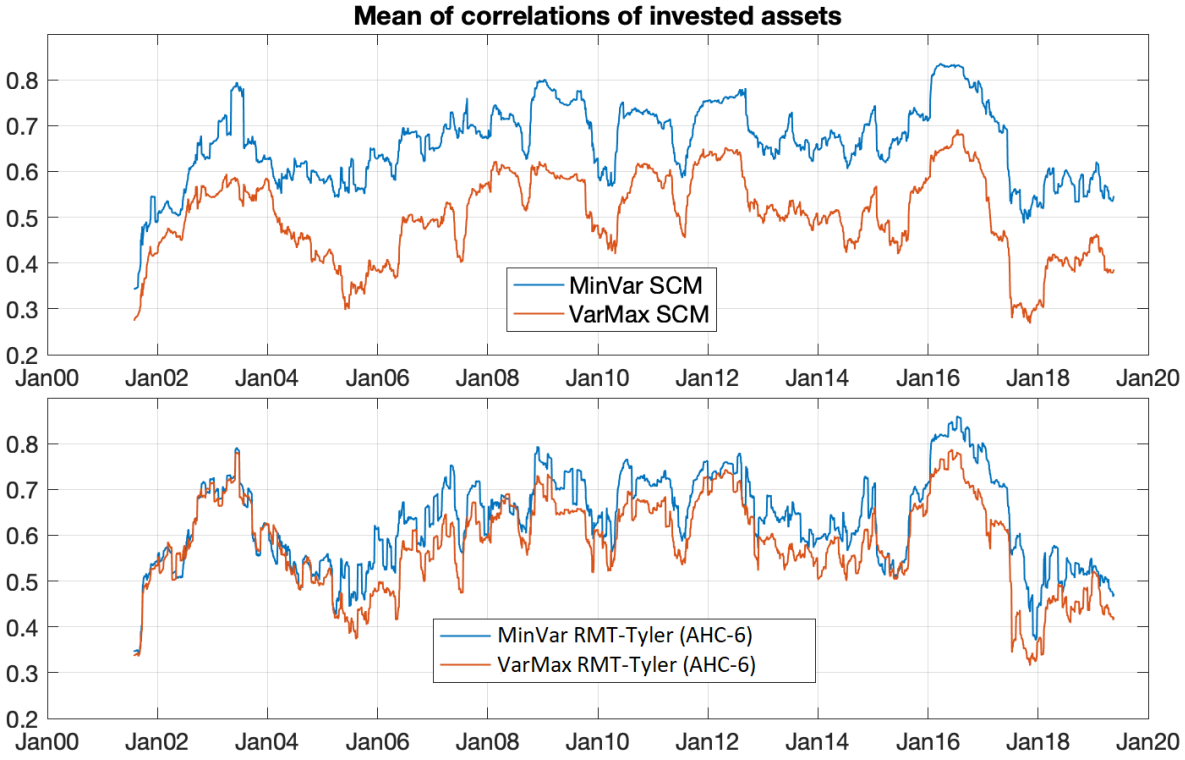


Figure 2.5: Average correlation of the invested assets for the VarMax and MinVar portfolios combined with either SCM or RMT-Tyler (AHC-6) method. VarMax SCM weights are assigned to the less correlated assets if compared to the SCM MinVar weights and the difference is reduced in the RMT-Tyler(AHC-6) case. The same conclusion can be drawn for the RMT-Tyler(AP) and RMT-Tyler(AHC) cases.

2.5.3 US Variety Maximum (VarMax) portfolios results

Figure 2.6 shows the evolution of the VarMax portfolios wealth on the US universe. As for the EU VarMax portfolios, the "RMT-Tyler (AP)" clearly outperforms the conventional ones. Moreover, whitening homogeneous groups of data instead of the whole data set improves even more the results especially for the "RMT-Tyler (AP)". Regarding the other methods, "RMT-SCM" is the only one that outperform significantly "SCM", but shows weaker performances than our proposed method does; "LW" and "RIE" are quite similar to "SCM".

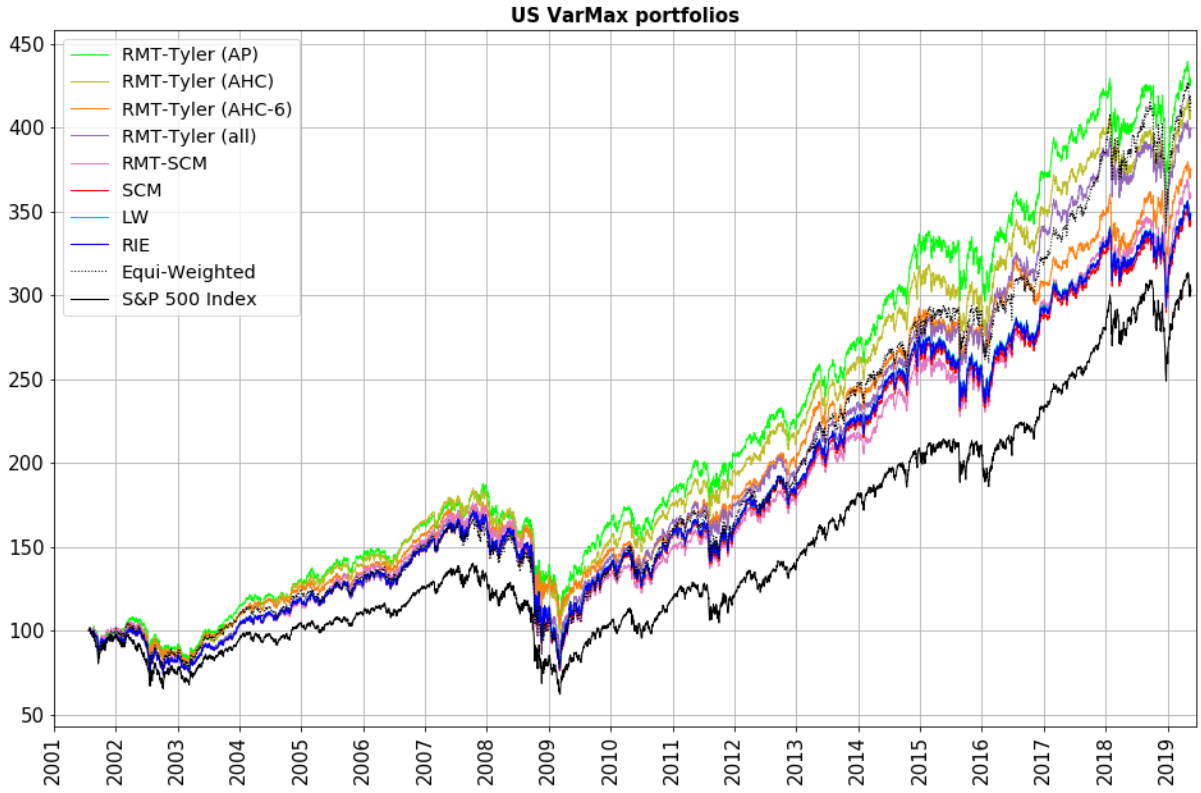


Figure 2.6: VarMax portfolios wealth from July 2001 to May 2019 on the US universe. The proposed "RMT-Tyler (AP)" (green line) leads to improved performances vs the "RMT-Tyler (AHC)" (brown), the "RMT-Tyler (AHC-6)" (orange), the "RMT-SCM" (pink), the "LW" (cyan), the "RIE" (blue) and the "SCM" (red), as shown in Table 2.4.

Table 2.3 reports the US VarMax portfolios statistics. Here again, the indicators related to the proposed technique show an improvement if compared to the classical techniques.

US VarMax Portfolios	Ann. Ret.	Ann. Vol.	Ratio Ret/Vol	Max Drawdown	$\mathcal{VR}$ (avg)
RMT-Tyler (AP)	8.76%	11.11%	0.79	42.82%	1.51
RMT-Tyler (AHC)	8.57%	11.53%	0.74	46.57%	1.55
RMT-Tyler (AHC-6)	7.98%	10.79%	0.74	41.5%	1.52
RMT-Tyler (all)	8.49%	12.09%	0.70	49.27%	1.53
Equi-Weighted	*8.92%*	*13.83%*	*0.65*	*53.70%*	*1.25*
RMT-SCM	8.03%	13.13%	0.61	56.53%	1.34
LW	7.85%	13.02%	0.60	54.32%	1.46
RIE	7.86%	13.23%	0.59	55.17%	1.47
SCM	7.80%	13.27%	0.59	55.47%	1.46
S&P 500 Index	*7.21%*	*14.18%*	*0.51*	*55.71%*	

Table 2.3: Performance numbers for the US VarMax portfolios with 0.07% of fees from July 2001 to May 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

2.5.4 US Minimum Variance (MinVar) portfolios results

Figure 2.7 shows the evolution of the MinVar portfolios wealth on the US universe. In this case, the classical method outperform the "RMT-Tyler"-based techniques, especially the "RMT-SCM". However, the whitening process applied by group improves the results compared to the whitening process on the whole data and the "RMT-Tyler (AP)" still clearly stands out from others.

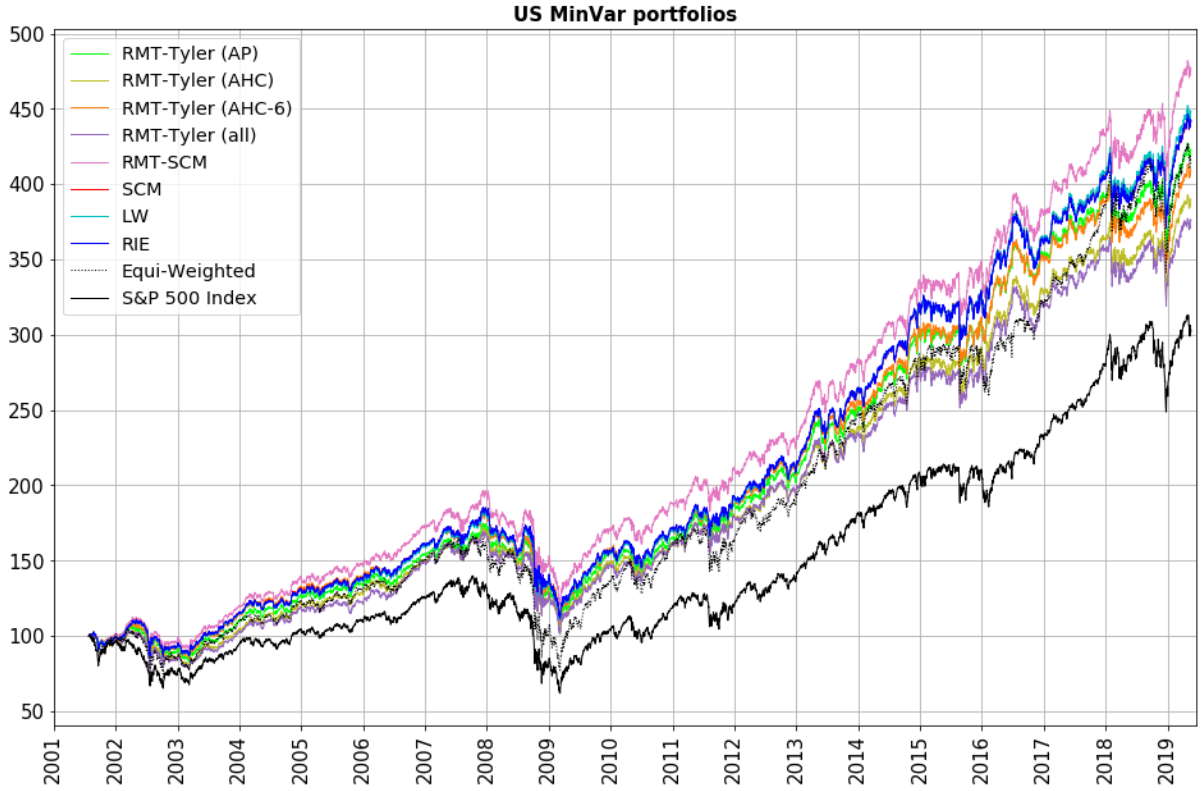


Figure 2.7: MinVar portfolios wealth from July 2001 to May 2019 on the US universe. The "RMT-SCM" (purple line) outperform the "RMT-Tyler (AP)" (green), the "RMT-Tyler (AHC)" (brown), the "RMT-Tyler (AHC-6)" (orange), the "LW" (cyan), the "RIE" (blue) and the "SCM" (red). However, the "RMT-Tyler (AP)" improves the results compared to the others "RMT-Tyler"-based techniques, as shown in Table 2.4.

Table 2.4 reports the US MinVar portfolios statistics. As previously stated, the classical method outperform the "RMT-Tyler"-based techniques, even if the indicators related of the "RMT-Tyler (AP)" show an improvement if compared to the others "RMT-Tyler"-based techniques.

US MinVar Portfolios	Ann. Ret.	Ann. Vol.	Ratio Ret/Vol	Max Drawdown	$\mathcal{VR}$ (avg)
RMT-SCM	9.28%	10.34%	0.90	38.81%	1.18
LW	8.96%	10.57%	0.85	39.73%	1.30
RIE	8.89%	10.61%	0.84	39.69%	1.27
SCM	8.89%	10.62%	0.84	39.78%	1.27
RMT-Tyler (AP)	8.65%	10.63%	0.81	39.35%	1.45
RMT-Tyler (AHC-6)	8.46%	10.58%	0.80	40.66%	1.45
RMT-Tyler (AHC)	8.18%	10.55%	0.78	40.60%	1.43
RMT-Tyler (all)	7.98%	10.62%	0.75	40.45%	1.46
Equi-Weighted	*8.92%*	*13.83%*	*0.65*	*53.70%*	*1.25*
S&P 500 Index	*7.21%*	*14.18%*	*0.51*	*55.71%*	

Table 2.4: Performance numbers for the US MinVar portfolios with 0.07% of fees from July 2001 to May 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

2.6 Conclusion

In this paper, we have shown that when the covariance matrix is estimated with the Tyler M-estimator and RMT, the Maximum Variety and the Minimum Variance Portfolio allocation processes lead to improved performances with respect to several classical estimators. Moreover, we have proposed to extend the first results in [168] by considering the case of non-homogeneous asset returns while keeping a multi-factor model where the error term is a multivariate and correlated elliptical symmetric noise. Indeed, the underlying assumption of the whitening process is that asset returns are homogeneous in distribution, which is unrealistic for financial time series of returns.

To deal with this point, we have first grouped the assets within homogeneously distributed classes before processing. Applying the whitening process on homogeneous groups of data rather than the whole data set improves even more the results. Moreover, we have also questioned the ability of classification methods (AP algorithm and AHC) to improve the estimation of the covariance matrix of financial assets using the Tyler M-estimator and the RMT. The withening process using the AP algorithm have been tested on both the Maximum Variety and Minimum Variance portfolios and prove its superiority to produce higher performances compared to AHC for both EU and US universes.

This paper has focused on both the Maximum Variety and Minimum Variance portfolios but can be applied on other allocation framework involving covariance matrix estimation (and/or model order selection). Finally, the main factors identified by the whitening process can also be used and offer many possible avenues for future research, such as creating dynamic factor portfolios or reducing the dimension of the covariance matrix when $T < m$.

Chapter 3

Frequency causality measures and VAR models: an improved subset selection method suited to parsimonious systems

This chapter is based on two conferences and one submitted article in collaboration with Christophe Chorro, Emmanuelle Jay and Philippe De Peretti.

VAR Estimation Impacts on Frequency Causality Measures,
C. Chorro, E. Jay, P. De Peretti, T. Soler,
2nd International Conference on Econometrics and Statistics (EcoSta), June 2018, Hong-Kong

VAR Estimation Impacts on Frequency Causality Measures,
C. Chorro, E. Jay, P. De Peretti, T. Soler,
12th International Conference on Computational and Financial Econometrics, December 2018, Pisa, Italy

Abstract

Finding causal relationships in large dimensional systems is of key importance in a number of fields. Granger non-causality tests have become standard tools, but they only detect the direction of the causality, not its strength. To overcome this point, in the frequency domain, several measures have been introduced such as the Direct Transfer Function (DTF), the Partial Directed Coherence measure (PDC) or the Generalized Partial Directed Coherence measure (GPDC). Since these measures are based on a two-step estimation, consisting in i) estimating a Vector AutoRegressive (VAR) in the time domain and ii) using the VAR coefficients to compute measures in the frequency domain, they may suffer from cascading errors. Indeed, a flawed VAR estimation will translate into large biases in coherence measures. Our goal in this paper is twofold. First, using Monte Carlo simulations, we quantify these biases. We show that the two-step procedure results in highly inaccurate coherence measures, mostly due to the fact that non-significant coefficients are kept, especially in parsimonious systems. Based on this idea, we next propose a new methodology (mBTS-TD) based on VAR reduction procedures, combining the modified-Backward-in-Time selection method (mBTS) and the Top-Down strategy (TD). We show that our mBTS-TD method outperforms the classical two-step procedure. At last, we apply our new approach to recover the topology of a weighted financial network in order to identify through the local directed weighted clustering coefficient the most systemic assets and exclude them from the investment universe before allocating the portfolio to improve the return/risk ratio.

3.1 Introduction

Vector AutoRegressive (VAR) models are popular models used for analyzing multivariate time series. They have been widely used in many fields, such as macroeconomics [132], finance [146], or even neuroscience [206]. Their use comes from their simplicity and straightforward theoretical framework for understanding the dynamical structure of systems, capturing complex temporal relationships among time series. In dynamical systems admitting a VAR representation, it is often of interest to capture and quantify complex internal dynamics. These complex interactions can be estimated by Granger non-causality tests [77]. Given an information set, non-causality tests check whether or not adding past values of univariate or multivariate series significantly reduces the forecast error variance. Nevertheless, if causality is detected (rejection of the non-causality hypothesis), Granger's tests do not give any information about the strength of this causality. They only assess the existence and direction of causal relationships. But in many applications such as weighted graphs or networks, the quantification of causal strength is of great importance. To deal with this point, the concept of Granger causality has been extended in the frequency domain by considering several indicators measuring the causal strength or coupling similarities. The most popular measures are the Direct Transfer Function measure (DTF) [174, 175], the Partial Directed Coherence measure (PDC) [176, 177], and the Generalized Partial Directed Coherence measure (GPDC) [178]. Such indicators are computed using a two-step approach: first estimate a VAR model of lag p, then switching to the frequency domain using Fourier transform of the estimated VAR coefficients to compute the indicator of interest. In such an approach, it is obvious that a flawed estimation of the VAR model will translate into inaccurate measures of the DTF, PDC, or GPDC. We will refer to this aspect as cascading errors.

A common way to estimate a VAR is to rely on a suitable estimation method, and then to use information criteria such as the Akaike Information Criterion (AIC) [142, 143, 207] or the Bayesian Information Criterion (BIC) [145] to select the correct lag. This procedure raises two issues: the first concerns the ability of the information criteria to correctly approximate the true lag p of the underlying data generating process [146, 141], especially for small samples. The second one refers to the significance of individual VAR parameters. By construction, in VAR models, time series depend on all lagged variables in the system. This assumption is very strong and unrealistic in most applications. Indeed, most systems will admit parsimonious structures with only a few significant coefficients. Said differently, in a multivariate system, it is unusual for all time series to be mutually dependent at each lag. Thus, if information criteria are used solely to estimate p, then when computing coherence measures, non-significant coefficients are likely to be used, thus biasing the causality measures.

Our contribution to the literature on coherence measures is twofold. First, we estimate the impact of cascading errors on the accuracy of computed coherence measures within a standard VAR estimation. We implement Monte Carlo simulations using a system with five time series, with $p = 3$. This system has been analyzed with PDC and DTF methods in [177, 208, 209]. It admits 12% non-zero coefficients with only one on the third lag. This toy model is quite stringent, and likely to be found in real systems such as macroeconomics [189] and finance [188, 210, 211]. Through simulations, we also compute cascading errors for several sample sizes and residual correlation

matrices (non-diagonal matrices) to present a realistic framework. We show that standard VAR estimation leads to highly biased coherence measures, since non-significant coefficients appear in the coherence measures, as mentioned above. Therefore, it is straightforward to investigate if more advanced subset selection method can solve the issue raised. Accordingly, our major second contribution checks whether or not more advanced VAR model selection methods leading to more parsimonious representations could lower or even suppress the cascading errors. In the literature, three types of procedures may be considered. The first type reduces the number of VAR coefficients by adding/deleting parameters using information criteria. These include Top-Down (TD), Bottom-Up (BU) [141], or a more recent method based on the BU approach called modified Backward-in-Time Selection (mBTS) [188, 151]. The second type of procedures is based on hypothesis testing, either at an individual level, i.e. on each coefficient taken separately such as t-test (hereafter TT) or for a group of coefficients such as likelihood ratio and Wald tests [141, 149, 150]. Finally, the third procedure relies on shrinkage methods, such as Lasso (Least Absolute Shrinkage and Selection Operator) [186, 189]. In this latter, selection of variables and VAR coefficients estimation are conducted simultaneously. We evaluate each approach and propose a new and extended method by combining the mBTS method and the TD strategy (mBTS-TD). We show, in the framework of coherence measures, that our approach outperforms the competing ones.

Based on this result, we apply our methodology combining mBTS-TD and GPDC measure to build financial networks. This application emphasizes the advantages of our methodology. Firstly, the mBTS-TD method inherently provides a parsimonious structure without using network dimensionality reduction tools such as a significant threshold, the Minimum Spanning Tree [69] or Planar Maximally Filtered Graphs [128]. Secondly, the strength of the causal relationship allows us to recover the network topology in a more precise way. In such a financial network, we propose a new portfolio allocation method by excluding the most systemic assets, identified using the local directed weighted clustering coefficient [127]. We therefore obtain financial portfolio as performing as possible, where the non-systemic assets are equally allocated [212]. Related performance measures are compared with those obtained using a classical VAR estimation to compute the GPDC or allocating the whole universe.

This paper is set out as follows: in section 3.2, we introduce the econometric methodology (VAR model and Granger non-causality tests) and coherence measures; in section 3.3, we estimate the cascading errors and show that the use of standard VAR estimation leads to large errors in the coherence measures; in section 3.4, we begin by introducing advanced VAR estimation procedures and ascertaining their efficiency when used to compute coherence measures; in section 3.5, we apply our methodology on financial time series and finally, section 3.6 concludes and discusses our results.

3.2 Econometric methodology

In this section, we first introduce Vector AutoRegressive models (VAR) and model order identification. Then we present the concept of Granger causality and coherence measures, paying particular attention to Generalized Partial Directed Coherence (GPDC). Furthermore, we discuss the reasons why focusing on coherence measures rather than Granger

causality can improve information about causal strength. Finally, we address the accuracy of the coherence measure, since this may present cascading errors due to its first step in the VAR estimation.

3.2.1 Non-causality test in Vector AutoRegressive Models

Let $\mathbf{x}(t) = (x_1(t), \ldots, x_m(t))'$ be a zero-mean m-dimensional stationary process admitting the following VAR(p) representation (see [141] section 2 for classical stability and stationarity conditions):

$$\mathbf{x}(t) = \mathbf{A}_1\mathbf{x}(t-1) + \cdots + \mathbf{A}_p\mathbf{x}(t-p) + \boldsymbol{\epsilon}(t), \quad t \in \mathbb{Z} \tag{3.1}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $(m \times m)$ coefficient matrices, p is the model order, and $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \ldots, \epsilon_m(t))'$ is a $(m \times 1)$ vector of white noises with $\mathrm{E}[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}'(s)] = 0$ for $t \neq s$ and $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$.

The coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ describe the temporal relationships within the m time series in the system. The concept of causality is therefore directly related to these coefficients. These coefficient matrices also play a fundamental role when making forecasts. The structure of $\boldsymbol{\Sigma}_\epsilon$ reveals the contemporaneous or instantaneous effects between the time series.

In this paper, the VAR coefficients are estimated using the Least Squares estimator (LS), either in a multivariate (LS) or in univariate (OLS) environment (equation by equation) [141]. In addition to estimating the VAR coefficients, the model order p must also be estimated. This step is crucial for the accuracy of the VAR estimate. The lag order p is chosen to minimize an information criterion such as AIC [142, 143] or BIC [145]. In this paper, we choose AIC and BIC by using the two estimators of $\boldsymbol{\Sigma}_\epsilon(p)$ to investigate the role of the penalty factor in model order selection. For VAR(p) in (3.1), AIC and BIC are defined as follows:

$$\mathrm{AIC}(p) = \ln\left(\det\left(\widehat{\boldsymbol{\Sigma}}_\epsilon(p)\right)\right) + \frac{2}{T}pm^2$$

$$\mathrm{BIC}(p) = \ln\left(\det\left(\widehat{\boldsymbol{\Sigma}}_\epsilon(p)\right)\right) + \frac{\ln T}{T}pm^2$$

$$\mathrm{AIC}_{un}(p) = \ln\left(\det\left(\widetilde{\boldsymbol{\Sigma}}_\epsilon(p)\right)\right) + \frac{2}{T}pm^2$$

$$\mathrm{BIC}_{un}(p) = \ln\left(\det\left(\widetilde{\boldsymbol{\Sigma}}_\epsilon(p)\right)\right) + \frac{\ln T}{T}pm^2$$

where T is the number of observations, $\widetilde{\boldsymbol{\Sigma}}_\epsilon(p)$ and $\widehat{\boldsymbol{\Sigma}}_\epsilon(p)$ are the unbiased and the maximum likelihood estimates of $\boldsymbol{\Sigma}_\epsilon(p)$ for the VAR(p) in (3.1), given by

$$\widehat{\boldsymbol{\Sigma}}_\epsilon(p) = \frac{\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}'(t)}{T} \tag{3.2}$$

$$\widetilde{\boldsymbol{\Sigma}}_\epsilon(p) = \frac{\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}'(t)}{T - pm - 1} \tag{3.3}$$

The main difference between AIC and BIC is the increase in the BIC penalty factor compared to AIC.

After estimating the VAR, the most common way to assess complex interactions is to use Granger non-causality tests. The concept of non-causality defined by Granger [77] is based on the idea that, if a time series $x_k(t)$ causes another time series $x_j(t)$, then the past of $x_k(t)$ will significantly decrease the forecast error in $x_j(t)$. Let $\mathbf{x}(t)$ be a m-dimensional stationary process admitting the VAR(p) representation defined in (3.1). Granger non-causality can be tested by using a Wald multiple restrictions test [141] on the VAR coefficients. This test, jointly tests whether a set of coefficients are non-significant. For example, if a time series $x_k(t)$ does not *Granger-cause* $x_j(t)$, then $a_{kj}(1) = \ldots = a_{kj}(p) = 0$, where $a_{kj}(1), \ldots, a_{kj}(p)$ are the elements k, j of the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$. The general null hypothesis is given by $\mathrm{H}_0 : \boldsymbol{C\beta} = \boldsymbol{c}$, where $\boldsymbol{C}$ is a $(q \times m^2 p)$ matrix called the restriction matrix of the VAR coefficients (1 for tested coefficients and 0 otherwise). Moreover, q denotes the number of restrictions, $\boldsymbol{\beta}$ is a $(m^2 p \times 1)$ vector with $\boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$, and $\boldsymbol{c}$ is a $(q \times 1)$ vector with $\boldsymbol{c} = \mathbf{0}_q$ for Granger non-causality.

The Wald statistic is therefore

$$\Gamma = (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})' \left[\boldsymbol{C} \left((\boldsymbol{X}\boldsymbol{X}')^{-1} \otimes \widetilde{\boldsymbol{\Sigma}}_\epsilon \right) \boldsymbol{C}' \right]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) \tag{3.4}$$

where

- $\widehat{\boldsymbol{\beta}}$ is the $(m^2 p \times 1)$ vector of the estimated VAR coefficients $\boldsymbol{\beta}$,
- $\boldsymbol{X}$ is the $(mp \times T)$ matrix with $\boldsymbol{X}(t) = \mathrm{vec}(\mathbf{x}(t-1), \ldots, \mathbf{x}(t-p))$,
- $\widetilde{\boldsymbol{\Sigma}}_\epsilon$ is the unbiased estimator of $\boldsymbol{\Sigma}_\epsilon(p)$ defined in (3.3).

Under the null hypothesis H_0: $\Gamma \sim \chi^2(q)$. This result is valid only asymptotically and for the VAR model assumptions defined in (3.1) with $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \sigma^2 \boldsymbol{I}_m$. H_0 is not rejected (non-causality) for a given probability α if $\Gamma \leq \chi^2_\alpha(q)$, where $\chi^2_\alpha(q)$ is the quantile of the distribution. Nevertheless, if the null hypothesis H_0 is rejected, it means that a time series $x_k(t)$ *Granger-causes* another time series $x_j(t)$. This test does not provide any information about the causal strength. To deal with this aspect, several measures called coherence measures have been proposed in the frequency domain.

3.2.2 Coherence measures

Coherence measures describe the connectivity between times series in the frequency domain and are often used in the neurosciences to understand functional connectivity patterns between different brain regions. The most popular coherence measures are the Directed Coherence measure (DC) [173], the Partial Directed Coherence measure (PDC) [176, 177], the Direct Transfer Function measure (DTF) [174, 175], and the Generalized Partial Directed Coherence measure (GPDC) [178]. These measures are based on a two-step approach: first the VAR coefficients are estimated, and then the measure is computed using the transfer function matrix or its inverse matrix on the VAR coefficients. The DC, introduced by Saito and Harashima [173] for bivariate cases, describes whether and how two time series are functionally connected. The three other measures PDC, GPDC, and DTF can be applied to multivariate cases. The PDC introduced by Baccalá and Sameshima [176, 177] provides a frequency domain representation of Granger Causality. It is a generalization to the multivariate case of the DC, based on the Partial Coherence that describes the mutual interaction between two time series when the

effects of all others have been subtracted. In other words, it quantifies only the direct connections between time series. Baccalá and Sameshima [178] have extended their measure, called GPDC, by taking into account the variance of white noise, so that it is more accurate with finite time series samples and leads to a scale invariant measure. Finally, the DTF was introduced by Kamiński and Blinowska [174, 175]. It describes the causal influence, but it does not distinguish direct and indirect relationships, whereas PDC and GPDC provide the multivariate relationships from a partial perspective. In this paper we focus on PDC and GPDC measures because in a multivariate environment the distinction between indirect and direct relationships is of great importance.

For two time series $x_j(t)$ and $x_k(t)$ the GPDC is defined so as to exhibit the causality from k to j at each frequency f as follows:

$$\omega_{jk}(f) = \frac{\frac{1}{\sigma_{jj}}\tilde{a}_{jk}(f)}{\sqrt{\sum_{n=1}^{m}\frac{1}{\sigma_{nn}^2}\tilde{a}_{nk}(f)\tilde{a}_{nk}^*(f)}} \tag{3.5}$$

where

- f are the discrete frequencies[1] lying in $\left[-\frac{1}{2};\frac{1}{2}\right]$,
- $\tilde{a}_{jk}(f)$ is the discrete Fourier transform of the coefficients $a_{jk}(1),\ldots,a_{jk}(p)$ defined by

$$\tilde{a}_{jk}(f) = \begin{cases} 1 - \sum_{l=1}^{p} a_{jk}(l)e^{-2i\pi fl}, & \text{if } j = k \\ -\sum_{l=1}^{p} a_{jk}(l)e^{-2i\pi fl}, & \text{otherwise} \end{cases}$$

- σ_{jj}^2 is the j-th element of the diagonal of $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$.

The GPDC $\omega_{jk}(f)$ at frequency f represents the relative strength of interaction with respect to a given signal source. Note that the GPDC is represented as a power spectral density, i.e., $|\omega_{jk}(f)|^2$. Moreover, given that the VAR coefficients $a_{jk}(l)$ are real numbers, their Discrete Transform Fourier has Hermitian symmetry $\tilde{a}_{jk}(f) = \tilde{a}_{jk}(-f)^*$. The spectrum is symmetric at the frequency $f = 0$, i.e., $|\tilde{a}_{jk}(f)| = |\tilde{a}_{jk}(-f)|$, so it is possible to represent only a half-period of the spectrum ($f \in \left[0;\frac{1}{2}\right]$). Finally, the GPDC satisfies the following properties :

$$0 \leqslant |\omega_{jk}(f)|^2 \leqslant 1 \tag{3.6}$$

[1]For a discrete time series sampled at frequency f_e, its Fourier Transform will reveal information for frequencies lying in $\left[-\frac{f_e}{2};\frac{f_e}{2}\right]$. In our case $f_e = 1$, we can therefore choose the interval $\left[-\frac{1}{2};\frac{1}{2}\right]$ with a step of $\frac{1}{F-1}$, where F is the number of frequencies.

$$\sum_{n=1}^{m} |\omega_{nk}(f)|^2 = 1, \forall k = 1, \ldots, m. \tag{3.7}$$

In such an approach, it is obvious that a flawed estimation of the VAR model will translate into inaccurate measures of GPDC, but it is also true for all coherence measures that directly use VAR coefficients in the transfer function matrix or its inverse.

The VAR estimation may be incorrectly performed for many well-known reasons, such as incorrect model order selection, small sample size, or correlated residuals. Moreover, an obvious error that is often omitted when computing coherence measures and also generates cascading errors is the estimation of zero coefficients due to VAR estimation in a multivariate environment. In other words, in a multivariate system it is unusual for all time series to be mutually dependent (parsimonious model). Some coefficients are therefore equal to zero, and the estimation of these zero coefficients (non-causal terms) inevitably biases the non-zero ones (causal terms). This is particularly true for PDC and GPDC, which compute only direct causality, and due to their normalization property (3.7), the errors made in the non-causal terms have a direct impact on the accuracy of the causal terms. Thus, in the presence of parsimonious VAR models and by combining the well-known estimation errors, high cascading errors and spurious causalities become likely.

3.3 Impacts of standard VAR estimation on GPDC

In this section, we use Monte Carlo simulations to illustrate the way estimation of VAR model in a multivariate environment can impact GPDC accuracy when the underlying system is parsimonious. The simulations are conducted by varying both the sample size and the residual correlation matrix in order to cover more realistic examples, but also to highlight standard VAR estimation errors on the GPDC. Firstly, the system and settings are presented. Then we focus on the errors in the VAR coefficients, and finally on the GPDC errors by separating the causal and non-causal terms to determine the part that will be most impacted by cascading errors.

3.3.1 System and error measures

In the simulation study, we analyze five time series generated by the VAR(3) model used in [177, 208, 209]. This system admits a parsimonious structure with 12% of the m^2p coefficients being non-zero. It has only one coefficient on the third lag. Hence, this system is in principle not suitable for a standard VAR estimation where the m^2p coefficients are estimated.

The VAR model [177, 208, 209] is as follows:

$$\begin{aligned}
x_1(t) &= 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + \epsilon_1(t) \\
x_2(t) &= 0.5x_1(t-2) + \epsilon_2(t) \\
x_{3(t)} &= -0.4x_1(t-3) + \epsilon_3(t) \\
x_4(t) &= -0.5x_1(t-2) + 0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + \epsilon_4(t) \\
x_5(t) &= -0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + \epsilon_5(t)
\end{aligned} \tag{S}$$

The causal structure of S is shown in Fig. 3.1 and the theoretical GPDC in Fig. 3.2.

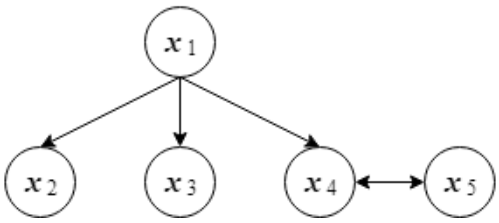


Figure 3.1: Causal structure of (S). In this system, the time series $\boldsymbol{x}_1$ causes $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4$, while $\boldsymbol{x}_4$ and $\boldsymbol{x}_5$ are causing each other.

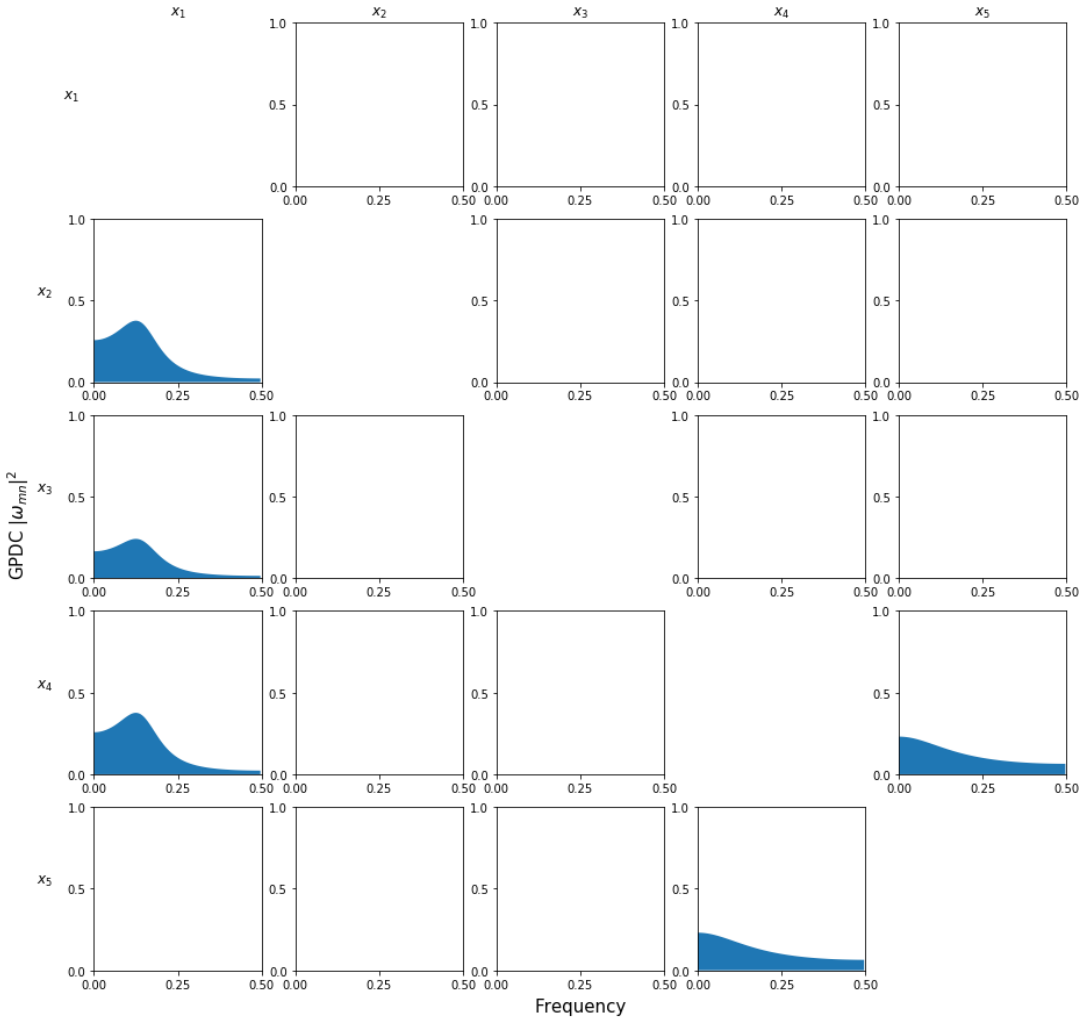


Figure 3.2: Theoretical GPDC of (S) (column k causes row j). This is computed using the true coefficients and the identity matrix for the residual correlation matrix ($\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \boldsymbol{I}_5$). Interpreting the GPDC: $\boldsymbol{x}_1$ causes $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4$. In contrast, $\boldsymbol{x}_1$ causes $\boldsymbol{x}_5$ indirectly via $\boldsymbol{x}_4$, but as GPDC quantifies only direct interactions, the causality values for all frequencies are equal to zero. If $j = k$, the GPDC represents the part that is not explained by other signals. Since it is quite difficult to interpret, the diagonal is not reported here.

The impacts of the different VAR models are evaluated through 1000 simulations of (S). For each simulation, (S) is generated by using a sample size T and a multivariate Gaussian distribution for the white noise with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$. The initial values used to generate (S) are set to zero. For T and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, we use the following settings:

- Four sample sizes: $T = \{128, 256, 512, 1024\}$
- Four $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ matrices, the identity and three symmetric Toeplitz matrices with $\rho \in \{0.25, 0.50, 0.75\}$. A $(m \times m)$ symmetric Toeplitz matrix $(\mathbf{Tp})$ has the form $(\mathbf{Tp})_{jk} = \rho^{|j-k|}$ for $j, k = 1, \ldots, m$, with $\rho \in \mathbb{R} : |\rho| < 1$. The four matrices are as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.25 & 0.06 & 0.02 & 0 \\ 0.25 & 1 & 0.25 & 0.06 & 0.02 \\ 0.06 & 0.25 & 1 & 0.25 & 0.06 \\ 0.02 & 0.06 & 0.25 & 1 & 0.25 \\ 0 & 0.02 & 0.06 & 0.25 & 1 \end{pmatrix}$$

(1) Id (2) Tp_1

$$\begin{pmatrix} 1 & 0.50 & 0.25 & 0.13 & 0.06 \\ 0.50 & 1 & 0.50 & 0.25 & 0.13 \\ 0.25 & 0.50 & 1 & 0.50 & 0.25 \\ 0.13 & 0.25 & 0.50 & 1 & 0.50 \\ 0.06 & 0.13 & 0.25 & 0.50 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.75 & 0.56 & 0.42 & 0.32 \\ 0.75 & 1 & 0.75 & 0.56 & 0.42 \\ 0.56 & 0.75 & 1 & 0.75 & 0.56 \\ 0.42 & 0.56 & 0.75 & 1 & 0.75 \\ 0.32 & 0.42 & 0.56 & 0.75 & 1 \end{pmatrix}$$

(3) Tp_2 (4) Tp_3

The increase in ρ indicates that the error processes are more strongly correlated, and allows us to see the estimation impacts when the system approaches a structural VAR model. To build the covariance matrices, we use the symmetric Toeplitz matrix [213, 210] because it provides a flexible framework to generate a positive-definite covariance matrix. The symmetric Toeplitz matrix structure used depends on just one parameter ρ. Moreover, it reflects the stationarity of auto-regressive systems (if $|\rho| < 1$).

To compare and quantify the cascading errors of the different VAR models, the relative L_2-norm error is used for the VAR coefficients, whereas the L_2-norm error is computed for GPDC. The L_2-norm error is used instead of the relative one because, for non-causal terms, the theoretical GPDC is null for all frequencies, resulting in a null denominator for the relative error. The two error metrics are defined as follows:

- Relative L_2-norm error of VAR coefficients: $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_2 / \|\boldsymbol{A}\|_2$ where $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$
- L_2-norm error of GPDC: $\||\widehat{\boldsymbol{\omega}}_{jk}|^2 - |\boldsymbol{\omega}_{jk}|^2\|_2$ on each pair $j \neq k$, where $\boldsymbol{\omega}_{jk}$ is the vector containing each value of $\omega_{jk}(f)$ for all discrete frequencies f.

Finally, to assess the errors through the 1000 simulations, we report for the VAR coefficients the median of the relative L_2-norm error and for the GPDC the sum of the median of the L_2-norm error on both the causal terms and the non-causal terms.

3.3.2 Estimation errors

Here, we evaluate the impact of standard VAR estimations through 1000 Monte Carlo simulations of (S) on both the VAR coefficient and the GPDC errors. Fig. 3.4 shows the first results for the simulated data, focusing on the estimation errors of the VAR coefficients. For this purpose, five VAR models are estimated in a multivariate environment, either by setting the order of the model at $p = 3$ (true model order), or by determining it using the four information criteria defined in section 3.2. Note that the Tp_1 results are not reported because they are quite similar to the identity matrix, but they are available upon request.

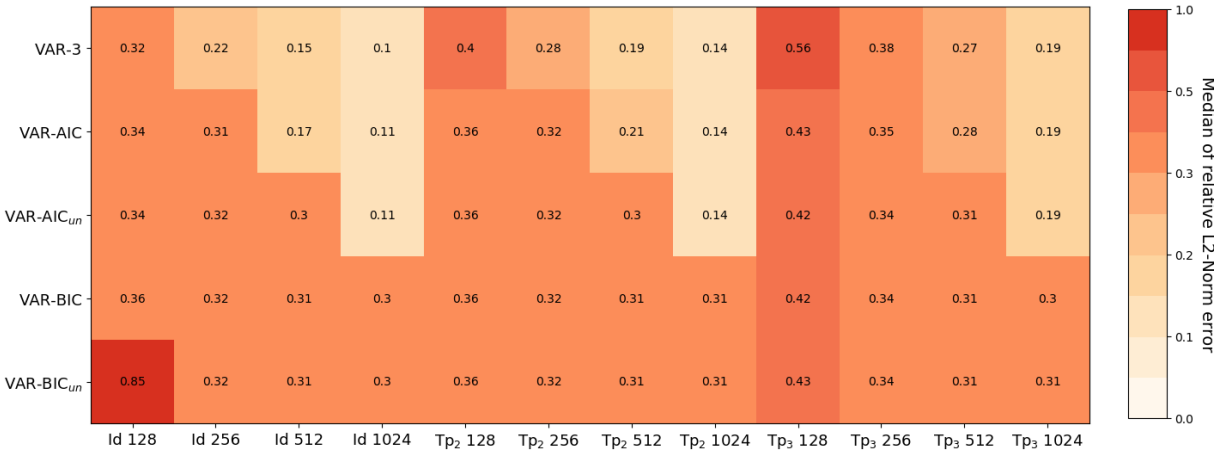


Figure 3.4: Median over 1000 simulations of relative L_2-norm error on the coefficients estimated using the five standard VAR estimations. The increase in the median indicates a deterioration in the estimation of coefficients.

On the basis of these first empirical results presented in Fig. 3.4, the following conclusions can be drawn:

- The accuracy of the estimated coefficients improves significantly when the sample size increases, but this improvement is less prominent when using BIC criteria. For the VAR-3 or the two VAR-AIC, the errors are almost reduced by a factor of three between $T = 128$ and $T = 1024$. Moreover, the performance for all models deteriorates when the residuals are strongly correlated.

- In VAR estimation the selection of the model order is crucial for accuracy. The four information criteria rarely find the true model order for small sample sizes ($T = 128, 256$) and correlated residuals (Tp_2 and Tp_3). For the smallest sample sizes, they select the true lag order $p = 3$ in 7% of cases and otherwise they have a lag order $p = 2$, except for BIC_{un}, which finds a model order $p = 1$ for Id 128 in 70% of simulations. Thus, in these cases, VAR models using information criteria must have higher errors than VAR-3. Nevertheless, for three cases Tp_2 128, Tp_3 128, and Tp_3 256 it is in fact VAR-3 that has the worst results. Indeed, as (S) has only one non-zero coefficient on the third lag, the VAR-3 estimates twenty-four zero coefficients, which significantly increases the error. For a parsimonious system like (S), the true model order can become an over-fitted model, especially for small sample sizes. However, the model order must not be too seriously undervalued, as happens with VAR-BIC_{un} for Id128, otherwise a compensation effect will occur in

the values of the coefficients, overestimating them by significantly increasing the errors.

- On this system, the AIC criterion provides better results for the LS estimation, because it is the least restricted and so can find the true model order for large sample sizes ($T = 512, 1024$). However, it can also under-fit the model to yield lower errors than VAR-3 for small sample sizes and correlated residuals.

The following two Figs. 3.5 and 3.6 show the GPDC errors on causal (Fig. 3.5) and non-causal terms (Fig. 3.6), using the same models and settings as before.

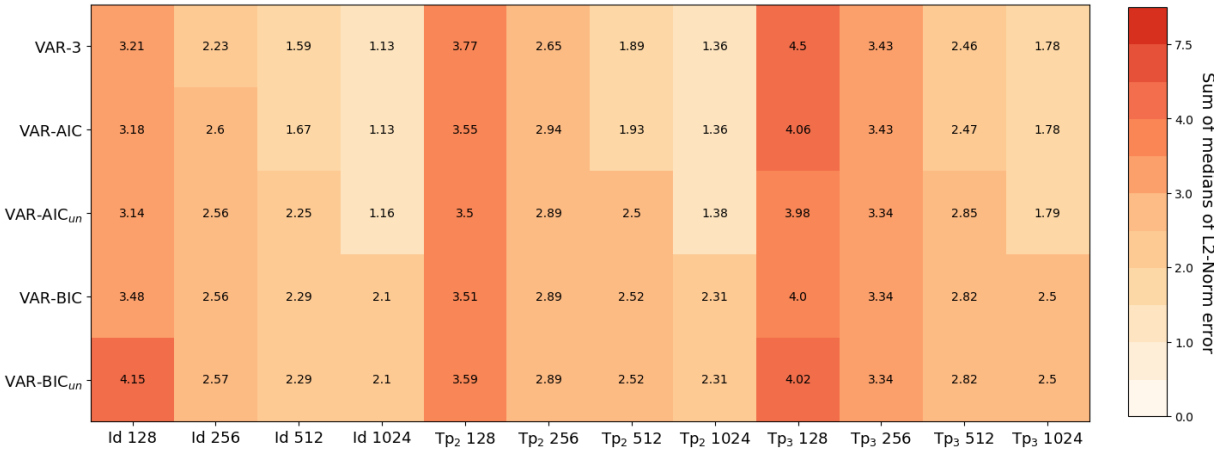


Figure 3.5: Sum of medians over 1000 simulations of L_2-norm error on the causal GPDC (3.5) estimated using the five standard VAR estimations. The increase in the median indicates a deterioration in the estimation of the causal GPDC.

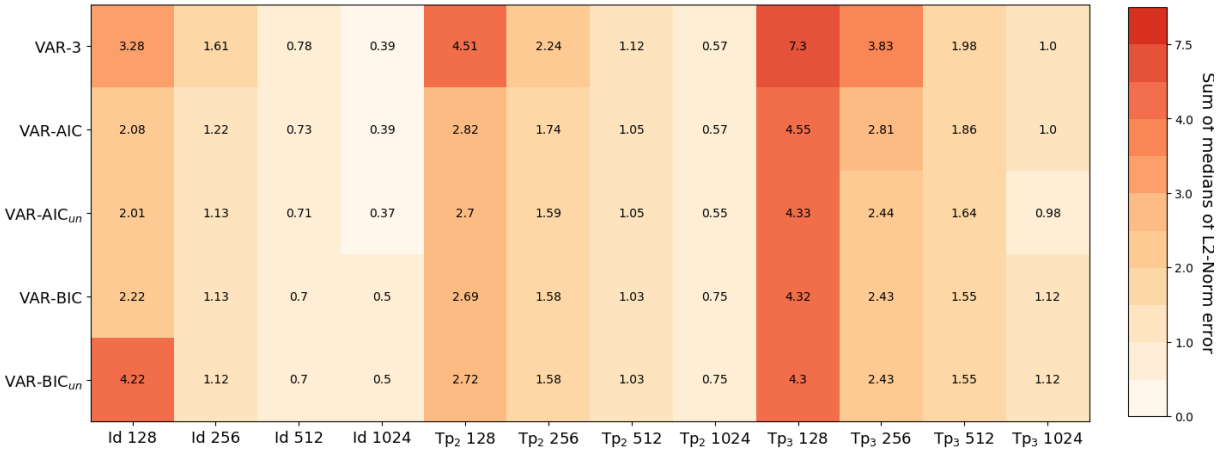


Figure 3.6: Sum of medians over 1000 simulations of L_2-norm error on the non-causal GPDC (3.5) estimated using the five standard VAR estimations. The increase in the median indicates a deterioration in the estimation of the non-causal GPDC.

For causal terms (Fig. 3.5), the same conclusions can be drawn as for the errors in the coefficients, whereas for non-causal terms, the results are quite different (Fig. 3.6). For non-causal terms (zero coefficients), VAR-3 presents the worst errors for all settings, except for $T = 1024$, where errors are similar to the other VAR models. These errors in VAR-3 support the idea of an over-fitted model because non-causal terms are

directly related to the estimation of zero coefficients. In contrast, VAR models using the most restricted information criteria (AIC_{un}, BIC, and BIC_{un}) become the best models for all cases, due to the non-estimation of the third lag. Nevertheless, if the model is under-fitted (VAR-BIC_{un}), as noted with the errors in the coefficients, the errors will also be very high (4.22 for Id128) due to compensation effects in the coefficients. We also note that, in the case where the VAR model contains many zero coefficients (88% of zero coefficients), the errors in non-causal terms can be higher than those in the causal terms. Moreover, due to the normalization property defined in (3.7), the errors in the non-causal terms have a direct impact on the accuracy of the causal strength. Thus, in the presence of parsimonious VAR models, the standard VAR estimation with all possible information criteria is in fact not well suited.

In this section, we have confirmed that standard VAR estimation provides high cascading errors and reveals spurious causalities. Indeed, a VAR estimation in a multivariate environment does not take into account the parsimonious structure of the underlying model, and better suited methods are needed. Thus, in order to correct the cascading errors in the GPDC and to adopt a more parsimonious structure, we use searching procedures with parameter constraints, or shrinkage methods, called subset selection methods.

3.4 Improving GPDC estimation accuracy

Standard VAR estimation is therefore not well-suited to estimate parsimonious data generating processes, since such processes may have only a few non-zero coefficients. The significance of individual coefficients must then be assessed prior to computing GPDC functions. To deal with this point, it is possible to use subset selection methods. In the literature, at least three procedures have been considered: first, procedures based on information criteria to add or delete coefficients such as Bottom-Up strategy, Top-Down strategy [141] and modified Backward-in-Time Selection [151]; second, procedures based on hypothesis testing, such as t-test for individual coefficients and multivariate tests [141, 149, 150], i.e. that jointly test the coefficients (e.g. likelihood ratio and Wald test); finally, procedures based on shrinkage methods such as Lasso Regression [186]. However, for the second family, the multivariate approach aims at testing the non-significance of a set of coefficients, and is therefore not of particular interest in our setting (such a drawback is also found in the classical Bottom-Up strategy [141]) explaining why we only focus on the individual t-test (TT) in our study. The results are assessed through VAR coefficients errors and GPDC accuracy as in the previous section by comparing our mBTS-TD method to the mBTS method, the TD strategy, the TT procedure (described in 1.4.2.2), and the Lasso method (described in 1.4.2.2). Finally, we check error distributions and the identification of the true causal structure of (S).

3.4.1 Proposed method: mBTS-TD

The proposed method first uses the modified Backward-in-Time Selection (mBTS) to estimate the VAR coefficients. The main advantage of the mBTS method is that only terms that improve the prediction of the equation are included, and this allowing us to work with high-dimensional systems like $K = 20$ in [151]. As already shown in [184], the mBTS method dramatically improves GPDC accuracy. Nevertheless, a drawback with the mBTS method is that the maximum lag p_{max} is fixed *a priori*. If it is too

small, the internal dynamics of the system are not completely modeled and the model is under-fitted. In this case, the coefficients are over-estimated due to the compensation effects causing possibly large errors as remarked for example in the last line of Fig. 3.4. On the other hand, if p_{max} is too large, undesirable lagged variables may appear in the model by revealing spurious causalities. We therefore propose to combine the mBTS method with the Top-Down strategy (TD). Indeed, the advantage of the TD strategy is that all coefficients in each equation are tested, but it is very sensitive to the initial VAR estimation which determines the model order and thus can amplify the errors if the initial VAR estimation is not carried out properly. The two reasons to combine the mBTS method and TD strategy are: firstly, to be less dependent on the choice of p_{max} so that we may set its values high enough to capture all possible connections; and secondly, as a further consequence, so that we may produce a more parsimonious model when p_{max} is set at a high value. Hereafter, we define the mBTS method and the TD strategy.

The modified Backward-in-Time Selection (mBTS) is a Bottom-Up strategy (BU) introduced by Vlachos and Kugiumtzis [151]. It is based on Dynamic Regression models [188], which estimate each equation separately. Unlike the TD strategy, the mBTS method adds progressively lagged variables, starting from the first lag for all variables, and moving backward in time.

First, a maximum order p_{max} is fixed, and this provides the vector $(1 \times mp_{max})$ of all lagged variables for the j-th equation of the VAR(p) model in (3.1):

$$\boldsymbol{v} = (x_1(t-1), \ldots, x_1(t-p_{max}), \ldots, x_m(t-1), \ldots, x_m(t-p_{max}))$$

An explanatory vector $\boldsymbol{\vartheta}$ is built from $\boldsymbol{v}$ by progressively adding only the most significant lagged variable at each step.

For the j-th equation of the VAR(p) model in (3.1), the mBTS algorithm is as follows:

1. Start with an empty vector $\boldsymbol{\vartheta} = \emptyset$, the information criterion IC^{old} initialized to the variance of the j-th series, and $\boldsymbol{\tau} = (1, \ldots, 1)'$ the $(m \times 1)$ lag order vector of the variables.

2. Compute IC_n^{new} relative to the m dynamic regression models formed by the m candidate explanatory vectors $\boldsymbol{\vartheta}_n^{\text{cand}}$, where $\boldsymbol{\vartheta}_n^{\text{cand}} = (\boldsymbol{\vartheta}, x_n(t-\tau_n))$, $\forall n \in \{1, \ldots, m\}$.

3. Select the variable according to the IC value:

   - If $\min\{IC^{\text{old}}, IC_1^{\text{new}}, \ldots, IC_m^{\text{new}}\} = IC^{\text{old}}$, then $\boldsymbol{\tau} = \boldsymbol{\tau} + \mathbf{1}_m$.
   - If $\min\{IC^{\text{old}}, IC_1^{\text{new}}, \ldots, IC_m^{\text{new}}\} = IC_n^{\text{new}}$, then $IC^{\text{old}} = IC_n^{\text{new}}$, $x_n(t-\tau_n)$ is added to the explanatory vector $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}, x_n(t-\tau_n))$ and only τ_n is increased by one.

4. Repeat steps 2 and 3 until $\boldsymbol{\tau} = (p_{max}, \ldots, p_{max})'$.

Finally, the Top-Down strategy (TD) tests the VAR coefficients separately in the m equations. The goal is to eliminate the non-significant coefficients for each equation by evaluating the information criterion. The order of the tested terms is arbitrary, but as in [141], the largest lag p is tested first for all variables from $x_m(t-p)$ to $x_1(t-p)$, then the lag $p-1$ with the same order of variables, and the process is iterated until $p=1$.

For the j-th equation obtained with the mBTS algorithm, the TD strategy is applied as follows:

1. Start with the vector $\boldsymbol{\vartheta}$ and the information criterion IC^{old} obtained with the mBTS algorithm.
2. Sort the vector $\boldsymbol{\vartheta}$ from the largest to the smallest lag p and for all series from x_m to x_1.
3. Compute IC_n^{new} by deleting the n-th element in the vector $\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}_n^{\text{cand}} = \boldsymbol{\vartheta} \setminus \{\boldsymbol{\vartheta}_n\}$.
4. Delete the variable according to the IC_n^{new} value:
   - If $\min\{IC^{\text{old}}, IC_n^{\text{new}}\} = IC^{\text{old}}$, then $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}$.
   - If $\min\{IC^{\text{old}}, IC_n^{\text{new}}\} = IC_n^{\text{new}}$, then $IC^{\text{old}} = IC_n^{\text{new}}$ and $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_n^{\text{cand}}$.
5. Repeat steps 3 and 4 $\forall(\boldsymbol{\vartheta}_n)_{n\in[1,|\boldsymbol{\vartheta}|]}$, where $|.|$ denotes the cardinality of the vector in this case.

3.4.2 Comparison with standard VAR

In this part, subset selection methods are compared to each other, and also to the standard VAR estimation presented in the previous section. The impacts are evaluated on the accuracy of both the VAR coefficients and the GPDC through 1000 Monte Carlo simulations on (S). Twenty nine different VAR models are estimated. First, the five VAR models presented in section 3.3, and then the six subset methods with the four information criteria. For the mBTS and mBTS-TD methods we set $p_{max} = 6$, but in section 3.4.3 we will check the robustness of these methods by testing the stability with respect to p_{max}. The t-test procedure (TT) is used for both a significance level of 5% and 1% (see in 1.4.2.2). The Lasso tuning parameter is estimated using 5-fold cross-validation (see in 1.4.2.2). Moreover, only the best information criterion is reported for each method, and as previously, we do not report the Tp_1 results. Fig. 3.7 shows the Monte Carlo simulation results for the estimation errors in the VAR coefficients.

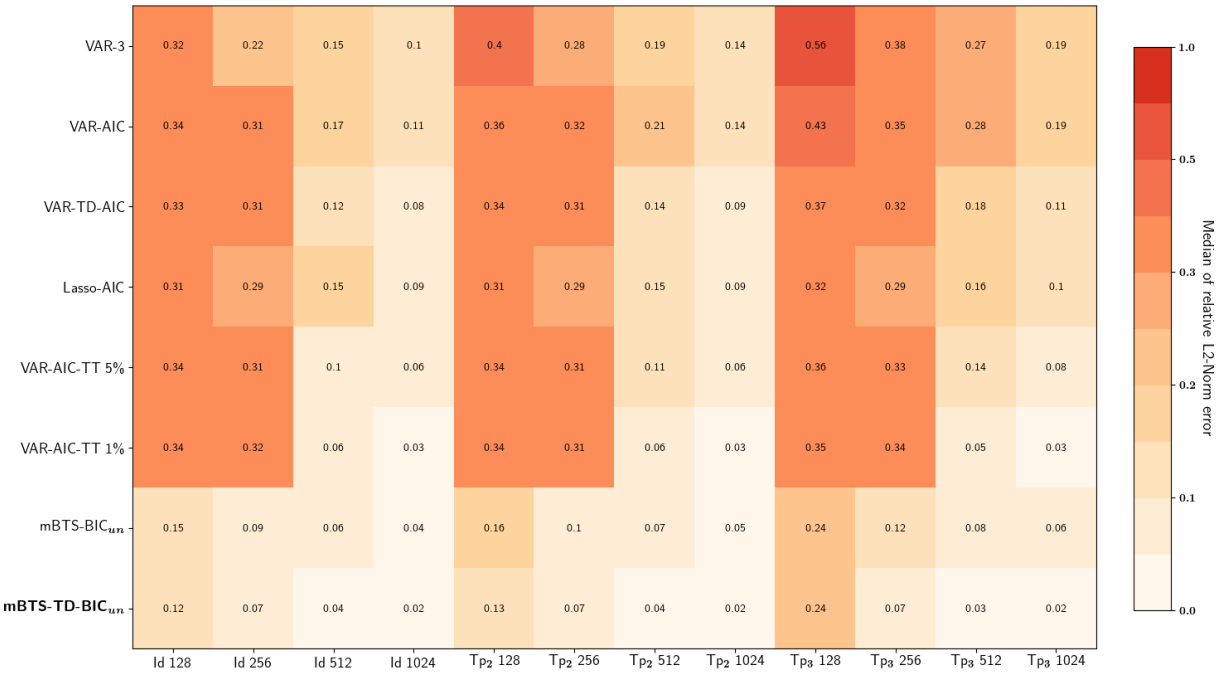


Figure 3.7: Median over 1000 simulations of the relative L_2-norm error on the coefficients estimated using subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the coefficients.

For estimation errors on the VAR coefficients, all subset selection methods provide better results than VAR-AIC for all the settings proposed, and they also behave better than VAR-3 for almost all settings. However, the two methods using mBTS, and in particular the combination of mBTS and TD, clearly stand out from the others. There is a considerable gap between these methods and the others. The errors are at least divided by two for small sample sizes and can be divided by five for the largest ones. Moreover, by adding the TD strategy to a first suitable VAR estimation that respects the parsimonious structure, such as mBTS, the improvement can be significant, especially when residuals are correlated. Nevertheless, TD may also confirm its drawback of being very sensitive to the first estimation (VAR-AIC), because it does not provide better results than VAR-AIC, even if it is more stable in the presence of correlated residuals. Lasso-AIC also provides good results compared to VAR-AIC or VAR-3, but its errors are at least twice as great as with mBTS methods. The TT procedures (especially TT 1%) are better than TD and Lasso for $T \geq 512$, but are quite similar for the two smaller sample sizes with higher errors than the mBTS methods. mBTS-TD can ensure highly stable errors across the different covariance matrices, in particular for sample sizes larger than 256, where the errors are the same. For example, with $T = 256$ the errors are equal to 0.07 whatever the covariance matrix. Finally, to conclude regarding these first results, subset selection methods are better suited to modelling parsimonious structures for the estimation of coefficients. When adding (mBTS) or deleting (TD) parameters as in mBTS-TD, the information criterion with the highest penalty factor, viz., BIC$_{un}$, provides better results than the less restricted ones.

Figs. 3.8 and 3.9 show GPDC errors in causal (Fig. 3.8) and non-causal terms (Fig. 3.9) of (S), confirming previous results for the VAR coefficients.

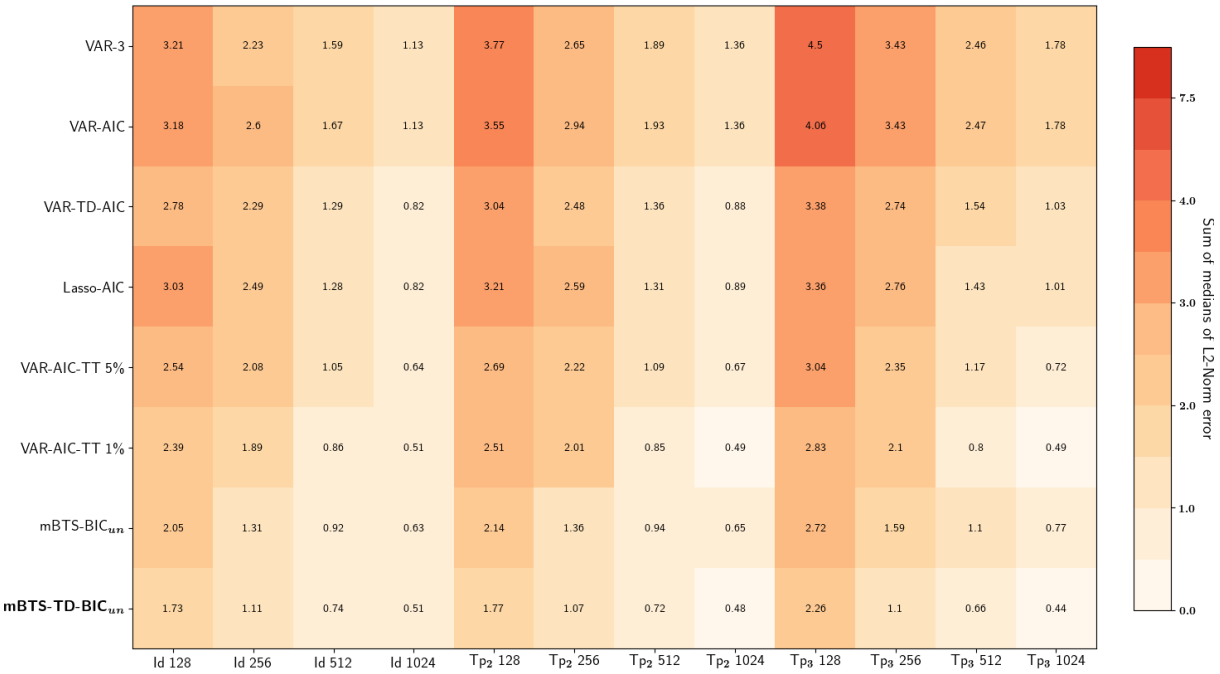


Figure 3.8: Sum of medians over 1000 simulations of the L_2-norm error in the causal GPDC (3.5), estimated using the subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the causal GPDC.

In Fig. 3.8, the results for the GPDC causal terms are similar to the errors in the coefficients. This confirms the idea that subset selection methods are better suited than standard VAR. mBTS-TD performs better than the others for all settings, and also provides stable errors.

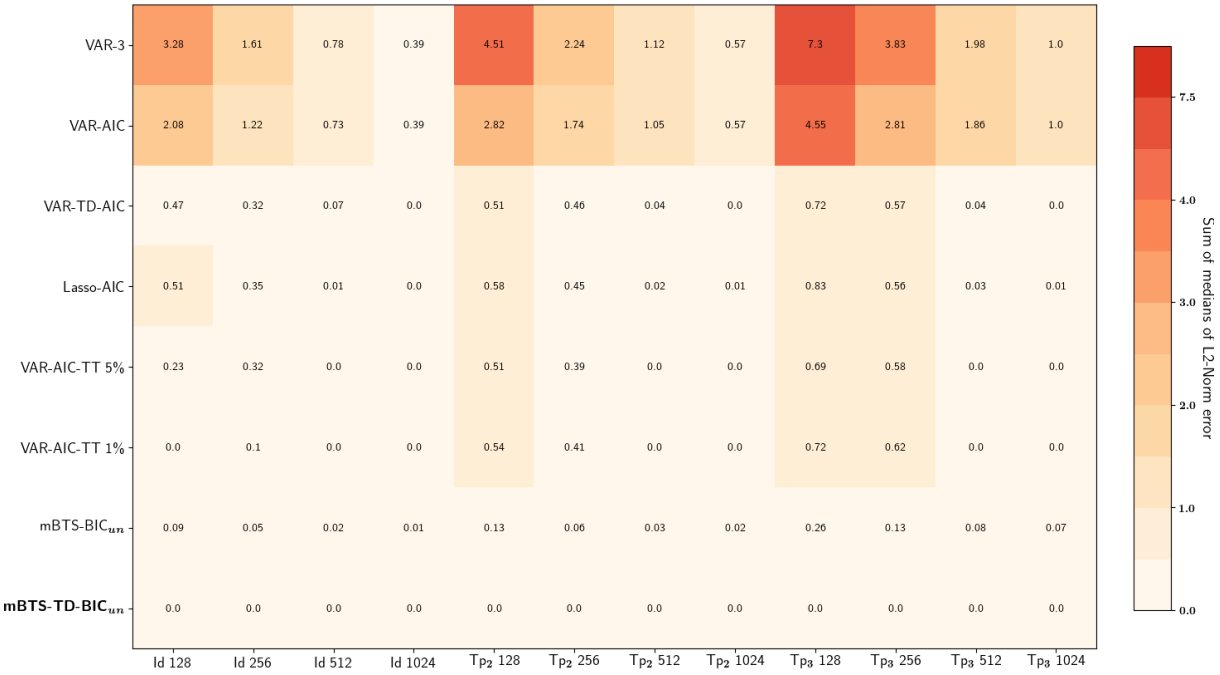


Figure 3.9: Sum of medians over 1000 simulations of the L_2-norm error in the non-causal GPDC (3.5), estimated using the subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the non-causal GPDC.

For the errors in the non-causal terms shown in Fig. 3.9, the results are clear. Each subset selection method perfectly plays out its role in modelling only the most significant coefficients (causal terms). The six subset selection methods greatly reduce errors compared to VAR-AIC or VAR-3, and can provide a sum of medians of the L_2-norm errors close to zero for all non-causal terms. However, even in this case, the mBTS-TD method provides the most interesting results. This is the only method that presents a sum of medians close to zero for all settings, and it significantly improves on the mBTS method for sample sizes $T \leq 256$ and correlated residuals.

3.4.3 Robustness checks: error distributions and causal structure identification

Having used our preliminary analysis in section 3.4.2 to identify the three methods TT 1%, mBTS and mBTS-TD that seem most suitable for computing the GPDC, we extend here the comparison. In contrast to in section 3.4.2, where medians of the L_2-norm error were used to compare the methods, in this section, we first evaluate the two methods through the L_2-norm error distributions for the causal and non-causal GPDC, then focus on identification of the true causal structure of (S) using the F-Measure (FM) and Hamming Distance (HD) as in [184].

GPDC error distributions

In Table 3.1, we report the average value and standard deviation of the L_2-norm error distributions for the causal GPDC and in Fig. 3.10 we provide an example of the L_2-norm error distribution with $T = 256$ and $\mathbf{\Sigma_\epsilon} = \mathrm{Tp}_2$. Table 3.2 and Fig. 3.11 exhibit the same results for the non-causal terms.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 3	**mBTS-TD BIC_{un} 3**	mBTS BIC_{un} 6	**mBTS-TD BIC_{un} 6**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	0.687 (*0.350*)	0.544 (*0.356*)	0.449 (*0.303*)	**0.389** (*0.293*)	0.480 (*0.322*)	**0.417** (*0.312*)	0.492 (*0.326*)	**0.429**$^{+}$ (*0.318*)
Id 256	0.546 (*0.294*)	0.406 (*0.294*)	0.302 (*0.206*)	**0.251** (*0.187*)	0.317 (*0.217*)	**0.266** (*0.199*)	0.327 (*0.223*)	**0.274**$^{+}$ (*0.206*)
Id 512	0.378 (*0.217*)	0.239 (*0.216*)	0.203 (*0.133*)	**0.167** (*0.123*)	0.212 (*0.140*)	**0.175** (*0.131*)	0.218 (*0.144*)	**0.180**$^{+}$ (*0.135*)
Id 1024	0.238 (*0.122*)	0.123 (*0.095*)	0.140 (*0.094*)	**0.113** (*0.085*)	0.145 (*0.098*)	**0.117** (*0.089*)	0.147 (*0.099*)	**0.119**$^{+}$ (*0.091*)
Tp_2 128	0.764 (*0.404*)	0.575 (*0.403*)	0.475 (*0.348*)	**0.407** (*0.347*)	0.505 (*0.364*)	**0.439** (*0.368*)	0.524 (*0.371*)	**0.457**$^{+}$ (*0.375*)
Tp_2 256	0.614 (*0.334*)	0.423 (*0.319*)	0.309 (*0.216*)	**0.247** (*0.196*)	0.327 (*0.229*)	**0.262** (*0.208*)	0.336 (*0.233*)	**0.271**$^{+}$ (*0.214*)
Tp_2 512	0.447 (*0.253*)	0.242 (*0.231*)	0.212 (*0.151*)	**0.168** (*0.132*)	0.220 (*0.156*)	**0.175** (*0.139*)	0.226 (*0.160*)	**0.181**$^{+}$ (*0.143*)
Tp_2 1024	0.291 (*0.152*)	0.119 (*0.097*)	0.143 (*0.099*)	**0.108** (*0.084*)	0.147 (*0.101*)	**0.112** (*0.088*)	0.149 (*0.103*)	**0.114**$^{+}$ (*0.089*)
Tp_3 128	0.882 (*0.462*)	0.674 (*0.480*)	0.631 (*0.464*)	**0.565** (*0.486*)	0.652 (*0.466*)	**0.591** (*0.490*)	0.665 (*0.466*)	**0.604**$^{+}$ (*0.491*)
Tp_3 256	0.721 (*0.390*)	0.451 (*0.360*)	0.370 (*0.288*)	**0.276** (*0.273*)	0.386 (*0.300*)	**0.291** (*0.286*)	0.394 (*0.303*)	**0.300**$^{+}$ (*0.291*)
Tp_3 512	0.561 (*0.320*)	0.239 (*0.245*)	0.245 (*0.185*)	**0.161** (*0.142*)	0.254 (*0.191*)	**0.168** (*0.149*)	0.259 (*0.192*)	**0.173**$^{+}$ (*0.153*)
Tp_3 1024	0.393 (*0.223*)	0.127 (*0.121*)	0.178 (*0.133*)	**0.108** (*0.092*)	0.182 (*0.137*)	**0.111** (*0.097*)	0.184 (*0.138*)	**0.114**$^{+}$ (*0.100*)

Table 3.1: Causal GPDC: Average value and standard deviation in parentheses of the L_2-norm error distribution (1000 simulations) for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 3, 6, 9$. The lower average error is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 3, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

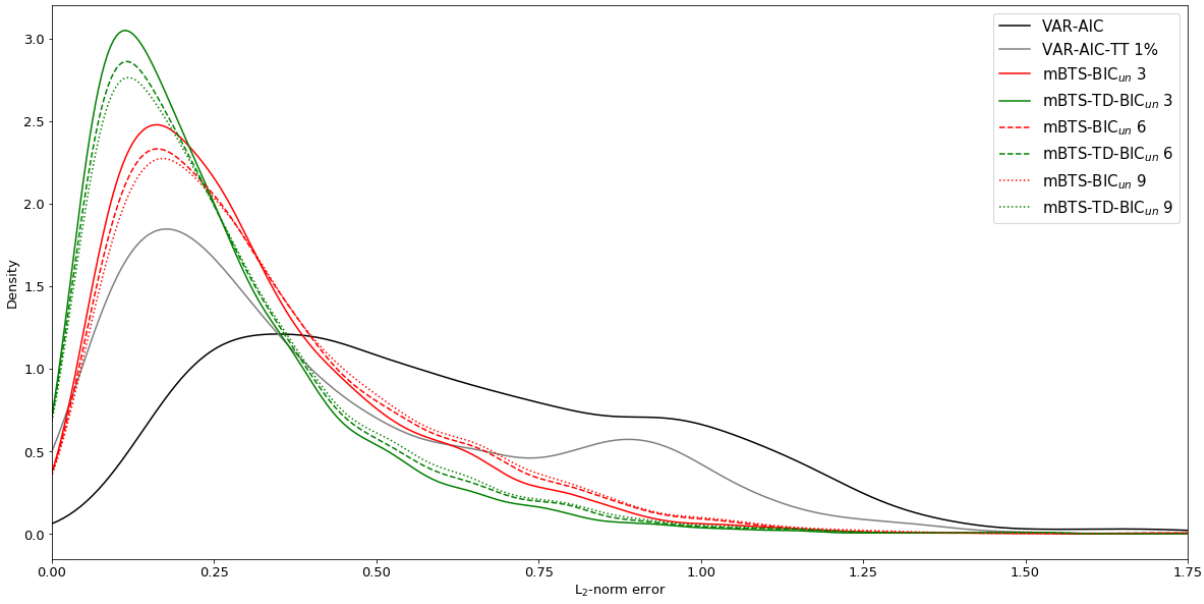


Figure 3.10: Causal GPDC. L_2-norm error distribution (1000 simulations) with $T = 256$ and $\mathbf{\Sigma_\epsilon} = Tp_2$ for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 3, 6, 9$.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 3	**mBTS-TD BIC_{un} 3**	mBTS BIC_{un} 6	**mBTS-TD BIC_{un} 6**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	0.195 (*0.202*)	0.055 (*0.208*)	0.035^{+} (*0.124*)	**0.020** (*0.115*)	0.048 (*0.150*)	**0.031** (*0.137*)	0.058 (*0.167*)	**0.040** (*0.156*)
Id 256	0.112 (*0.115*)	0.031 (*0.113*)	0.014 (*0.052*)	**0.005** (*0.041*)	0.019 (*0.063*)	**0.009** (*0.052*)	0.022 (*0.069*)	$\mathbf{0.012}^{+}$ (*0.059*)
Id 512	0.065 (*0.063*)	0.011 (*0.054*)	0.005 (*0.022*)	**0.001** (*0.015*)	0.007 (*0.026*)	**0.003** (*0.020*)	0.008 (*0.029*)	$\mathbf{0.004}^{+}$ (*0.023*)
Id 1024	0.033 (*0.028*)	0.002 (*0.013*)	0.002 (*0.011*)	**0** (*0.007*)	0.003 (*0.012*)	**0.001** (*0.008*)	0.003 (*0.013*)	$\mathbf{0.001}^{+}$ (*0.010*)
Tp_2 128	0.266 (*0.283*)	0.071 (*0.249*)	0.042^{+} (*0.156*)	**0.026** (*0.145*)	0.055 (*0.177*)	**0.037** (*0.165*)	0.065 (*0.190*)	**0.045** (*0.178*)
Tp_2 256	0.156 (*0.171*)	0.041 (*0.149*)	0.016 (*0.061*)	**0.006** (*0.048*)	0.020 (*0.069*)	**0.010** (*0.056*)	0.023 (*0.075*)	$\mathbf{0.012}^{+}$ (*0.062*)
Tp_2 512	0.096 (*0.095*)	0.015 (*0.075*)	0.008 (*0.030*)	**0.002** (*0.019*)	0.009 (*0.033*)	**0.003** (*0.023*)	0.011 (*0.036*)	$\mathbf{0.004}^{+}$ (*0.026*)
Tp_2 1024	0.047 (*0.040*)	0.002 (*0.014*)	0.003 (*0.013*)	**0** (*0.006*)	0.004 (*0.015*)	**0.001** (*0.009*)	0.005 (*0.016*)	$\mathbf{0.001}^{+}$ (*0.010*)
Tp_3 128	0.404 (*0.417*)	0.113 (*0.367*)	0.075 (*0.253*)	**0.050** (*0.246*)	0.086 (*0.265*)	**0.061** (*0.260*)	0.095 (*0.274*)	$\mathbf{0.069}^{+}$ (*0.269*)
Tp_3 256	0.256 (*0.267*)	0.059 (*0.216*)	0.029 (*0.112*)	**0.012** (*0.097*)	0.033 (*0.116*)	**0.015** (*0.099*)	0.036 (*0.120*)	$\mathbf{0.018}^{+}$ (*0.104*)
Tp_3 512	0.162 (*0.158*)	0.019 (*0.104*)	0.013 (*0.046*)	**0.002** (*0.027*)	0.014 (*0.049*)	**0.003** (*0.030*)	0.016 (*0.051*)	$\mathbf{0.004}^{+}$ (*0.033*)
Tp_3 1024	0.084 (*0.077*)	0.003 (*0.028*)	0.008 (*0.028*)	**0** (*0.012*)	0.009 (*0.029*)	**0.001** (*0.013*)	0.009 (*0.030*)	$\mathbf{0.001}^{+}$ (*0.014*)

Table 3.2: Non-causal GPDC: Average value and standard deviation in parentheses of the L_2-norm error distribution (1000 simulations) for the non-causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 3, 6, 9$. The lower average error is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 3, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

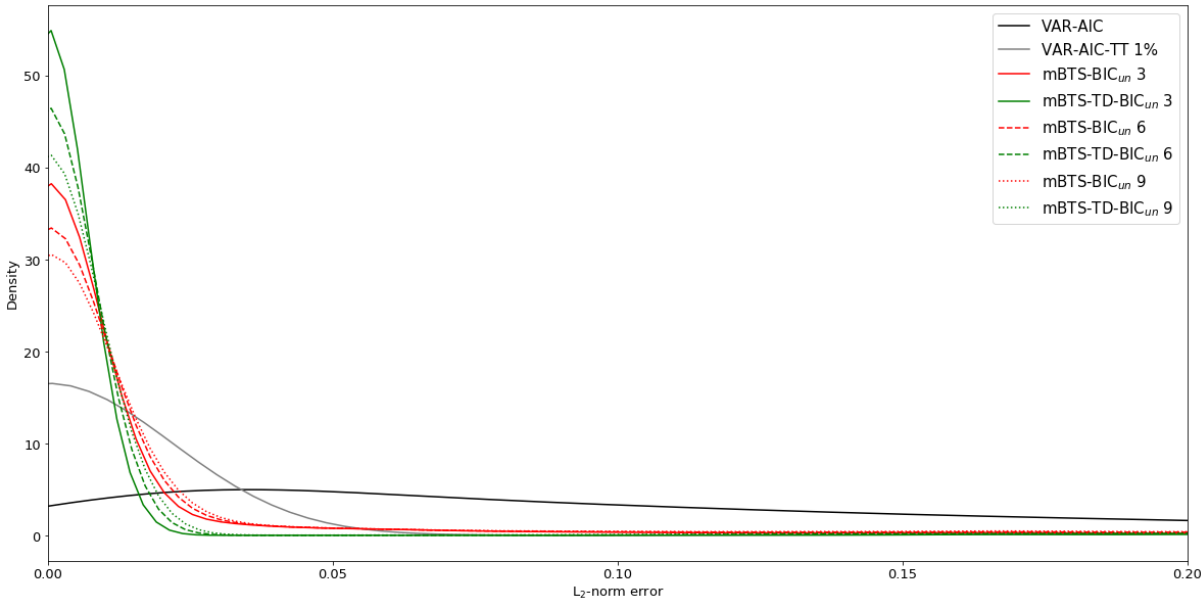


Figure 3.11: Non-causal GPDC. L_2-norm error distribution (1000 simulations) with $T = 256$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}} = Tp_2$ for the non-causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 3, 6, 9$.

The following conclusions can be drawn regarding both causal (Table 3.1 and Fig. 3.10) and non-causal (Table 3.2 and Fig. 3.11) terms:

- By taking into account only the same p_{max} for the two methods, denoted in bold in Tables 3.1 and 3.2, mBTS-TD clearly stands out from mBTS by providing lower average errors for each setting.
- Whatever p_{max} is selected, the mBTS-TD error distributions are more concentrated, with a lower fat tail.
- mBTS-TD-BIC$_{un}$9 always admits lower errors than mBTS-BIC$_{un}$3 and VAR-AIC-TT 1% for the causal GPDC, and only in two cases exhibits higher errors than mBTS-BIC$_{un}$3 for the non-causal terms denoted with a superscript symbol $^{+}$ in Tables 3.1 and 3.2.

Causal structure identification

To identify the true causal structure, we use the F-Measure (FM) and the Hamming Distance (HD) discussed in [184]. FM focuses on the identification of pairs of true causality, whereas HD focuses on the identification of all pairs. We consider the existence of causality between two time series $x_j(t)$ and $x_k(t)$ if $|\widehat{\boldsymbol{\omega}}_{jk}|^2 > 0.01$ at least at one frequency f. FM and HD are defined as follows:

$$\mathrm{FM} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}$$
$$\mathrm{HD} = \mathrm{FN} + \mathrm{FP}$$

where TP are the True Positives (causality correctly identified), FN are the False Negatives (causality not identified), and FP are the False Positives (wrongly identified causality). FM ranges from 0 to 1. If FM = 1, then there is a perfect identification of the pairs of true causality, whereas if FM = 0, then no true causality is detected. HD ranges from 0 to $m(m-1)$. If HD = 0, there is a perfect identification, whereas if HD = $m(m-1)$, all pairs are misclassified.

Tables 3.3 and 3.4 report the average value of FM (Table 3.3) and HD (Table 3.4) for each setting.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 3	**mBTS-TD BIC_{un} 3**	mBTS BIC_{un} 6	**mBTS-TD BIC_{un} 6**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	0.500	0.881	0.873	**0.941**	0.840	**0.909**	0.818	**0.885**$^+$
Id 256	0.568	0.884	0.925	**0.968**	0.898	**0.943**	0.881	**0.928**$^+$
Id 512	0.663	0.928	0.970	**0.985**	0.958	**0.977**	0.950	**0.971**$^+$
Id 1024	0.851	0.991	0.994	**0.997**	0.992	**0.996**	0.990	**0.995**$^+$
Tp2 128	0.475	0.863	0.861	**0.929**	0.830	**0.897**	0.808	**0.874**$^+$
Tp2 256	0.530	0.878	0.916	**0.965**	0.894	**0.945**	0.879	**0.930**$^+$
Tp2 512	0.581	0.924	0.952	**0.980**	0.941	**0.971**	0.930	**0.965**$^+$
Tp2 1024	0.751	0.990	0.988	**0.997**	0.984	**0.995**	0.981	**0.994**$^+$
Tp3 128	0.446	0.819	0.794	**0.878**	0.771	**0.853**	0.754	**0.833**$^+$
Tp3 256	0.479	0.873	0.875	**0.954**	0.857	**0.936**	0.842	**0.921**$^+$
Tp3 512	0.500	0.932	0.932	**0.986**	0.921	**0.978**	0.912	**0.971**$^+$
Tp3 1024	0.596	0.977	0.958	**0.995**	0.953	**0.993**	0.950	**0.992**$^+$

Table 3.3: Average value of the F-measure (FM) over 1000 simulations of the GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 3, 6, 9$. FM ranges from 0 to 1. If FM = 1 there is perfect identification of the pairs of true causality, whereas if FM = 0 no true causality is detected. The lower average value is highlighted for each setting and p_{max}. The superscript symbol $^+$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 3, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 3	**mBTS-TD BIC_{un} 3**	mBTS BIC_{un} 6	**mBTS-TD BIC_{un} 6**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	10.014	1.325	1.451	**0.626**	1.893	**0.997**	2.215	**1.291**$^+$
Id 256	7.596	1.311	0.808	**0.333**	1.134	**0.605**	1.351	**0.779**$^+$
Id 512	5.076	0.774	0.306	**0.150**	0.439	**0.237**	0.529	**0.302**$^+$
Id 1024	1.750	0.088	0.063	**0.028**	0.085	**0.037**	0.099	**0.046**$^+$
Tp2 128	10.995	1.540	1.586	**0.756**	2.011	**1.128**	2.340	**1.412**$^+$
Tp2 256	8.877	1.394	0.920	**0.363**	1.186	**0.582**	1.381	**0.758**$^+$
Tp2 512	7.226	0.820	0.499	**0.202**	0.626	**0.294**	0.751	**0.368**$^+$
Tp2 1024	3.320	0.106	0.118	**0.033**	0.165	**0.050**	0.194	**0.063**$^+$
Tp3 128	12.309	2.070	2.392	**1.279**	2.740	**1.579**	3.018	**1.846**$^+$
Tp3 256	10.877	1.444	1.406	**0.472**	1.643	**0.678**	1.852	**0.848**$^+$
Tp3 512	9.989	0.727	0.729	**0.146**	0.861	**0.229**	0.968	**0.300**$^+$
Tp3 1024	6.771	0.240	0.434	**0.050**	0.495	**0.070**	0.527	**0.077**$^+$

Table 3.4: Average value of Hamming Distance (HD) over 1000 simulations of the GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 3, 6, 9$. HD ranges from 0 to $m(m-1)$, where $m = 5$. If HD = 0 there is perfect identification, whereas if HD = 20 all pairs are misclassified. The lower average value is highlighted for each setting and p_{max}. The superscript symbol $^+$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 3, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

For the two measures FM (Table 3.3) and HD (Table 3.4), mBTS-TD provides on average a better identification of the true causal structure than mBTS for each setting by taking into account only the same p_{max} highlighted in boldface in Tables 3.3 and 3.4. As previously for the error distributions, mBTS-TD-$BIC_{un}9$ outperforms mBTS-$BIC_{un}3$ and VAR-AIC-TT 1%, denoted with a superscript symbol $^+$ in Tables 3.3 and 3.4. Note that VAR-AIC provides the worst results, whatever measure is taken into account.

We end this simulation study by comparing the computational efficiency of mBTS and mBTS-TD on (S) across the different sample sizes. The computation times of the two methods for one realization ($T = 1024$ and Tp_3) are quite similar, with 0.166 seconds for mBTS and 0.174 seconds for mBTS-TD. The computations are carried out using Python 3.7 with 2.70GHz CPU (Intel Xeon E-2176M) and 32Gb RAM.

To conclude regarding the accuracy of the GPDC results, subset selection methods are well suited when the underlying model admits a parsimonious structure. Each subset method improves the GPDC accuracy, and when we combine mBTS and TD methods, the cascading errors on GPDC are drastically reduced for both causal and non-causal terms. We can also add that, with $p_{max} = 9$, for mBTS-TD the errors on the GPDC are lower than for mBTS starting with the true lag $p_{max} = 3$. mBTS-TD therefore reduces dependence on the choice of p_{max} and produces a more parsimonious model when p_{max} is large. Fig. 3.12 compares examples of the GPDC estimated using mBTS-TD, VAR-3, and VAR-AIC with the theoretical GPDC.

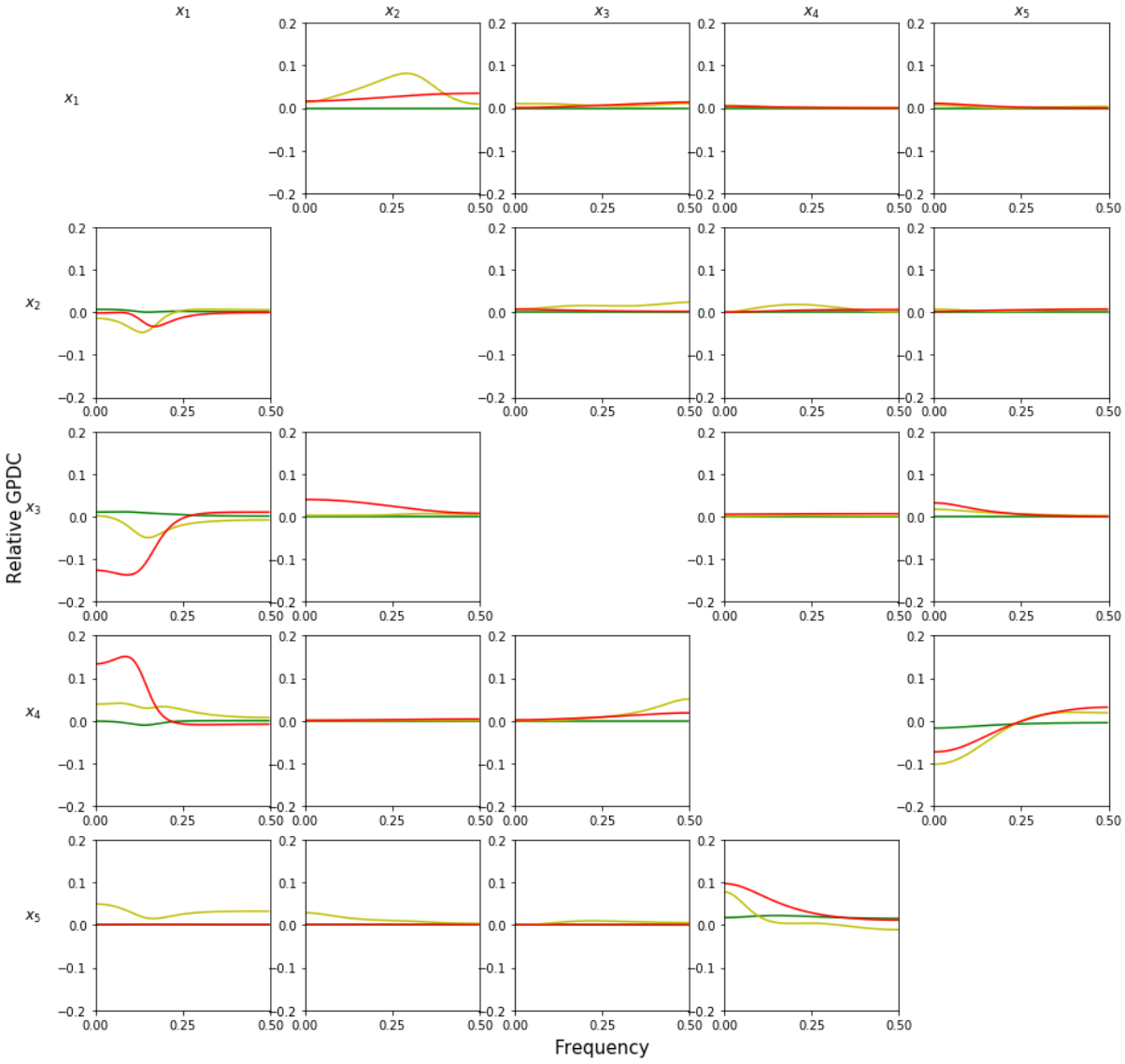


Figure 3.12: Relative GPDC compared to the theoretical GPDC of (S) for $T = 256$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \mathrm{Tp}_2$: VAR-2 in red, VAR-3 in yellow, and mBTS-TD-BIC$_{un}$ in green. Relative GPDC: $|\widehat{\boldsymbol{\omega}}_{jk}|^2 - |\boldsymbol{\omega}_{jk}|^2$ for each pair $j \neq k$, where $\boldsymbol{\omega}_{jk}$ is the vector containing each value of $\omega_{jk}(f)$ for all discrete frequencies f.

As found with Monte Carlo simulations, the differences in the causal part between the mBTS-TD method and the theoretical GPDC are very small. Moreover, in this example, none of the zero-coefficients are estimated by mBTS-TD, which makes it possible to have a GPDC equal to zero, like the theoretical GPDC for non-causal terms.

In addition, in Appendix 3.7.1 & 3.7.2, we provide results on two other systems, a VAR(4) model on $m = 5$ (model 1 in [214, 151]) and a high-dimensional parsimonious VAR(4) model on $m = 12$ with only 4% of the m^2p coefficients being non-zero (model used in [215]).

3.5 Financial application

In the literature, several classical approaches exist to model the links between the assets of a financial universe. The correlation matrix is often used [69, 71, 72, 216] to build weighted or binary networks. Unfortunately, this methodology suffers from two major drawbacks. First, these networks are undirected, only highlighting the existence of the relationship between assets not their directions. Second, the network dimension must also be reduced (using methods such as the Minimum Spanning Tree [69] or Planar Maximally Filtered Graphs [128]) otherwise the network is complete and difficult to use in practice for portfolio allocation. In a symmetric way, methods based on Granger non-causality tests in VAR models as in [217, 73] allow to retrieve a directed but unweighted network, remaining very sensitive to the underlying VAR processes. In [74], an alternative directed causal network is built, beyond VAR modeling, but focuses only on very short-dynamics.

In this section, we make use of the GPDC measure, estimated with our mBTS-TD method, to modeling financial markets dependency structures. This approach provides not only a precise network topology (taking into account both the direction and the strength of the relationship between assets via the GPDC) but also solve the dimensionality puzzle, via the mBTS-TD estimation process that intrinsically produces parsimonious causal structures. In a second step, we study on real data the empirical performances of financial portfolios obtained excluding the most systemic nodes of the incomplete GPDC financial network.

3.5.1 Building a GPDC financial network to identify systemic assets

A network $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ is a set of objects with $\mathbf{V}$ the set of nodes and $\mathbf{E}$ the set of edges between nodes. The edge (j, k) connects a pair of nodes j and k. The mathematical representation of a directed weighted network is given by the $m \times m$ adjacency matrix $\mathbf{Z} = (z_{jk})$,$z_{jk} \in \mathbb{R}^+$ if $(j, k) \in \mathbf{E}$ and 0 otherwise. In the sequel, let $\mathbf{Z}^{(1)} = (z_{jk}^{(1)})$ be the adjacency matrix built using the GPDCs in which are plugged the VAR coefficients, estimated using our mBTS-TD procedure, and $\mathbf{Z}^{(2,\gamma)} = (z_{jk}^{(2,\gamma)})$ the adjacency matrix built using the GPDCs based on VAR-AIC models. In this latter case, note that $\mathbf{Z}^{(2,\gamma)}$ depends on a threshold parameter γ, defined hereafter. Since our mBTS-TD procedure pre-filters the VAR by removing unnecessary coefficients, $z_{jk}^{(1)}$ can be directly defined as follows:

$$z_{jk}^{(1)} = \begin{cases} \max |\, \boldsymbol{\omega}_{jk} \,|^2, & \text{if } j \neq k \\ \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

where $\boldsymbol{\omega}_{jk}$ is the vector containing each value of $\omega_{jk}(f)$ for all discrete frequencies f ($f \in [0, \frac{1}{2}]$) as defined in (3.5). We use the maximal value in the $\boldsymbol{\omega}_{jk}$ vector in order to

take into account the most relevant information between the two assets, i.e. based on short (high-frequency) and long-term (low frequency) relationships.

When a classical VAR-AIC is used to compute the GPDCs, all VAR coefficients are involved, returning, in general, non-null GPDCs. The resulting complete weighted network is useless in practice. We thus apply a filter to each component of the associated $\boldsymbol{\omega}_{jk}$ vector and compute the vector $\boldsymbol{\omega}_{jk}^{\gamma}$ whose coordinates are given by:

$$\omega_{jk}^{\gamma}(f) = \begin{cases} \omega_{jk}(f), & \text{if } |\omega_{jk}(f)| \geq \gamma \\ \\ 0, & \text{otherwise} \end{cases}$$

with $\gamma \in \{0.01, 0.02, 0.03, 0.04, 0.05\}$.

Thus, an element of the adjacency matrix $\mathbf{Z}^{(2,\gamma)}$, is defined as:

$$z_{jk}^{(2,\gamma)} = \begin{cases} \max |\, \boldsymbol{\omega}_{jk}^{\gamma} \,|^2, & \text{if } j \neq k \\ \\ 0, & \text{otherwise} \end{cases} \tag{3.9}$$

Once built, the previous incomplete financial networks can help us to improve asset allocation strategies using the classical tools of network theory such as centrality measures or clustering coefficients as in [71, 72, 218, 219, 75, 76]. In order to show the potential of using both the mBTS-TD method and the GPDC to build a financial network, we propose to identify the most systemic assets with the local directed weighted clustering coefficient. Indeed, the local directed weighted clustering coefficient allows to identify the most embedded assets in the network and thus the most systemic ones. This tool introduced by Clemente et al. in [127] measures how a node is embedded into the network by quantifying its number of triangles out of all its possible triangles. Furthermore, it takes into account the strength of a node in the normalization factor (see also [126]). Starting from a directed weighted network with adjacency matrix $\mathbf{Z}$, we obtain the associated directed unweighted network with adjacency matrix $\mathbf{Z}^u$ defining $z_{jk}^u = 1$ if $z_{jk} \neq 0$, and 0 otherwise. Thus, the local directed weighted clustering coefficient for the asset (node) j is defined as follows:

$$h_j = \frac{\frac{1}{2}\left[(\mathbf{Z} + \mathbf{Z}')\,(\mathbf{Z}^u + \mathbf{Z}^{u\prime})^2\right]_{jj}}{s_j\,(d_j - 1) - 2s_j^{\leftrightarrow}}$$

where $d_j = (\mathbf{Z}^{u\prime} + \mathbf{Z}^u)_j\, \mathbf{1}_m$ and $s_j = (\mathbf{Z}' + \mathbf{Z})_j\, \mathbf{1}_m$ are respectively the total degree (total number of edges) and the total strength (case of weighted graph) of the asset j. $s_j^{\leftrightarrow} = \dfrac{(\mathbf{Z}\,\mathbf{Z}^u + \mathbf{Z}^u\,\mathbf{Z})_{jj}}{2}$ is the strength of bilateral edges between j and k. Note that h_j belongs to $[0,\, 1]$, a high value indicating that the asset j is heavily embedded in the network, and captures in particular in and out diffusion processes, and therefore spillover and feedback effects.

3.5.2 Building a diversified Equally Weighted portfolio

For a given investment universe, the above methodology allow us to identify the most systemic nodes of the GPDC network as the one associated with the greatest local directed weighted clustering coefficients computed from adjacency matrices $\mathbf{Z}^{(1)}$ or $\mathbf{Z}^{(2,\gamma)}$. To build and compare financial portfolios we only base our financial strategy on m non-systemic assets and we basically allocate all of them with the same weight $\frac{1}{m}$ resulting in an Equally Weighted (EW) portfolio. The interest of such an approach in our framework is to focus solely on the improvement resulting from the asset selection process. It does not require any additional estimation procedure (covariance matrix) nor complex optimization issues. What is more, the authors in [212] have shown that this method can even provide higher performances than more advanced ones.

3.5.3 Dataset description and Empirical performances

We consider national financial markets, each market (node) being represented by the MSCI ACWI (All Country World Index). We first apply two filters. First we remove the less liquid ones (Argentina, Czech Republic, Egypt, Greece, Hungary, Pakistan), and then those with no quotes since 2001 (Qatar, Saudi Arabia, United Arab Emirates). This universe of 40 assets (see Table 3.15 in Appendix 3.7.3) allows us to take both in account the differences in time delay between areas (feedback effects) as well as local discrepancy (e.g. macroeconomic differences). We use asset returns computed on a daily basis from January 2001, the 18th to October 2019, the 25th to build, every four weeks, a temporal network, using a rolling window of $T = 256$ working days, with a rebalancing period of four weeks. Since financial assets returns exhibit heteroskedasticity, we normalize each time series using a Generalized Auto-Regressive Condional Heteroskedastic (GARCH) [220] filter (see Appendix 3.7.4) to estimate the corresponding VAR process used to compute the GPDC measure and to identify the non-systemic assets.

The asset exclusion procedure using mBTS-TD for the VAR estimation and the GPDC measure ("mBTS-TD GPDC") is compared with those obtained using a classical VAR estimation ("VAR-AIC GPDC") with several thresholds $\gamma \in \{0.01, 0.02, 0.03, 0.04, 0.05\}$ and on the whole universe ("EW"), i.e. without any asset selection. In order to assess the potential of our methodology, we report several portfolio statistics computed over the whole period: the annualized return, the annualized volatility, the ratio between the annualized return and the annualized volatility and the maximum drawdown (largest decline in portfolio value). The portfolio generates better performances if it provides an higher return/volatility ratio and a lower maximum drawdown. The (EW) portfolios' performances are computed in USD currency, because if the asset returns are kept into local currency, hedging costs (selling the currency forward) have to be considered.

In Table 3.5, we provide the portfolios' results when ten assets are excluded representing 25% of the initial universe (we provide in Appendix 3.7.5 the results for 20%, 37% and 50%) and we only report the best threshold γ for the classical VAR estimation. For this case,"mBTS-TD GPDC" shows a significant improvement with respect to the other methods. It provides significant higher annualized return, similar annualized volatility (higher return/volatility ratio) and also a similar drawdown compared to (EW) without exclusion.

EW Portfolios 10 excluded assets	Annualized Return	Annualized Volatility	Ratio Return/Volatility	Max Drawdown
mBTS-TD GPDC	10.46%	16.63%	0.63	62.03%
VAR-AIC GPDC 0.03	9.51%	16.84%	0.56	62.11%
EW	9.35%	16.76%	0.56	61.90%
Non-selected assets (VAR-AIC GPDC 0.03)	8.70%	17.64%	0.49	61.48%
Non-selected assets (mBTS-TD GPDC)	5.89%	18.28%	0.32	61.63%

Table 3.5: Performance indicators for EW portfolios with 10 excluded assets from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility)

Given these results, we can consider that our methodology "mBTS-TD GPDC" succeeds in identifying the less performing/riskiest assets. To reinforce this aspect, Table 3.6 provides the first four order moments of systemic assets return distribution for our proposed methodology and the "VAR-AIC GPDC 0.03".

10 Non-Selected Assets return distribution	mBTS-TD GPDC	VAR-AIC GPDC 0.03
Mean	0.0003	0.0004
Standard Deviation	0.0156	0.0159
Skewness	-0.0180	-0.0114
Kurtosis	10.3154	10.8294

Table 3.6: Moments of out-of-sample asset return distribution for ten non-selected assets for "mBTS-TD GPDC" and "VAR-AIC GPDC 0.03" from January 2002 to October 2019.

We observe that the assets return distribution in the non-selected universe obtained using "mBTS-TD GPDC" provides better figures, in particular for the mean and the skewness. Indeed, the assets non-selected by the "mBTS-TD GPDC" have a lower average return and more negative skewness than in the "VAR-AIC GPDC 0.03" case, which confirms that "mBTS-TD GPDC" better identifies the less performing/riskiest assets than the standard VAR estimation. Regarding portfolios' performances, the "mBTS-TD GPDC" exclusion process takes full advantage of the related precise network topology combined with an intrinsic parsimonious causal structure. This paves the way for interesting results in the case of more complex assets allocation processes that will be the objective of a forthcoming study.

3.6 Conclusion

Retrieving complex interactions in multivariate systems admitting a VAR representation is of key importance in a number of fields. To this end, coherence measures have been introduced to quantify causal strength between variables. Nevertheless, we prove in this paper, through careful Monte Carlo simulations, that applying a naive approach first estimating a VAR model using LS, and then computing coherence measures, is highly inefficient especially when the underlying data generating processes are parsimonious. To overcome this problem, we apply classical subset selection methods, and show that they do improve coherence measures but not sufficiently. We therefore introduce a new subset selection method, namely the mBTS-TD one, and, still using Monte Carlo simulations, prove that it clearly outperforms its natural competitors, and allows us to dramatically reduce to so-called cascading errors in both the causal and non-causal structure of the system. Last, we have implemented our procedure in the financial domain making use of the GPDC measure estimated with the mBTS-TD strategy to model financial markets dependency structures. This approach provides us not only with a precise network topology (taking into account both the direction and the strength of the relationship between assets) but also solves the network dimension puzzle producing a parsimonious causal structure. We take advantage of this financial network identifying, via the local directed weighted clustering coefficient, the most systemic assets to exclude them, with profit, from our investment universe.

3.7 Appendix Chapter 3

3.7.1 GPDC results on Winterhalder et al. system

We extend our first results obtained with the system (S) by analyzing 4 times series generated by the VAR(5) model used in [214, 151]. This system also admits a parsimonious structure with 11% of the m^2p coefficients being non-zero and has only one coefficient on the fifth lag.

The VAR model [214] is as follows:

$$\begin{aligned} x_1(t) &= 0.8x_1(t-1) + 0.65x_2(t-4) + \epsilon_1(t) \\ x_2(t) &= 0.6x_2(t-1) + 0.6x_4(t-5) + \epsilon_2(t) \\ x_3(t) &= 0.5x_3(t-3) - 0.6x_1(t-1) + 0.4x_2(t-4) + \epsilon_3(t) \\ x_4(t) &= 1.2x_4(t-1) + 0.7x_4(t-2) + \epsilon_4(t) \end{aligned} \tag{S2}$$

3.7.1.1 S2: GPDC errors

Figs. 3.13 and 3.14 show GPDC errors in causal (Fig. 3.13) and non-causal terms (Fig. 3.14) for the (S2) system

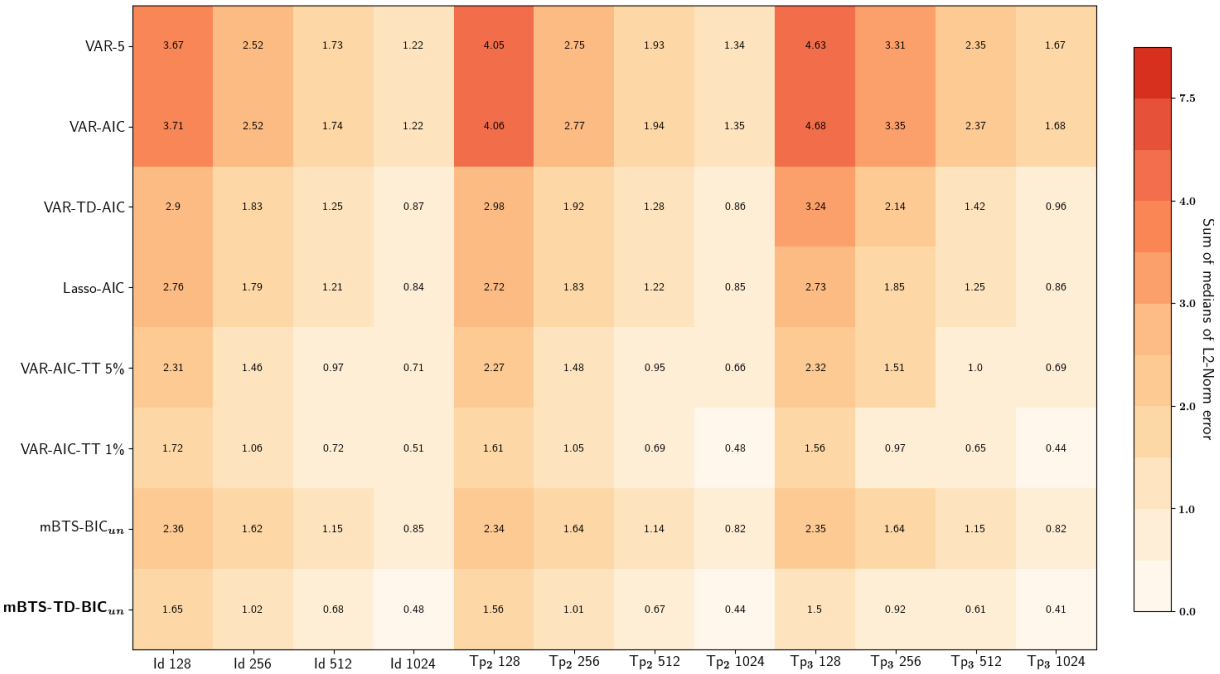


Figure 3.13: (S2): Sum of medians over 1000 simulations of the L_2-norm error in the causal GPDC (3.5), estimated using the subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the causal GPDC.

In Fig. 3.13, the results on the causal terms of (S2) are more contrasted. mBTS-TD and TT 1% provide the best results with a slight advantage for mBTS-TD.

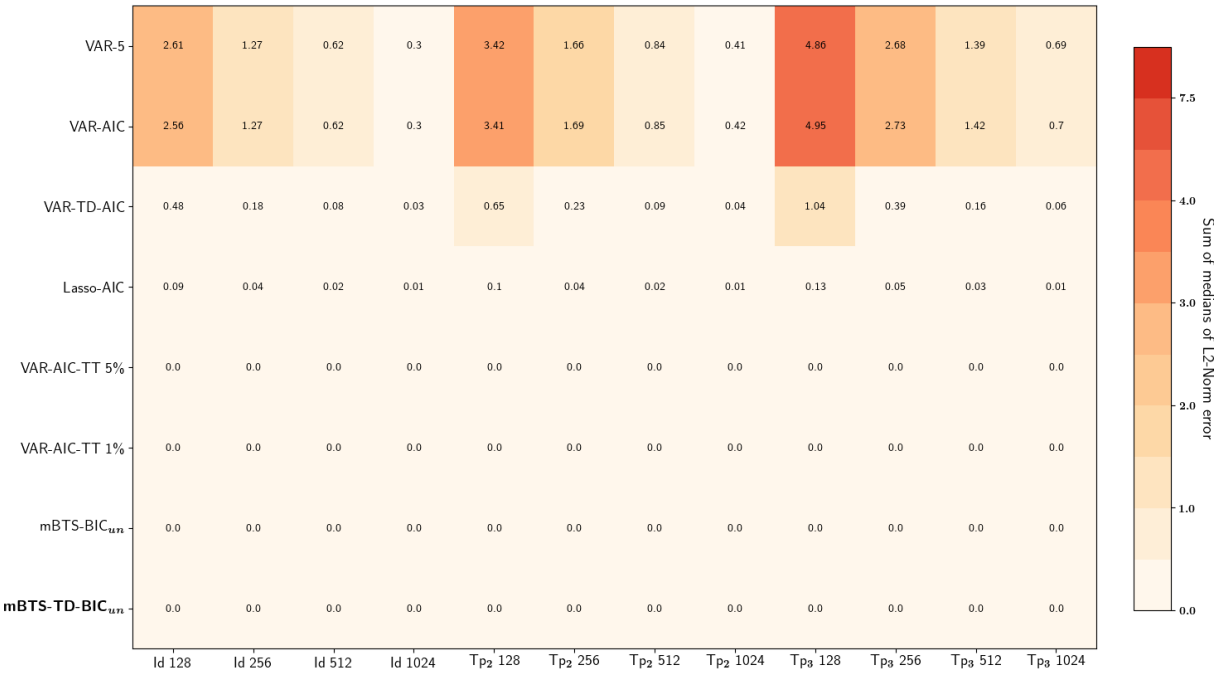


Figure 3.14: (S2): Sum of medians over 1000 simulations of the L_2-norm error in the non-causal GPDC (3.5), estimated using the subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the non-causal GPDC.

For the errors in the non-causal terms shown in Fig. 3.14, the results are clear. Each subset selection method perfectly plays out its role in modelling only the most significant coefficients (causal terms). The six subset selection methods greatly reduce errors compared to VAR-AIC or VAR-5, and can provide a sum of medians of the L_2-norm errors close to zero for all non-causal terms. However, in this case, mBTS-TD, mBTS and TT provide the most interesting results with a sum of medians equals to zero for all settings.

3.7.1.2 S2: GPDC error distributions

In Table 3.7, we report for the system (S2) the average value and standard deviation of the L_2-norm error distributions for the causal GPDC and in Fig. 3.15 we provide an example of the L_2-norm error distribution with $T = 256$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}} = \text{Tp}_2$. Table 3.8 and Fig. 3.16 exhibit the same results for the non-causal terms.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC$_{un}$ 5	**mBTS-TD BIC$_{un}$ 5**	mBTS BIC$_{un}$ 7	**mBTS-TD BIC$_{un}$ 7**	mBTS BIC$_{un}$ 9	**mBTS-TD BIC$_{un}$ 9**
Id 128	1.006 (*0.472*)	0.528 (*0.428*)	0.628 (*0.350*)	**0.457** (*0.321*)	0.658 (*0.373*)	**0.483** (*0.340*)	0.679 (*0.386*)	**0.499**$^{+}$ (*0.352*)
Id 256	0.659 (*0.266*)	0.309 (*0.219*)	0.430 (*0.226*)	**0.290** (*0.202*)	0.445 (*0.238*)	**0.299** (*0.211*)	0.455 (*0.243*)	**0.305**$^{+}$ (*0.216*)
Id 512	0.455 (*0.176*)	0.212 (*0.147*)	0.306 (*0.156*)	**0.196** (*0.134*)	0.312 (*0.162*)	**0.199** (*0.138*)	0.318 (*0.164*)	**0.203**$^{+}$ (*0.141*)
Id 1024	0.319 (*0.123*)	0.149 (*0.101*)	0.224 (*0.111*)	**0.135** (*0.091*)	0.228 (*0.114*)	**0.138** (*0.094*)	0.230 (*0.115*)	**0.139**$^{+}$ (*0.095*)
Tp$_2$ 128	1.085 (*0.478*)	0.495 (*0.395*)	0.627 (*0.362*)	**0.453** (*0.332*)	0.649 (*0.375*)	**0.469** (*0.343*)	0.666 (*0.384*)	**0.481**$^{+}$ (*0.351*)
Tp$_2$ 256	0.740 (*0.314*)	0.312 (*0.230*)	0.441 (*0.242*)	**0.291** (*0.209*)	0.453 (*0.250*)	**0.297** (*0.214*)	0.462 (*0.255*)	**0.302**$^{+}$ (*0.219*)
Tp$_2$ 512	0.511 (*0.210*)	0.204 (*0.147*)	0.303 (*0.157*)	**0.190** (*0.133*)	0.310 (*0.164*)	**0.195** (*0.138*)	0.315 (*0.166*)	**0.197**$^{+}$ (*0.140*)
Tp$_2$ 1024	0.352 (*0.140*)	0.140 (*0.097*)	0.215 (*0.106*)	**0.127** (*0.085*)	0.218 (*0.109*)	**0.129** (*0.088*)	0.221 (*0.111*)	**0.130**$^{+}$ (*0.090*)
Tp$_3$ 128	1.258 (*0.571*)	0.509 (*0.438*)	0.641 (*0.380*)	**0.446** (*0.331*)	0.662 (*0.396*)	**0.463** (*0.346*)	0.677 (*0.406*)	**0.476**$^{+}$ (*0.355*)
Tp$_3$ 256	0.897 (*0.401*)	0.306 (*0.249*)	0.443 (*0.255*)	**0.273** (*0.204*)	0.458 (*0.267*)	**0.282** (*0.212*)	0.466 (*0.272*)	**0.288**$^{+}$ (*0.218*)
Tp$_3$ 512	0.634 (*0.280*)	0.203 (*0.168*)	0.308 (*0.166*)	**0.177** (*0.129*)	0.316 (*0.174*)	**0.180** (*0.132*)	0.321 (*0.177*)	**0.183**$^{+}$ (*0.134*)
Tp$_3$ 1024	0.444 (*0.193*)	0.138 (*0.111*)	0.218 (*0.114*)	**0.119** (*0.083*)	0.221 (*0.117*)	**0.120** (*0.084*)	0.223 (*0.118*)	**0.121**$^{+}$ (*0.085*)

Table 3.7: (S2): Causal GPDC. Average value and standard deviation in parentheses of the L$_2$-norm error distribution (1000 simulations) for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC$_{un}$, and mBTS-TD-BIC$_{un}$ with p$_{max} = 5, 7, 9$. The lower average error is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC$_{un}$ 9, mBTS-BIC$_{un}$ 5, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

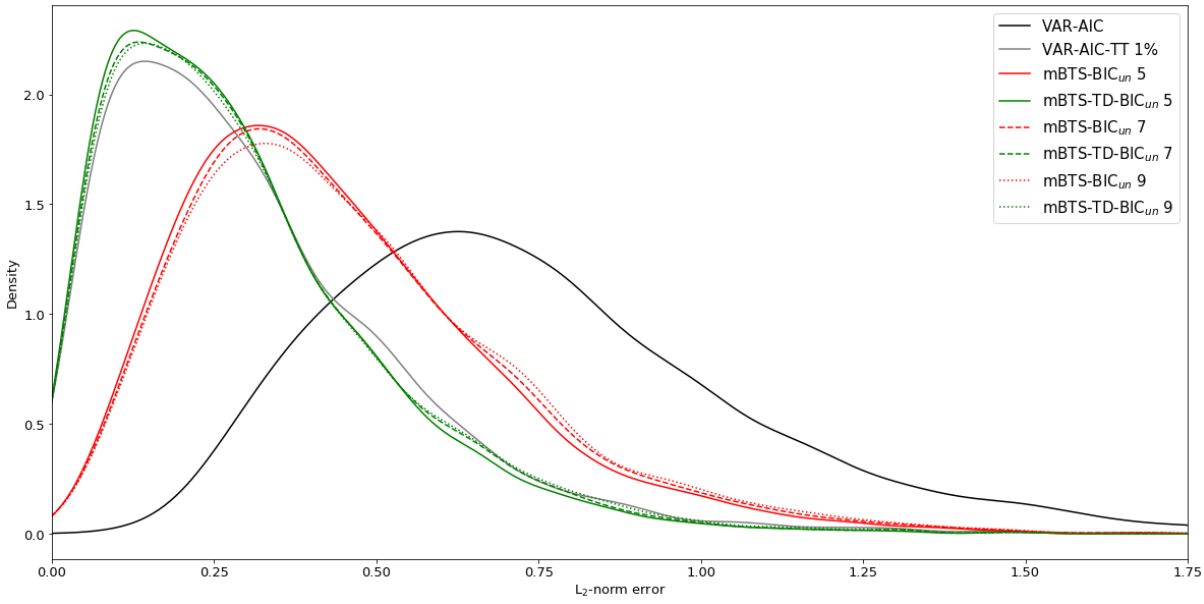


Figure 3.15: (S2): Causal GPDC. L$_2$-norm error distribution (1000 simulations) with $T = 256$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}} =$ Tp$_2$ for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC$_{un}$, and mBTS-TD-BIC$_{un}$ with p$_{max} = 5, 7, 9$.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 5	**mBTS-TD BIC_{un} 5**	mBTS BIC_{un} 7	**mBTS-TD BIC_{un} 7**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	0.387 (*0.289*)	0.023 (*0.125*)	0.010^{+} (*0.056*)	**0.006** (*0.046*)	0.014 (*0.077*)	**0.008** (*0.061*)	0.020 (*0.103*)	**0.011** (*0.082*)
Id 256	0.189 (*0.135*)	0.011 (*0.059*)	0.004 (*0.023*)	**0.002** (*0.016*)	0.006 (*0.035*)	**0.003** (*0.022*)	0.008 (*0.044*)	$\mathbf{0.004}^{+}$ (*0.030*)
Id 512	0.092 (*0.066*)	0.004 (*0.026*)	0.001 (*0.011*)	**0** (*0.008*)	0.002 (*0.014*)	**0** (*0.008*)	0.002 (*0.018*)	$\mathbf{0.001}^{+}$ (*0.010*)
Id 1024	0.045 (*0.032*)	0.002 (*0.013*)	0 (*0.004*)	**0** (*0.002*)	0.001 (*0.007*)	**0** (*0.004*)	0.001 (*0.008*)	$\mathbf{0}^{+}$ (*0.005*)
Tp_2 128	0.510 (*0.375*)	0.030 (*0.156*)	0.009^{+} (*0.053*)	**0.004** (*0.039*)	0.013 (*0.068*)	**0.007** (*0.052*)	0.017 (*0.091*)	**0.009** (*0.066*)
Tp_2 256	0.258 (*0.198*)	0.012 (*0.066*)	0.003 (*0.022*)	**0.001** (*0.017*)	0.004 (*0.027*)	**0.002** (*0.019*)	0.006 (*0.036*)	$\mathbf{0.002}^{+}$ (*0.021*)
Tp_2 512	0.128 (*0.097*)	0.005 (*0.033*)	0.002 (*0.012*)	**0** (*0.005*)	0.002 (*0.015*)	**0** (*0.007*)	0.003 (*0.019*)	$\mathbf{0.001}^{+}$ (*0.009*)
Tp_2 1024	0.063 (*0.047*)	0.002 (*0.015*)	0.001 (*0.009*)	**0** (*0.003*)	0.002 (*0.011*)	**0** (*0.004*)	0.002 (*0.012*)	$\mathbf{0}^{+}$ (*0.004*)
Tp_3 128	0.729 (*0.515*)	0.043 (*0.236*)	0.009 (*0.059*)	**0.004** (*0.044*)	0.013 (*0.074*)	**0.006** (*0.057*)	0.016 (*0.095*)	$\mathbf{0.009}^{+}$ (*0.075*)
Tp_3 256	0.410 (*0.314*)	0.018 (*0.105*)	0.002 (*0.020*)	**0.001** (*0.011*)	0.004 (*0.029*)	**0.001** (*0.017*)	0.005 (*0.039*)	$\mathbf{0.002}^{+}$ (*0.023*)
Tp_3 512	0.214 (*0.167*)	0.008 (*0.051*)	0.001 (*0.013*)	**0** (*0.006*)	0.002 (*0.016*)	**0** (*0.008*)	0.002 (*0.020*)	$\mathbf{0.001}^{+}$ (*0.010*)
Tp_3 1024	0.107 (*0.086*)	0.003 (*0.021*)	0.001 (*0.011*)	**0** (*0.002*)	0.001 (*0.012*)	**0** (*0.003*)	0.001 (*0.013*)	$\mathbf{0}^{+}$ (*0.004*)

Table 3.8: (S2): Non-causal GPDC. Average value and standard deviation in parentheses of the L_2-norm error distribution (1000 simulations) for the non-causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 5, 7, 9$. The lower average error is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 5, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

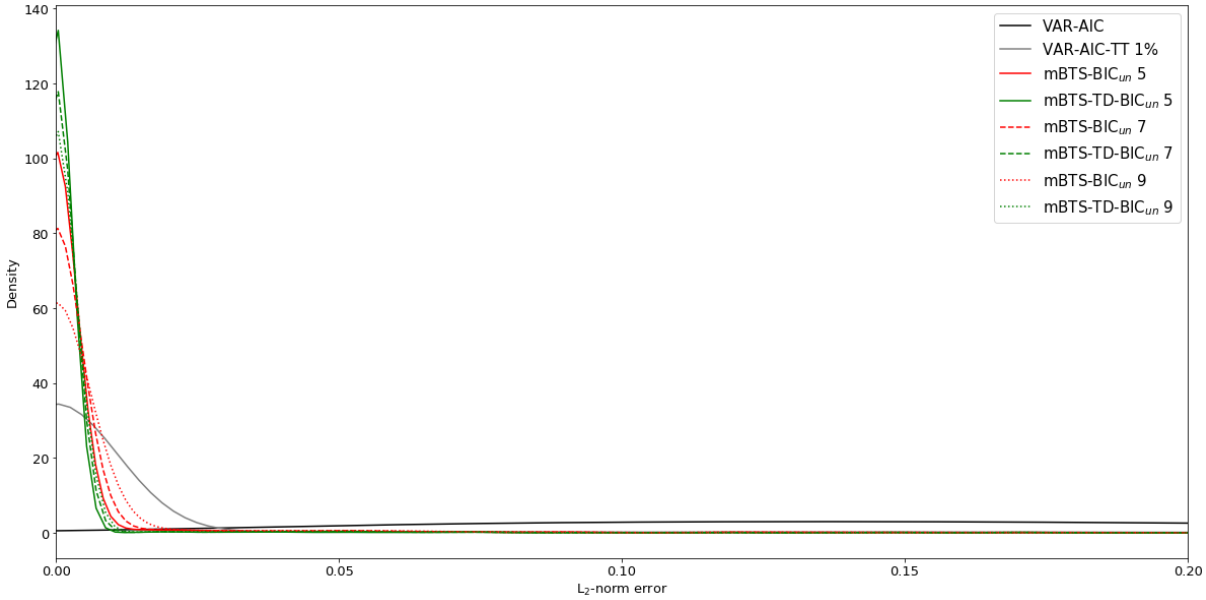


Figure 3.16: (S2): Non-causal GPDC. L_2-norm error distribution (1000 simulations) with $T = 256$ and $\mathbf{\Sigma_\epsilon} = Tp_2$ for the non-causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 5, 7, 9$.

For the system (S2), the same conclusions as for the system (S) can be drawn regarding both causal (Table 3.7 and Fig. 3.15) and non-causal (Table 3.8 and Fig. 3.16) terms:

- By considering only the same p_{max} for the two methods, denoted in bold in Tables 3.7 and 3.8, mBTS-TD clearly outperforms from mBTS by providing lower average errors for each setting.
- Whatever p_{max} is selected, the mBTS-TD error distributions are more concentrated, with a lower fat tail.
- mBTS-TD-BIC$_{un}$9 only in two cases has higher errors than mBTS-BIC$_{un}$5 and VAR-AIC-TT 1% for causal and non-causal terms, denoted with a superscript symbol $^{+}$ in Tables 3.7 and 3.8.

3.7.1.3 S2: Causal structure identification

Tables 3.9 and 3.10 report the average value of FM (Table 3.9) and HD (Table 3.10) for each setting.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC$_{un}$ 5	**mBTS-TD BIC$_{un}$ 5**	mBTS BIC$_{un}$ 7	**mBTS-TD BIC$_{un}$ 7**	mBTS BIC$_{un}$ 9	**mBTS-TD BIC$_{un}$ 9**
Id 128	0.511	0.949	0.953	**0.972**	0.944	**0.965**	0.935	**0.958**$^{+}$
Id 256	0.534	0.948	0.976	**0.986**	0.968	**0.981**	0.962	**0.977**$^{+}$
Id 512	0.612	0.972	0.994	**0.997**	0.991	**0.997**	0.989	**0.996**$^{+}$
Id 1024	0.776	0.988	0.999^{+}	**1**	0.997	**0.999**	0.995	**0.998**
Tp2 128	0.506	0.945	0.955	**0.978**	0.945	**0.970**	0.937	**0.964**$^{+}$
Tp2 256	0.521	0.952	0.980	**0.991**	0.974	**0.988**	0.969	**0.985**$^{+}$
Tp2 512	0.573	0.970	0.989	**0.996**	0.987	**0.995**	0.984	**0.994**$^{+}$
Tp2 1024	0.699	0.985	0.991	**1**	0.990	**0.999**	0.988	**0.999**$^{+}$
Tp3 128	0.503	0.942	0.958	**0.980**	0.950	**0.973**	0.943	**0.967**$^{+}$
Tp3 256	0.509	0.950	0.981	**0.992**	0.977	**0.990**	0.973	**0.987**$^{+}$
Tp3 512	0.533	0.965	0.989	**0.997**	0.986	**0.996**	0.984	**0.995**$^{+}$
Tp3 1024	0.605	0.982	0.992	**1**	0.990	**1**	0.989	**0.999**$^{+}$

Table 3.9: (S2): Average value of the F-measure (FM) over 1000 simulations of the GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC$_{un}$, and mBTS-TD-BIC$_{un}$ with $p_{max} = 5, 7, 9$. FM ranges from 0 to 1. If FM = 1 there is perfect identification of the pairs of true causality, whereas if FM = 0 no true causality is detected. The lower average value is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC$_{un}$ 9, mBTS-BIC$_{un}$ 5, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 5	**mBTS-TD BIC_{un} 5**	mBTS BIC_{un} 7	**mBTS-TD BIC_{un} 7**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	7.669	0.427	0.393	**0.229**	0.477	**0.287**	0.558	**0.350**$^+$
Id 256	6.977	0.435	0.199	**0.110**	0.261	**0.152**	0.314	**0.192**$^+$
Id 512	5.075	0.228	0.050	**0.021**	0.069	**0.026**	0.089	**0.034**$^+$
Id 1024	2.306	0.095	0.011$^+$	**0.003**	0.025	**0.010**	0.038	**0.013**
Tp2 128	7.822	0.463	0.379	**0.184**	0.465	**0.249**	0.540	**0.303**$^+$
Tp2 256	7.362	0.403	0.166	**0.075**	0.210	**0.096**	0.255	**0.121**$^+$
Tp2 512	5.964	0.251	0.087	**0.034**	0.104	**0.042**	0.127	**0.050**$^+$
Tp2 1024	3.443	0.125	0.069	**0.004**	0.082	**0.007**	0.095	**0.008**$^+$
Tp3 128	7.915	0.495	0.351	**0.164**	0.420	**0.218**	0.488	**0.271**$^+$
Tp3 256	7.729	0.419	0.151	**0.061**	0.185	**0.080**	0.219	**0.103**$^+$
Tp3 512	7.019	0.287	0.092	**0.028**	0.112	**0.035**	0.130	**0.043**$^+$
Tp3 1024	5.224	0.147	0.068	**0**	0.077	**0.001**	0.090	**0.007**$^+$

Table 3.10: (S2): Average value of Hamming Distance (HD) over 1000 simulations of the GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 5, 7, 9$. HD ranges from 0 to $m(m-1)$, where $m = 4$. If HD = 0 there is perfect identification, whereas if HD = 12 all pairs are misclassified. The lower average value is highlighted for each setting and p_{max}. The superscript symbol $^+$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 5, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

Here again, for the two measures FM (Table 3.9) and HD (Table 3.10), mBTS-TD provides on average a better identification of the true causal structure than mBTS for each setting by considering only the same p_{max} highlighted in boldface in Tables 3.9 and 3.10. As previously for the error distributions, mBTS-TD-$BIC_{un}9$ outperforms mBTS-$BIC_{un}5$ and VAR-AIC-TT 1%, denoted with a superscript symbol $^+$ in Tables 3.9 and 3.10.

3.7.2 GPDC results on a high-dimensional parsimonious system

Finally, we extend the first results obtained with the systems (S) and (S2) by analyzing a high-dimensional parsimonious VAR(4) model on $m = 12$ used in [215]. This system (S3) is even more parsimonious than the two other ones by admitting only 4% of the m^2p coefficients being non-zero.

The VAR model [215] is as follows:

$$
\begin{aligned}
x_1(t) &= 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + \epsilon_1(t)\\
x_2(t) &= -0.5x_1(t-2) + \epsilon_2(t)\\
x_3(t) &= 0.8x_1(t-4) - 0.5x_2(t-2) + \epsilon_3(t)\\
x_4(t) &= -1.2x_3(t-1) + 0.25\sqrt{2}x_4(t-1) + 0.65\sqrt{2}x_1(t-1) + \epsilon_4(t)\\
x_5(t) &= -0.25\sqrt{2}x_4(t-1) + 0.5\sqrt{6}x_6(t-1) + \epsilon_5(t)\\
x_6(t) &= -0.6\sqrt{2}x_4(t-2) + 0.8x_5(t-1) + \epsilon_6(t)\\
x_7(t) &= 0.8x_5(t-3) + 0.7x_6(t-1) + 0.8x_{10}(t-2) + 0.7x_1(t-2) + \epsilon_7(t)\\
x_8(t) &= 0.85x_6(t-2) + \epsilon_8(t)\\
x_9(t) &= 1.15x_1(t-1) - 0.9025x_6(t-2) + 0.7x_7(t-2) + \epsilon_9(t)\\
x_{10}(t) &= 0.5x_7(t-2) + \epsilon_{10}(t)\\
x_{11}(t) &= 0.8x_6(t-2) + 0.6x_9(t-3) + \epsilon_{11}(t)\\
x_{12}(t) &= -0.5x_7(t-3) + \epsilon_{12}(t)
\end{aligned}
\tag{S3}
$$

3.7.2.1 S3: GPDC errors

Figs. 3.17 and 3.18 show GPDC errors in causal (Fig. 3.17) and non-causal terms (Fig. 3.18), confirming previous results for the VAR coefficients.

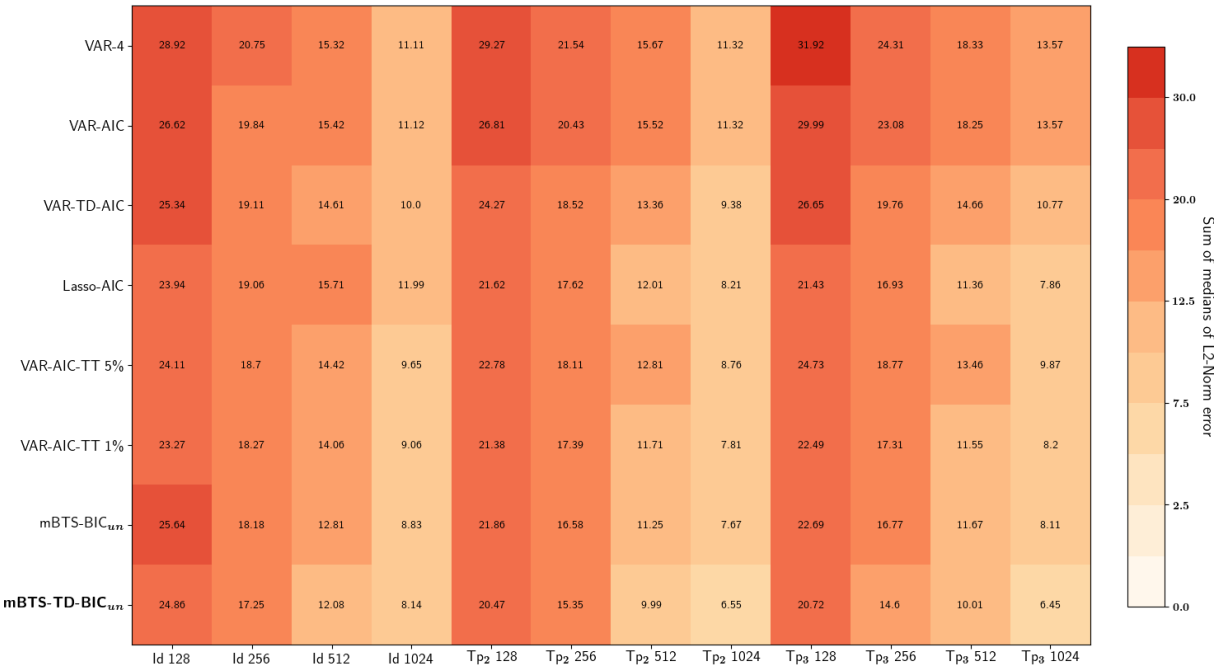


Figure 3.17: (S3): Sum of medians over 1000 simulations of the L_2-norm error in the causal GPDC (3.5), estimated using the subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the causal GPDC.

In Fig. 3.17, the results for the GPDC causal terms are quite similar to S and S2 systems. This confirms the idea that subset selection methods are better suited than standard VAR. However, for this high-dimensional system (S3) mBTS-TD performs better than the others for all settings except Id 128 (TT procedure and Lasso).

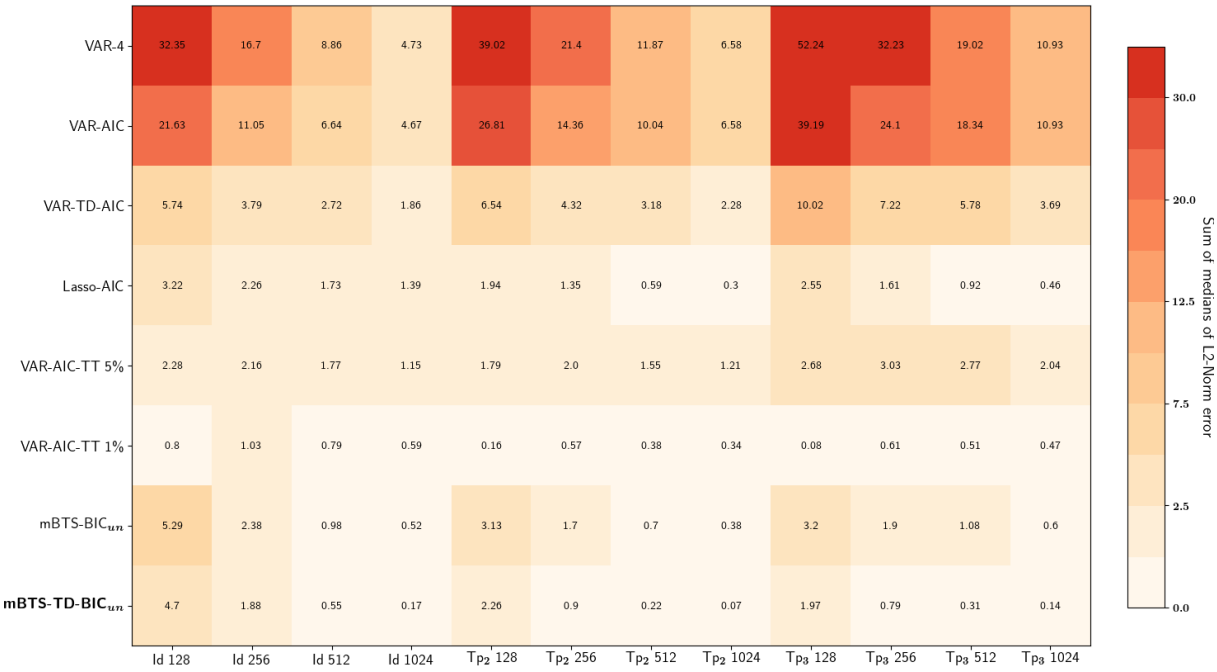


Figure 3.18: (S3): Sum of medians over 1000 simulations of the L_2-norm error in the non-causal GPDC (3.5), estimated using the subset selection methods and standard VAR estimations. An increase in the median implies a deterioration in the estimate of the non-causal GPDC.

For the errors in the non-causal terms shown in Fig. 3.18, the results are clear. The six subset selection methods greatly reduce errors compared to VAR-AIC or VAR-3. Moreover, even in this case, the mBTS-TD method provides the most interesting results. Indeed, even if for some settings TT 1% has better results, it is too restrictive and therefore also removes causal coefficients increasing the errors on causal terms (see Fig. 3.17).

3.7.2.2 S3: GPDC error distributions

In Table 3.11, we report for the system (S3) the average value and standard deviation of the L_2-norm error distributions for the causal GPDC and in Fig. 3.19 we provide an example of the L_2-norm error distribution with $T = 256$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}} = \text{Tp}_2$. Table 3.12 and Fig. 3.20 exhibit the same results for the non-causal terms.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC$_{un}$ 4	**mBTS-TD BIC$_{un}$ 4**	mBTS BIC$_{un}$ 6	**mBTS-TD BIC$_{un}$ 6**	mBTS BIC$_{un}$ 9	**mBTS-TD BIC$_{un}$ 9**
Id 128	1.339 (*0.892*)	1.196^{+} (*0.917*)	1.266 (*1.098*)	**1.228** (*1.099*)	1.272 (*1.098*)	**1.232** (*1.099*)	1.277 (*1.099*)	**1.236** (*1.099*)
Id 256	1.011 (*0.691*)	0.934 (*0.731*)	0.943 (*0.862*)	**0.908** (*0.868*)	0.947 (*0.861*)	**0.909** (*0.866*)	0.949 (*0.862*)	**0.910**$^{+}$ (*0.867*)
Id 512	0.792 (*0.553*)	0.724 (*0.597*)	0.696 (*0.672*)	**0.665** (*0.677*)	0.697 (*0.671*)	**0.666** (*0.677*)	0.699 (*0.671*)	**0.667**$^{+}$ (*0.677*)
Id 1024	0.582 (*0.424*)	0.488 (*0.468*)	0.492 (*0.505*)	**0.468** (*0.509*)	0.493 (*0.504*)	**0.468** (*0.507*)	0.494 (*0.504*)	**0.469**$^{+}$ (*0.508*)
Tp$_2$ 128	1.331 (*0.861*)	1.105 (*0.862*)	1.130 (*0.964*)	**1.073** (*0.968*)	1.137 (*0.965*)	**1.078** (*0.970*)	1.142 (*0.966*)	**1.082**$^{+}$ (*0.971*)
Tp$_2$ 256	1.021 (*0.658*)	0.887 (*0.686*)	0.870 (*0.769*)	**0.819** (*0.770*)	0.875 (*0.769*)	**0.822** (*0.769*)	0.877 (*0.770*)	**0.824**$^{+}$ (*0.770*)
Tp$_2$ 512	0.790 (*0.504*)	0.632 (*0.544*)	0.613 (*0.568*)	**0.564** (*0.573*)	0.616 (*0.567*)	**0.565** (*0.573*)	0.617 (*0.567*)	**0.566**$^{+}$ (*0.573*)
Tp$_2$ 1024	0.582 (*0.372*)	0.426 (*0.411*)	0.423 (*0.406*)	**0.380** (*0.415*)	0.427 (*0.407*)	**0.382** (*0.416*)	0.427 (*0.407*)	**0.382**$^{+}$ (*0.416*)
Tp$_3$ 128	1.458 (*0.935*)	1.152 (*0.886*)	1.158 (*0.961*)	**1.087** (*0.965*)	1.164 (*0.962*)	**1.093** (*0.965*)	1.169 (*0.964*)	**1.097**$^{+}$ (*0.967*)
Tp$_3$ 256	1.134 (*0.710*)	0.890 (*0.713*)	0.877 (*0.756*)	**0.796** (*0.755*)	0.883 (*0.756*)	**0.800** (*0.755*)	0.885 (*0.756*)	**0.801**$^{+}$ (*0.755*)
Tp$_3$ 512	0.917 (*0.544*)	0.644 (*0.564*)	0.647 (*0.567*)	**0.576** (*0.573*)	0.651 (*0.567*)	**0.578** (*0.574*)	0.652 (*0.567*)	**0.579**$^{+}$ (*0.574*)
Tp$_3$ 1024	0.706 (*0.419*)	0.481 (*0.447*)	0.470 (*0.423*)	**0.404** (*0.430*)	0.474 (*0.423*)	**0.406** (*0.433*)	0.475 (*0.424*)	**0.407**$^{+}$ (*0.433*)

Table 3.11: (S3): Causal GPDC. Average value and standard deviation in parentheses of the L_2-norm error distribution (1000 simulations) for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC$_{un}$, and mBTS-TD-BIC$_{un}$ with $p_{max} = 4, 6, 9$. The lower average error is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC$_{un}$ 9, mBTS-BIC$_{un}$ 4, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

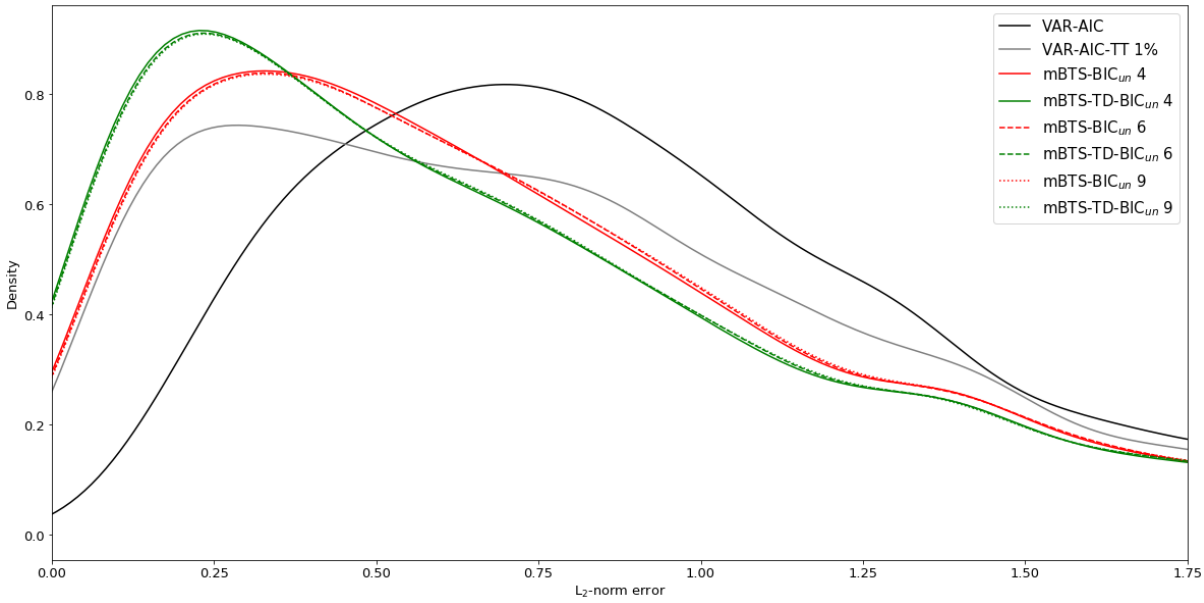


Figure 3.19: (S3): Causal GPDC. L_2-norm error distribution (1000 simulations) with $T = 256$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}} = \text{Tp}_2$ for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC$_{un}$, and mBTS-TD-BIC$_{un}$ with $p_{max} = 4, 6, 9$.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 4	**mBTS-TD BIC_{un} 4**	mBTS BIC_{un} 6	**mBTS-TD BIC_{un} 6**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	0.313 (*0.409*)	0.125 (*0.418*)	0.092 (*0.466*)	**0.088** (*0.479*)	0.093 (*0.466*)	**0.088** (*0.479*)	0.093 (*0.467*)	**0.089**$^{+}$ (*0.480*)
Id 256	0.150 (*0.189*)	0.077 (*0.243*)	0.059 (*0.318*)	**0.056** (*0.327*)	0.059 (*0.318*)	**0.056** (*0.327*)	0.059 (*0.318*)	**0.057**$^{+}$ (*0.327*)
Id 512	0.089 (*0.118*)	0.048 (*0.145*)	0.034 (*0.178*)	**0.033** (*0.185*)	0.034 (*0.178*)	**0.033** (*0.185*)	0.035 (*0.178*)	**0.033**$^{+}$ (*0.184*)
Id 1024	0.059 (*0.075*)	0.028 (*0.087*)	0.019 (*0.088*)	**0.018** (*0.092*)	0.019 (*0.088*)	**0.018** (*0.092*)	0.019 (*0.088*)	**0.018**$^{+}$ (*0.092*)
Tp_2 128	0.367 (*0.422*)	0.120 (*0.408*)	0.077 (*0.383*)	**0.070** (*0.390*)	0.078 (*0.385*)	**0.071** (*0.391*)	0.079 (*0.386*)	**0.071**$^{+}$ (*0.392*)
Tp_2 256	0.190 (*0.220*)	0.080 (*0.251*)	0.051 (*0.257*)	**0.046** (*0.263*)	0.051 (*0.258*)	**0.046** (*0.264*)	0.052 (*0.258*)	**0.046**$^{+}$ (*0.264*)
Tp_2 512	0.130 (*0.148*)	0.052 (*0.161*)	0.027 (*0.141*)	**0.024** (*0.147*)	0.028 (*0.141*)	**0.024** (*0.147*)	0.028 (*0.141*)	**0.024**$^{+}$ (*0.147*)
Tp_2 1024	0.082 (*0.096*)	0.032 (*0.098*)	0.016 (*0.078*)	**0.013** (*0.080*)	0.016 (*0.078*)	**0.013** (*0.080*)	0.016 (*0.078*)	**0.013**$^{+}$ (*0.080*)
Tp_3 128	0.499 (*0.508*)	0.163 (*0.491*)	0.088 (*0.398*)	**0.078** (*0.400*)	0.091 (*0.402*)	**0.080** (*0.403*)	0.092 (*0.403*)	**0.081**$^{+}$ (*0.404*)
Tp_3 256	0.299 (*0.307*)	0.114 (*0.333*)	0.060 (*0.272*)	**0.052** (*0.273*)	0.062 (*0.275*)	**0.053** (*0.275*)	0.062 (*0.276*)	**0.053**$^{+}$ (*0.275*)
Tp_3 512	0.221 (*0.215*)	0.081 (*0.232*)	0.041 (*0.183*)	**0.034** (*0.184*)	0.042 (*0.184*)	**0.035** (*0.185*)	0.042 (*0.185*)	**0.035**$^{+}$ (*0.185*)
Tp_3 1024	0.137 (*0.150*)	0.055 (*0.158*)	0.028 (*0.123*)	**0.022** (*0.127*)	0.028 (*0.125*)	**0.023** (*0.128*)	0.028 (*0.125*)	**0.023**$^{+}$ (*0.127*)

Table 3.12: (S3): Non-causal GPDC. Average value and standard deviation in parentheses of the L_2-norm error distribution (1000 simulations) for the causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 4, 6, 9$. The lower average error is highlighted for each setting and p_{max}. The superscript symbol $^{+}$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 4, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

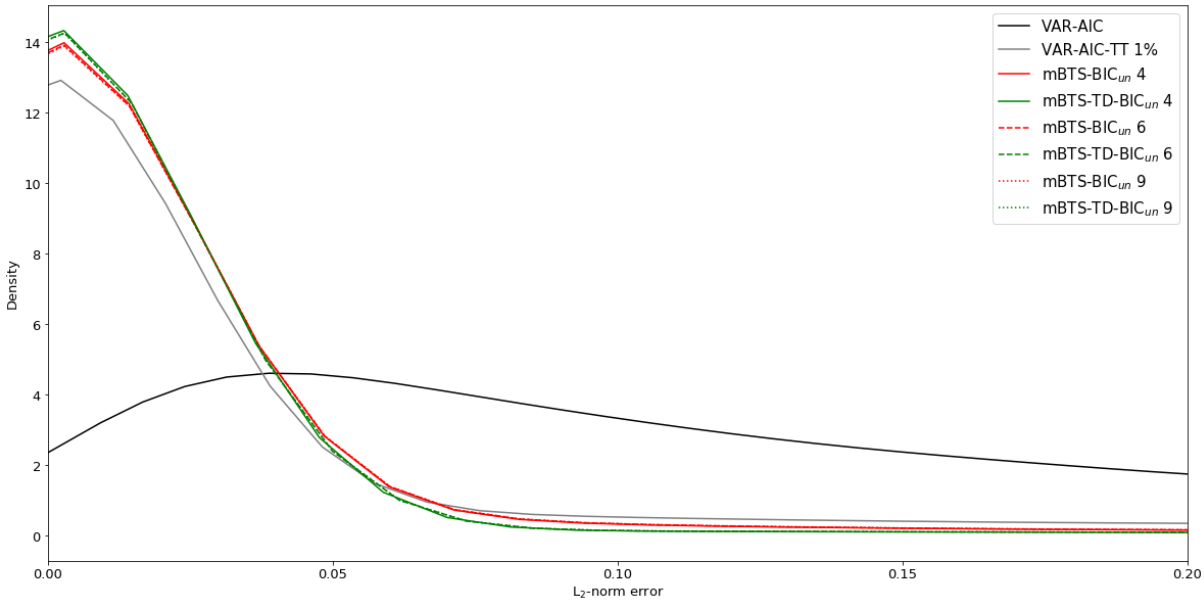


Figure 3.20: (S3): Non-causal GPDC. L_2-norm error distribution (1000 simulations) with $T = 256$ and $\mathbf{\Sigma_\epsilon} = Tp_2$ for the non-causal GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 4, 6, 9$.

For the system (S3), the same conclusions as for the system (S) can be drawn regarding both causal (Table 3.11 and Fig. 3.19) and non-causal (Table 3.12 and Fig. 3.20) terms:

- By taking into account only the same p_{max} for the two methods, denoted in bold in Tables 3.11 and 3.12, mBTS-TD clearly stands out from mBTS by providing lower average errors for each setting.
- Whatever p_{max} is selected, the mBTS-TD error distributions are more concentrated, with a lower fat tail.
- mBTS-TD-BIC$_{un}$9 admits always lower errors than mBTS-BIC$_{un}$4 and VAR-AIC-TT 1% for the causal and non-causal GPDC (except for Id 128), denoted with a superscript symbol $^+$ in Tables 3.11 and 3.12.

3.7.2.3 S3: Causal structure identification

Tables 3.13 and 3.14 report the average value of FM (Table 3.13) and HD (Table 3.14) for each setting.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC$_{un}$ 4	**mBTS-TD BIC$_{un}$ 4**	mBTS BIC$_{un}$ 6	**mBTS-TD BIC$_{un}$ 6**	mBTS BIC$_{un}$ 9	**mBTS-TD BIC$_{un}$ 9**
Id 128	0.332	0.670	0.749	**0.784**	0.745	**0.781**	0.743	**0.779**$^+$
Id 256	0.395	0.692	0.804	**0.835**	0.801	**0.833**	0.799	**0.831**$^+$
Id 512	0.487	0.727	0.834	**0.854**	0.832	**0.853**	0.832	**0.853**$^+$
Id 1024	0.576	0.798	0.870	**0.878**	0.869	**0.877**	0.868	**0.877**$^+$
Tp2 128	0.314	0.680	0.769	**0.816**	0.765	**0.813**	0.762	**0.811**$^+$
Tp2 256	0.361	0.693	0.814	**0.854**	0.810	**0.852**	0.809	**0.851**$^+$
Tp2 512	0.394	0.724	0.856	**0.887**	0.852	**0.885**	0.851	**0.885**$^+$
Tp2 1024	0.478	0.782	0.888	**0.913**	0.885	**0.912**	0.884	**0.912**$^+$
Tp3 128	0.298	0.643	0.747	**0.796**	0.741	**0.791**	0.739	**0.788**$^+$
Tp3 256	0.319	0.659	0.789	**0.837**	0.784	**0.833**	0.782	**0.832**$^+$
Tp3 512	0.323	0.678	0.818	**0.863**	0.813	**0.859**	0.811	**0.858**$^+$
Tp3 1024	0.377	0.710	0.840	**0.882**	0.837	**0.880**	0.836	**0.881**$^+$

Table 3.13: (S3): Average value of the F-measure (FM) over 1000 simulations of the GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC$_{un}$, and mBTS-TD-BIC$_{un}$ with p$_{max} = 4, 6, 9$. FM ranges from 0 to 1. If FM = 1 there is perfect identification of the pairs of true causality, whereas if FM = 0 no true causality is detected. The lower average value is highlighted for each setting and p_{max}. The superscript symbol $^+$ indicates the lowest average error among mBTS-TD-BIC$_{un}$ 9, mBTS-BIC$_{un}$ 4, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

	VAR-AIC	VAR-AIC TT 1%	mBTS BIC_{un} 4	**mBTS-TD BIC_{un} 4**	mBTS BIC_{un} 6	**mBTS-TD BIC_{un} 6**	mBTS BIC_{un} 9	**mBTS-TD BIC_{un} 9**
Id 128	83.038	19.508	12.421	**10.164**	12.680	**10.351**	12.849	**10.473**$^+$
Id 256	63.810	18.365	9.794	**7.931**	9.951	**8.051**	10.053	**8.124**$^+$
Id 512	44.219	15.695	8.227	**7.072**	8.348	**7.137**	8.397	**7.155**$^+$
Id 1024	30.943	10.646	6.242	**5.827**	6.334	**5.890**	6.360	**5.897**$^+$
Tp2 128	90.667	18.788	11.653	**8.717**	11.953	**8.890**	12.129	**9.033**$^+$
Tp2 256	73.893	18.233	9.349	**6.955**	9.568	**7.108**	9.690	**7.167**$^+$
Tp2 512	64.573	15.962	7.034	**5.292**	7.239	**5.406**	7.308	**5.419**$^+$
Tp2 1024	45.914	11.714	5.286	**4.020**	5.439	**4.040**	5.502	**4.052**$^+$
Tp3 128	98.307	21.922	13.062	**9.814**	13.426	**10.097**	13.610	**10.242**$^+$
Tp3 256	89.386	21.190	10.911	**7.889**	11.238	**8.116**	11.379	**8.191**$^+$
Tp3 512	87.870	19.868	9.265	**6.604**	9.554	**6.807**	9.663	**6.855**$^+$
Tp3 1024	69.402	17.108	7.958	**5.587**	8.189	**5.693**	8.231	**5.683**$^+$

Table 3.14: (S3): Average value of Hamming Distance (HD) over 1000 simulations of the GPDC, estimated using VAR-AIC, VAR-AIC-TT 1%, mBTS-BIC_{un}, and mBTS-TD-BIC_{un} with $p_{max} = 4, 6, 9$. HD ranges from 0 to $m(m-1)$, where $m = 12$. If HD = 0 there is perfect identification, whereas if HD = 132 all pairs are misclassified. The lower average value is highlighted for each setting and p_{max}. The superscript symbol $^+$ indicates the lowest average error among mBTS-TD-BIC_{un} 9, mBTS-BIC_{un} 4, VAR-AIC-TT 1% to underline the efficiency of the mBTS-TD approach even for a large p_{max}.

Here again for the two measures FM (Table 3.13) and HD (Table 3.14), mBTS-TD provides on average a better identification of the true causal structure than mBTS for each setting by taking into account only the same p_{max} highlighted in boldface in Tables 3.13 and 3.14. As previously for the error distributions, mBTS-TD-$BIC_{un}9$ outperforms mBTS-$BIC_{un}4$ and VAR-AIC-TT 1%, denoted with a superscript symbol $^+$ in Tables 3.13 and 3.14.

3.7.3 Country indices dataset

Index	Currency
Australia	AUD
Austria	EUR
Belgium	EUR
Brazil	BRL
Canada	CAD
Chile	CLP
China	HKD
Colombia	COP
Denmark	DKK
Finland	EUR
France	EUR
Germany	EUR
Hong Kong	HKD
India	INR
Indonesia	IDR
Ireland	EUR
Israel	ILS
Italy	EUR
Japan	JPY
Korea	KRW
Malaysia	MYR
Mexico	MXN
Netherlands	EUR
New Zealand	NZD
Norway	NOK
Peru	PEN
Philippines	PHP
Poland	PLN
Portugal	EUR
Russia	RUB
Singapore	SGD
South Africa	ZAR
Spain	EUR
Sweden	SEK
Switzerland	CHF
Taiwan	TWD
Thailand	THB
Turkey	TRY
United Kingdom	GBP
United States	USD

Table 3.15: Country equity indices in the MSCI ACWI (All Country World Index)

3.7.4 GARCH model

Let $x(t)$ be a zero-mean stationary process admitting the following GARCH(1,1) representation:

$$x(t) = \sqrt{h(t)}\,\epsilon(t)$$
$$h(t) = \beta_0 + \beta_1\, x^2(t-1) + \beta_2\, h(t-1)$$

where $h(t)$ is the conditional variance, β_0, β_1 and β_2 are the coefficients and $\epsilon(t)$ is the white noise with $\epsilon(t) \sim \mathcal{N}(0,\, 1)$. Moreover, the parameters are estimated by maximizing the conditional log-likelihood.

In order to remove the heteroskedasticity, $x(t)$ is standardized as follows:

$$\tilde{x}(t) = \frac{x(t)}{\sqrt{h(t)}}$$

3.7.5 EW portfolio results with additional exclusion levels

In Table 3.16, Table 3.17 and Table 3.18, we provide the performance indicators when eight, fifteen and twenty assets are excluded respectively. For all exclusion levels tested, our proposed methodology outperforms either the classical VAR estimation or when the allocation is applied on the whole universe.

EW Portfolios 8 excluded assets	Annualized Return	Annualized Volatility	Ratio Return/Volatility	Max Drawdown
mBTS-TD GPDC	9.99%	16.68%	0.60	62.65%
VAR-AIC GPDC 0.02	9.56%	16.70%	0.57	61.74%
EW	9.35%	16.76%	0.56	61.90%
Non-selected assets (VAR-AIC GPDC 0.02)	8.28%	18.32%	0.45	63.12%
Non-selected assets (mBTS-TD GPDC)	6.61%	18.45%	0.36	58.99%

Table 3.16: Performance indicators for EW portfolios with 8 excluded assets from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility)

EW Portfolios 15 excluded assets	Annualized Return	Annualized Volatility	Ratio Return/Volatility	Max Drawdown
mBTS-TD GPDC	10.12%	16.78%	0.60	61.62%
VAR-AIC GPDC 0.04	9.39%	16.63%	0.56	60.16%
EW	9.35%	16.76%	0.56	61.90%
Non-selected assets (mBTS-TD GPDC)	7.96%	17.59%	0.45	62.47%
Non-selected assets (VAR-AIC GPDC 0.04)	7.35%	18.43%	0.40	66.32%

Table 3.17: Performance indicators for EW portfolios with 15 excluded assets from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility)

EW Portfolios 20 excluded assets	Annualized Return	Annualized Volatility	Ratio Return/Volatility	Max Drawdown
mBTS-TD GPDC	10.44%	17.13%	0.61	61.49%
VAR-AIC GPDC 0.03	9.86%	16.71%	0.59	60.82%
EW	9.35%	16.76%	0.56	61.90%
Non-selected assets (VAR-AIC GPDC 0.03)	8.69%	17.47%	0.50	63.02%
Non-selected assets (mBTS-TD GPDC)	8.20%	17.14%	0.48	62.50%

Table 3.18: Performance indicators for EW portfolios with 20 excluded assets from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility)

Chapter 4

Asset selection process: A new perspective from frequency causality measure and clustering coefficient

This chapter is based on a working paper in collaboration with Christophe Chorro, Emmanuelle Jay and Philippe De Peretti.

Abstract

Modeling the structure and dynamics of financial markets is of major importance for handling the asset allocation problem. In classical portfolio allocation strategies (Mean-Variance, Equal Risk Contribution, Minimum Variance, Maximum Variety, etc.) the dependency structure among financial assets is not taken into account, leading to incomplete risk assessments on the investment universe. To overcome this point, the network approach allows to identify the relationships between assets and make it possible to remove systemic/influenced assets before allocating portfolios. However, common approaches to recover the network topology (e.g. the sample correlation matrix or Granger non-causality tests) lead to partial information never providing a directed weighted network. The goal of this paper is twofold. First, we recover the financial network topology using the GPDC measure (Generalized Partial Directed Coherence) in which are plugged the Vector Autoregressive (VAR) coefficients estimated combining the modified-Backward-in-Time selection method and the Top-Down strategy (mBTS-TD strategy). This approach leads to parsimonious estimation of the VAR model (no estimation of non-significant coefficients) reducing by construction the network dimension and the GPDC measure assesses the direction and the strength of causal relationships providing a precise network topology. Then, from this network we exclude the most systemic/influenced assets in order to reduce the systemic risk and thus obtain a well-diversified universe. To this end, we use the local directed weighted clustering coefficient, based on the in and the out information spreading patterns, neglecting the middleman and cycle ones. We show that our dynamic pre-selection procedure significantly contributes to improving portfolio performances of classical allocation strategies when compared to a network based on the Granger non-causality or applied on the whole universe.

4.1 Introduction

Since the seminal paper by Markowitz [1] which introduced optimal portfolio construction with the so-called mean-variance strategy, portfolio allocation has been a widely studied problem. The mean-variance strategy allocates assets by maximizing the expected return for a given risk level to obtain the optimal weights. Nevertheless, due to the estimation errors on the input variables, especially for the expected return, this strategy can lead to poor out-of-sample performances [6, 7, 5, 8]. To overcome these drawbacks, risk-based allocation strategies have been introduced because they focus solely on the covariance matrix estimation. The most well-known risk-based allocation methodologies are the Equal Risk Contribution portfolio (ERC) [29], the Minimum Variance portfolio (MinVar) [37] and the Maximum Variety portfolio (VarMax) [30]. Another alternative to the mean-variance strategy is the naive Equally Weighted portfolio (EW) that does not require any estimation procedure and that can provide higher performances than more advanced approaches as remarked in[28]. However, whatever their allocation characteristics, these strategies do not captured the dependency structure among financial assets, leading to incomplete risk assessments of investment universes. They must be applied to initial universes as diversified as possible to reduce the interconnectedness and improve the return/risk profile.

Over the last decades the asset allocation problem has been tackled form a new perspective, using network theory to represent the dependence among financial assets. This new approach provides useful insights in the portfolio selection process. Indeed, following the seminal work of Mantegna [69] which modelled financial markets as complex systems using network theory, many works have emerged. In particular, the work of Pozzi et al [71] shows that a network approach can be used in order to build a well-diversified portfolio and thus control the risk. They start from the correlation matrix computed between all pairs of assets and use the Minimum Spanning Trees [69] or Planar Maximally Filtered Graphs [128] to reduce the network dimension. Then, they identify the central (highly connected) and peripheral (poorly connected) assets of the network through hybrid centrality indices combining the most common centrality/peripherality measures (degree, betweenness centrality, eccentricity, closeness and eigenvector centrality). Finally, the authors find that investing in peripheral assets leads to a more diversified portfolio and increases the return/volatility ratio relative to central assets. In the same way, the work of Peralta et al [72] theoretically proves a negative relationship between the mean-variance optimal weights and the centrality of assets, i.e. optimal portfolios overweight low-central assets. Next, they empirically show that central assets tend to have "value" characteristics (large market capitalization, undervalued, financially risky). As in [71], the authors find that investing in assets with low-centrality improves portfolio performances. Moreover, recent alternative works [75, 76] use the clustering coefficient [126, 127] that allows to measure how much a node is embedded into the network and thus assess the systemic risk of an asset. They propose an optimization problem directly based on the network structure. Instead of using the covariance matrix to obtain the optimal weights of the Global Minimum Variance portfolio, they derive, from the clustering coefficients an interconnectedness matrix taking into account for each asset both systemic risk and individual volatility. Each of these works confirmed the usefulness of using a network approach to improve the portfolio selection process.

However, in most cases, the correlation matrix of assets is used to build the network leading to two major drawbacks. First, it leads to undirected networks (weighted or unweighted), i.e. only the relationship existence is known but the relationship direction is unknown. By considering this kind of networks, we can easily assume that we are missing information on the dependency structure. Second, it leads to complete networks where dimension reduction tools must be used such as the Minimum Spanning Tree [69] or the Planar Maximally Filtered Graphs [128], otherwise the most common centrality measures cannot be applied. An alternative way to build the network is to use Granger non-causality tests [77] as in [70, 73, 74] to identify significant causal relations among assets. In contrast to networks based on sample correlation matrix, networks based on Granger-non causality tests provide directed networks. However, here again, the Granger non-causality tests have two drawbacks. First, even if causality is detected (rejection of non-causality hypothesis), Granger's tests do not give any information about the strength of this causality and therefore do not allow to build a weighted network. Second, the Vector AutoRegressive (VAR) models estimation may fail for many well-known reasons, such as incorrect model order selection, small sample size, or correlated residuals [146, 141].

This paper contributes to the literature on asset allocation problems in the sense that it brings a precise estimate of the financial market dependency structure and then a dynamic pre-selection of assets is applied to obtain a well-diversified universe (without systemic or influenced assets) before allocating portfolios (EW, ERC, MinVar and VarMax).
To recover the network topology, we use both a parsimonious VAR estimation (mBTS-TD strategy) and the Generalized Partial Directed Coherence measure (GPDC) [178] as in [221]. The mBTS-TD strategy combines two subset selection methods (the modified Backward-in-Time Selection (mBTS) [151] and the Top Down strategy (TD) [141]) to estimate the VAR and has two advantages: first, by not estimating non-significant coefficients of the underlying VAR model, it reduces or even suppresses cascading errors on the GPDC; second, by providing a parsimonious structure, the use of network dimension reduction tools is not necessary. Then, the GPDC measure is used to identify the causal relationship and to quantify the causal strength leading to a precise network topology (directed-weighted network).
The dynamic pre-selection of assets excludes from the initial universe, the assets that interact too much, hence contributing to a possible contagion. Such assets are then seen as risky in the network sense. In order to identify these assets, we use the in and out triangle patterns of the local directed weighted clustering coefficient [127]. The local directed weighted clustering can be divided into four types of triangles in, out, cycle and middleman [126] giving rise to completely different interpretations. The out triangle pattern allows to identify the assets that strongly cause the other ones. They can therefore be considered as systemic assets because they are too influential within the network. The in triangle pattern identifies the assets that are strongly caused within the network and therefore easily influenced if a market shock occurs. This pre-selection reduces asset allocation errors being complementary to the use of the covariance matrix that does not quantify if an asset is too systemic or too influenced.
To examine the performances of our dynamic pre-selection method, we consider a global equity universe composed of 40 country indices within the MSCI ACWI (MSCI All Country World Index). This universe allows us to take both in account the differences

in time delay between areas and global/regional macroeconomic issues. Our proposed methodology is compared to asset selections based on Granger causality networks and to strategies allocating the whole universe, i.e. without applying any selection. Finally, we conclude the empirical analysis, using a cleaned and robust covariance matrix estimation (whitening procedure) [222, 223] instead of the classical Sample Covariance Matrix (SCM) in the allocation optimization problem to see if we can improve even more the results.

This paper is set out as follows: in section 4.2, we briefly describe the portfolio allocation strategies and the covariance matrix estimation; in section 4.3, we define the asset selection procedure (mBTS-TD strategy, GPDC measure and the local directed weighted clustering coefficient); in section 4.4, we present the empirical results; in section 4.5 we draw some conclusions.

4.2 Portfolio allocation

Portfolio allocation is a widely studied problem since the seminal work of Markowitz [1]. In this section, we detail the allocation strategies used for the empirical investigation. We first introduce the Equally Weighted portfolio (EW) [28] and then the most well-known risk-based allocation methodologies that depend solely on the covariance matrix of asset returns: the Equal Risk Contribution portfolio (ERC) [29], the Minimum Variance portfolio (MinVar) [37] and the Maximum Variety portfolio (VarMax) [30].

Here, we focus only on "Long only" portfolios, i.e. all the quantities invested in assets are necessarily greater than or equal to 0. This choice is motivated for two reasons. First, when short selling the assets, borrowing costs have to be taken into account to compute portfolio performances. What is more, these costs are not uniform among assets being dependent on their liquidity. Second, when building a long/short portfolio the size of the two legs must be defined as a constraint in the optimization process, otherwise the result obtained will not be realistic with possible strong leverage effects.

4.2.1 Equally Weighted portfolio (EW)

The Equally Weighted process aims to allocate all assets with the same weight $\frac{1}{m}$ where m is the number of assets in the investment universe. It does not require estimation of the covariance matrix nor complex optimization issues. What is more, this method can even provide higher performances than more advanced ones as remarked in [28].

4.2.2 Equal Risk Contribution portfolio (ERC)

The Equal Risk Contribution portfolio (ERC) [29] allocates assets according to their contribution to the risk of the portfolio. Let $\sigma_p = (\mathbf{w}' \, \mathbf{\Sigma} \, \mathbf{w})^{1/2}$ be the risk of the portfolio, where $\mathbf{w}$ is the m-vector of weights and $\mathbf{\Sigma}$ is the $m \times m$ covariance matrix of the m assets returns. Then the marginal contribution of asset j associated to σ_p is defined as follows:

$$\partial_{w_j} \sigma_p = \frac{\partial_{\sigma_p}}{\partial_{w_j}} = \frac{(\mathbf{\Sigma} \, \mathbf{w})_j}{\sigma_p}$$

According to [29], the risk of the portfolio can be expressed as the sum of the total risk contribution:

$$\sigma_p = \sum_{j=1}^{m} \sigma_j(w)$$

where $\sigma_j(w) = w_j \, \partial_{w_j} \sigma_p$ is the risk contribution of asset j.

The final portfolio is obtained equalizing all risk contributions ($\sigma_j(w) = \sigma_k(w) \; \forall \, j, k$) by minimizing the total squared differences between the risk contributions of all pairs of assets with respect to the weight vector $\mathbf{w}$ to get the solution $\mathbf{w}^*_{erc}$:

$$\mathbf{w}^*_{erc} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{j=1}^{m} \sum_{k=1}^{m} \Big(w_j \left(\mathbf{\Sigma} \, \mathbf{w} \right)_j - w_k \left(\mathbf{\Sigma} \, \mathbf{w} \right)_k \Big)^2 \,, \quad \text{s.t. } \mathbf{w}' \, \mathbf{1}_m = 1 \text{ and } \mathbf{w} \geq \mathbf{0}_m \quad (4.1)$$

where the constraints $\mathbf{w}' \, \mathbf{1}_m = 1$ and $\mathbf{w} \geq \mathbf{0}_m$ denote respectively no leverage effect and long-only strategies.

4.2.3 Minimum Variance portfolio (MinVar)

The Minimum Variance portfolio (MinVar) [37] is obtained by minimizing the variance of the final portfolio with respect to the weight vector $\mathbf{w}$ to get the solution $\mathbf{w}^*_{mv}$. The optimization problem is defined as follows:

$$\mathbf{w}^*_{mv} = \underset{\mathbf{w}}{\operatorname{argmin}} \, \mathbf{w}' \, \mathbf{\Sigma} \, \mathbf{w} \,, \quad \text{s.t. } \mathbf{w}' \, \mathbf{1}_m = 1 \text{ and } \mathbf{w} \geq \mathbf{0}_m \quad (4.2)$$

In this paper, we use an optimization algorithm to obtain the weights because we focus solely on long-only strategies. Nevertheless, for unconstrained weights, a closed form expression is available (see 2.3).

4.2.4 Maximum Variety portfolio (VarMax)

The Maximum Variety (VarMax) process, also called the Maximum Diversified Portfolio in [30] allocates assets by maximizing the Variety Ratio ($\mathcal{VR}$) of the portfolio. The Variety Ratio ($\mathcal{VR}$) quantifies the degree of diversification of a portfolio and is directly related to the number of independent factors within a portfolio [30, 52]. The Variety Ratio ($\mathcal{VR}$) of the portfolio is defined as follows:

$$\mathcal{VR}\left(\mathbf{w}, \mathbf{\Sigma}\right) = \frac{\mathbf{w}' \, \boldsymbol{\sigma}}{\left(\mathbf{w}' \, \mathbf{\Sigma} \, \mathbf{w}\right)^{1/2}} \,, \quad (4.3)$$

where $\boldsymbol{\sigma}$ is the m-vector of the square roots of the diagonal element of $\mathbf{\Sigma}$, i.e. $\sigma_j = \sqrt{\mathbf{\Sigma}_{jj}}$, representing the standard deviation of the returns of the m assets.

The final portfolio is obtained by maximizing the above diversification ratio with respect to the weight vector $\mathbf{w}$ to get the solution $\mathbf{w}^*_{vm}$:

$$\mathbf{w}^*_{vm} = \underset{\mathbf{w}}{\operatorname{argmax}} \, \mathcal{VR}\left(\mathbf{w}, \mathbf{\Sigma}\right) \,, \quad \text{s.t. } \mathbf{w}' \, \mathbf{1}_m = 1 \text{ and } \mathbf{w} \geq \mathbf{0}_m \quad (4.4)$$

4.2.5 Covariance matrix estimation

In order to get solutions for (4.1), (4.2) and (4.4) the covariance matrix $\mathbf{\Sigma}$ has to be estimated. The covariance matrix estimation is also a largely studied problem in many fields and several estimators exist to reduce estimation errors such as Ledoit-Wolf's shrinkage [58], Eigenvalue clipping [55], Rotational Invariant Estimators [63] or hybrid robust methods [118, 117, 222, 223, 115, 62]. In this paper, we estimate the covariance matrix with the Sample Covariance Matrix (SCM) and with a more recent method using a cleaned and robust covariance matrix estimation (whitening procedure) [222, 223], precisely described and studied in chapter 2.

4.3 Financial networks and asset selection

The use of financial networks has been widely studied over the last two decades and can help to deal with the asset allocation problem. The knowledge of the dependency structure of the market allows to characterize the interactions between assets and thus determine their position within the network (e.g. central, peripheral, systemic, etc.). Indeed, with centrality/peripherality measures or clustering coefficients, we can distinguish the peripheral or least systemic assets in the network as in [71, 72, 75, 76]. Identifying these assets and excluding them from the investment universe reduces portfolio risk and results, in general, to a more diversified portfolio [71, 72].

In this section, we propose to modeling financial markets dependency structures with the Generalized Partial Directed Coherence measure (GPDC) [176, 177, 178], estimated with our mBTS-TD strategy (modified Backward-in-Time Selection [151] and the Top-Down strategy (TD) [141]) as in [221]. This methodology has several advantages when compared to networks based on the sample correlation matrix [71, 72, 75, 76] or the Granger non-causality test [217, 73, 74]. First, it provides a parsimonious structure via the mBTS-TD strategy solving the network dimensionality, contrary to the correlation matrix approach which requires dimension reduction (the Minimum Spanning Tree [69] or Planar Maximally Filtered Graphs [128]). Second, it provides a precise network topology taking into account both the direction and the strength of the relationship between assets via the GPDC.

Once the financial network is built, we propose a dynamic method to select the assets in the universe before allocating the portfolio. We exclude from the universe the assets that present major risks due to their systemic position or those that are already too influenced to withstand a regional/global crisis. In order to identify these assets, we define an indicator based on the in and out triangle patterns of the local directed weighted clustering coefficient [127]. By construction, these two measures identify either the most influential (out) or the most influenced (in) assets within the network.

In the following section, we define how the financial network is obtained from the GPDC measure and we describe our asset selection methodology.

4.3.1 GPDC financial network

As in chapter 3, the temporal relationships among financial time series are captured through a VAR model. Let $\mathbf{x}(t) = (x_1(t), \ldots, x_m(t))'$ be a zero-mean m-dimensional stationary process admitting the following VAR(p) representation (see 1.11 for classical stability and stationarity conditions):

$$\mathbf{x}(t) = \mathbf{A}_1\mathbf{x}(t-1) + \cdots + \mathbf{A}_p\mathbf{x}(t-p) + \boldsymbol{\epsilon}(t), \quad t \in \mathbb{Z} \tag{4.5}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $m \times m$ coefficient matrices, p is the model order, and $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \ldots, \epsilon_m(t))'$ is the m-vector of white noises with $\mathrm{E}[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}'(s)] = 0$ for $t \neq s$ and $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$.

We estimate the VAR coefficients $\mathbf{A}_1, \ldots, \mathbf{A}_p$ with our proposed mBTS-TD strategy described and studied in chapter 3. This estimation provides a parsimonious VAR (no estimation of non-significant coefficients) by taking into account that a time series x_j does not depend on all lagged variables in the system. The mBTS method and the TD strategy are complementary methods. Indeed, the mBTS includes only the coefficients that improve the prediction of the equation and allows to work with high-dimensional systems [151]. For the TD strategy, it allows to be less dependent on the choice of the maximal lag p_{max} when initializing mBTS method so that we may set its values high enough to capture all possible connections preserving at the same time the tractability of the model when p_{max} is set at a high value. Hereafter, we define the mBTS-TD strategy that solves the precision/parsimony dilemma.

First, a maximum order p_{max} is fixed, and this provides the mp_{max}-vector of all lagged variables for the j-th equation of the VAR(p) model (4.5):

$$\boldsymbol{v} = (x_1(t-1), \ldots, x_1(t-p_{max}), \ldots, x_m(t-1), \ldots, x_m(t-p_{max}))'$$

An explanatory vector $\boldsymbol{\vartheta}$ is built from $\boldsymbol{v}$ by progressively adding only the most significant lagged variable at each step.

For the j-th equation of the VAR(p) model (4.5), the mBTS-TD algorithm is as follows:

1. Start with an empty vector $\boldsymbol{\vartheta} = \emptyset$, the unbiased BIC criterion [145] (IC^{old}) initialized to the variance of the j-th series, and $\boldsymbol{\tau} = (1, \ldots, 1)'$ the $m \times 1$ lag order vector of the variables.

2. Compute IC_n^{new} relative to the m dynamic regression models formed by the m candidate explanatory vectors $\boldsymbol{\vartheta}_n^{\text{cand}}$, where $\boldsymbol{\vartheta}_n^{\text{cand}} = (\boldsymbol{\vartheta}, x_n(t-\tau_n)) \quad \forall n \in \{1, \ldots, m\}$.

3. Select the variable according to the IC value:

   - If $\min\{IC^{\text{old}}, IC_1^{\text{new}}, \ldots, IC_m^{\text{new}}\} = IC^{\text{old}}$, then $\boldsymbol{\tau} = \boldsymbol{\tau} + \mathbf{1}_m$.
   - If $\min\{IC^{\text{old}}, IC_1^{\text{new}}, \ldots, IC_m^{\text{new}}\} = IC_n^{\text{new}}$, then $IC^{\text{old}} = IC_n^{\text{new}}$, $x_n(t-\tau_n)$ is added to the explanatory vector $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}, x_n(t-\tau_n))$ and only τ_n is increased by one.

4. Repeat steps 2 and 3 until $\boldsymbol{\tau} = (p_{max}, \ldots, p_{max})'$.

Then, the TD strategy (TD) tests the VAR coefficients separately in the m equations. For the j-th equation obtained with the mBTS algorithm, the TD strategy is applied as follows:

1. Start with the vector $\boldsymbol{\vartheta}$ and the information criterion IC^{old} obtained with the mBTS algorithm.
2. Sort the vector $\boldsymbol{\vartheta}$ from the largest to the smallest lag p and for all series from x_m to x_1.
3. Compute IC_n^{new} by deleting the n-th element in the vector $\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}_n^{\text{cand}} = \boldsymbol{\vartheta} \setminus \{\boldsymbol{\vartheta}_n\}$.
4. Delete the variable according to the IC_n^{new} value:
   - If $\min\{IC^{\text{old}}, IC_n^{\text{new}}\} = IC^{\text{old}}$, then $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}$.
   - If $\min\{IC^{\text{old}}, IC_n^{\text{new}}\} = IC_n^{\text{new}}$, then $IC^{\text{old}} = IC_n^{\text{new}}$ and $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_n^{\text{cand}}$.
5. Repeat steps 3 and 4 $\forall(\boldsymbol{\vartheta}_n)_{n\in[1,|\boldsymbol{\vartheta}|]}$, where $|.|$ denotes the cardinality of the vector in this case.

Once the VAR coefficients are estimated, we compute the GPDC measure. It provides a frequency domain representation of Granger Causality [77] and quantifies only the direct connections between time series. For two time series $x_j(t)$ and $x_k(t)$ the GPDC is defined so as to exhibit the causality from k to j at each frequency f as follows:

$$\omega_{jk}(f) = \frac{\frac{1}{\sigma_{jj}}\tilde{a}_{jk}(f)}{\sqrt{\sum_{n=1}^{m}\frac{1}{\sigma_{nn}^2}\tilde{a}_{nk}(f)\tilde{a}_{nk}^*(f)}}$$

where

- f are the discrete frequencies[1] lying in $\left[-\frac{1}{2};\frac{1}{2}\right]$,
- $\tilde{a}_{jk}(f)$ is the discrete Fourier transform of the coefficients $a_{jk}(1),\ldots,a_{jk}(p)$ defined by

$$\tilde{a}_{jk}(f) = \begin{cases} 1 - \sum_{l=1}^{p} a_{jk}(l)e^{-2i\pi fl}, & \text{if } j = k \\ -\sum_{l=1}^{p} a_{jk}(l)e^{-2i\pi fl}, & \text{otherwise} \end{cases}$$

- σ_{jj}^2 is the j-th element of the diagonal of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$.

[1]For a discrete time series sampled at frequency f_e, its Fourier Transform will reveal information for frequencies lying in $\left[-\frac{f_e}{2};\frac{f_e}{2}\right]$. In our case $f_e = 1$, we can therefore choose the interval $\left[-\frac{1}{2};\frac{1}{2}\right]$ with a step of $\frac{1}{F-1}$, where F is the number of frequencies.

Finally, we recover the financial network topology. A network $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ is a set of objects with $\mathbf{V}$ the set of nodes and $\mathbf{E}$ the set of edges between nodes. The edge (j, k) connects a pair of node j and k. The mathematical representation of a directed weighted network is given by the $m \times m$ adjacency matrix $\mathbf{Z}^w = (z_{jk}^w)$, $z_{jk}^w \in \mathbb{R}^+$ if $(j, k) \in \mathbf{E}$ and 0 otherwise. For two assets j and k, the adjacency matrix $\mathbf{Z}^{GPDC}$ based on the GPDC measure is defined as follows:

$$z_{jk}^{GPDC} = \begin{cases} \max |\boldsymbol{\omega}_{jk}|^2, & \text{if } j \neq k \\ 0, & \text{otherwise} \end{cases}$$

where $\boldsymbol{\omega}_{jk}$ is the vector containing each value of $\omega_{jk}(f)$ for all discrete frequencies f. We use the maximum value in the $\boldsymbol{\omega}_{jk}$ vector in order to take into account the most relevant information between the two assets, i.e. whether the relationship is short term (high-frequency) or long term (low-frequency).

4.3.2 Asset selection methodology

After recovering the financial network topology, we propose a dynamic pre-selection of financial assets from an initial investment universe. This dynamic pre-selection is based on the local directed weighted clustering coefficient [127]. This tool introduced by Clemente et al. in [127] measures how a node is embedded into the network by quantifying its number of triangles out of all its possible triangles (see also Fagiolo [126]). Furthermore, this tool takes into account the strength of a node in the normalization factor unlike Fagiolo [126]. For an asset j, it is defined as follows:

$$h_j = \frac{\frac{1}{2}\left[\left(\mathbf{Z}^w + \mathbf{Z}^{w\prime}\right)\left(\mathbf{Z}^u + \mathbf{Z}^{u\prime}\right)^2\right]_{jj}}{s_j\left(d_j - 1\right) - 2s_j^{\leftrightarrow}}$$

where $\mathbf{Z}^u$ is the unweighted version of $\mathbf{Z}^w$ ($z_{jk}^u = 1$ if $z_{jk}^w \neq 0$, and 0 otherwise), $d_j = \left(\mathbf{Z}^{u\prime} + \mathbf{Z}^u\right)_j \mathbf{1}_m$ and $s_j = \left(\mathbf{Z}^{w\prime} + \mathbf{Z}^w\right)_j \mathbf{1}_m$ are respectively the total degree (total number of edges) and the total strength (for the case of weighted network) of the asset j. $s_j^{\leftrightarrow} = \dfrac{\left(\mathbf{Z}^w\,\mathbf{Z}^u + \mathbf{Z}^u\,\mathbf{Z}^w\right)_{jj}}{2}$ is the strength of bilateral edges between j and k. Note that $h_j \in [0, 1]$, a high value indicating that the asset j is heavily embedded in the network. Moreover, in the case of a directed unweighted network, the local directed weighted clustering coefficient [127] is equivalent to the Fagiolo coefficient [126].

Moreover, the local directed clustering coefficient can be classically divided into four types of triangles (in[2], out[3], cycle[4] and middleman[5]) [126] giving rise to different financial interpretations. Having in mind to reduce the risk and improve portfolio performances, we want to identify the assets that present major risks due to their systemic position or those that are too influenced. Hence, we only focus on the in and out triangle patterns. The out triangle pattern allows to identify the assets that strongly cause the other ones.

[2]In: a triangle in which there are two edges pointing toward j.

[3]Out: a triangle in which there are two edges starting from j.

[4]Cycle: a triangle in which every edge has the same direction.

[5]Middleman: a triangle in which j has two edges of different directions and with an edge between k and l without forming a cycle.

They can therefore be considered as systemic assets because they are highly influential within the network. The in triangle pattern identifies the assets that are strongly caused within the network and therefore easily influenced if a market shock occurs. The two specific triangle patterns are:

$$h_j^{out} = \frac{\frac{1}{2}\left[\mathbf{Z}^w\left(\mathbf{Z}^u + \mathbf{Z}^{u\prime}\right)\mathbf{Z}^{u\prime}\right]_{jj}}{s_j^{out}\left(d_j^{out} - 1\right)}$$

$$h_j^{in} = \frac{\frac{1}{2}\left[\mathbf{Z}^{w\prime}\left(\mathbf{Z}^u + \mathbf{Z}^{u\prime}\right)\mathbf{Z}^u\right]_{jj}}{s_j^{in}\left(d_j^{in} - 1\right)}$$

where $d_j^{out} = \mathbf{Z}_j^u \mathbf{1}_m$ and $s_j^{out} = \mathbf{Z}_j^w \mathbf{1}_m$ are respectively the out-degree (number of edges starting from j) and the out-total strength. $d_j^{in} = \mathbf{Z}_j^{u\prime} \mathbf{1}_m$ and $s_j^{in} = \mathbf{Z}_j^{w\prime} \mathbf{1}_m$ are respectively the in-degree (number of edges pointing towards j) and the in-strength.

In this paper, we propose to build an indicator essentially based on the out and in triangle patterns in order to select only the assets that have an out equals to 0 (the least systemic assets) or an in also equals to 0 (the least influenced assets). By combining these two triangle patterns, we remove from the universe both systemic and influenced assets, which are therefore more unstable and risky. We also exclude assets with both an out and an out-total strength equal to 0, as they are highly influenced and assets with a local directed clustering coefficient equals to 1, whatever their level for the out and in because they are too embedded in the network. Thus, the selected universe $\mathbf{V}^*$ is given by:

$$\mathbf{V}^* = \{v_j \,|\, h_j^{in} = 0 \text{ and } h_j < 1\} \cup \{v_j \,|\, h_j^{out} = 0 \text{ and } s_j^{out} > 0 \text{ and } h_j < 1\}$$

where $v_j \in \mathbf{V}$ is the node of the network and representing the asset j.

To summarize, our investment universe $\mathbf{V}^*$ reflects the following financial choices:

- $h_j^{out} = 0$: the least influential asset is probably the least systemic asset and therefore the one with the least chance of triggering a crisis. However, an asset with $s_j^{out} = 0$ is not retained because it is too much influenced ($h_j^{in} >> 0$ or $s_j^{in} >> 0$), unless it fulfills $h_j^{in} = 0$ being thus considered as a fairly detached asset when problems occur,
- $h_j^{in} = 0$: the least influenced asset is probably the one that is the most detached from current financial market problems and can therefore quickly become detached from a regional/global problem due to its economic characteristics,
- An asset with $h_j = 1$ is excluded because it is too embedded in the network and thus too systemic.

This asset selection procedure has several advantages. First, it focuses essentially to unstable assets (systemic or influenced) in order to reduce the portfolio risk. Second, by construction it adapts itself to the number of connections in the network and therefore removes only the most embedded assets for each period. Indeed, if the network is very disconnected, no asset will be removed. Hence, it is not necessary to either set a number of excluded assets or set an exclusion threshold on the local directed weighted clustering coefficient as in chapter 3.

4.4 Empirical Analysis

In this section, we apply the above methodology to exclude systemic and influenced assets from the initial investment universe. From this selected universe, the performances of our dynamic pre-selection are tested on four portfolio allocation strategies: EW, ERC, MinVar and VarMax, described in 4.2. The portfolio performances are also compared to alternative strategies, where the assets are selected using a financial network based on Granger non-causality tests[6] (GC) defined in (1.14), or where the whole universe is allocated without any selection.

4.4.1 Dataset description

The investment universe is composed of the $m = 40$ most liquid and still alive since 2001 national indices (network nodes) belonging to the MSCI All Country World Index (MSCI ACWI) already used in the last chapter and summarized in Table 4.16a. This universe allows us to take in account both the differences in time delay between areas (feedback effects) and global/regional macroeconomic issues. The asset returns are computed on a daily basis from January 18th, 2001 to October 25th, 2019. Every four weeks, we build a temporal network using a rolling window of $T = 256$ working days. Before computing the GPDCs into which the VAR coefficients are plugged, we normalize each time series using a Generalized Auto-Regressive Condional Heteroskedastic (GARCH) [220] (described in the last chapter 3.7.3) to remove the heteroskedasticity in asset returns. Then, we apply the asset pre-selection on temporal networks. Finally, for risk-based allocation strategies (ERC (4.1), MinVar (4.2) or VarMax (4.2)), we estimate the covariance matrix on the asset returns converted into USD currency, to take into account the risk exposure of foreign currency investments. The portfolios' performances are therefore computed in USD currency[7].

For all allocation strategies stated in section 4.2, the asset selection procedure using mBTS-TD for the VAR estimation and the GPDC measure ("GPDC") is compared with those obtained using a classical VAR estimation and the GC test (1.14) with a probability $\alpha = 1\%$ or $\alpha = 5\%$ ("GC 1%" or "GC 5%") and on the whole universe, i.e. without any selection. In order to assess the potential of our methodology, we report several portfolio statistics computed over the whole period: the annualized return, the annualized volatility, the ratio between the annualized return and the annualized volatility and the maximum drawdown (largest decline in portfolio value). The portfolio generates better performances if it provides an higher return/volatility ratio and a lower maximum drawdown. Moreover, to be as realistic as possible in the investment process, the portfolio performances are "net of transactions fees", i.e. we consider 0.07% of fees for any change in weighting at each rebalancing date. Indeed, the portfolio turnover is important in portfolio allocation and has to be integrated into the performance computation, as a very high level of turnover can significantly reduce it: for a given rebalancing date t the portfolio turnover at time t is $\kappa_t = \sum_{j=1}^{m} |w_{j,t} - w_{j,t-1}|$ and the portfolio performance is therefore decreased by $\kappa_t \times 0.07\%$.

[6] The VAR coefficients are estimated using the Least Squares (LS) (1.12) and the model order p is estimated with the AIC criterion (1.2.2) [142, 143].

[7] If the asset returns are kept in local currency to compute the portfolios' performances, hedging costs (selling the currency forward) have to be considered.

4.4.2 EW portfolios

Here, we examine the potential of our proposed methodology for the Equally Weighted portfolio (EW). This allocation strategy invests in all assets and does not require any estimation procedure for the covariance matrix nor complex optimization issues. It therefore allows us to focus solely on the improvement resulting from the asset selection process. We first report the portfolio performances for the three asset selection procedures ("GPDC", "GC 1%" and "GC 5%"). Then, we focus on the asset return distribution between rebalancing dates (i.e. selected or non-selected asset returns) to understand where the out-performance of our methodology is coming from. We also provide the intertemporal number of non-selected assets in our selection process. Finally, as an example of asset selection, we depict in Fig. 4.3 the network based on the GPDC method obtained for the national indices in the window covering the period of November 7th, 2014 to October 30th, 2015 (commodity crisis).

Fig. 4.1 and Table 4.1 show the evolution of the EW portfolios and some statistics on the overall portfolios' performances respectively. For this purpose, we have added to the three methods ("GPDC", "GC 1%" and "GC 5%"), the EW portfolio allocated to the assets not selected by the "GPDC" method, in order to see whether these assets are indeed both less performing and riskier.

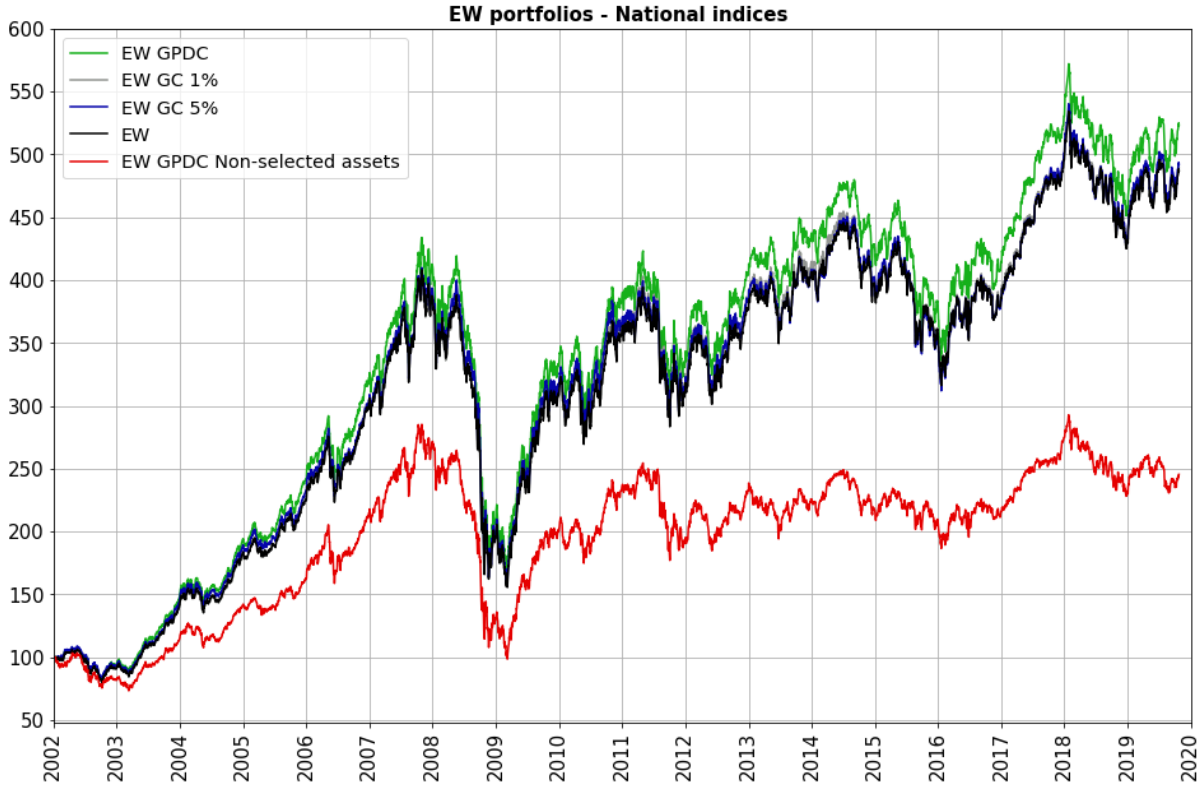


Figure 4.1: EW portfolios' performances on national indices with 0.07% of fees from January 2002 to October 2019. The proposed "EW GPDC" (green line) leads to improved performances vs the "EW GC 1%" (grey), the "EW GC 5%" (blue) and the "EW" (black). Moreover, the "EW GPDC Non-selected assets" (red) provides the worst performances by far.

EW Portfolios National indices	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
EW GPDC	9.75%	16.95%	0.58	61.05%
EW GC 1%	9.34%	16.76%	0.56	61.21%
EW	9.31%	16.76%	0.56	61.92%
EW GC 5%	9.37%	16.93%	0.55	61.24%
EW GPDC Non-selected assets	5.18%	17.49%	0.30	65.50%

Table 4.1: Performance indicators for EW portfolios on national indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

The proposed GPDC asset selection method shows an improvement with respect to the other methods, especially for the annualized return, the return/volatility ratio and the maximum drawdown. The other methods "GC 1%" and "GC 5%" do not provide a significant improvement compared to EW applied to all assets. Furthermore, when focusing on the non-selected assets by the GPDC process, we see clearly that all indicators are deteriorated with a return/volatility ratio less than at least 0.25 and a maximum drawdown more than 3.58% compared to the other methods. The annual returns between "EW GPDC" and "EW" may seem close (9.75% vs 9.31%), but note that this gap (0.40%) can absorb the management fees for the less expensive fund shares.

Given these results, we can consider that the GPDC asset selection process succeeds in identifying the less risky assets. To illustrate this purpose, Table 4.2 provides the four moments of asset return distribution for the selected and non-selected assets. Looking at Table 4.2, the asset return distribution of the selected assets provides better figures, in particular for the skewness which is positive (0.0065 vs -0.1683), i.e. the returns of the selected assets are more positive than those of the non-selected assets. However, the kurtosis is a little higher for selected assets, although extreme values are more concentrated in the right tail (positive returns).

GPDC Asset return distribution	Selected Assets	Non-Selected Assets	All Assets
Mean	0.0005	0.0002	0.0004
Standard Deviation	0.0156	0.0164	0.0158
Skewness	0.0065	-0.1683	-0.0393
Kurtosis	10.3942	9.7714	10.2511

Table 4.2: Moments of asset return distribution for the selected and non-selected assets by the GPDC process based on national indices from January 2002 to October 2019.

Fig. 4.2 plots the number of non-selected assets for the three asset selection procedures. The "GC 1%" process is the most stable with a standard deviation of the number of assets selected equal to 2.95. However, it does not eliminate more than two assets on average except during the 2008 crisis where it eliminates 20. As we have seen in Table 4.1, the "GC 1%" process is very close to the EW portfolio without selection. The other two methods very actively select assets by removing on average 10 assets. Again, even if "GC 5%" removes 10 assets on average, the performance is similar to the EW portfolio without

selection. Regarding the "GPDC" procedure, as we have seen in Table 4.1, it significantly improves the results while removing the same number of assets as "GC 5%". Moreover, when we look at the standard deviation of the number of assets selected, the "GPDC" procedure appears more stable over time than "GC 5%" (3.86 vs 5.16).

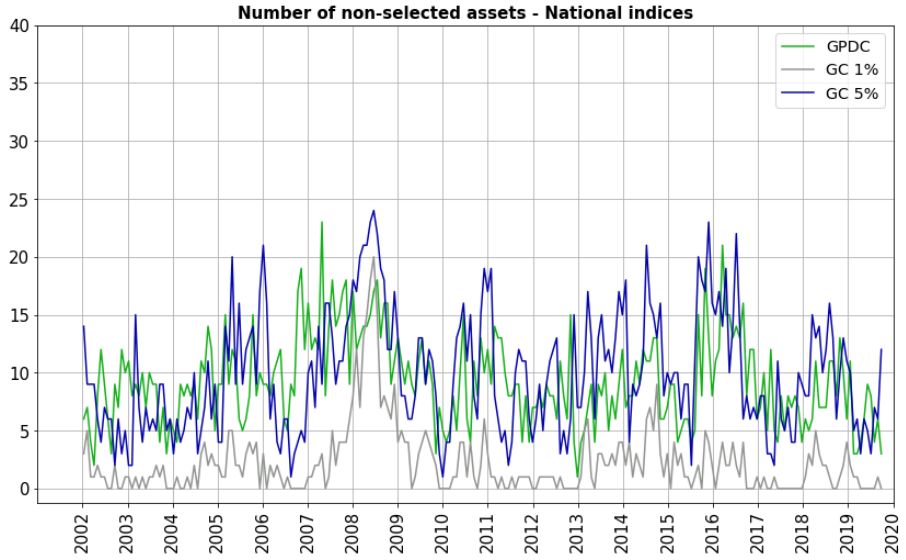


Figure 4.2: Number of non-selected assets for the three asset selection procedures on the national indices ($m = 40$) from January 2002 to October 2019.

Finally in Fig. 4.3, we depict a network based on the GPDC method obtained during the commodity crisis in the window covering the period of November 7th, 2014 to October 30th, 2015. As previously explained, the nodes here represent national indices (Table 4.16a) and the weighted edge (j,k) measures the relationship intensity (causal strength) between indices j and k. The nodes in green are the selected assets while those in white are not selected because they are considered too risky. This example perfectly illustrates our proposed methodology, i.e. estimating the GPDC measure with a parsimonious estimation of the VAR (mBTS-TD strategy), then selecting the assets that are the least nested in the network in order to reduce the systemic risk present in the portfolio and thus improve performances. In this Fig. 4.3, we clearly see that most of the non-selected assets have economic activities directly related to the production of commodities and/or the use of these commodities. Indeed, our procedure removes the main countries most dependent on raw material exports both in South America (Chile, Colombia and Peru) and Mexico, but also in Australia and Europe (Poland and Russia). Furthermore, it also removes the countries dependent on raw material imports especially in Asia (South Korea, India, Hong Kong and Taiwan). Comparing these two portfolios over the next four weeks, the portfolio of non-selected assets underperforms by 0.70% the portfolio with the selected assets.

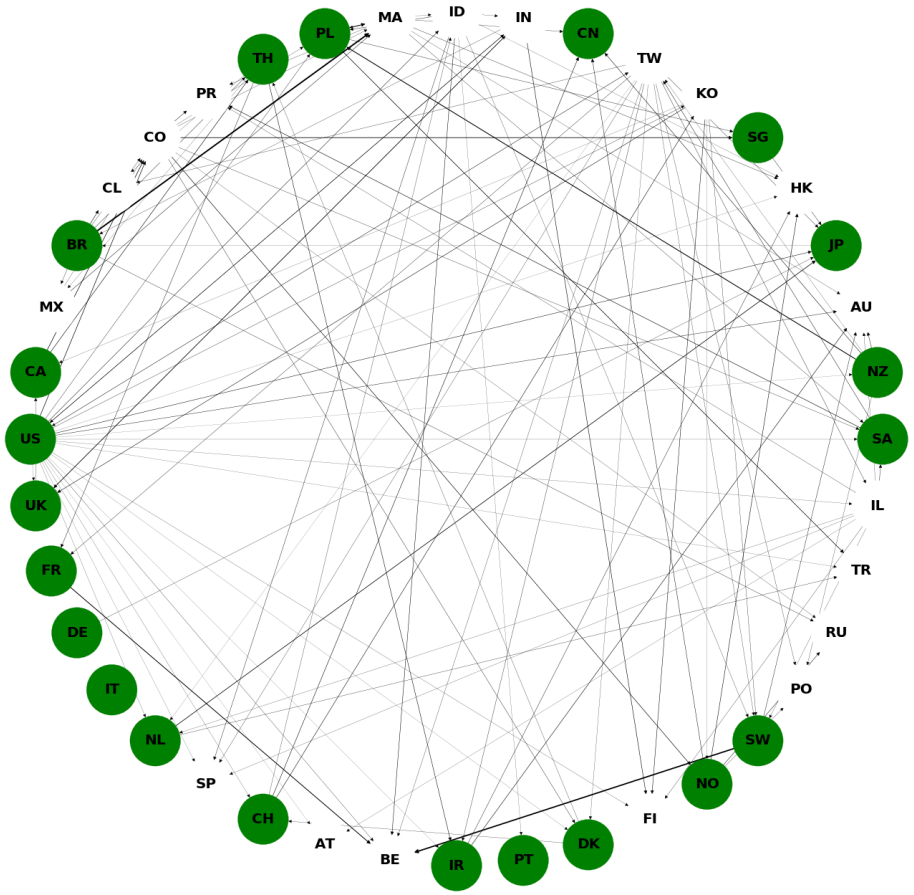


Figure 4.3: Directed weighted network of national indices based on the GPDC measure and estimated from November 7th, 2014 to October 30th, 2015 (commodity crisis). In green are the selected assets according to the method described in section 4.2 using essentially the triangle patterns (in and out) of the local directed weighted clustering coefficient. Conversely, the assets in white are not retained. The width of the directed edge represents the causal strength. Asset abbreviations are in Table 4.16a.

In this section, we have confirmed that our methodology using both a parsimonious estimation of the VAR to compute the GPDC measure and a selection process based on the in/out clustering coefficient significantly improves the performances of the EW portfolio. Indeed, our methodology succeeds in identifying the worst performing/riskiest assets in order to remove them from the investment universe. After these promising results, a natural question arises: is it possible to confirm these improvements for more sophisticated allocation processes, namely the ERC strategy that also invests in all assets but allocates them according to their risk contribution, or the MinVar/VarMax strategies that both select and weight the assets?
In what follows, we first answer this question in the naive case where the Sample Covariance Matrix (SCM) is plugged in the optimization algorithm at work in these three approaches.

4.4.3 ERC SCM portfolios

Now, we examine our proposed methodology for the ERC portfolio. In focusing on this strategy, we first seek to determine whether our asset selection is also consistent with risk-

based approaches. We then investigate if being systemic or influenced inevitably leads to high-volatility asset exclusions that further bias this type of strategies on the low-volatility anomaly. Note that in this universe the volatility levels are very heterogeneous, ranging from 15% for Malaysia to 42% for Russia over the whole observation period (January 2002 to October 2019). In Fig. 4.4 and Table 4.3, we provide the ERC portfolios' performances and performance indicators on the whole period.

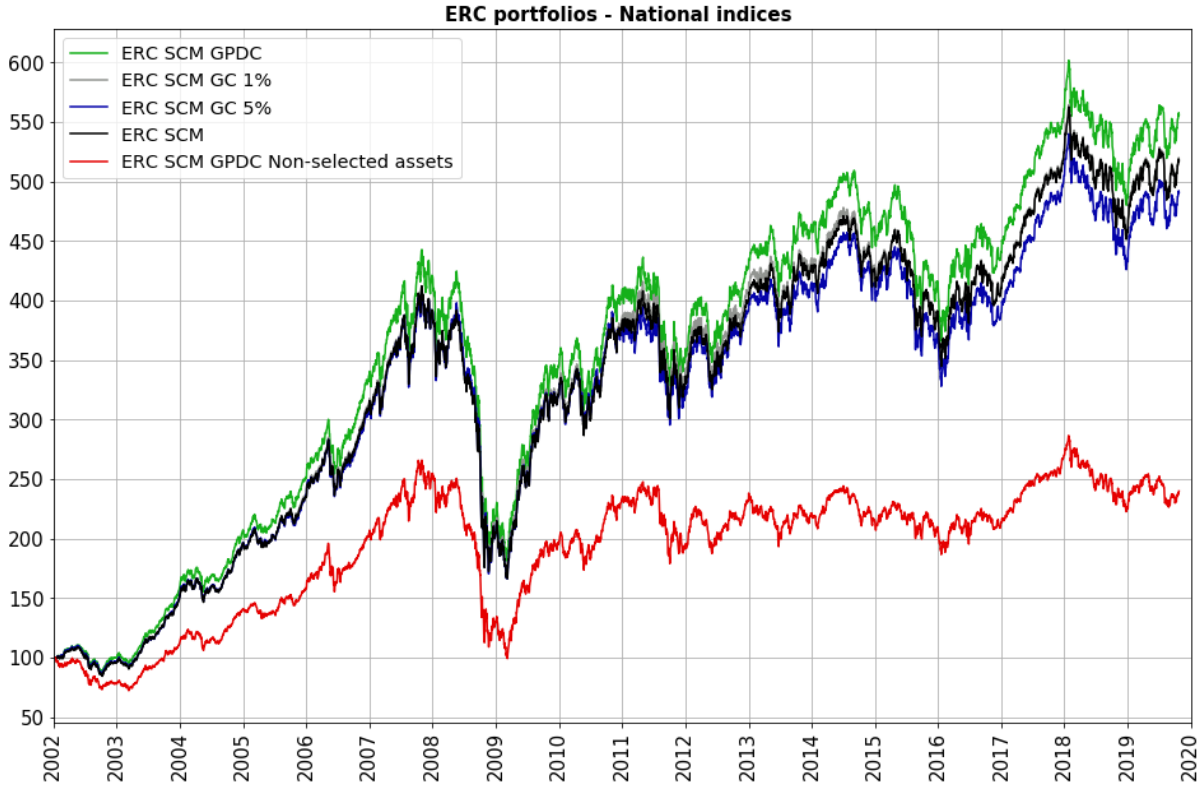


Figure 4.4: ERC portfolios' performances on national indices with 0.07% of fees from January 2002 to October 2019. The proposed "ERC SCM GPDC" (green line) leads to improved performances vs the "ERC SCM GC 1%" (grey), the "ERC SCM GC 5%" (blue) and the "ERC SCM" (black). Moreover, the "ERC SCM GPDC Non-selected assets" (red) provides the worst performances by far.

ERC Portfolios National Indices	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
ERC SCM GPDC	10.13%	15.26%	0.66	59.21%
ERC SCM GC 1%	9.70%	15.07%	0.64	58.85%
ERC SCM	9.68%	15.08%	0.64	59.65%
ERC SCM GC 5%	9.36%	15.32%	0.61	59.26%
ERC SCM GPDC Non-selected assets	5.04%	16.21%	0.31	62.68%

Table 4.3: Performance indicators for ERC portfolios on national indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

Based on these results, our approach appears to be perfectly suited to the ERC strategy. The proposed GPDC asset selection method shows an improvement with respect

to the other methods providing a higher annualized return, higher return/volatility ratio and the second maximum drawdown. Moreover, the annualized volatility has slightly increased confirming that our approach removes both low-volatility and high-volatility assets without creating a low-volatility bias. To illustrate this purpose, Fig. 4.5 plots the standard deviations of the invested assets versus the resulting weights obtained for "ERC SCM GPDC" (on the top graph) and "ERC SCM" (on the bottom graph). Considering all periods, the asset selection removes on average 24% of the 1st quartile of volatility (low) and 20% of the 4th quartile (high).

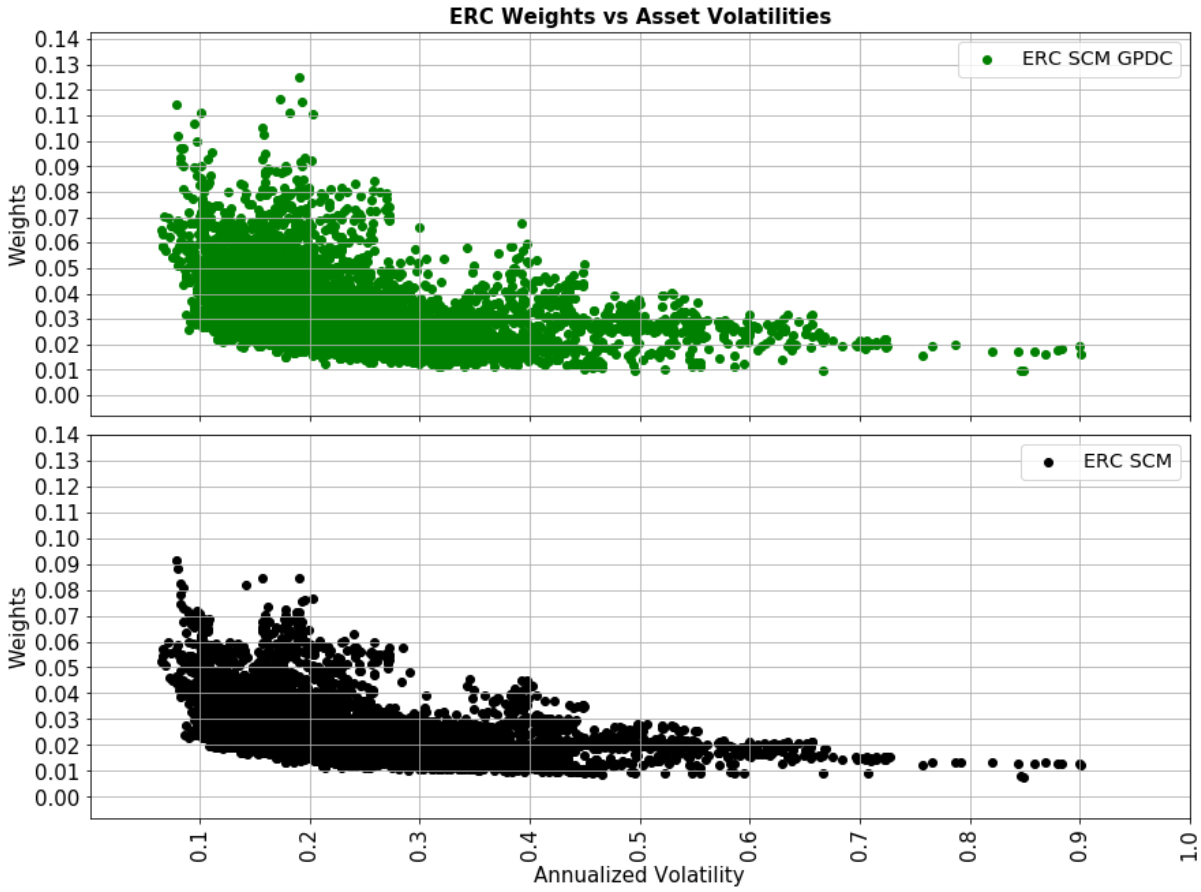


Figure 4.5: "ERC SCM GPDC" and "ERC SCM" weights versus asset volatilities.

4.4.4 MinVar SCM portfolios

After testing our methodology on the ERC strategy, we focus on the MinVar portfolio. This strategy selects and weights the assets in order to obtain the portfolio with minimum volatility. However, we know that minimizing the portfolio volatility leads to choosing the assets having the lowest volatilities, with a small number of assets in the portfolio. For this strategy, we want to see if removing some low-volatility assets that are highly embedded in the network improves performances. We provide in Fig. 4.6 and Table 4.4 the evolution of the MinVar portfolios and some statistics on the overall portfolio performances.

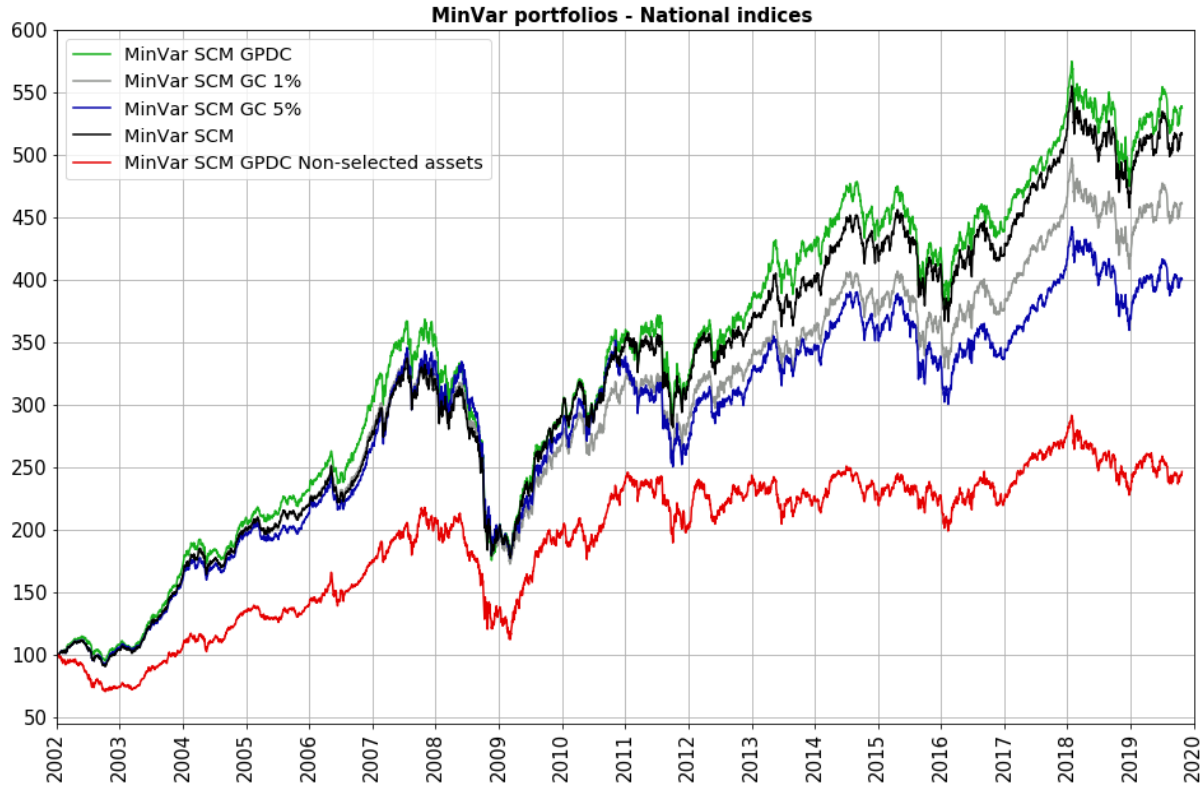


Figure 4.6: MinVar portfolios' performances on national indices with 0.07% of fees from January 2002 to October 2019. The proposed "MinVar SCM GPDC" (green line) leads to improved performances vs the "MinVar SCM GC 1%" (grey), the "MinVar SCM GC 5%" (blue) and the "MinVar SCM" (black). Moreover, the "MinVar SCM GPDC Non-selected assets" (red) provides the worst performances by far.

MinVar Portfolios **National Indices**	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
MinVar SCM	9.67%	10.73%	0.90	47.36%
MinVar SCM GPDC	9.92%	11.24%	0.88	52.33%
MinVar SCM GC 1%	8.98%	10.82%	0.83	49.61%
MinVar SCM GC 5%	8.12%	11.72%	0.69	48.63%
MinVar SCM GPDC Non-selected assets	5.21%	14.06%	0.37	48.39%

Table 4.4: Performance indicators for MinVar portfolios on national indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

For the MinVar portfolio, our methodology provides a higher annualized return, but also a higher annualized volatility compared to "MinVar SCM" and therefore a lower return/volatility ratio. It also has the highest maximum drawdown (52%). When we look at the annual performances of the different methods, we see that our method has an annual return that is equivalent or higher for 13 out of 18 years compared to "MinVar SCM". However, it completely misses the 2008 financial crisis by losing 10% more than the "MinVar SCM" (-44% vs -38%). The method especially misses the big rebounds that occurred after the strong declines. We know that the MinVar strategy leads to highly

concentrated portfolios, especially in highly volatile and hence highly correlated markets. Moreover, we also know that our asset selection procedure removes on average 24% of low-volatility assets (1st quartile) leading to even more concentrated portfolios for "MinVar SCM GPDC". To highlight this point, we plot in Fig. 4.7 the Herfindahl-Hirschman[8] index (HHI) for "MinVar SCM GPDC" and "MinVar SCM". We clearly notice that the HHI for the "MinVar SCM GPDC" is lower than "MinVar SCM", increasing the idiosyncratic risk and thus making it more complicated to manage drastic market changes as in 2008, 2015 and 2018.

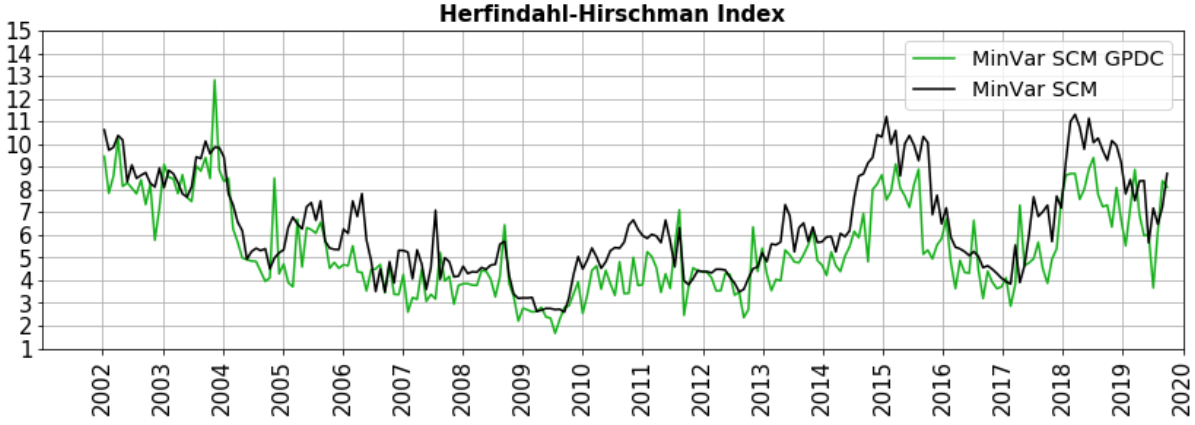


Figure 4.7: Herfindahl-Hirschman index for "MinVar SCM GPDC" (green line) and "MinVar SCM" (black line). "MinVar SCM GPDC" has on average one less asset in the portfolio than "MinVar SCM" (5.35 vs 6.36)

4.4.5 VarMax SCM portfolios

Finally, we test our asset selection methodology on the VarMax portfolio. It is probably the allocation strategy that requires the most asset selection before building the portfolio. Indeed, to obtain the most diversified portfolio, it will search for the least correlated assets but the correlation does not reflect the interconnectedness. An asset with low correlation may turn out to be potentially highly systemic since it carries a new risk premium (e.g. US subprime crisis, Italy/Spain sovereign debt crisis, etc.) or strongly influenced and overreacting to every market shock. The purpose of testing our methodology on the VarMax portfolio is to see if the asset selection improves portfolio performances by not proposing some low correlated assets that are highly nested in the network and low performing without drastically reducing the Variety ratio ($\mathcal{VR}$) defined in 4.3. In Fig. 4.8 and Table 4.5, we provide VarMax portfolios' performances and performance indicators over the whole period.

[8]The Herfindahl-Hirschman index (HHI) is the sum of the square of portfolio weights. Here, we use the inverse of the HHI where 1 represents the minimum value (only one asset in the portfolio) and m the maximum value (EW portfolio).

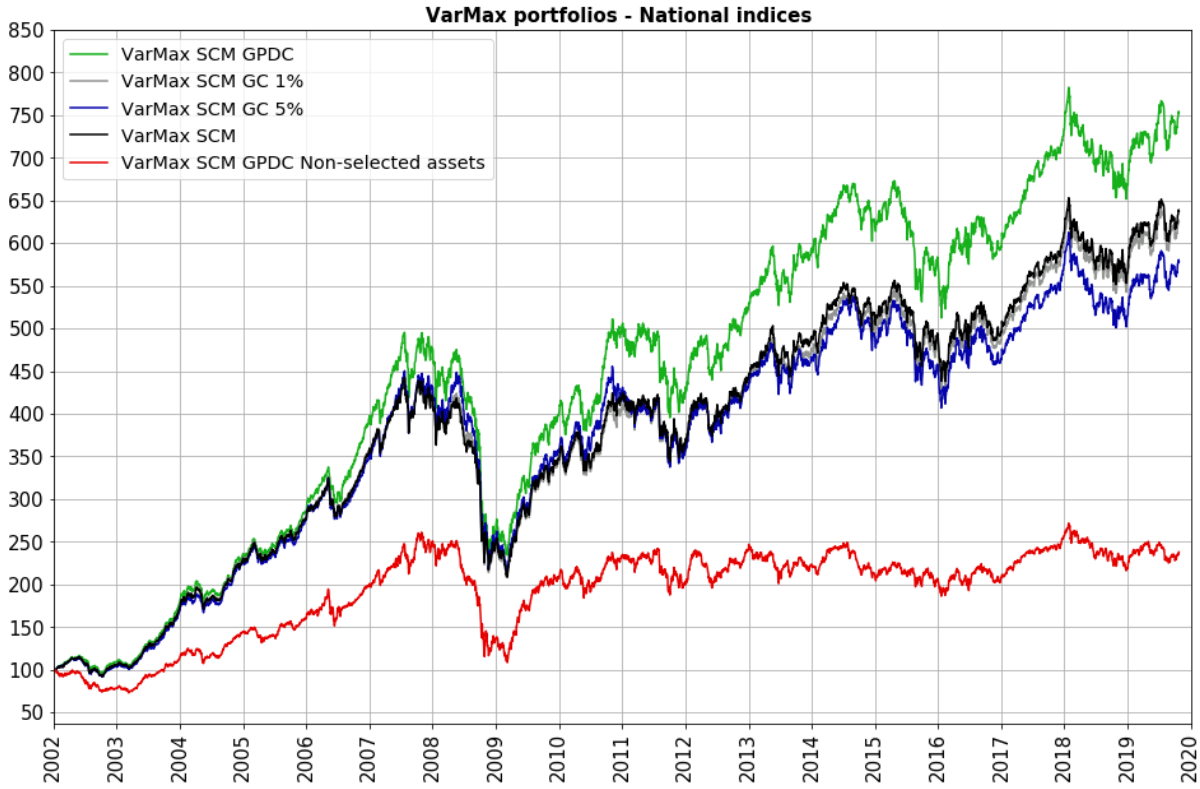


Figure 4.8: VarMax portfolios' performances on national indices with 0.07% of fees from January 2002 to October 2019. The proposed "VarMax SCM GPDC" (green line) leads to improved performances vs the "VarMax SCM GC 1%" (grey), the "VarMax SCM GC 5%" (blue) and the "VarMax SCM" (black). Moreover, the "VarMax SCM GPDC Non-selected assets" (red) provides the worst performances by far.

VarMax Portfolios National Indices	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
VarMax SCM GPDC	12.02%	12.78%	0.94	53.61%
VarMax SCM	10.98%	12.20%	0.90	52.87%
VarMax SCM GC 1%	10.86%	12.35%	0.88	52.83%
VarMax SCM GC 5%	10.37%	13.11%	0.79	52.60%
VarMax SCM GPDC Non-selected assets	5.00%	15.41%	0.32	58.40%

Table 4.5: Performance indicators for VarMax portfolios on national indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

As expected, our methodology applied on the VarMax portfolio significantly improves the annualized return (more than 1%) and the the return/volatility ratio (0.94 vs 0.90) compared to "VarMax SCM". To confirm that our methodology does not significantly reduce portfolio diversification, we plot in Fig. 4.9 the $\mathcal{VR}$ for "VarMax SCM GPDC" and "VarMax SCM". Given these results, our approach does not distort the ratio profile, remaining relatively close to "VarMax SCM" but sometimes decreasing sharply as in 2002, 2004, 2006-2008, 2010 and 2016. For these different years, our method provides a higher annual return for 5 out of 7 years (2002, 2006, 2007, 2010 and 2016) compared to "VarMax SCM".

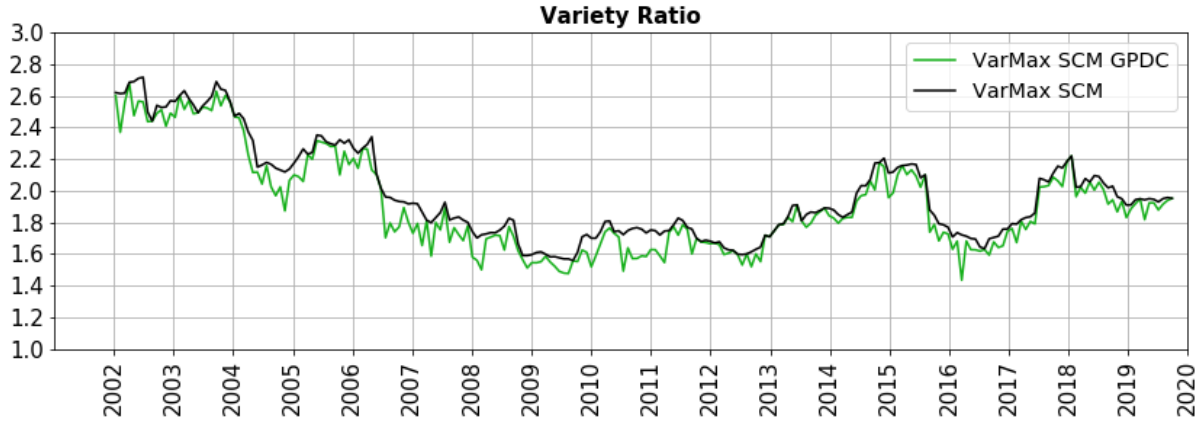


Figure 4.9: Variety ratio ($\mathcal{VR}$) defined in 4.3 for "VarMax SCM GPDC" (green line) and "VarMax SCM" (black line). "VarMax SCM GPDC" has on average a slightly lower ratio than "VarMax SCM" (1.91 vs 1.97)

To conclude, regarding the portfolio performances, our proposed asset selection method has improved annualized returns for all allocation strategies. The most significant improvement is for the VarMax strategy with a increase on the ratio return/volatility of 0.04. The only time that our methodology fails in terms of return/volatility ratio is on the MinVar portfolio mainly due to a high portfolio concentration. For ERC, MinVar and VarMax portfolios, the covariance matrix has been estimated using the SCM that is optimal under Normal assumptions. However, most of financial time series of returns exhibit fat tails and asymmetry that are hardly compatible with the Gaussian hypothesis. In order to reduce covariance matrix estimation errors, we use our cleaned and robust covariance matrix estimation (whitening procedure). The whitening procedure has already confirmed good results on VarMax and MinVar strategies in chapter 2.

4.4.6 Whitening procedure

In this section, we conclude the empirical analysis by using a cleaned and robust covariance matrix estimation (whitening procedure) [222, 223] instead of the Sample Covariance Matrix (SCM) in order to see if we can improve even more the results. We compare our proposed asset selection procedure "GPDC" with a robust covariance matrix estimation "WH". For this purpose, we only compare "GPDC WH" and "GPDC SCM" with those obtained on the whole universe, i.e. without applying any selection "WH" nor "SCM". We provide in Table 4.6 performance indicators computing over the whole period (January 2002 to October 2019) for ERC, MinVar, VarMax portfolios.

Portfolios **National Indices**	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
ERC WH GPDC	10.32%	15.32%	0.67	59.28%
ERC SCM GPDC	10.13%	15.26%	0.66	59.21%
ERC WH	9.81%	15.14%	0.65	59.75%
ERC SCM	9.68%	15.08%	0.64	59.65%
MinVar WH GPDC	11.23%	11.88%	0.95	52.68%
MinVar WH	10.65%	11.38%	0.94	49.78%
MinVar SCM	9.67%	10.73%	0.90	47.36%
MinVar SCM GPDC	9.92%	11.24%	0.88	52.33%
VarMax WH GPDC	11.98%	12.52%	0.96	53.00%
VarMax SCM GPDC	12.02%	12.78%	0.94	53.61%
VarMax SCM	10.98%	12.20%	0.90	52.87%
VarMax WH	10.83%	12.10%	0.90	52.15%

Table 4.6: GPDC asset selection and whitening procedure. Performance indicators for ERC, MinVar and VarMax portfolios on national indices combining with 0.07% of fees from January 2002 to October 2019. For each portfolio, the results are ranked in descending order according to the ratio (Return / Volatility).

When combining our proposed asset selection method and a robust covariance matrix, the portfolio performances improve even more. Indeed, for all allocation strategies the return/volatility ratio is higher than "GPDC SCM". When we used the MinVar portfolio with the SCM, our asset selection method had a lower return/volatility ratio than the portfolio applied on all assets. Now, using both our proposed asset selection method and the whitening procedure for the covariance matrix estimation, the portfolio "MinVar GPDC WH" provides the highest return/volatility ratio. As seen in chapter 2, using a robust approach does flatten the volatility differences between assets, allowing to improve the diversification ratio and reduce the portfolio concentration. The HHI of "MinVar GPDC WH" is now equal to 8.29.

Furthermore, in Appendix 4.6.1, to check the robustness of our methodology, we provide results on another universe composed of $m = 44$ GICS [193] sector indices. In this universe as well, the results are promising in improving portfolio performances for all allocation strategies.

4.5 Conclusion

In asset allocation problems, assessing the dependency structure among financial assets is of major importance to control the risk within investment universes and improve portfolio performances. For this purpose, the network approach provides useful insights to characterize the interactions between assets and thus determine their position (systemic or influenced). However, common approaches (the sample correlation matrix or Granger non-causality tests) to modeling the interactions lead to incomplete information never providing a directed weighted network. In this paper, to recover the topology, we use the GPDC measure in which are plugged the VAR coefficients estimated via the mBTS-TD strategy. This approach has two advantages: first, the mBTS-TD strategy solves the network dimensionality issue; second, the GPDC measure assesses both the direction and the strength of the relationships providing a precise network topology. Then, to characterize the relationships and reduce asset connections within the initial universe, we propose a dynamic pre-selection method based on the in and out triangle patterns of the local directed weighted clustering coefficient. This method allows to remove the most unstable assets (systemic and influenced) and adapts to the numbers of connections in the network removing only the most embedded assets without setting a number of asset exclusions or a threshold on the local directed weighted clustering coefficient.

Our dynamic pre-selection method applied on classical allocation strategies such as Equally Weighted, Equal Risk Contribution, Minimum Variance and Maximum Variety leads to significantly improve portfolio performances with respect to a network based on Granger non-causality or on the whole universe. Indeed, on 40 national equity indices the pre-selection process succeeds in identifying the worst performing/riskiest assets to remove them before allocating portfolios. Finally, we have associated our process with our whitening procedure [222, 223] to estimate the covariance matrix and again, the results are improved compared to the use of the classical Sample Covariance Matrix.

4.6 Appendix Chapter 4

4.6.1 Sector universe portfolio results

The second universe consists of $m = 44$ GICS [193] sector indices representing four geographical areas (MSCI Emerging Markets, MSCI Europe, MSCI Japan, MSCI United States) summarized in Table 4.16b. The sector indices universe allows us to focus more on economic activity interactions within or between the geographical areas.

EW portfolios - Sector indices

We examine our proposed methodology for the EW portfolio. As for the national indices universe, we report the portfolios' performances. Then, we focus on the asset return distribution between rebalancing dates and on the number of non-selected assets. Finally, we depict in Fig. 4.12 the networks based on GPDC method obtained for the sector indices in the window covering the period of November 16th, 2007 to November 7th, 2008 (financial crisis). In Fig. 4.10 and Table 4.7, we provide the evolution of the EW portfolios and some statistics on the overall portfolio performances.

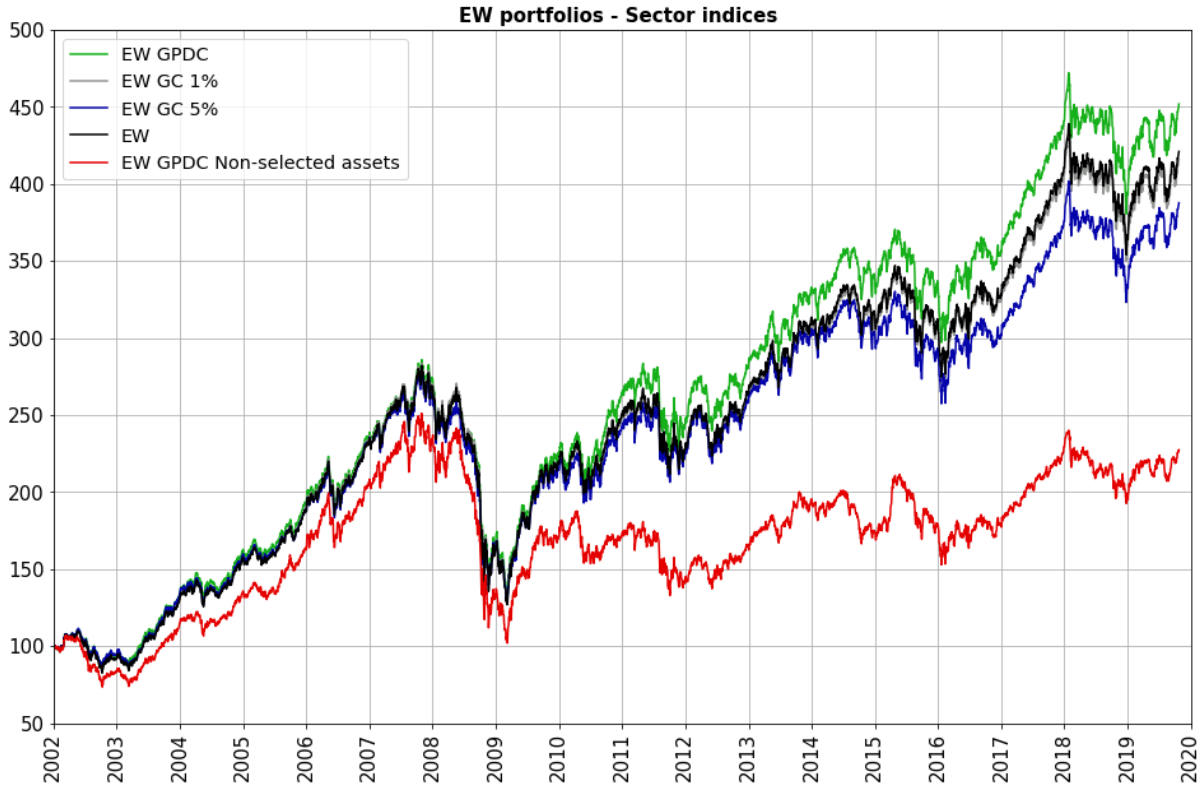


Figure 4.10: EW portfolios' performances on sector indices with 0.07% of fees from January 2002 to October 2019. The proposed "EW GPDC" (green line) leads to improved performances vs the "EW GC 1%" (grey), the "EW GC 5%" (blue) and the "EW" (black). Moreover, the "EW GPDC Non-selected assets" (red) provides the worst performances by far.

EW Portfolios Sector Indices	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
EW GPDC	8.85%	13.76%	0.64	54.03%
EW	8.42%	13.86%	0.61	55.06%
EW GC 1%	8.34%	13.81%	0.60	55.07%
EW GC 5%	7.92%	13.74%	0.58	54.33%
EW GPDC Non-selected assets	4.73%	16.10%	0.29	59.34%

Table 4.7: Performance indicators for EW portfolios on sector indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

As for the national indices, the proposed GPDC asset selection method shows an improvement with respect to the other methods. The other two methods "GC 1%" and "GC 5%" do not improve the EW portfolio, especially with the procedure using "GC 5%" where the return/volatility ratio is lower. When we focus on the non-selected assets by the GPDC process, here again we see clearly that all indicators are deteriorated with a return/volatility ratio less than at least 0.29 and a maximum drawdown more than 4% compared to the other methods.

Once again, we can consider that the GPDC asset selection process succeeds in identifying the less risky and less performing assets. The Table 4.8 provides the four moments of asset return distribution for the selected and non-selected assets. The asset return distribution of the selected assets (Table 4.8) provides better figures for all moments, especially for the skewness and the kurtosis.

GPDC Asset return distribution	Selected Assets	Non-Selected Assets	All Assets
Mean	0.0004	0.0003	0.0004
Standard Deviation	0.0138	0.0156	0.0141
Skewness	0.0356	0.0094	0.0281
Kurtosis	8.4141	9.3456	8.8036

Table 4.8: Moments of asset return distribution for the selected and non-selected assets by the GPDC process based on sector indices from January 2002 to October 2019.

The number of non-selected assets for the three asset selection procedures is plotted in Fig. 4.11. The "GC 1%" deletes only one asset on average and only six during the 2008 crisis with similar results compared to the EW portfolio without selection. The other two methods are more actives in selecting assets with an average of more than 10 non-selected assets. Based on Table 4.7 and Fig. 4.11, it is clear that for sector indices, the "CG 5%" method fails to identify the worst performing/riskiest assets. On the contrary, the GPDC procedure not only improves portfolio results, but also, as in the previous universe, provides greater stability in terms of the number of assets selected. The standard deviation of the number of assets selected for the "GPDC" procedure is 3.6 vs 5.2 for "GC 5%".

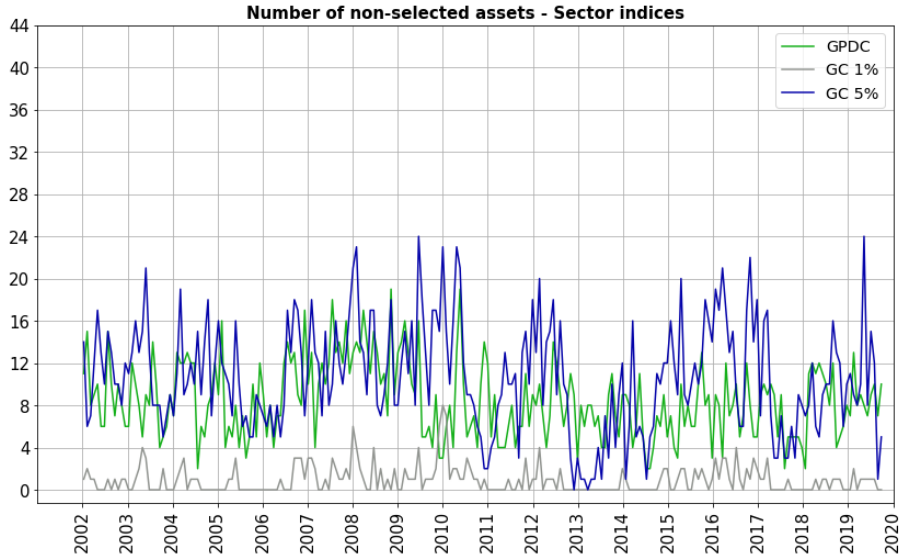


Figure 4.11: Number of non-selected assets for the three asset selection procedure in the sector indices universe ($m = 44$) from January 2002 to October 2019.

In Fig. 4.12, we depict a network based on the GPDC method obtained for the financial crisis in the window covering the period of November 16th, 2007 to November 7th, 2008. Here, the nodes represent sector indices (Table 4.16b). This example also perfectly illustrates our proposed methodology. In this Fig. 4.12, we see that the non-selected assets are for the most part sectors directly involved in the financial crisis (financial and real estate sectors) or suffering the consequences of an economic crisis, such as the consumer discretionary (automobile, restaurants, etc.), industrial and materials sectors. Indeed, our procedure removes the financial and real estate sectors in the US. It also removes consumer discretionary in three areas (EM, JP and US), industrial for EM and JP, and finally materials for EU, JP and US. During the December 1, 2008 trading session, which was one of the largest daily percentage losses, our methodology was successful in removing the U.S. financial and real estate sectors, which fell 15% and 18% respectively. In comparing these two portfolios over the next four weeks, we note that the portfolio of non-selected assets underperforms by 1% the portfolio with the selected assets.

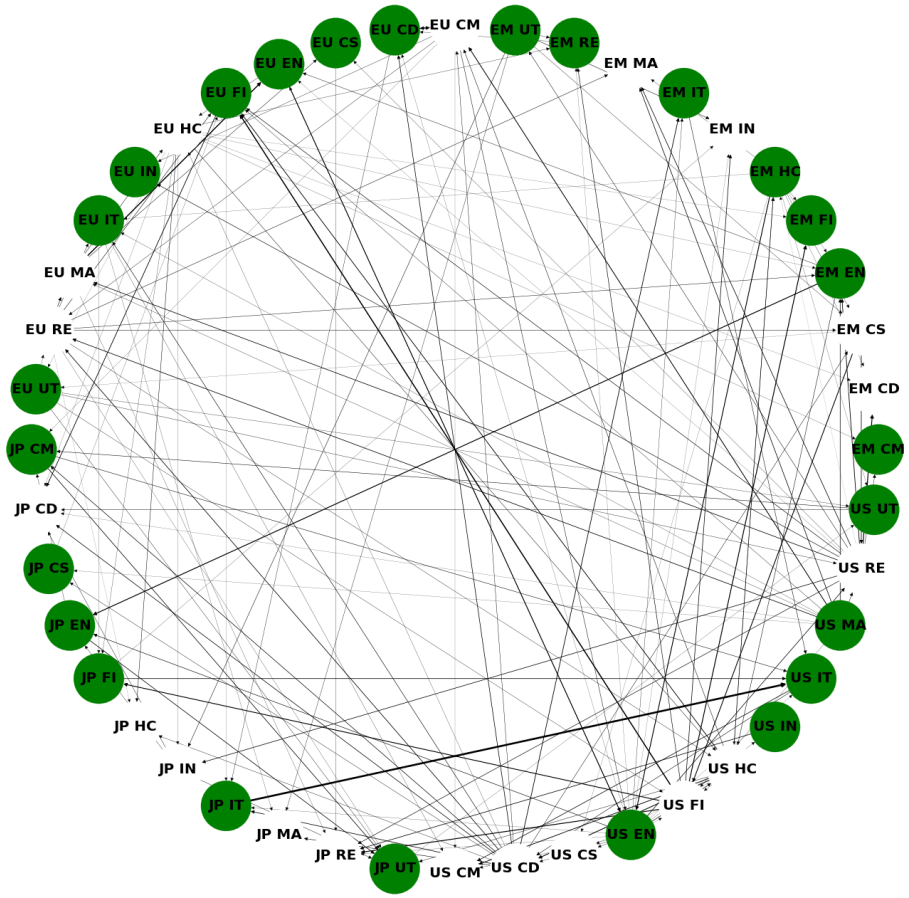


Figure 4.12: Directed weighted network of sector indices based on the GPDC measure and estimated from November 16th, 2007 to November 7th, 2008 (financial crisis). In green are the selected assets according to the method described in section 4.2 using essentially the triangle patterns (*in* and *out*) of the local directed weighted clustering coefficient. Conversely, the assets in white are not retained. The width of the directed edge represents the causal strength. Asset abbreviations are in Table 4.16b.

In this section, we have confirmed that our methodology improves the performances of the EW portfolio. Here again, our methodology succeeds in identifying the worst performing/riskiest assets in order to remove them from the portfolio. In the next three sections, we test our methodology on ERC, MinVar and VarMax strategies using the SCM.

4.6.1.1 ERC SCM portfolios - Sector indices

We provide in Fig. 4.13 and Table 4.9 the evolution of the ERC portfolios and performance indicators on the overall period.

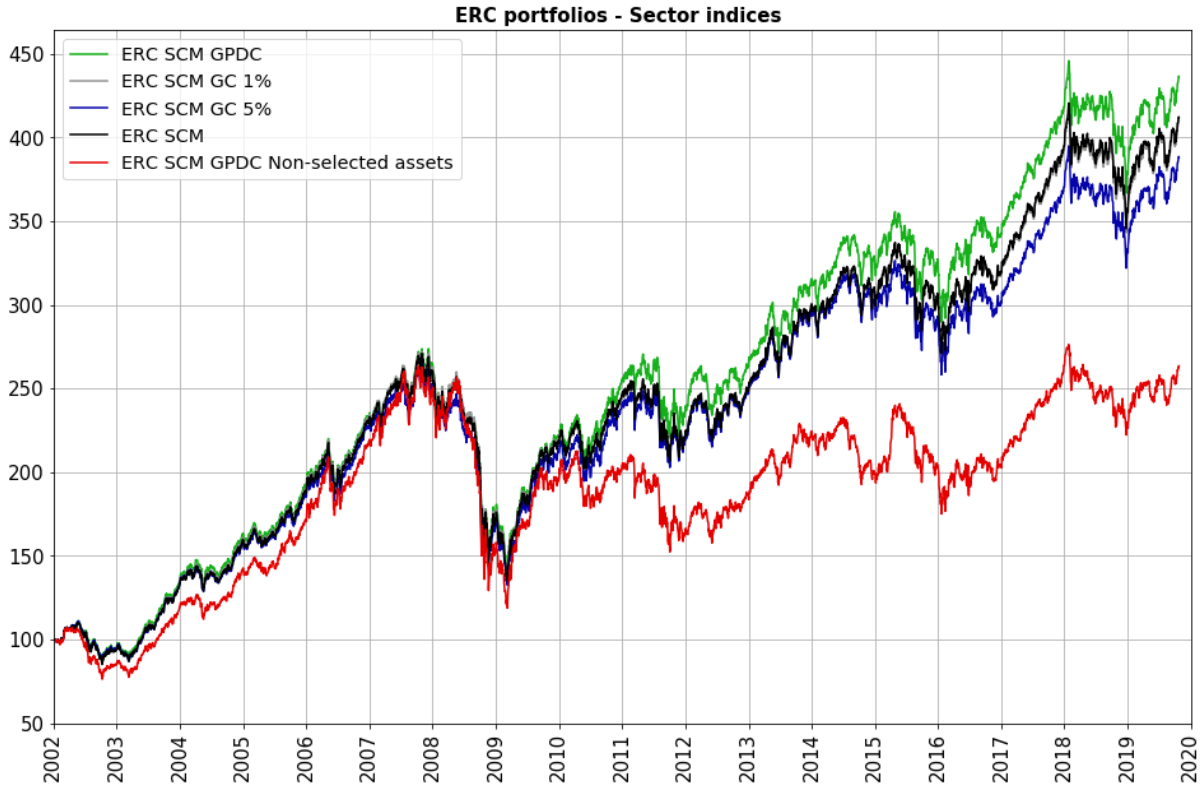


Figure 4.13: ERC portfolios' performances on sector indices with 0.07% of fees from January 2002 to October 2019. The proposed "ERC SCM GPDC" (green line) leads to improved performances vs the "ERC SCM GC 1%" (grey), the "ERC SCM GC 5%" (blue) and the "ERC SCM" (black). Moreover, the "ERC SCM GPDC Non-selected assets" (red) provides the worst performances by far.

ERC Portfolios Sector Indices	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
ERC SCM GPDC	8.64%	12.29%	0.70	49.10%
ERC SCM	8.29%	12.44%	0.67	50.17%
ERC SCM GC 1%	8.26%	12.41%	0.67	50.08%
ERC SCM GC 5%	7.93%	12.41%	0.64	49.97%
ERC SCM GPDC Non-selected assets	5.60%	15.03%	0.37	54.89%

Table 4.9: Performance indicators for ERC portfolios on sector indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

For ERC strategy, the dynamic pre-selection shows an improvement with respect to the other methods. For this universe, all indicators related to the "GPDC" are improved (higher annualized return, lower annualized volatility, higher return/volatility ratio and a lower maximum drawdown). Considering all periods, the asset selection removes on average 16% of the 1st quartile of volatility (low) and 21% of the 4th quartile (high). In this universe, the method tends to remove more high volatility assets, slightly biasing the universe on the low-volatility anomaly. However, in this type of universe the most systemic or influenced assets are more related to cyclical sectors assimilated to the value factor with generally higher betas and higher volatilities.

4.6.1.2 MinVar SCM portfolios - Sector indices

In Fig. 4.14 and Table 4.10, we provide the MinVar portfolio performances and performance indicators on the whole period.

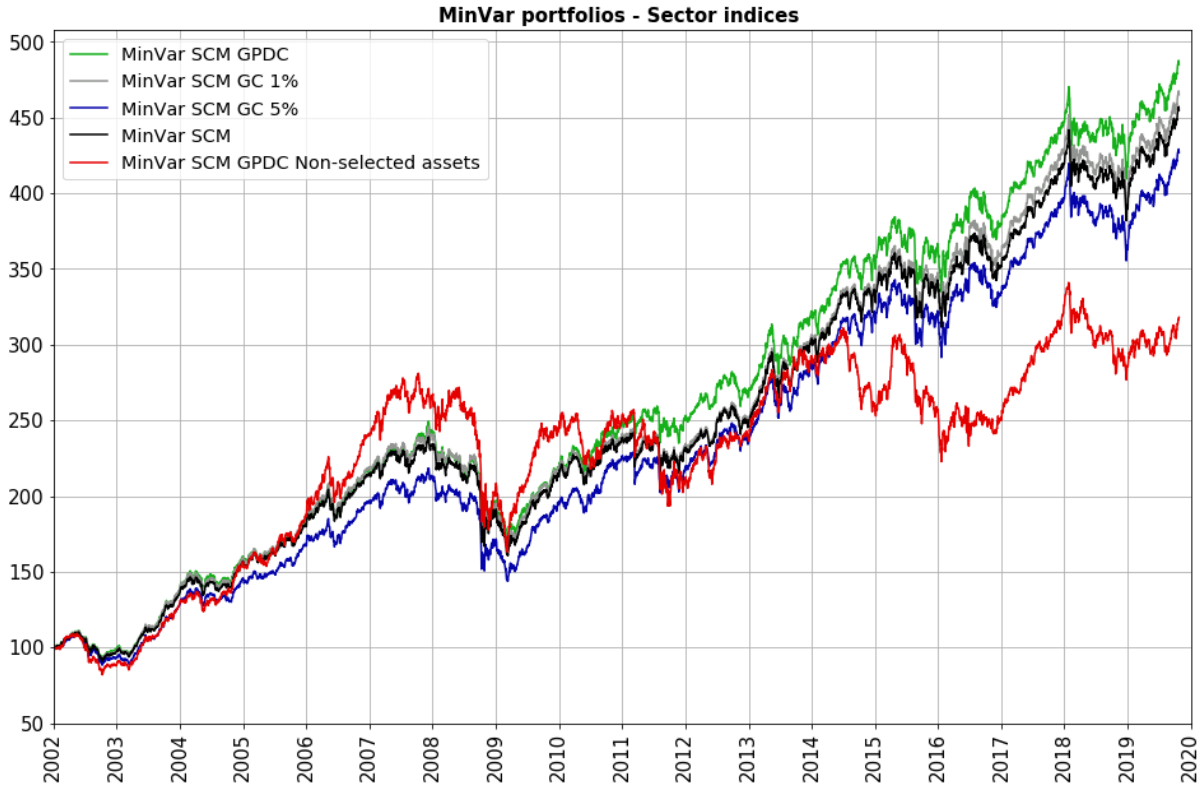


Figure 4.14: MinVar portfolios' performances on sector indices with 0.07% of fees from January 2002 to October 2019. The proposed "MinVar SCM GPDC" (green line) leads to improved performances vs the "MinVar SCM GC 1%" (grey), the "MinVar SCM GC 5%" (blue) and the "MinVar SCM" (black). Moreover, the "MinVar SCM GPDC Non-selected assets" (red) provides the worst performances by far.

MinVar Portfolios **Sector Indices**	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
MinVar SCM GPDC	9.29%	9.52%	0.97	32.48%
MinVar SCM GC 1%	9.03%	9.48%	0.95	33.34%
MinVar SCM	8.89%	9.49%	0.94	32.82%
MinVar SCM GC 5%	8.50%	9.67%	0.88	34.13%
MinVar SCM GPDC Non-selected assets	6.72%	13.19%	0.51	42.00%

Table 4.10: Performance indicators for MinVar portfolios on sector indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

Unlike national indices, the proposed asset selection shows a significant improvement for MinVar strategy where all indicators are similar or improved (higher annualized return, similar annualized volatility, higher return/volatility ratio and a lower maximum drawdown) when compared to other methods. In this universe, MinVar portfolios appear less concentrated with an HHI equal to 7.80 for "MinVar SCM GPDC" against 9.41 for "MinVar SCM".

4.6.1.3 VarMax SCM portfolios - Sector indices

We provide in Fig. 4.15 and Table 4.11 the evolution of the MinVar portfolios and some statistics on the overall portfolio performances for the case of sector indices.

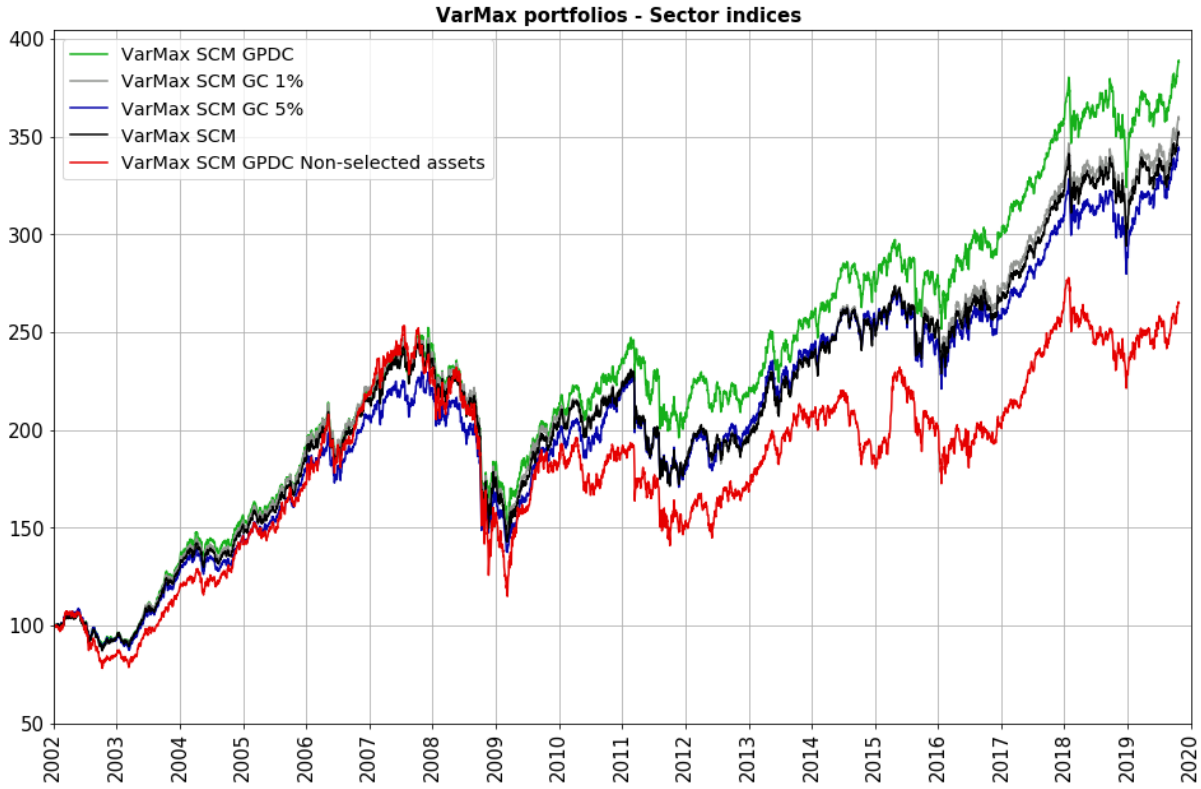


Figure 4.15: VarMax portfolios' performances on sector indices with 0.07% of fees from January 2002 to October 2019. The proposed "VarMax SCM GPDC" (green line) leads to improved performances vs the "VarMax SCM GC 1%" (grey), the "VarMax SCM GC 5%" (blue) and the "VarMax SCM" (black). Moreover, the "VarMax SCM GPDC Non-selected assets" (red) provides the worst performances by far.

VarMax Portfolios **Sector Indices**	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
VarMax SCM GPDC	7.93%	11.10%	0.71	39.88%
VarMax SCM GC 1%	7.45%	11.34%	0.66	41.31%
VarMax SCM	7.32%	11.37%	0.64	41.65%
VarMax SCM GC 5%	7.18%	11.32%	0.63	40.22%
VarMax SCM GPDC Non-selected assets	5.63%	15.01%	0.38	54.67%

Table 4.11: Performance indicators for VarMax portfolios on sector indices with 0.07% of fees from January 2002 to October 2019. The results are ranked in descending order according to the ratio (Return / Volatility).

As for national indices universe, the asset selection applied on VarMax strategy shows a significant improvement with respect to the other methods. Indeed, all indicators are improved (higher annualized return, lower annualized volatility, higher return/volatility ratio and a lower maximum drawdown). Note that the dynamic pre-selection does not distort the $\mathcal{VR}$ ratio. "VarMax SCM GPDC" has on average a $\mathcal{VR}$ ratio of 1.97 vs 2.02 for "VarMax SCM".

For sector indices universe, our proposed asset selection method has improved portfolio performances for all allocation strategies. Here again, the most significant improvement is for the VarMax strategy with a increase on the ratio return/volatility of 0.05.

4.6.1.4 Whitening procedure - Sector indices

We provide in Table 4.12 performance indicators for ERC, MinVar and VarMax portfolios where the covariance matrix is estimated using the whitening procedure.

Portfolios **Sector Indices**	Ann. Return	Ann. Volatilty	Ratio Return/Volatilty	Max Drawdown
ERC WH GPDC	8.76%	12.20%	0.72	48.62%
ERC SCM GPDC	8.64%	12.29%	0.70	49.10%
ERC WH	8.42%	12.32%	0.68	49.22%
ERC SCM	8.29%	12.44%	0.67	50.17%
MinVar SCM GPDC	9.29%	9.52%	0.97	32.48%
MinVar SCM	8.89%	9.49%	0.94	32.82%
MinVar WH GPDC	8.63%	9.95%	0.87	35.09%
MinVar WH	8.18%	9.93%	0.82	31.98%
VarMax WH GPDC	8.04%	10.53%	0.76	38.84%
VarMax WH	7.68%	10.57%	0.73	36.74%
VarMax SCM GPDC	7.93%	11.10%	0.71	39.88%
VarMax SCM	7.32%	11.37%	0.64	41.65%

Table 4.12: GPDC asset selection and whitening procedure. Performance indicators for ERC, MinVar and VarMax portfolios on sector indices with 0.07% of fees from January 2002 to October 2019. For each portfolio, the results are ranked in descending order according to the ratio (Return / Volatility).

For this universe, by combining our proposed methodology and the whitening procedure to estimate the covariance matrix, we improve even more the portfolio performances for ERC and VarMax portfolios. Nevertheless, the whitening procedure applied on MinVar portfolio fails to improve the performances, even if the selection asset process "MinVar GPDC WH" improves the indicators when compared to "MinVar WH".

4.6.2 Country and sector dataset

Index	Currency	Abr.
Australia	AUD	AU
Austria	EUR	AT
Belgium	EUR	BE
Brazil	BRL	BR
Canada	CAD	CA
Chile	CLP	CL
China	HKD	CN
Colombia	COP	CO
Denmark	DKK	DK
Finland	EUR	FI
France	EUR	FR
Germany	EUR	DE
Hong Kong	HKD	HK
India	INR	IN
Indonesia	IDR	ID
Ireland	EUR	IR
Israel	ILS	IL
Italy	EUR	IT
Japan	JPY	JP
Korea	KRW	KO
Malaysia	MYR	MA
Mexico	MXN	MX
Netherlands	EUR	NL
New Zealand	NZD	NZ
Norway	NOK	NO
Peru	PEN	PR
Philippines	PHP	PL
Poland	PLN	PO
Portugal	EUR	PT
Russia	RUB	RU
Singapore	SGD	SG
South Africa	ZAR	SA
Spain	EUR	SP
Sweden	SEK	SW
Switzerland	CHF	CH
Taiwan	TWD	TW
Thailand	THB	TH
Turkey	TRY	TR
United Kingdom	GBP	UK
United States	USD	US

(a) Country equity indices representing 40 countries of the MSCI ACWI (All Country World Index).

Index	Currency	Abr.
EM Communication Services	USD	EM CM
EM Consumer Discretionary	USD	EM CD
EM Consumer Staples	USD	EM CS
EM Energy	USD	EM EN
EM Financials	USD	EM FI
EM Health Care	USD	EM HC
EM Industrials	USD	EM IN
EM Information Technology	USD	EM IT
EM Materials	USD	EM MA
EM Real Estate	USD	EM RE
EM Utilities	USD	EM UT
EU Communication Services	EUR	EU CM
EU Consumer Discretionary	EUR	EU CD
EU Consumer Staples	EUR	EU CS
EU Energy	EUR	EU EN
EU Financials	EUR	EU FI
EU Health Care	EUR	EU HC
EU Industrials	EUR	EU IN
EU Information Technology	EUR	EU IT
EU Materials	EUR	EU MA
EU Real Estate	EUR	EU RE
EU Utilities	EUR	EU UT
JP Communication Services	JPY	JP CM
JP Consumer Discretionary	JPY	JP CD
JP Consumer Staples	JPY	JP CS
JP Energy	JPY	JP EN
JP Financials	JPY	JP FI
JP Health Care	JPY	JP HC
JP Industrials	JPY	JP IN
JP Information Technology	JPY	JP IT
JP Materials	JPY	JP MA
JP Real Estate	JPY	JP RE
JP Utilities	JPY	JP UT
US Communication Services	USD	US CM
US Consumer Discretionary	USD	US CD
US Consumer Staples	USD	US CS
US Energy	USD	US EN
US Financials	USD	US FI
US Health Care	USD	US HC
US Industrials	USD	US IN
US Information Technology	USD	US IT
US Materials	USD	US MA
US Real Estate	USD	US RE
US Utilities	USD	US UT

(b) Sector equity indices representing Emerging Markets (EM), Europe (EU), Japan (JP) and United States (US) sectors.

Figure 4.16: List of assets for both universes (currency and abbreviation).

Chapter 5

Conclusion

In this thesis, we addressed different aspects of portfolio allocation issues allowing the improvement of the covariance matrix estimation, proposing a method to identify more precisely the dependency structure among financial assets and finally defining a dynamic network indicator to detect unstable assets (systemic/influenced). The join use of these approaches provides a complete asset selection methodology with promising results on the performances of risk-based allocation strategies such as the Equal Risk Contribution (ERC), Minimum Variance (MinVar) and Variety Maximum (VarMax) portfolios.

In the following we summarize the main results of this thesis and suggest several research avenues or extensions.

Chapter 2 contributes to the literature of covariance matrix estimation in the sense that it extends recent works to asset returns that are globally non-homogeneously distributed but grouped into homogeneous distributed classes. Under this hypothesis, although asset returns are still modelled according to a multi-factor model embedded in correlated elliptical and symmetric noise, we propose to distinguish p groups of assets that are homogeneous in law. The whitening procedure proposed in [163, 164, 165] which combines the robust Tyler-M estimator and the RMT results is then applied at the level of each group. For determining the number of groups p, two classification methods are investigated: the Ascending Hierarchical Clustering (AHC) method that requires the number of groups to be either fixed *a priori* or determined using a predefined criterion, and the Affinity Propagation (AP) method that self-determines the number of groups. Applying such covariance matrix estimation on the MinVar and VarMax portfolios improved the overall performances with respect to several classical estimators such as the Ledoit & Wolf (LW) shrinkage estimator, the Eigenvalue clipping method, and the optimal Rotational Invariant Estimator (RIE). In addition, we show the superiority of the AP algorithm to produce higher performances for both the European universe (composed of 43 industrial sub-sector, factor and country equity indices) and the US universe (composed of 30 industrial sub-sector and factor equity indices).

Chapter 3 focuses on complex interactions in multivariate systems to recover accurate financial network topology. To this end, the Generalized Partial Directed Coherence measure (GPDC) is used since it assesses both the direction and strength of relationships. However, as it is closely related to the estimation of the VAR coefficients, it cannot be used if the VAR model is not estimated in a parsimonious way, i.e. excluding non-significant

coefficients. To overcome this issue, we proposed an extended subset selection method (mBTS-TD) combining the modified Backward-in-Time Selection method (mBTS) and the Top-Down strategy (TD) that aims at removing non-significant coefficients even for high-dimensional systems. Through Monte Carlo simulations on different classical toy systems, we prove that the mBTS-TD method clearly outperforms the classical ones (Lasso method, t-test procedure, TD strategy, and mBTS method) in identifying causal and non-causal interactions. Then, the GPDC measure is applied to build financial causal network providing not only a precise network topology (directed weighted network), but also solving the network dimension puzzle (mBTS-TD method). And finally, computing the local directed weighted clustering coefficient of such networks allow us to remove the most systemic assets and improve the Equally Weighted (EW) portfolio performances.

In chapter 4, we proposed a dynamic pre-selection method based on the out and in triangle patterns of the local directed weighted clustering coefficient to identify and exclude the most unstable assets (systemic and influenced) to reduce systemic risk in the initial investment universe. Such an adaptive asset selection process follows the number of connections in the network and therefore removes the most embedded assets without imposing any exclusion threshold (chapter 3). The empirical study carried out on the EW, ERC, MinVar and VarMax portfolios shows that this dynamic asset selection strategy implemented on the GPDC financial network improves the overall financial performances when compared to Granger-based approaches or when allocating the whole universe. Indeed, the asset selection process succeeds in identifying for two universes (national or sector indices belonging to the MSCI All Country World Index) the worst performing/riskiest assets before allocating portfolios. What is more, when we associate our dynamic pre-selection method and the whitening procedure presented in chapter 2, the results are even improved.

As a continuation of this thesis, several extensions could be considered. In chapter 2, we used a straightforward extension of the consistency theorem (2.7) [163, 164, 165] in the case where the scatter matrix $\mathbf{C}$ is block-Toeplitz structured. In fact, we based our approach on homogeneous groups of assets with a strong independence hypothesis between groups. In this case, the almost sure convergence in spectral norm of the Tyler-M estimate, under the RMT regime, can be simply obtained using arguments at a group level. A natural question is to see if the consistency theorem still holds if we impose a complex covariance structure between groups to cope with more realistic financial situations. Another interesting way is to generalize this whitening procedure in the case $m > T$, by using either the main factors identified by the whitening process on a similar sub-universe or hybrid robust shrinkage covariance matrix estimators [116, 117, 115, 62, 224]. It would also be interesting to focus on other clustering methods for identifying homogeneous groups (spectral clustering [225]) and also to examine recent covariance matrix estimation methods (lassoing eigenvalues [226] and linear pooling [227]). From a more practical viewpoint, the main factors identified can also be used to create dynamic factor portfolios, and what is more, such a whitening procedure can be very useful to other finance applications (risk measures, correlation-based network, clustering methods, etc.) or in other fields such as radar and hyperspectral fields.

Concerning the parsimonious VAR model and the GPDC estimation accuracy, there are also several research avenues to be pursued. The case of non-Gaussian noises should be addressed for subset VAR models. In [211], Qiu et al proposed an elliptical VAR model based on the Yule-Walker equation [210] and where scatter matrices are estimated with the Q_n estimator [228]. We believe that robust matrix covariance estimators such as the Tyler-M estimator can provide promising results, even if matrix inversion problems will arise due to the "pure" collinearity between the lagged variables. In addition, although the mBTS-TD method behaves rather well for correlated noises (instantaneous interactions), we are interested in studying both parsimonious estimates in the case of structural VAR and/or examining more carefully the iGPDC (instantaneous GPDC) [229]. The case of non-stationary data should also be addressed in future research, which could be useful for trends in time series such as in macroeconomics. Furthermore, the mBTS-TD method and the GPDC network-based approach can also be very useful for the field of neuroscience to understand functional connectivity patterns between different brain regions.

Finally, based on the last chapter, it would also be interesting to apply this dynamic pre-selection procedure on nonlinear measures such as the PMIME [73]. We also believe that our methodology can be used to track market indices by only identifying the most central assets, which can be very useful for passive strategies. The current method can be widely extended to different asset classes, such as foreign exchange markets, sovereign bonds, but also multi-asset portfolios.

Bibliography

[1] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[2] H. M. Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly*, 3:111–133, 1956.

[3] V. K. Chopra and W. T. Ziemba. The effects of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11, 1993.

[4] R. C. Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323 – 361, 1980.

[5] M. Best and R. R. Grauer. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, 4:315–342, 1991.

[6] S. D Hodges and R. A Brealey. Portfolio selection in a dynamic and uncertin world, in James H. Lorie and R. A. Brealey (eds.). *Modern Developments in Investment Management*, pages 2nd ed. Hinsdale, IL: Dryden Press, 1978.

[7] Richard O. Michaud. The markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

[8] R. Litterman. *Modern Investment Management: An Equilibrium Approach*. Wiley, 2003.

[9] C. B. Barry. Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, 29(2):515–522, 1974.

[10] V. S. Bawa, S. Brown, and R. Klein. *Estimation Risk and Optimal Portfolio Choice*. 1979.

[11] D. Jobson, R. B. Korkie, and V. Ratti. Improved estimation for markowitz portfolios using james-stein estimators, 1979. volume 41, pages 279–92, 1979.

[12] J. D. Jobson and B. Korkie. Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554, 1980.

[13] P. Jorion. International portfolio diversification with estimation risk. *The Journal of Business*, 58(3):259–278, 1985.

[14] P. Jorion. Bayes-stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292, 1986.

[15] L. Pástor. Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223, 2000.

[16] L. Pástor and R. F. Stambaugh. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335 – 381, 2000.

[17] D. Goldfarb and G. Iyengar. Robust portfolio selection problem. *Mathematics of Operations Research*, 28:1–38, 2003.

[18] I. Garlappi, R. Uppal, and T. Wang. Portfolio selection with parameter and model uncertainty : A multi-prior approach. *The Review of Financial Studies*, 20:41–81, 2007.

[19] R. Kan and G. Zhou. Optimal portfolio choice with parameter uncertainty. *The Journal of Financial and Quantitative Analysis*, 42(3):621–656, 2007.

[20] A. C. MacKinlay and L. Pástor. Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Review of Financial Studies*, 13(4):883–916, 2000.

[21] M. J. Best and R. R. Grauer. Positively weighted minimum-variance portfolios and the structure of asset expected returns. *The Journal of Financial and Quantitative Analysis*, 27(4):513–537, 1992.

[22] L. Chan, J. Karceski, and J. Lakonishok. On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies*, 12:937–974, 1999.

[23] O. Ledoit and M. Wolf. Honey, i shrunk the sample covariance matrix. problems in mean-variance optimization. *Journal of Portfolio Management*, 30:110–119, 2004.

[24] F. Frost and J. Savarino. For better performance constrain portfolio weights. *Journal of Portfolio Management*, 15:29–34, 1988.

[25] V. K. Chopra. Improving optimization. *Journal of investing*, 8:51–59, 1993.

[26] R. Jagannathan and T. Ma. Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1684, 2003.

[27] B. Scherer. Can robust optimization build better portfolios? *Journal of Asset Management*, 7:374–387, 2007.

[28] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification : How inefficient is the $1/n$ portfolio strategy? *The review of financial studies*, 22:1915–1953, 2009.

[29] S. Maillard, T. Roncalli, and J. Teiletche. The properties of equally weighted risk contributions portfolios. *Journal of Portfolio Management*, 36:60–70, 2010.

[30] Y. Choueifaty and Y. Coignard. Toward maximum diversification. *Journal of Portfolio Management*, 35(1):40–51, 2008.

[31] S. Benartzi and R. H. Thaler. Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1):79–98, March 2001.

[32] H. Windcliff and P. P. Boyle. The 1/n pension investment puzzle. *North American Actuarial Journal*, 8(3):32–45, 2004.

[33] G. Huberman and W. Jiang. Offering versus choice in 401(k) plans: Equity exposure and number of funds. *The Journal of Finance*, 61(2):763–801, 2006.

[34] R. A. Haugen and N. L. Baker. The efficient market inefficiency of capitalization–weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):35–40.

[35] R. G. Clarke, H. de Silva, and S. Thorley. Minimum-variance portfolios in the u.s. equity market. *The Journal of Portfolio Management*, 33(1):10–24, 2006.

[36] T. Chow, J. Hsu, V. Kalesnik, and B. Little. A survey of alternative equity index strategies. *Financial Analysts Journal*, 67(5):37–57, 2011.

[37] R. Clarke, H. De Silva, and S. Thorley. Minimum variance, maximum diversification, and risk parity: an analytic perspective. *Journal of Portfolio Management*, June 2012.

[38] B. Scherer. A note on the returns from minimum variance investing. *Journal of Empirical Finance*, 18:652–660, 2011.

[39] R. Clarke, H. de Silva, and S. Thorley. Minimum-variance portfolio composition. *The Journal of Portfolio Management*, 37(2):31–45, 2011.

[40] E. Qian. Risk parity portfolios: Efficient portfolios through true diversification. *Panagora Asset Management*, September 2005.

[41] E. Qian. On the financial interpretation of risk contribution: Risk budgets do add up. *Journal of Investment Management*, 4:1–11, 2006.

[42] D. G. Booth and Eugene F. Fama. Diversification returns and asset contributions. *Financial Analysts Journal*, 48(3):26–32, 1992.

[43] R. Fernholz, R. Garvy, and J. Hannon. Diversity-weighted indexing. *The Journal of Portfolio Management*, 24(2):74–82, 1998.

[44] E. Qian. Risk parity and diversification. *Journal of Investing*, 20:119–127, 2011.

[45] T. Griveau-Billion, J.-C Richard, and T. Roncalli. A fast algorithm for computing high-dimensional risk parity portfolios. *SSRN Electronic Journal*, 11 2013.

[46] W. Lee. Risk-based asset allocation: a new answer to an old question. *Journal of Portfolio Management*, 37:11–28, 2011.

[47] N. Baltas. 3 - trend-following, risk-parity and the influence of correlations. In Emmanuel Jurczenko, editor, *Risk-Based and Factor Investing*, pages 65 – 95. Elsevier, 2015.

[48] S. Darolles, C. Gouriéroux, and Emmanuelle Jay. 5 - robust portfolio allocation with systematic risk contribution restrictions. In Emmanuel Jurczenko, editor, *Risk-Based and Factor Investing*, pages 123 – 146. Elsevier, 2015.

[49] E. Jurczenko and J. Teiletche. 6 - risk-based investing but what risk(s)? In Emmanuel Jurczenko, editor, *Risk-Based and Factor Investing*, pages 147 – 171. Elsevier, 2015.

[50] B. Bruder, N. Kostyuchyk, and T. Roncalli. Risk parity portfolios with skewness risk: An application to factor investing and alternative risk premia. *Econometric Modeling: Capital Markets - Asset Pricing eJournal*, 2016.

[51] O. Blin, F. Ielpo, J. Lee, and J. Teiletche. 12 - a macro risk-based approach to alternative risk premia allocation. In Emmanuel Jurczenko, editor, *Factor Investing*, pages 285 – 316. Elsevier, 2017.

[52] Y. Choueifaty, T. Froidure, and J. Reynier. Properties of the most diversified portfolio. *Journal of investment strategies*, 2(2):49–70, 2013.

[53] L. Theron and G. van Vuuren. The maximum diversification investment strategy: A portfolio performance comparison. *Cogent Economics & Finance*, 6(1):1427533, 2018.

[54] W. F. Sharpe. Mutual fund performance. *The Journal of Business*, 39(1):119–138, 1966.

[55] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Physycal Review Letters*, 83(1468), 1999.

[56] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud. Random Matrix Theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3(03):391–397, 2000.

[57] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley. Collective behavior of stock price movements: A Random Matrix Theory approach. *Physica A*, 299:175–180, 2001.

[58] O. Ledoit and M. Wolf. Improved estimation of covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.

[59] M. Potters, J. P. Bouchaud, and L. Laloux. Financial applications of Random Matrix Theory: old laces and new pieces. *Acta Physica Polonica B*, 36(9), 2005.

[60] V. Demiguel and F. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57:560–577, 06 2009.

[61] J. P. Bouchaud and M. Potters. Financial applications of random matrix theory: a short review. *The Oxford handbook of Random Matrix Theory, Oxford University Press*, 2011.

[62] L. Yang, R. Couillet, and M. R. McKay. A robust statistics approach to minimum variance portfolio optimization. *IEEE Transactions on Signal Processing*, 63(24):6684–6697, Aug 2015.

[63] J. Bun, J. P. Bouchaud, and M. Potters. Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 666:1–165, 2017.

[64] B. Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36:394–394, 1963.

[65] E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.

[66] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223 – 236, 2001.

[67] G. Frahm, M. Junker, and A. Szimayer. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286, 07 2003.

[68] M. Junker and A. May. Measurement of aggregate risk with copulas. *The Econometrics Journal*, 8(3):428–454, 2005.

[69] R. N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems*, 11:193–197, 1999.

[70] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sector. *Journal of Financial Economics*, 104:535–559, 2012.

[71] F. Pozzi, T. Di Matteo, and T. Aste. Spread of risk across financial markets: better to invest in the peripheries. *Scientific Reports*, 3, 2013.

[72] G. Peralta and A. Zareei. A network approach to portfolio selection. *Journal of Empirical Finance*, 38:157 – 180, 2016.

[73] A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks. Financial networks based on granger causality: A case study. *Physica A: Statistical Mechanics and its Applications*, 482:65 – 73, 2017.

[74] H. Gatfaoui and P. de Peretti. Flickering in information spreading precedes critical transitions in financial markets. *Scientific Reports*, 9, 2019.

[75] G. Clemente, R. Grassi, and A. Hitaj. Asset allocation: new evidence through network approaches. *Annals of Operations Research*, pages 1–20, 2019.

[76] G. P. Clemente, R. Grassi, and A. Hitaj. Smart network based portfolios. *arXiv preprint*, 2019.

[77] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[78] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 197–206. University of California Press, 1956.

[79] W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 361–379. University of California Press, 1961.

[80] B. Efron and C. Morris. Stein's paradox in statistics. *Scientific American - SCI AMER*, 236:119–127, 05 1977.

[81] L. R. Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics*, 8(3):586–597, 05 1980.

[82] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411, 2004.

[83] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[84] B. Rosenberg. Extra-Market components of covariance in security markets. *Journal of Financial and Quantitative Analysis*, pages 263–274, March 1974.

[85] S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Finance*, 13(3), 1976.

[86] R. Grinold, A. Rudd, and D. Stefek. Global factors: Fact or fiction? *Journal of Portfolio Management*, pages 79–88, Fall 1989.

[87] E. F. Fama and K. R. French. The cross-section of expected stock returns. *Journal of Finance*, 1992.

[88] E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

[89] E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, April 2015.

[90] J. Fan, Y. Fan, and J. Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147:186–197, 2008.

[91] E. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

[92] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1(4):457–483, 1967.

[93] Z. Burda, J. Jurkiewicz, M Nowak, G. Papp, and I. Zahed. Free lévy matrices and financial correlations. *Physica A: Statistical Mechanics and its Applications*, 343:694–700, 2004.

[94] Z. Burda, J. Jurkiewicz, and B. Waclaw. Spectral moments of correlated wishart matrices. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 71:026111, 03 2005.

[95] D. Bartz and K.-R. Müller. Covariance shrinkage for autocorrelated data. volume 2, 12 2014.

[96] J. Vinogradova, R. Couillet, and W. Hachem. Statistical inference in large antenna arrays under unknown noise pattern. *IEEE Transactions on Signal Processing*, 61(22):5633–5645, Nov 2013.

[97] R. Couillet. Robust spiked random matrices and a robust G-MUSIC estimator. *Journal of Mult. Analysis*, 140:139–161, 2015.

[98] J. Bun, R. Allez, J. P. Bouchaud, and M. Potters. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory*, 62(12):7475–7490, 2016.

[99] J. Bun. *Application of Random Matrix Theory to HighDimensional Statistics*. PhD thesis, Paris-Saclay University, 2016.

[100] O. Ledoit and S. Péché. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1):233–264, 2011.

[101] O. Ledoit and M. Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies*, 30:4349–4388, 12 2017.

[102] O. Ledoit and M. Wolf. Numerical implementation of the QuEST function. *Computational Statistics & Data Analysis*, 115(C):199–223, 2017.

[103] P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[104] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *5st Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233, 1967.

[105] J. W. Tukey. A survey of sampling from contemned distributions. *Contributions to Probability ans Statistics Essays in Honor of Harold Hotelling*, pages 448–485, 1960.

[106] J. W. Tukey. The future of data analysis. *The annals of Mathematical Statistics*, 33(1):1–67, 1962.

[107] R. A. Maronna. Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, 4(1):51–67, January 1976.

[108] D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *The annals of Statistics*, 15(1):234–251, 1987.

[109] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, 32(4):419–430, 1970.

[110] S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.

[111] K. T. Fang and T. W. Anderson. *Statistical inference in elliptically contoured and related distributions*. Allerton Pr, 1990.

[112] K. T. Fang, S. Kotz, and K. W Ng. *Symmetric multivariate and related distributions*. Chapman & Hall, 1990.

[113] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Complex Elliptically Symmetric distributions: Survey, new results and applications. *IEEE Transactions on Signal Processing*, 60(11):5597–5625, Nov 2012.

[114] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal. Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1):34–48, Jan 2008.

[115] F. Pascal, Y. Chitour, and Y. Quek. Generalized robust shrinkage estimator and its application to STAP detection problem. *IEEE Transactions on Signal Processing*, 62(21), November 2014.

[116] Y. Abramovich and N. K. Spencer. Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering. In *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, volume 3, April 2007.

[117] Y. Chen, A. Wiesel, and A. O. Hero. Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Transactions on Signal Processing*, 59(9), September 2011.

[118] R. Couillet, F. Pascal, and J. W. Silverstein. Robust estimates of covariance matrices in the large dimensional regime. *IEEE Transactions on Information Theory*, 60(11), September 2014.

[119] R. Couillet and M. R. McKay. Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, 131, 2014.

[120] G. Bonanno, g. Caldarelli, F. Lillo, and R. N Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130, 2003.

[121] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertesz. Dynamic asset trees and black monday. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):247–252, 2003.

[122] V. Boginski, S. Butenko, and P. M. Pardalos. Statistical analysis of financial networks. *Computational statistics & data analysis*, 48(2):431–443, 2005.

[123] M. Tumminello, T. Di Matteo, T. Aste, and R. N. Mantegna. Correlation based networks of equity returns sampled at different time horizons. *The European Physical Journal B*, 55(2):209–217, 2007.

[124] T. Di Matteo, F. Pozzi, and T. Aste. The use of dynamical networks to detect the hierarchical organization of financial market sectors. *The European Physical Journal B*, 73(1):3–11, 2010.

[125] M. Billio, L. Frattarolo, H. Gatfaoui, and P. De Peretti. Clustering in dynamic causal networks as a measure of systemic risk on the euro zone. `https://halshs.archives-ouvertes.fr/halshs-01339826v1`, MAy 2016.

[126] G. Fagiolo. Clustering in complex directed networks. *Physical review. E*, 76:026107, 2007.

[127] G.P. Clemente and R. Grassi. Directed clustering in weighted networks: A new perspective. *Chaos, Solitons & Fractals*, 107:26 – 38, 2018.

[128] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426, 2005.

[129] K. Baba, R. Shibata, and M. Sibuya. Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, 46(4):657–664, 2004.

[130] D. Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R. N. Mantegna, and E. Ben-Jacob. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLOS ONE*, 5(12):1–14, 12 2010.

[131] G.-J. Wang, C. Xie, and H. E. Stanley. Correlation structure and evolution of world stock markets: Evidence from pearson and partial correlation-based networks. *Computational Economics*, 51(3):607–635, 2018.

[132] C. A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

[133] J. F. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77:304–313, 1982.

[134] J. F. Geweke. Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association*, 79(388):907–915, 1984.

[135] A. Zaremba and T. Aste. Measures of causality in complex datasets with application to financial data. *Entropy*, 16:2309–2349, 2014.

[136] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai. Granger causality network reconstruction of conductance-based integrate-and-fire neuronal systems. *PLOS ONE*, 9(2):1–17, 2014.

[137] E. Siggiridou, C. Koutlis, A. Tsimpiris, V. K. Kimiskidis, and D. Kugiumtzis. Causality networks from multivariate time series and application to epilepsy. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2015, Milan, Italy, August 25-29, 2015*, pages 4041–4044. IEEE, 2015.

[138] A. Porta and L. Faes. Wiener-granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proceedings of the IEEE*, 104:282–309, 2016.

[139] A. Sheikhattar, S. Miran, J. Liu, J. B. Fritz, S. A. Shamma, P. O. Kanold, and B. Babadi. Extracting neuronal functional network dynamics via adaptive granger causality analysis. *Proceedings of the National Academy of Sciences*, 115(17):E3869–E3878, 2018.

[140] E. Siggiridou, C. Koutlis, A. Tsimpiris, and D. Kugiumtzis. Evaluation of granger causality measures for constructing networks from multivariate time series. *Entropy*, 21(11):1080, 2019.

[141] H. Lütkepohl. *New introduction to multiple time series analysis*. Springer, 2005.

[142] H. Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, 1973.

[143] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[144] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195, 1979.

[145] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

[146] R. S. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley-Interscience, 2nd edition, 2005.

[147] A. Hatemi-J and R. S. Hacker. Can the lr test be helpful in choosing the optimal lag order in the var model when information criteria suggest different lag orders? *Applied Economics*, 41(9):1121–1125, 2009.

[148] S. B. Bruns and D. Stern. Lag length selection and p-hacking in granger causality testing: prevalence and performance of meta-regression models. *Empirical Economics*, 56(3):797–830, 2019.

[149] R. Brüggemann and H. Lütkepohl. Lag selection in subset var models with an application to a u.s. monetary system. Econometric Society World Congress 2000 Contributed Papers 0821, Econometric Society, 2000.

[150] R. Brüggemann. *Model reduction methods for vector autoregressive processes*. Springer, 01 2004.

[151] E. Siggiridou and D. Kugiumtzis. Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model. *IEEE Transactions on Signal Processing*, 64:1759–1773, 2016.

[152] T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.

[153] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948.

[154] L. Barnett, A. B. Barrett, and A. K. Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.*, 103:238701, Dec 2009.

[155] A. Montalto, S. Stramaglia, L. Faes, G. Tessitore, R. Prevete, and D. Marinazzo. Neural networks with non-uniform embedding and explicit validation phase to assess granger causality. *Neural Networks*, 71:159 – 171, 08 2015.

[156] D. Kugiumtzis. Direct-coupling information measure from nonuniform embedding. *Phys. Rev. E*, 87:062918, Jun 2013.

[157] E.Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, Inc., 2011.

[158] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[159] F. Diebold and K. Yilmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, 2014.

[160] T. Kuzubaş, I. Ömercikoğlu, and B. Saltoglu. Network centrality measures and systemic risk: An application to the turkish financial crisis. *Physica A: Statistical Mechanics and its Applications*, 405:203–215, 07 2014.

[161] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 06 1998.

[162] R. Couillet, F. Pascal, and J. W. Silverstein. The random matrix regime of maronna's M-estimator with elliptically distributed samples. *Journal of Multivariate Analysis*, 2015.

[163] E. Terreaux, J. P. Ovarlez, and F. Pascal. Robust model order selection in large dimensional Elliptically Symmetric noise. *arXiv preprint, https://arxiv.org/abs/1710.06735*, 2017.

[164] E. Terreaux, J. P. Ovarlez, and F. Pascal. New model order selection in large dimension regime for Complex Elliptically Symmetric noise. In *25th European Signal Processing Conference (EUSIPCO)*, pages 1090–1094, Aug 2017.

[165] E. Terreaux, J. P. Ovarlez, and F. Pascal. A Toeplitz-Tyler estimation of the model order in large dimensional regime. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr 2018.

[166] E. Eberlein and U. Keller. Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299, 1995.

[167] N. Bingham and R. Kiesel. Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance*, 2:241 – 250, 2002.

[168] E. Jay, E. Terreaux, J. P. Ovarlez, and F. Pascal. Improving portfolios global performance with robust covariance matrix estimation: Application to the maximum variety portfolio. In *26th European Signal Processing Conference (EUSIPCO)*, Sept 2018.

[169] Robert M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory*, 2(3):155–239, 2006.

[170] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27, 1974.

[171] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007.

[172] E. Pereda, R. Q. Quiroga, and J. Bhattacharya. Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, 77(1-2):1–37, 2005.

[173] Y. Saito and H. Harashima. Tracking of information within multichannel EEG record - causal analysis in EEG. In N. Yamaguchi and K. Fujisawa, editors, *Recent Advances in EEG and EMG Data Processing*, pages 133–146. Elsevier, New York, 1981.

[174] M. J. Kamiński and K. J. Blinowska. A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, 65(3):203–210, 1991.

[175] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 85(2):145–157, 2001.

[176] K. Sameshima and L. A. Baccalá. Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods*, 94(1):93–103, 1999.

[177] L. A. Baccalá and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–474, 2001.

[178] L. A. Baccalá, K. Sameshima, and D. Y. Takahashi. Generalized partial directed coherence. In *2007 15th International Conference on Digital Signal Processing*, pages 163–166, 2007.

[179] K. J. Blinowska. Review of the methods of determination of directed connectivity from multichannel data. *Medical & Biological Engineering & Computing*, 49(5):521–529, 2011.

[180] L. Faes, S. Erla, and G. Nollo. Measuring connectivity in linear multivariate processes: Definitions, interpretation, and practical analysis. *Computational and Mathematical Methods in Medicine*, 2012, 2012.

[181] M. R. Gevers and B. D. O. Anderson. Representations of jointly stationary stochastic feedback processes. *International Journal of Control*, 33(5):777–809, 1981.

[182] L. A. Baccalá, K. Sameshima, A. Valle G. Ballester, and C. Timo-Iaria. Studying the interaction between brain structures via directed coherence and granger causality. *Applied Signal Processing*, 5, 01 1998.

[183] B. Schelter, J. Timmer, and M. Eichler. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of neuroscience methods*, 179(1):121–130, April 2009.

[184] E. Siggiridou, V. K. Kimiskidis, and D. Kugiumtzis. Dimension reduction of frequency-based direct granger causality measures on short time series. *Journal of Neuroscience Methods*, 289:64–74, 2017.

[185] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038, 1963.

[186] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288, 1996.

[187] H. Zou and t. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.

[188] I. Vlachos and D. Kugiumtzis. Backward-in-Time Selection of the Order of Dynamic Regression Prediction Model. *Journal of Forecasting*, 32(8):685–701, 2013.

[189] N.-J. Hsu, H.-L. Hung, and Y.-M. Chang. Subset selection for vector autoregressive processes using lasso. *Computational Statistics and Data Analysis*, 52(7):3645–3657, 2008.

[190] L. Leistritz, B. Pester, A. Doering, K. Schiecke, F. Babiloni, L. Astolfi, and H. Witte. Time-variant partial directed coherence for analysing connectivity: a methodological study. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1997):20110616, 2013.

[191] D. Huang, A. Ren, J. Shang, Q. Lei, Y. Zhang, Z. Yin, J. Li, K. V. Von Deneen, and L. Huang. Combining partial directed coherence and graph theory to analyse effective brain networks of different mental tasks. *Frontiers in Human Neuroscience*, 10, 2016.

[192] J. A. Gaxiola-Tirado, R. Salazar-Varas, and D. Gutiérrez. Using the partial directed coherence to assess functional connectivity in electroencephalography data for brain-computer interfaces. *IEEE Transactions on Cognitive and Developmental Systems*, 10(3):776–783, 2018.

[193] MSCI. GICS. https://www.msci.com/gics.

[194] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 1967.

[195] L. Yang, R. Couillet, and M. R. McKay. Minimum variance portfolio optimization in the spiked covariance model. In *Proc. IEEE 7th Computational Advances in Multi-Sensor Adaptive Processing Workshop (CAMSAP))*, December 2017.

[196] P. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.

[197] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

[198] D. Melas, R. Suryanarayanan, and S. Cavaglia. Efficient replication of factor returns, July 2009. MSCI Barra Research Paper No. 2009-23.

[199] E. Jay, P. Duvaut, S. Darolles, and A. Chrétien. Multi-factor models: examining the potential of signal processing techniques. *IEEE Signal Processing Magazine*, 28(5), September 2011.

[200] S. Darolles, P. Duvaut, and E. Jay. *Multi-factor models and signal processing techniques: Application to quantitative finance*. John Wiley & Sons, 2013.

[201] S. Darolles, C. Gouriéroux, and E. Jay. Robust portfolio allocation with risk contribution restrictions. In *Forum GI - Paris*, March 2013.

[202] S. Kritchman and B. Nadler. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10):3930–3941, Oct 2009.

[203] R. Couillet, M. S. Greco, J. P. Ovarlez, and F. Pascal. RMT for whitening space correlation and applications to radar detection. In *IEEE CAMSAP*, pages 149–152, Dec 2015.

[204] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet. A subspace estimator for fixed rank perturbations of large random matrices. *Journal of Multivariate Analysis*, 114:427–447, 2013.

[205] J. H. Jr. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244, 1963.

[206] P. Valdés-Sosa, J. M. Sanchez, A. Castellanos, M. Vega-Hernández, J. Bosch, L. Melie-Garcia, and E. Canales-Rodríguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360:969–981, 2005.

[207] E. J. Bedrick and C.-L. Tsai. Model selection for multivariate regression in small samples. *Biometrics*, 50(1):226–231, 1994.

[208] L. A. Baccalá, C. S. N. de Brito, D. Y. Takahashi, and K. Sameshima. Unified asymptotic theory for all partial directed coherence forms. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371:20120158, 2013.

[209] K. Sameshima and Baccalá L. A. *Methods in brain connectivity inference through multivariate time series analysis*. CRC Press, 2014.

[210] F. Han, H. Lu, and H. Liu. A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16:3115–3150, 2015.

[211] H. Qiu, S. Xu, F. Han, H. Liu, and B. Caffo. Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. *Proceedings of the International Conference on Machine Learning. International Conference on Machine Learning*, 37:1843–1851, 2015.

[212] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812, May 2009.

[213] S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.

[214] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, D. Klan, R. Bauer, J. Timmer, and H. Witte. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Process.*, 85(11):2137–2160, November 2005.

[215] C. Zou, C. Ladroue, S. Guo, and J.Feng. Identifying interactions in the time and frequency domains in local and global networks - a granger causality approach. *BMC Bioinformatics*, 11:337, 06 2010.

[216] Y. Li, X.-F. Jiang, Y. Tian, S.-P Li, and B. Zheng. Portfolio optimization based on network topology. *Physica A: Statistical Mechanics and its Applications*, 515:671 – 681, 2019.

[217] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535 – 559, 2012.

[218] F. Ren, Y.-N. Lu, S.-P. Li, X.-F Jiang, L.-X Zhong, and T.Qiu. Dynamic portfolio strategy using clustering approach. *PLOS ONE*, 12(1):1–23, 01 2017.

[219] X. Guo, Z. Xue, H. Zhang, and T. Tian. Development of stock correlation networks using mutual information and financial big data. *PLOS ONE*, 13(4):1–16, 04 2018.

[220] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[221] C. Chorro, E. Jay, P. D. Peretti, and T. Soler. Frequency causality measures and Vector AutoRegressive (VAR) models: An improved subset selection method suited to parsimonious systems. *Working paper*, 2019.

[222] E. Jay, T. Soler, E. Terreaux, J. P. Ovarlez, F. Pascal, P. De Peretti, and C. Chorro. Improving portfolios global performance using a cleaned and robust covariance matrix estimate. *Soft Computing*, March 2020.

[223] E. Jay, T. Soler, J. P. Ovarlez, P. D. Peretti, and C. Chorro. Robust covariance matrix estimation and portfolio allocation: The case of non-homogeneous assets. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8449–8453, May 2020.

[224] E. Ollila, D. P. Palomar, and F. Pascal. Shrinking the eigenvalues of m-estimators of covariance matrix. *IEEE Transactions on Signal Processing*, 69:256–269, 2021.

[225] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 12 2007.

[226] D. E. Tyler and M. Yi. Lassoing eigenvalues. *Biometrika*, 107:397–414, 06 2020.

[227] E. Raninen, D. E. Tyler, and E. Ollila. Linear pooling of sample covariance matrices, 2020.

[228] P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993.

[229] L. Faes and G. Nollo. Extended causal modeling to assess partial directed coherence in multiple time series with significant instantaneous interactions. *Biological cybernetics*, 103:387–400, 10 2010.

List of Figures

List of Tables

Résumé

Cette thèse aborde les problèmes d'allocation de portefeuille en étudiant les estimateurs robustes de la matrice de covariance et la dépendance dynamique entre les actifs financiers afin d'améliorer la performance globale des stratégies d'allocation basées sur le risque. De nos jours, il est bien établi que la solution proposée par Markowitz conduit à de mauvaises performances, en raison des erreurs d'estimation sur les paramètres, notamment pour le rendement espéré. Malgré plusieurs extensions de la stratégie moyenne-variance au cours des dernières décennies, la plupart des praticiens préfèrent des modèles plus simples et plus robustes tels que le portefeuille à variance minimale (MinVar), le portefeuille de contribution à risque égal (ERC) et le portefeuille de diversification maximale (MDP), où la matrice de covariance est le seul paramètre. Cependant, deux problèmes principaux demeurent : premièrement, l'estimateur empirique "classique" (*Sample Covariance Matrix* ou SCM) n'est pas optimal dans un cadre non gaussien et sur des échantillons de petite taille ; deuxièmement, la matrice de covariance ne capture pas la structure de dépendance entre actifs (effets de contagion et de rétroaction), conduisant à des évaluations incomplètes des risques de l'univers d'investissement.
La première partie de cette thèse porte sur l'estimation de la matrice de covariance. Nous développons un estimateur robuste et non bruité adapté à des hypothèses plus réalistes sur les rendements des actifs, basé sur le M-estimateur de Tyler et la théorie des matrices aléatoires (RMT). Cet estimateur est adapté aux distributions non gaussiennes (distributions elliptiques) et montrons que les actifs doivent être de préférence classés en groupes homogènes avant d'appliquer la méthodologie proposée.
La deuxième partie est consacrée à l'évaluation de la dépendance dynamique entre les actifs en utilisant la mesure généralisée de cohérence partielle dirigée (GPDC) pour prendre en compte à la fois la direction et la force des relations causales entre les actifs. Néanmoins, une estimation naïve du modèle vecteur autorégressif (VAR) conduit à de mauvais résultats pour la GPDC. Pour capturer avec précision les schémas de diffusion, nous proposons une estimation parcimonieuse (mBTS-TD) du VAR (suppression des coefficients non significatifs) en combinant deux méthodes de sous-sélection, la méthode modified Backward-in-Time Selection (mBTS) et la stratégie Top-Down (TD).
Enfin, dans la dernière partie, nous dérivons du coefficient de clustering un indicateur adapté au nombre de connexions dans le réseau afin d'éliminer les actifs les plus instables (systémiques et influencés) avant d'allouer les portefeuilles. De plus, une étude empirique est réalisée qui montre qu'en combinant les différents résultats des chapitres, nous arrivons à améliorer significativement les performances des stratégies d'allocation.

Mots-clés: Allocation de portefeuille, Séries temporelles multivariées, Matrice de covariance, Matrices aléatoires, Modèle à facteurs, Distributions elliptiques, Vecteur autorégressif, Méthodes de sous-sélection, Mesures de causalité, Mesures de causalité fréquentielle, Réseaux financiers, Coefficient de clustering

Chapter I

Introduction

I.1 Motivations

Le principal défi de la gestion d'actifs a toujours été de déterminer les allocations entre actifs financiers afin de maximiser la valeur d'un investissement tout en gérant efficacement les risques du portefeuille. En 1952, Markowitz [1, 2] a proposé une solution au problème de sélection d'actifs dans un cadre moyenne-variance, en supposant que les investisseurs ne se soucient que du rendement espéré (la moyenne) et du niveau de risque (la variance) de leur portefeuille. Les pondérations optimales entre les actifs sont donc obtenues en maximisant le rendement espéré pour un niveau de risque donné, ou en minimisant le risque pour un rendement espéré donné. Tous ces différents portefeuilles forment la frontière efficiente qui représente le meilleur rendement espéré pour un niveau de risque donné ou vice versa. Pour un investisseur, elle représente le compromis entre le risque et le rendement espéré lors de l'allocation du portefeuille. De plus, la frontière efficiente illustre aussi les avantages de la diversification, car un portefeuille bien diversifié peut réduire le risque tout en préservant le même niveau de rendement espéré, voire augmenter le rendement du portefeuille sans en augmenter le risque. Malgré son cadre théorique puissant, cette stratégie souffre d'importants problèmes en pratique. Tout d'abord, les paramètres ne sont pas connus *a priori* et doivent être estimés, entraînant des erreurs d'estimation, surtout pour le rendement espéré. Par exemple, Chopra et Ziemba dans [3] ont montré que les erreurs d'estimation sur la moyenne sont environ dix fois supérieures à celles sur la variance, et vingt fois supérieures pour les covariances. Deuxièmement, la solution moyenne-variance est très sensible aux paramètres, là encore, principalement pour les rendements espérés [4, 5]. Si ces limitations sont ignorées, le portefeuille moyenne-variance obtenu sera très concentré avec des pondérations extrêmes, une composition instable dans le temps et de mauvaises performances [6, 7, 8].

Pour remédier à ces limites, plusieurs extensions sont apparues dans la littérature au cours des cinquante dernières années. Ces extensions peuvent être divisées en deux classes. La première classe comprend l'approche bayésienne, qui permet d'estimer des paramètres inconnus en réduisant les erreurs d'estimation, comme la distribution prédictive des rendements [9, 10], l'approche Bayes-Stein basée sur des estimateurs de rétrécissement (shrinkage) [11, 12, 13, 14], ou des modèles d'évaluation d'actifs fournissant des distributions antérieures informatives pour les rendements futurs [15, 16]. La seconde classe comprend des approches plus hétérogènes, telles que des règles robustes d'allocation de portefeuille utilisant des paramètres bornés ou des intervalles de confiance

[17, 18, 19], des restrictions sur les moments basées sur des modèles à facteurs [20], l'estimation de la matrice de covariance [21, 22, 23], ou des contraintes d'optimisation spécifiques [24, 25, 26]. Bien que ces extensions réduisent la sensibilité du portefeuille à l'estimation des paramètres, elles augmentent cependant la complexité de calcul tout en ne garantissant pas de meilleurs performances [27, 28]. De plus, la plupart des praticiens préfèrent des modèles plus simples et plus robustes dans lesquels les rendements espérés sont mis de côté et où l'estimation de la matrice de covariance est le seul objectif.

Les stratégies d'allocation alternatives les plus connues sont le portefeuille équipondéré (EW) [28] et les stratégies d'allocation basées sur le risque telles que le portefeuille à variance minimale (MinVar) [3], le portefeuille de contribution à risque égal (ERC) [29] et le portefeuille de diversification maximale [1] (MDP) [30].

Le portefeuille EW est la manière la plus simple d'allouer des portefeuilles, puisque les actifs sont alloués avec les mêmes poids sans aucune estimation de paramètres ni d'optimisation complexe. Dans le cadre moyenne-variance, ce portefeuille est optimal seulement si les actifs ont les mêmes rendements espérés, variances et covariances. Cependant, si les niveaux de risque sont très hétérogènes, cette stratégie conduit à une mauvaise diversification des risques, puisque même si les actifs ont le même poids dans le portefeuille, la contribution au risque total du portefeuille est plus élevée pour les actifs risqués que pour les actifs faiblement risqués. Néanmoins, malgré ses défauts, le portefeuille EW est largement utilisé en pratique par les investisseurs, comme le montrent [31, 32, 33] et peut même surpasser en termes de performance diverses extensions de la stratégie moyenne-variance [28].

Comme suggéré dans [3], la façon la plus simple de mettre de côté le rendement espéré de la stratégie moyenne-variance est de supposer que tous les actifs ont le même rendement espéré. Sous cette hypothèse, le portefeuille optimal est le portefeuille MinVar. Cette stratégie minimise la variance du portefeuille final et sa solution est unique. Dans [34, 35, 36, 37], les auteurs ont montré que les portefeuilles MinVar améliorent les rendements avec des volatilités plus faibles par rapport aux stratégies pondérées par les capitalisations boursières, comme par exemple le S&P 500 ou le CAC 40. Depuis la crise financière de 2007-2008, les portefeuilles MinVar ont été largement utilisés par les investisseurs offrant des performances supérieures aux stratégies factorielles traditionnelles (dividende, croissance, momentum, value, etc.), et renforçant ainsi le concept d'anomalie de faible volatilité[2] sur les marchés [38, 39]. Cependant, dans la pratique, la stratégie MinVar conduit à des portefeuilles très concentrés, en particulier, quand les marchés sont très volatils et donc fortement corrélés, ce qui nécessitent l'intégration de contraintes individuelles afin de réduire le risque idiosyncratique.

La stratégie de parité des risques a été utilisée pour la première fois par la société de gestion d'actifs Bridgewater dans les années 1990. La stratégie initiale allouait les actifs en proportion de l'inverse de leur volatilité, sans tenir compte des covariances.

[1] Dans les chapitres suivants, le portefeuille le MDP sera appelé portefeuille "variété maximale" (VarMax).

[2] L'anomalie de faible volatilité est l'observation que les actions à faible volatilité peuvent offrir des rendements plus élevés que les actions à forte volatilité, ce qui remet en question les hypothèses sur le risque et le rendement.

Dans [40, 41], l'auteur a introduit le concept de budget de risque en utilisant à la fois les variances et covariances, étendant ainsi la stratégie initiale. Cette stratégie, désormais plus connue sous le nom de portefeuille ERC grâce à [29], se situe à mi-chemin entre les portefeuilles EW et MinVar, répartissant les actifs en fonction de leur contribution au risque du portefeuille. Elle préserve les avantages du portefeuille EW en investissant dans tous les actifs, mais améliore la diversification des risques lorsque ceux-ci sont hétérogènes dans l'univers d'investissement. De plus, il est maintenant bien établi que la diversification des risques peut améliorer le rendement des portefeuilles [42, 43, 44]. Cependant, il convient de noter que pour des univers en grande dimension, l'algorithme ERC est coûteux en temps de calcul et ne converge pas toujours [45]. Depuis ces travaux, différentes approches de la parité des risques et/ou extensions de l'ERC ont été proposées en tenant compte soit de l'exposition au marché, de la "value-at-risk", de "l'expected shortfall", ou encore du risque systématique, etc. [46, 47, 48, 49, 50, 51].

Choueifaty et Coignard dans [30] ont proposé une stratégie alternative, basée directement sur la diversification de portefeuille afin de réduire les expositions aux risques communs, offrant une alternative efficiente aux portefeuilles pondérés par les capitalisations boursières [30, 37, 52, 53]. Le portefeuille de diversification maximale (MDP) maximise le ratio de la moyenne pondérée des volatilités des actifs sur la volatilité totale du portefeuille. Pour des portefeuilles "long-only"[3], l'objectif de la maximisation du ratio de diversité ($\mathcal{DR}$) est d'acheter les risques les plus indépendants possibles au sein d'un univers pour réduire la volatilité du portefeuille. En effet, si l'on considère deux actifs indépendants avec les mêmes volatilités, le $\mathcal{DR}$ est égal à $\sqrt{2}$, réduisant ainsi la volatilité du portefeuille de $\sqrt{2}$ et de $\sqrt{m}$ pour m actifs indépendants. De plus, si le rendement moyen des actifs augmente proportionnellement avec la volatilité, alors le portefeuille MDP est le portefeuille tangent à la frontière efficiente [35], présentant le ratio de Sharpe le plus élevé [54]. Le portefeuille MDP posséde aussi plusieurs propriétés d'invariance intéressantes [52] : invariant par duplication, c'est-à-dire que si un actif est dupliqué dans l'univers, alors l'allocation MDP sera inchangée ; invariant au levier, c'est-à-dire que la pondération reste inchangée quelle que soit la politique de l'entreprise en matière de levier ; invariant par combinaison linéaire positive (invariance po-li-co), c'est-à-dire que le MDP reste inchangé si une combinaison linéaire positive d'actifs de l'univers est ajoutée en tant que nouvel actif. Contrairement au MDP, le portefeuille EW ne vérifie pas ces propriétés d'invariance, le portefeuille MinVar satisfait uniquement l'invariance par duplication, et l'ERC uniquement l'invariance au levier. Cependant, comme pour le portefeuille MinVar, la stratégie MDP souffre d'une forte concentration de portefeuille nécessitant des contraintes individuelles [37]. De plus, ces stratégies présentent un ordre naturel de volatilité de portefeuille, où le portefeuille MinVar est sans surprise le moins volatil, le portefeuille MDP est le deuxième moins volatil, suivi du portefeuille ERC, et enfin le portefeuille EW [29].

Malgré le fait que ces stratégies d'allocation se concentrent uniquement sur l'estimation de la matrice de covariance, cette étape est cruciale et ne doit pas être négligée afin d'obtenir des portefeuilles stables avec de meilleures performances [55, 56, 57, 58, 23, 59, 60, 61, 62, 63]. L'estimation d'un paramètre tel que la matrice de covariance à l'aide d'une méthode statistique donnée est une tâche complexe. La valeur du paramètre doit être aussi proche que possible de sa valeur théorique. En modélisation financière, l'estimateur

[3]Portefeuille "long-only" : toutes les quantités investies sont nécessairement supérieures ou égales à 0.

de la matrice de covariance le plus largement utilisé est la "Sample Covariance Matrix" (SCM). La SCM est l'estimateur optimal dans le cas où les échantillons multivariés sont gaussiens, coïncidant dans ce cas avec l'estimateur du maximum de vraisemblance (MLE). Soit $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_T) \in \mathbb{R}^{m \times T}$ la matrice des observations où $\forall t \in [1, T]$, $\mathbf{r}_t$ est un vecteur de m variables gaussiennes indépendantes de moyenne nulle et de matrice de covariance $\mathbf{\Sigma}$. Alors, lorsque $T \to \infty$ pour un m fixé, la loi des grands nombres garantit :

$$\left\| \frac{1}{T} \mathbf{R}\,\mathbf{R}' - \mathbf{\Sigma} \right\| \xrightarrow{a.s.} 0$$

Toutefois, cet estimateur biaisé présente deux inconvénients majeurs : premièrement, l'estimation devient inexacte lorsque le nombre d'observations T n'est pas trop grand par rapport aux variables m (échantillons de petite taille) ; deuxièmement, le manque de robustesse pour les distributions non gaussiennes (asymétrie et queues de distribution épaisses). Ces problèmes sont maintenant bien identifiés dans de nombreux domaines tels que le traitement du signal et la finance, confirmant les très mauvaises performances de l'estimateur SCM. De plus, il est aussi bien connu que les rendements des actifs ne sont pas gaussiens [64, 65, 66, 67, 68] présentant généralement des faits stylisés tels que de l'asymétrie, des queues de distribution épaisses et de la dépendance dans les queues de distribution, entraînant donc de grandes erreurs d'estimation. L'hypothèse faite sur la distribution sous-jacente joue un rôle fondamental dans la précision de l'estimation et doit donc conditionner le choix de l'estimateur.

Un autre aspect essentiel à prendre en compte dans les problèmes d'allocation de portefeuille est la dépendance dynamique entre les actifs financiers, y compris les effets de contagion et de rétroaction. En effet, selon leur importance dans le marché, leur fragilité économique, leur activité commerciale et/ou leur position géographique, les actifs ne réagiront pas de manière identique à un choc de marché. Ignorer le caractère systémique ou influencé d'un actif conduit nécessairement à des évaluations incomplètes des risques sur l'univers d'investissement. Il est donc essentiel d'évaluer les relations de cause à effet entre les actifs, puisque la matrice de covariance utilisée comme seul paramètre dans les stratégies d'allocation mentionnées ci-dessus, ne quantifie le risque qu'en fonction des volatilités et des similitudes de comportement. Au cours des deux dernières décennies et depuis les travaux fondateurs de Mantegna [69], l'utilisation de la théorie des réseaux pour représenter la structure de dépendance des marchés (réseau financier) a joué un rôle important dans la littérature sur les problèmes d'allocation de portefeuille [70, 71, 72, 73, 74, 75, 76]. Cette approche fournit des indications utiles pour la sélection d'actifs permettant de comprendre les interactions complexes. Cependant, les approches couramment utilisées pour récupérer la topologie du réseau, telles que la matrice de corrélation ou les tests de non-causalité à la Granger [77], conduisent à des informations partielles, ne fournissant jamais d'indications à la fois sur la direction et la force des relations causales. Afin de déterminer une topologie de réseau précise, il est nécessaire d'utiliser des mesures de dépendance qui évaluent à la fois la direction et la force des relations causales.

I.2 Limites des approches existantes

L'estimation des matrices de covariance est un problème classique en statistiques multivariées. Comme nous l'avons déjà mentionné, l'estimation des matrices de covariance est conditionnée à la fois par le nombre de variables m par rapport au nombre d'observations T, et par la distribution multivariée sous-jacente des variables. Lorsque les variables sont gaussiennes et que $T >> m$, l'estimateur SCM $\widehat{\mathbf{\Sigma}}_{scm}$ (1.1) converge presque sûrement vers la matrice de covariance théorique $\mathbf{\Sigma}$. Cependant, en modélisation financière, ces hypothèses ne sont pas remplies. En pratique, peu d'observations historiques sont utilisées (1 ou 2 ans de rendements quotidiens), afin de ne conserver que les informations les plus récentes dans le processus d'optimisation, et les rendements des actifs ne sont pas gaussiens (asymétrie, queues de distribution épaisses, et dépendance dans les queues de distribution). Par conséquent, plusieurs approches ont été proposées pour traiter ces problèmes, telles que l'estimateur de Ledoit & Wolf (LW) [58], l'estimation de la matrice de covariance basée sur les modèles à facteurs [22, 90], la méthode de découpage des valeurs propres ("Eigenvalue clipping") [55, 56] utilisant la distribution des valeurs propres de Marčenko-Pastur [92], et l'estimateur invariant par rotation (RIE) [98, 63, 99].

- estimateur de LW (1.3) : particulièrement adapté aux rendements des actifs car il utilise les informations du marché (MEDAF [83]) dans le processus de rétrécissement (shrinkage),
- modèles à facteurs (1.4) : capturer les facteurs de risque communs et les utiliser pour réduire les problèmes de dimension ($K < m$),
- découpage des valeurs propres ("Eigenvalue clipping") (1.6) : identifier la partie signale (ou facteur) de la matrice de covariance grâce à la borne supérieure de la loi de Marčenko-Pastur [92], et ainsi obtenir un estimateur débruité de la matrice de covariance des observations.
- estimateur invariant par rotation (RIE) (1.7) : utilisation du chevauchement entre les vecteurs propres théoriques et de l'échantillon pour corriger l'estimation.

Néanmoins, comme ces estimateurs opèrent sur la SCM, ils ne sont pas adaptés aux distributions non gaussiennes. Pour gérer les distributions non gaussiennes, les M-estimateurs [105, 106, 103, 104, 107] et notamment le M-estimateur de Tyler[108] offrent des alternatives intéressantes à la SCM, grâce à leurs propriétés de robustesse pour les distributions elliptiques. Cependant, ces estimateurs ne sont pas exempts des problèmes liés à la taille de l'échantillon et nécessitent généralement que $m << T$, conduisant également à combiner les méthodes présentées ci-dessus pour débruiter l'estimation [116, 117, 118, 115, 119, 62]. De plus, l'estimation d'une matrice de covariance suppose que les observations proviennent d'une seule distribution multivariée (gaussienne ou non), ce qui semble assez restrictif pour la modélisation des rendements.

Afin de modéliser la structure et la dynamique des marchés financiers par le biais des réseaux financiers, il est nécessaire de retrouver la topologie du réseau le plus précisément possible. Pour obtenir des informations exhaustives sur les relations entre actifs et pour capturer avec précision les schémas de diffusion, les mesures utilisées doivent prendre en compte l'existence, la direction et la force des interactions, sinon il est légitime de supposer qu'il manque des informations sur la structure de dépendance. Portant, les

mesures les plus fréquemment utilisées pour récupérer la topologie du réseau ne sont pas pleinement satisfaisantes. Les réseaux basés sur la corrélation nécessitent des outils de réduction de dimension [69, 128], sinon ils ne sont pas adaptés en pratique et la structure hiérarchique ne peut être retrouvée. Ils conduisent également à des réseaux non dirigés (pondérés/non pondérés) et ne permettent pas de capturer les diffusions d'informations entre actifs (effets de contagion et de rétroaction). En effet, si par exemple un actif fait défaut, le schéma de contagion ne sera pas identifiable. À cette fin, les tests de non-causalité à la Granger peuvent aider à capturer les schémas de diffusion entre les actifs (réseaux dirigés). Ils sont cependant basés sur l'estimation d'un modèle VAR qui peut échouer pour de nombreuses raisons bien connues (ordre du modèle p incorrect [141, 146, 147, 148], systèmes en grande dimension [149, 150, 151, 140] et des bruits blancs corrélés [132]). Donc, si une attention toute particulière n'est portée à l'estimation du VAR, les relations temporelles ne seront pas capturées avec précision. Qui plus est, bien qu'ils saisissent les schémas de diffusion, ils ne quantifient pas la force des relations, ce qui est problématique pour établir l'intensité des interactions et donc le risque systématique des actifs. À cet effet, des mesures de causalité non linéaires telles que le transfert d'entropie (TE) peuvent être une alternative possible, mais il est difficile d'obtenir des estimations précises avec des systèmes en grande dimension et des échantillons de petite taille [135, 140], nécessitant là aussi des méthodes de réduction de dimension [156, 155, 73].

Enfin, une fois le réseau financier construit, les mesures de centralité ou le coefficient de clustering sont généralement utilisés pour sélectionner un nombre fixe d'actifs (les plus périphériques ou les moins systémiques) indépendamment du niveau de connectivité dans le réseau [124, 71, 72]. L'indicateur utilisé doit donc s'adapter au nombre de connexions dans le réseau et ne retirer que les actifs les plus imbriqués à chaque période, car selon les cycles de marché, les actifs sont plus ou moins connectés les uns aux autres, et il n'est donc pas pertinent de toujours éliminer le même nombre d'actifs [70, 125, 73, 74]. Si le réseau est très déconnecté, peu ou pas d'actifs devront être éliminés, à l'inverse, si le réseau est très connecté, un nombre beaucoup plus important devra être éliminé.

I.3 Objectifs

Dans cette thèse, nous abordons les questions d'allocation de portefeuille à travers des approches distinctes mais complémentaires visant à améliorer la performance globale des stratégies d'allocation basées sur le risque.

Le premier objectif est de développer un estimateur robuste et débruité de la matrice de covariance adapté à des hypothèses plus réalistes sur le rendements des actifs. Cet estimateur doit évidemment être adapté aux distributions non gaussiennes qui jouent un rôle fondamental dans la précision de l'estimation, mais doit également tenir compte du fait que les rendements des actifs ne sont pas homogènes en distribution.

Le second objectif est d'évaluer la dépendance dynamique entre les actifs financiers afin de retrouver la topologie du réseau et identifier les actifs les plus imbriqués/systémiques. Les mesures de dépendance utilisées doivent être adaptées aux systèmes en grande dimension, mais doivent également quantifier la force causale pour capturer avec précision les schémas de diffusion.

Le troisième objectif est de développer un indicateur de réseau adapté au nombre de connexions du réseau et de ne retirer que les actifs les plus instables (systémiques et influencés) à chaque période, afin de réduire le risque systémique au sein de l'univers d'investissement avant d'allouer les portefeuilles.

Enfin, le dernier objectif de cette thèse est de combiner ces différentes approches pour proposer une méthodologie complète d'allocation de portefeuille, où le risque systémique a été réduit par rapport à l'univers d'investissement initial et où la matrice de covariance des actifs restants est estimée avec un estimateur robuste et débruité.

I.4 Contributions principales

I.4.1 Chapitre 2 : Estimation robuste de la matrice de covariance

Afin de refléter le comportement très spécifique des rendements, nous les modélisons à l'aide d'un modèle multi-facteurs avec un bruit additif multivarié corrélé et non-Gaussien, en supposant que le nombre de facteurs inconnus K détermine la partie non diversifiable (risques communs) des rendements et que le bruit additif appartient aux distributions elliptiques et est corrélé selon une structure de Toeplitz. Dans ce cadre, nous appliquons le M-estimateur de Tyler (1.10) et les résultats de la RMT (1.6), de manière à filtrer la partie bruit des observations et donc d'estimer la matrice de covariance engendrée par le sous-espace des K vecteurs propres liés aux K plus fortes valeurs propres (facteurs).

Le M-estimateur de Tyler appliqué au modèle (1.29) s'avère être l'estimateur "le plus robuste" [108, 114] pour la "vraie" matrice de covariance $\mathbf{C}$, et également indépendant de la distribution τ_t. De plus, le *Théorème de consistance* trouvé dans [163, 164, 165] montre que si $\mathbf{C}$ admet une structure de Toeplitz, alors l'application d'un opérateur de rectification de Toeplitz $\mathcal{T}(.)$ sur $\widehat{\mathbf{C}}_{tyl}$ fournit un estimateur de $\mathbf{C}$ qui converge presque sûrement en norme spectrale sous le régime RMT, c'est-à-dire lorsque $T, m \to \infty$ et que $m/T \to c \in]0, \infty[$. Ce théorème assure la convergence de la matrice de covariance des observations vers la matrice de covariance du bruit additif, indépendamment de la variance du bruit. Une fois que la matrice de covariance $\widetilde{\mathbf{C}}_{tyl} = \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}\right)$ est estimée, les observations peuvent être débruitées et nous pouvons ainsi appliquer les résultats de la RMT dans les mêmes conditions que celles stipulées par la théorie, c'est-à-dire une distribution gaussienne multivariée non corrélée avec une variance unitaire. Ensuite, il est possible d'identifier les K plus grandes valeurs propres de la matrice de covariance déterminées grâce à la borne supérieure de la loi de Marčenko-Pastur [92], et ainsi obtenir un estimateur robuste et débruité de la matrice de covariance des observations (processus de débruitage).

Toutefois, le processus de débruitage proposé ci-dessus est réalisé en supposant implicitement que les rendements des actifs sont tirés d'une loi multivariée unique et sont donc homogènes en distribution. En effet, la structure de Toeplitz caractérise un processus stationnaire, qui, dans notre cas, signifierait que les rendements des actifs seraient tous issus d'un même processus et qui plus est spatialement stationnaire[4]. La

[4]Par spatialement stationnaire, nous entendons la stationnarité de la structure de dépendance des

stationnarité temporelle des rendements est généralement observable, mais pas la stationnarité spatiale entre les actifs. Dans le cas où les actifs ont une distribution hétérogène, l'hypothèse de structure de Toeplitz sur la matrice de covariance $\mathbf{C}$ est difficile à vérifier. Sous cette hypothèse, si l'ordre d'observation des actifs change, les covariances entre deux actifs ne devraient pas changer, alors qu'en pratique cette hypothèse n'est pas vérifiée. Compte tenu de ce phénomène, nous supposons maintenant que les rendements des actifs peuvent être distribués de manière non homogène, ce qui étend les résultats présentés dans [168] pour ne plus être dépendant de l'ordre d'observation. Nous proposons donc de diviser les m actifs en $p < m$ groupes, chacun composé d'actifs $\{m_q\}_{q=1}^{p}$ (avec $\sum_{q=1}^{p} m_q = m$), et formé pour être composé d'actifs ayant des distributions similaires. Sous cette hypothèse de rendements non homogènes, nous proposons de former des groupes d'actifs avant d'appliquer le processus de débruitage. Les groupes sont construits à l'aide de la méthode de classification hiérarchique ascendante (AHC) qui exige que le nombre de groupes soit fixé *a priori* ou déterminé à l'aide d'un critère prédéfini (critère de Caliński-Harabasz (CH) [170]), et de la méthode "Affinity Propagation" (AP) [171] qui détermine elle-même le nombre de groupes.

Des tests empiriques sont effectués sur deux univers d'actifs (univers d'actions européennes et américaines), tous deux alloués soit en Variété Maximale[5] (VarMax) ou en Minimum Variance (MinVar). Les résultats sont comparés à ceux obtenus avec plusieurs estimateurs classiques tels que Ledoit & Wolf (1.3), méthode "Eigenvalue clipping" (1.6), et RIE (1.7).

I.4.2 Chapitre 3 : Mesures de causalité dans le domaine fréquenciel et estimation parcimonieuse du VAR

La représentation VAR (1.11) permet, soit dans le domaine temporel, soit dans le domaine fréquentiel, de définir les interactions entre les séries temporelles. Son utilisation dans de nombreux domaines vient de son cadre théorique simple qui permet de comprendre la structure dynamique des systèmes en capturant les relations complexes des séries temporelles. Cependant, dans le domaine temporel, les tests de non-causalité à la Granger ne fournissent que des informations partielles sur ces interactions, puisqu'ils n'évaluent que l'existence et la direction des relations causales sans information sur la force causale. En neurosciences, plusieurs mesures de connectivité dans le domaine des fréquences ont été développées pour résoudre ce problème. Il convient de distinguer deux types de mesures de connectivité : premièrement, les mesures de couplage telles que les mesures de cohérence et de cohérence partielle [172], respectivement liées à la corrélation croisée ou à la corrélation croisée partielle ; deuxièmement, les mesures de causalité fréquentielle capables de quantifier la force des relations causales, étendant ainsi le concept de causalité à la Granger, comme la mesure de cohérence dirigée (DC) [173], la mesure de fonction de transfert direct (DTF) [174, 175], la mesure de cohérence partielle dirigée (PDC) [176, 177], et la mesure généralisée de la cohérence partielle dirigée (GPDC) [178]. Toutefois, ces mesures ne fournissent pas les mêmes informations [172, 179, 180]. La PDC et la GPDC sont des mesures de causalité directe, tandis que la DC et la DTF mesurent à la fois les causalités directes et indirectes. La différence entre la causalité directe et indirecte est illustrée sur la Fig. 1.4.

actifs.

[5]Variété Maximale (VarMax) faisant référence au portefeuille de diversification maximale [30].

Ici, nous nous focalisons seulement sur la mesure GPDC, car contrairement à la DC/DTF, elle représente la force relative d'une interaction par rapport à une source de signal donnée et est donc plus adapté pour capturer les schémas de diffusion entre les séries temporelles. Qui plus est, la GPDC résout les défauts principaux de la PDC [183] : i) elle n'est pas affectée lorsque plusieurs signaux sont émis par une source donnée ; ii) elle est invariante à la variance du bruit blanc ; iii) elle permet d'interpréter la force causale absolue. Sur la Fig. 1.6, nous donnons un exemple de la mesure GPDC pour la structure causale définie à la Fig. 1.4.

Cependant, bien que la GPDC soit une mesure puissante pour détecter et quantifier les relations causales dans les systèmes multivariés par rapport aux tests de non-causalité à la Granger, elle nécessite une estimation précise du modèle VAR. En effet, comme les coefficients du VAR sont directement utilisés pour calculer la GPDC, le problème est intrinsèquement lié à l'estimation du VAR et il est évident qu'une estimation incorrecte entraînera à la fois des causalités fallacieuses et des forces causales imprécises (erreurs en cascade). Un modèle VAR classique (modèle VAR non restreint) suppose qu'une série temporelle dépend de toutes les variables retardées du système. Cette hypothèse est très forte, voire déraisonnable dans le cas de modélisation de systèmes multivariés qui admettent des structures plus parcimonieuses avec seulement quelques coefficients non nuls, car dans un système multivarié, il est inhabituel que toutes les séries temporelles soient mutuellement dépendantes à chaque retard. L'estimation classique du VAR qui se limite à déterminer l'ordre du modèle "optimal" p, conduit à l'estimation de coefficients non significatifs et à de forts biais sur la mesure GPDC. Il est donc essentiel de procéder à une estimation précise des coefficients VAR, en particulier pour ceux qui sont non significatifs, sinon, des causalités fallacieuses apparaîtront, biaisant les "vraies" causalités, en raison à la fois de l'effet de compensation dans l'estimation du VAR, mais aussi de la propriété de normalisation de la GPDC (1.34). L'estimation classique du VAR n'est donc pas adaptée aux modèles parcimonieux, induisant des erreurs en cascade dans la GPDC, qui peuvent être décuplées pour des systèmes en grande dimension et/ou des échantillons de petite taille [184]. À cette fin, des estimations restrictives des VAR (méthodes de sous-sélection) peuvent aider à estimer uniquement les coefficients significatifs afin de réduire les erreurs en cascade tant sur la partie causale que non causale.

Pour déterminer la meilleure estimation possible du VAR, il faut idéalement tester tous les sous-ensembles possibles du VAR et sélectionner le modèle optimal pour un critère donné. En pratique, cette procédure est difficilement réalisable, car même pour de petits m et p, le nombre de possibilités est énorme (2^{m^2p} possibilités). Pour faire face à ce "fléau de la dimension", des procédures alternatives ont été développées. Dans la littérature relative aux modèles VAR, trois procédures peuvent être envisagées pour supprimer les coefficients non significatifs : premièrement, des procédures basées sur un critère d'information pour ajouter ou supprimer des coefficients (stratégie Bottom-Up (BU), stratégie Top-Down (TD) [141] et la méthode modified Backward-in-Time Selection (mBTS) [151]) ; ensuite, des procédures utilisant des tests d'hypothèses, comme le t-test (TT), le rapport de vraisemblance et le test de Wald [141, 149, 150] ; et enfin, des procédures basées sur des méthodes de rétrécissement (shrinkage) comme la régression Ridge [185], Lasso [186], et Elastic-Net [187].

Le problème ici est évidemment d'utiliser la méthode de sous-ensemble la plus efficace en termes de précision de l'estimation des coefficients VAR, mais aussi de pouvoir travailler sur des systèmes en grande dimension, afin d'évaluer la structure et la dynamique des marchés financiers. Selon cette idée, nous proposons de combiner la méthode mBTS avec la stratégie TD (mBTS-TD). Nous utilisons d'abord la méthode mBTS pour estimer les coefficients VAR, car elle inclut un à un seulement les termes qui améliorent la prédiction de l'équation à partir du modèle VAR "nul", et permet ainsi de travailler avec des systèmes en grande dimension comme $K = 20$ dans [151]. De plus, comme cela a déjà été montré dans [184], la méthode mBTS améliore considérablement la précision de la GPDC. Toutefois, un retard maximal p_{max} doit être fixé *a priori*, conduisant s'il est trop faible à un modèle sous-dimensionné où la dynamique interne du système n'est pas complètement capturée, et inversement si p_{max} est trop grand, des variables retardées indésirables apparaîtront probablement dans le modèle, révélant des causalités fallacieuses. Ainsi, pour être moins dépendant du choix de p_{max}, nous utilisons également la stratégie TD. Bien qu'elle soit très sensible à l'estimation initiale du VAR, dans notre cas, elle est déjà appliquée à un modèle parcimonieux, nous permettant seulement de tester la significativité des variables dans la direction opposée afin de produire si nécessaire un modèle plus parcimonieux lorsque p_{max} est élevé.

En utilisant des simulations de Monte Carlo, nous comparons notre méthode de sélection de sous-ensemble étendue (mBTS-TD) aux méthodes de sous-ensemble classiques (TD, Lasso, TT et mBTS). Toutes les méthodes de sous-ensemble améliorent la précision par rapport à l'estimation classique du VAR et sont donc mieux adaptées. Néanmoins, la méthode mBTS-TD se distingue clairement des quatre autres méthodes en réduisant considérablement les erreurs en cascade, tant sur la partie causale que non causale. Qui plus est, nous montrons également que la méthode mBTS-TD offre une plus grande précision sur la GPDC que les méthodes mBTS et TT, quelle que soit la valeur de p_{max} choisie. Par ailleurs, un travail récent a aussi montré [140] que les mesures linéaires (non-causalité à la Granger et GPDC) pouvaient offrir de bonnes performances par rapport aux mesures non linéaires (TE [152] et PMIME [156]) sur des systèmes non linéaires, en particulier lorsque les VAR sont estimés avec des méthodes de sous-ensemble.

Enfin, nous utilisons la mesure GPDC, estimée avec la méthode mBTS-TD pour modéliser la structure de dépendance des marchés financiers. Cette approche nous permet non seulement d'obtenir une topologie de réseau précise tenant compte à la fois de la direction et de la force des relations entre actifs via la GPDC, mais résout également le problème de la dimension via mBTS-TD, qui produit une structure causale parcimonieuse. Pour finir, nous appliquons sur le réseau financier obtenu le coefficient de clustering [127] afin d'exclure les actifs les plus systémiques, améliorant ainsi les performances du portefeuille EW. À notre connaissance, nous sommes les premiers à appliquer la GPDC aux réseaux financiers, même si la PDC ou la GPDC ont déjà été appliquées dans le domaine des neurosciences [190, 191, 192].

I.4.3 Chapitre 4 : Réseaux financiers basés sur la GPDC et sélection d'actifs

Dans ce dernier chapitre, afin d'obtenir une méthodologie complète pour les problèmes de sélection d'actifs, nous proposons un indicateur dynamique pour identifier dans les réseaux les actifs présentant des risques majeurs en raison de leur influence (systémique), ou au contraire, ceux trop influencés pour se détacher rapidement d'un choc de marché. Ignorer ces actifs dans le processus de sélection conduit inévitablement à sous-estimer les risques du portefeuille, en particulier pour les stratégies internationales ou multi-actifs. Ainsi, le fait d'identifier et de retirer les actifs instables de l'univers avant d'allouer les portefeuilles améliore la diversification des risques dans le sens où les actifs restants sont moins interconnectés. Une telle présélection peut également réduire les erreurs d'allocation d'actifs puisqu'elle est complémentaire à l'utilisation de la matrice de covariance qui ne quantifie pas les schémas de diffusion. Par exemple, une stratégie diversifiée recherchera les actifs les moins corrélés, mais les corrélations ne reflètent pas les interconnexions et un actif faiblement corrélé peut s'avérer potentiellement très systémique comportant une nouvelle prime de risque (la crise des subprimes aux états-Unis, la crise des dettes souveraines en Italie et en Espagne, etc.). Même si les mesures de centralité ou le coefficient de clustering sont des outils utiles pour identifier ces actifs dans la théorie des réseaux [124, 70, 71, 125, 72, 73, 74], ils ne permettent pas d'éliminer un nombre variable d'actifs selon la structure du réseau. Dans la sélection de portefeuille, ces mesures sont utilisées pour sélectionner/supprimer un nombre fixe d'actifs (les plus périphériques ou les moins systémiques) et ensuite allouer le portefeuille quel que soit le niveau de connectivité dans le réseau [124, 71, 72]. Mais, en fonction des événements économiques/politiques ou des cycles de marché, les actifs sont plus ou moins connectés entre eux et, selon ces cycles, il n'est pas raisonnable de toujours éliminer le même nombre d'actifs [70, 125, 73, 74]. Si le réseau est faiblement connecté, peu ou pas d'actifs devront être éliminés, à l'inverse, si le réseau est très connecté, un nombre beaucoup plus important devra être éliminé.

À partir du coefficient de clustering pondéré et dirigé [127], nous proposons de développer un indicateur dynamique essentiellement basé sur les schémas de diffusion sortants (1.25) et entrants (1.24), en négligeant les schémas indirects (cycles ou middleman). Le schéma sortant identifie les effets causaux ($h_j^{out} \neq 0$) et peut donc être directement assimilé à des actifs systémiques, tandis que le schéma entrant identifie les actifs causés ($h_j^{in} \neq 0$) et donc influencés ou qui réagiront fortement si un choc de marché se produit. Cette procédure de sélection d'actifs présente plusieurs avantages. Premièrement, elle se concentre essentiellement sur les actifs instables (systémiques ou influencés) afin de réduire le risque systémique au sein de l'univers. Deuxièmement, par construction, elle s'adapte au nombre de connexions dans le réseau et ne retire donc que les actifs les plus imbriqués à chaque période. En effet, si le réseau est très déconnecté, aucun actif ne sera retiré. Par conséquent, il n'est pas nécessaire de fixer un nombre d'actifs à exclure ou de fixer un seuil d'exclusion sur le coefficient de clustering.

Le processus de sélection dynamique d'actifs proposé est appliqué aux réseaux financiers basés sur la GPDC (chapitre 3) et examiné sur deux univers d'actions. Le premier univers est composé d'indices nationaux appartenant au MSCI ACWI (Indice mondial) et le second est composé d'indices sectoriels GICS [193] au sein de quatre

zones géographiques (MSCI Marchés émergents, MSCI Europe, MSCI Japon, MSCI États-Unis). Ces deux univers nous permettent de nous pencher sur des caractéristiques différentes telles que le temps de propagation entre les zones (effets de rétroaction) et les effets macroéconomiques mondiaux/régionaux pour l'univers national et les questions d'activité économique au sein ou entre les zones géographiques pour l'univers sectoriel.

La méthodologie est appliquée sur les portefeuilles EW, ERC, MinVar et VarMax[6], en utilisant d'abord l'estimation SCM pour les stratégies d'allocation basées sur la matrice de covariance. En ce qui concerne les portefeuilles EW, ERC et VarMax, la présélection dynamique améliore sensiblement les performances des portefeuilles par rapport à un réseau basé sur la non-causalité à la Granger ou appliqué sur l'ensemble de l'univers. Comme attendu, pour les deux univers, l'amélioration la plus significative concerne la stratégie VarMax. Le processus de sélection d'actifs réussit à identifier les actifs les moins corrélés qui sont soit les moins performants (influencés), soit les plus risqués (systémiques) sans réduire significativement la diversification du portefeuille. Quant au portefeuille MinVar, les résultats sont plus contrastés. La méthodologie ne parvient pas à améliorer le ratio rendement/volatilité sur l'univers des pays, augmentant encore plus le risque idiosyncratique et compliquant donc la gestion des changements brutaux du marché comme en 2008, 2015 et 2018. Enfin, lorsque l'on associe la sélection dynamique d'actifs à la procédure de débruitage (chapitre 2), les résultats sont encore meilleurs que la simple utilisation de la SCM (5 fois sur 6), le seul échec étant pour le portefeuille MinVar appliqué à l'univers sectoriel. Cette étude empirique souligne qu'en combinant tous les résultats des différents chapitres, nous pouvons améliorer de manière significative plusieurs stratégies d'allocation de portefeuille.

[6]Le portefeuille VarMax faisant référence au portefeuille de diversification maximale [30].