Sapienza Università di Roma

Faculty of Information Engineering, Informatics, and Statistics

A Unified Framework for Distance Based Pairs Trading via Graph Laplacians

Master Degree in Data Science

Professors:
Juri Marcucci
Daniele Bianchi

Students:
Michelangelo Saveriano

Academic Year 2021/2022

Contents

Introduction

Pairs trading is a popular investment strategy that in the last 20 years has gained interest among researches for a variety of reasons: i) the simplicity and versatility of the underlying idea, ii) the market neutrality and its low correlation with other common risk factors (Fama and French 2014) and iii) the high, though declining, excess returns that yields. As presented by Gatev, Goetzmann, and Rouwenhorst 2006 the strategy can be split into two steps: i) first find two securities whose prices move together, and ii) then, when they diverge, buy the undervalued asset and sell the overvalued one.

The pairs trading concept can be easily extended noticing that we can use as security any financial instrument (stocks, options, futures, ETFs) as well as any portfolio of financial instruments. Depending on the number of assets traded different types of pairs trading can be defined. From now on we will use the classification provided by Krauss 2015: i) in the univariate framework one security is traded against another security; ii) in the quasi-multivariate framework, a security is traded against a basket of securities; iii) in the fully multivariate framework, groups of securities are traded against other groups of securities. In our work we focus mainly on quasi-multivariate pairs trading because in the literature it has shown higher and more robust excess returns compared to univariate approaches.

As presented above pairs trading strategies inherently require the finding of reliable, long-lasting and, possibly, economically meaningful relationships between assets by which one can extract the idiosyncratic, potentially mean-reverting, components. So far such relationships were recovered mainly using Person correlations (M. Perlin 2007, M. S. Perlin 2009, C. W. Chen et al. 2017), industry-sector information (Dunis et al. 2010), statistical tests (Caldeira and Moura 2013) or a combination of these (Vidyamurthy 2004). However recent advances in the graph learning and network estimation fields give us new possibilities in the form of a constrained sparse inverse covariance matrix (Miranda Cardoso, Ying, and Daniel Perez Palomar 2020, Kalofolias 2016).

We also found that, due to differences in the dataset chosen and the methods used, it is often not possible to compare different works in order to understand which approach is preferable.

Therefore, our contribution is threefold:

1. First, we present a new unified framework, which extends the original procedure proposed by Gatev, Goetzmann, and Rouwenhorst 2006, where relationships between stocks are represented via a graphical structure. We also provide an extensive explanation of how to adapt prior works to the new framework. We argue that the use of a graph not only allows new approaches, i.e. network estimation, graph signals processing; but also gives us a common formal foundation to understand what kinds of relationships are preferable for pairs trading.

2. Second, we introduce two novel pairs formation approaches based on network estimation and provide justification for the usage of those methods when extracting the idiosyncratic component.

3. Third, in order to show the differences between the various techniques proposed in the literature and assess the performance of our pairs formation methods, we back-tested a total of 780 approaches on a small dataset, containing some of the S&P500 stocks for the period from 1995 to 2022, and then we verified the performance of a handful of configurations on a larger, more varied dataset, the CRSP

database from 1962 to 2022. We also defined and tested a simpler statistical arbitrage technique as a benchmark to see the benefits of using a pairs trading strategy. Moreover, an in-depth analysis of the performance over time is carried out to verify the loss in the profitability of the trading strategies and to assess the effect of certain market conditions, like high volatility periods, on the strategies α.

Surprisingly, we found no statistically significant differences in the excess returns between the benchmark strategy and the pairs trading ones, they all achieve an average monthly excess return of 2.1% (28% annualized) on the CRSP dataset from 1962-06 to 2022-10. The same consideration can be done about the 5+1-factors α's generated, depending on the trading rule chosen. They all achieve similar monthly α's of about 1.4% ($\sim$ 18% annualized). Some pairs trading strategies, however, show considerably higher risk-adjusted metrics such as an annual Sharpe ratio of 2.42 (0.6 higher than the benchmark) and an annual Sortino ratio of 5.15 (1.84 higher than the benchmark). This suggests that the information coming from a comoving portfolio of correlated assets might be used to lower the risk of a trading strategy without decreasing its excess returns.

To accommodate the growing concerns about the decline in profitability we analyze the evolution of the excess returns over time proving that they continued to plummet until 2020. We also show that this is mostly true for large capitalization, high liquidity firms, while it has limited effect on low liquidity, small cap.

Chapter 1

Pairs Trading

Since the seminal paper by Gatev, Goetzmann, and Rouwenhorst 2006 several streams of research have been created. The ones focusing on the pairs formation are of particular interest to us. Among them we identify 2 main approaches:1) the **Distance approach**, 2) the **Cointegration approach**.

1.1 Distance approach

Introduced by Gatev, Goetzmann, and Rouwenhorst 2006 (hereafter GGR) pairs are selected based on a distance metric, originally the sum of squared distance (SSD) between normalized price time series has been used. Let P_{it}, P_{j_t} be the normalized cumulative return of stocks i, j at time t, we define the sum of squared distance as:

$$SSD_{ij} = \frac{1}{T}\sum_{t=1}^{T}(P_{it} - P_{jt})^2 \tag{1.1}$$

Assuming a universe of n stocks, $n(n-1)/2$ SSDs are evaluated and the k pairs (usually 5 or 20) with smaller distances are selected for trading. Then, a position is opened when prices diverge by more than two historical standard deviations, estimated during the pairs formation period. According to the authors, such simple trading rule yields average annualized excess returns of up to 11% in the U.S. equity markets between 1962-2002. Do and Faff 2010, 2012 applied the same methodology on a larger period, until 2009, noticing a decline in profitability, proving that it is mostly due to the increasing number of non-convergent pairs and showing that the original technique had become largely unprofitable when accounting for transaction costs.

In order to improve the original GGR's methodology H. Chen et al. 2017 proposed the usage of the Pearson correlation on returns, instead of the SSD, for the creation of the pair. Trades are based on the return divergence between stock i and its comover j:

$$D_{ij} = \beta(R_j - R_f) - (R_i - R_f) \tag{1.2}$$

where D_{ij} is the return divergence, R_i is the monthly return of stock i, R_j is the return of its comover, R_f is the risk-free rate and β is a regression coefficient. H. Chen et al. 2017 tested two different approaches for the creation of R_j: in the first case they used a univariate approach selecting the return of the most highly correlated stock, in the second they used a quasi-multivariate approach creating a portfolio of the 50 most highly correlated stocks weighted uniformly. In this setting stocks that have shown the highest (lowest) return divergence in the last month are bought (sold). By using correlations instead of raw distances and a large portfolio of comoving stocks H. Chen et al. 2017 are able to achieve average monthly raw returns of 1.70 percent, almost twice as high as those of GGR.

A similar quasi-multivariate approach has been tested by M. Perlin 2007 on the Brazilian financial market in which stock returns are approximated by a portfolio containing the 5 most highly correlated stocks. Although H. Chen et al. 2017 and M. Perlin 2007 use different trading rules they reach the same conclusion: quasi-multivariate pairs trading results in higher and more robust annual excess returns than simple univariate pairs trading. This can be explained by saying that a divergence in the returns of a stock and its comoving portfolio is more likely caused by an idiosyncratic movement of the stock and therefore ideally reversible.

The use of correlations is better than raw distances since it penalizes less highly volatile stocks however it is still far from optimal as pointed out by Krauss 2015. A rational pairs trader should be looking for spreads exhibiting strong and frequent divergences from the equilibrium state. However, the distance based pairs creation technique discourages the first, penalizing spread volatility, and doesn't guarantee the latter.

1.2 Cointegration approach

In the cointegration approach the convergence of some hypothetical pairs spread is assessed through cointegration tests, mainly Engle-Granger or the Johansen method.

The widely most cited work using the cointegration approach is Vidyamurthy 2004, who presents a theoretical framework for the creation of univariate pairs trading strategy: first we perform a preselection of potentially cointegrated pairs, then we test them for tradability, finally, we design trading rules using nonparametric methods. In order to find potentially cointegrated pairs Vidyamurthy 2004 first expresses each stock i returns as a combination of k factors:

$$r_{it} = \boldsymbol{\beta}_i^T \boldsymbol{f} t + \epsilon_{it} \tag{1.3}$$

where $\boldsymbol{\beta}_i$ denotes the $k \times 1$ factors coefficients, $\boldsymbol{f}_t$ contains the $k \times 1$ factor returns and ϵ_{it} is the idiosyncratic return component; then he defines a distance between stocks based on the absolute value of the Pearson correlation of the common factor returns. The pairs with the highest correlation have a higher probability of being cointegrated therefore they are tested for tradability. In standard cointegration tests (e.g. Engle-Granger test) the idiosyncratic component is tested through a unit root test. However, arbitrageurs are more interested in spreads exhibiting strong mean-reversion property thus Vidyamurthy 2004 proposes a test based on the zero-crossing frequency which can be thought as a good proxy for that (Krauss 2015).

Girma and Paulson 1999 propose an application of the univariate cointegration approach for the future market proving the cointegration of crude oil and its end product (i.e. unleaded gasoline and heating oil) and the stationarity of the spreads according to Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) test statistics. As stated by Krauss 2015 "This direction of research is promising, since there is a clear fundamental reason for the cointegration relationship between crude oil and its end products". Such relationship is lost in the stock market, where researchers either rely on industry groups to restrict pairs formation (Dunis et al. 2010) or test all possible pairs (Caldeira and Moura 2013). However both the methods have problems: the first is unable to capture inter-industries relationships, the second suffers a large amount of false positives. Caldeira and Moura 2013 evaluate the univariate cointegration approach on the 50 most liquid stocks of the Brazilian stock index IBovespa. They tested, through Engle-Granger and Johansen method at the five percent significance level, 1,225 pairs of which only 94, on average, were classified as cointegrated; of the 94 which passed the tests roughly 61 pairs were false positives therefore a large share of the cointegrated pairs were misclassified. In order to reduce the number of false positives one can: i) implement better pairs preselection techniques, ii) control for the familywise error rate as in Cummins and Bucca 2012, iii) use bayesian approaches like the ones presented by Peters et al. 2011 or Gatarek, Hoogerheide, and Dijk 2014.

1.3 Problems

As anticipated before the approaches presented are not trouble-free, and we identify the following main issues.

Heterogeneity of the approaches

As we have seen a variety of strategies have been tested on a variety of different datasets however most of them require several steps and differ in more than one aspect from one another therefore it is impossible to understand which part would lead to higher profit (lower risk) and which is useless, if not counterproductive. Examples of this are H. Chen et al. 2017 and M. Perlin 2007 share the use of a portfolio made by the top k most highly correlated stocks as comoving assets however the reasons behind the choice of k were not provided while its optimal value cannot be said due to the different datasets used.

This leads us to the definition of a unified, modular framework able to fit all the different approaches proposed in the literature.

Market factors during pairs formation

Since the pairs trading main objective is to find couples of investment assets able to show cointegration property researchers often underlook and neglect the contribution of common risk factors, like Fama and French 2014, during the pairs formation step which we show are able to decrease considerably the risk while preserving the excess returns.

In order to be able to take into account common risk factors we expand the original two-step procedure, proposed by GGR, including in our framework a data preprocessing step at the beginning of the chain.

Chapter 2

Our Framework

Here we introduce our framework for quasi-multivariate pairs trading. In order to overcome the problems presented before it has to possess the following properties: i) generality, to be able to represent both approaches already present in the literature as well as new approaches that might be developed in the future, ii) modularity, a certain degree of separation between different phases is needed if we want to optimize different parts of the strategy independently of each other.

To meet the former requirements we propose a framework with the following three steps:

1. **data preprocessing**: in this phase we perform any filtering needed to meet tradability requirements together with the removal of possible common market factors;

2. **spreads evaluation**: here, in order to pursue a quasi-multivariate approach, we compute for each asset its comoving portfolio and the difference between returns and the returns of the estimated portfolio;

3. **trading**: finally we compute both the long and the short trading signals starting from the spreads.

2.1 Data Preprocessing

This phase takes in input stocks prices and produces as output the data that has to be used in the next phase to evaluate the spreads. As stated before, the purpose of this step is twofold: i) remove unwanted or problematic stocks that don't satisfy tradability criteria, ii) remove the effect of common components.

Let $\mathbf{p}_t$ be the $M \times 1$ vector containing all the stocks prices at time t, we can now filter it removing the unwanted entries hence obtaining the clean $N \times 1$ price vector $\mathbf{p}^*_t$. We can now use the clean price vector $\mathbf{p}^*_t$ to compute the stocks log-returns as

$$\mathbf{r}_t = log(\mathbf{p}^*_t) - log(\mathbf{p}^*_{t-1}) \tag{2.1}$$

Stock Filtering

Since in our application we performed the stock filtering a priori we skip it here and we will describe how it has been done later, however, we believe that such a step is crucial in a production environment where stricter tradability criteria have to be met.

Common Risk Factors

Arguably one of the objectives of pairs trading is to extract for each stock i its idiosyncratic component. The return divergences, captured by such components, are assumed to be motivated by idiosyncratic movement and hence possibly reversible.

Although a number of approaches have been tested to recover the idiosyncratic component, among the others: univariate (GGR), multivariate (H. Chen et al. 2017, M. Perlin 2007, M. S. Perlin 2009) and PCA/ETF based (Avellaneda and J. H. Lee 2010), many neglect to use common market risk-factors (Fama and French 1993, Fama and French 2014) when approximating stock returns. We decided to proceed differently: rather than evaluating the idiosyncratic component on the returns we use the Fama/French 5 factors (2x3) (Fama and French 2014) residuals instead. To do that we assume the returns to follow a linear factor model

$$\mathbf{r}_t = B\mathbf{f}_t + \epsilon_t \tag{2.2}$$

where $\mathbf{f}_t$ contains the F factors returns at time t and B is the $N \times F$ factor loading matrix. Through cross-validation we found that the coefficients matrix B that minimizes the elastic-net loss (Zou and Hastie 2005) with parameter $\alpha \approx 3 \cdot 10^{-3}$:

$$\underset{B}{\text{minimize}} \ \frac{1}{2t} \sum_{\tau=0}^{t} \|\mathbf{r}_t - B\mathbf{f}_t\|_2^2 + \frac{\alpha}{2}\|B\|_1 + \frac{\alpha}{4}\|B\|_2^2 \tag{2.3}$$

reaches lower out-of-sample MSE with respect to standard OLS.

Configurations Tested

In order to evaluate if different preprocessing steps lead to any difference in the results we tested the following two configurations:

- **Returns**: no preprocessing is performed therefore the output of the step, in this case, are the raw log-returns;

- **Residuals**: with this configuration the contribution of Fama/French 5 factors is removed as explained above hence the output are the residuals of the linear factor model.

For notation purposes from now on we will refer to the output of the data preprocessing step as $\mathbf{r}_t$ independently of the configuration chosen.

2.2 Spreads

The spreads evaluation step takes in input the $\mathbf{r}_t$ vector and produces as output the estimated idiosyncratic components, also known as spreads. In order to extract the peculiar components of each stock usually a partial dependence structure is estimated. Such structure can be captured by a weighted graph, generally represented as the triple $G = \langle \mathcal{V}, \mathcal{E}, A \rangle$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the node (stock) set, $\mathcal{E} \subseteq \{\{u, v\} : u, v \in \mathcal{V}\}$ is the edge set and A is the $N \times N$ weighted adjacency matrix. We say that $A_{ij} \neq 0$ if and only if $\{i, j\} \in \mathcal{E}$ hence there is a partial dependence between stock i and stock j.

The idiosyncratic component of stock i is in general represented as the difference between the stock returns (alternatively, prices) and the returns of its comoving portfolio. Here we assume the comoving portfolios to be identified by the weighted adjacency matrix A, thus the following mathematical formalization:

$$\mathbf{s}_t = \mathbf{r}_t^T - \mathbf{r}_t^T A = \mathbf{r}_t^T (I - A) \tag{2.4}$$

If we assume, without loss of generality, that the degree matrix, in-degree in the case of directed graph, is exactly equal to the identity $D = diag(\mathbf{1}^T A) = I$ we can rewrite the equation above as:

$$\mathbf{s}_t = \mathbf{r}_t^T (D - A) = \mathbf{r}_t^T L \tag{2.5}$$

where L is the graph Laplacian. The construction of the comoving portfolios, i.e. the pairs formation problem in the quasi-multivariate framework, can therefore be formulated as a graph reconstruction problem.

Since the reconstruction of the graph Laplacian is crucial in our framework we tested 2 different classes of approaches +1 benchmark.

Correlation Based

This kind of approach is the most common in the distance based literature (H. Chen et al. 2017, M. Perlin 2007, M. S. Perlin 2009) where stock returns are approximated through a weighted sum of the K most highly correlated stocks.

Let K be the size of the comoving portfolio, we define $\mathcal{V}_i$ the set of the K stocks most highly correlated with stock i, we can then define the Laplacian as:

$$L_{ij} = \begin{cases} 1 & i = j \\ -\frac{1}{K} & j \in \mathcal{V}_i \\ 0 & otherwise \end{cases} \tag{2.6}$$

We will refer to this class of approaches as **Corr-K**.

Please notice that this procedure does not necessarily produce a symmetric Laplacian (undirected graph), hence if a symmetric Laplacian (undirected graph) is demanded a different procedure can be used instead: let ρ the Pearson correlation matrix and $\hat{\rho}$ an arbitrary threshold, we can define the adjacency matrix as:

$$A_{ij} = \begin{cases} 0 & i = j \\ 1 & \rho_{ij} \geq \hat{\rho} \\ 0 & otherwise \end{cases} \tag{2.7}$$

The prices we have to pay in this case are: i) the node degrees are not equal to 1 anymore therefore the resulting Laplacian should be normalized in order to fit in our framework, ii) the absence of disconnected nodes is no longer guaranteed.

Sparse Inverse Covariance

Another family of approaches we have not found in the literature is based on the return inverse covariance matrix. The square root of such matrix is used in decorrelation transformation, i.e. whitening transformation, therefore it is ideal if we aim at extracting the idiosyncratic independent component for each stock. Moreover, there is a growing literature that show how to impose Laplacian constraints onto the inverse covariance in order to guarantee properties such as sparsity and positive semi-definiteness. Please notice that in the case of a positive semi-definite matrix its square root is both unique and easy to compute through the square root of its eigenvalues.

Here we're going to present what approaches and configurations we tested while in the next chapter we'll discuss the various approaches in mathematical details. We tested two different models:

- **LGRMF**: here we assume the data generating process to be a zero-mean Improper Gaussian Markov Random Field (IGMRF) and we model the rank-deficient precision matrix as a graph Laplacian, hence the name Laplacian constrained Gaussian Markov Random Field (LGMRF) (Ying et al. 2020);

- **SGS**: in this case we assume that the underlying signals in the graph are smooth, this is also known as Smooth Graph Signal (SGS) representation (Dong et al. 2016, Kalofolias 2016).

Despite being based on different assumptions both the methods require the minimization of the Dirichlet energy, i.e. given T observations of $\mathbf{r}_t$, $X = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_T], X \in \mathbb{R}^{T \times N}$ we define the Dirichlet energy as:

$$\text{D.E.} = \frac{1}{2} \sum_{i,j} A_{ij} \|\mathbf{r}_{*,i} - \mathbf{r}_{*,j}\|^2 = tr(XLX^T) = tr(XL^{\frac{1}{2}} L^{\frac{1}{2}} X^T) \tag{2.8}$$

We can define as $\Delta = XL^{\frac{1}{2}}$ the reconstruction error matrix, i.e. the idiosyncratic return matrix. Being the Laplacian symmetric it is trivial to show that minimize the Dirichlet energy is equivalent to minimize the reconstruction error:

$$\text{D.E.} = tr(XL^{\frac{1}{2}}L^{\frac{1}{2}}X^T) = tr(\Delta\Delta^T) = \|\Delta\|_F^2 \tag{2.9}$$

Due to this property of the Dirichlet energy and the regularization involved in the Laplacian estimation, which we will discuss in more details in the following chapter, we argue that the approaches based on a sparse reconstruction of the inverse covariance matrix presented above constitute a viable solution that has long been overlooked.

In this scenario equation 2.5 can be rewritten as $\mathbf{r}_t = \mathbf{r}_t^T L^{\frac{1}{2}}$, however, we found that the difference between L and $L^{\frac{1}{2}}$ was almost negligible despite the loss in sparsity therefore we tested both the usage of the Laplacian root as well as the Laplacian.

During our analysis we found a strong correlation between a stock spreads and its returns/residuals. In order to reduce the dependency between the two we propose a rescaling procedure: we scale the comoving portfolio returns/residuals by a factor $\beta_{ii} = \mathbb{C}ov(\mathbf{r}_{ti}, [A\mathbf{r}_t]_i)/\mathbb{V}ar([A\mathbf{r}_t]_i)$. Equation 2.4 can thus be written as $\mathbf{s}_t = \mathbf{r}_t^T - \mathbf{r}_t^T \beta A$. Ideally we could embed the β coefficients matrix into a new weighed adjacency $A^* = \beta A$, but unfortunately such a matrix would not be symmetric anymore. A possible workaround is to rescale the degree matrix instead of the comoving portfolio:

$$\mathbf{s}_t = \mathbf{r}_t^T \beta(\beta^{-1} - A) = \mathbf{r}_t^T \beta L^* \tag{2.10}$$

where $L^* = \beta^{-1} - A$, this matrix is not a Laplacian because $\mathbf{1}^T L^* \neq \mathbf{0}$, moreover, if the scale of the spreads is neglected the β term can be removed from the calculation. We will refer to this procedure interchangeably as Laplacian adjustment or adjusted Laplacian.

Identity

This approach represents a benchmark since we use an identity matrix as a degenerate Laplacian, i.e. $L = I$. In this case we assume a completely disconnected graph hence the output of the data preprocessing step is assumed to already be the idiosyncratic component:

$$\mathbf{s}_t = \mathbf{r}_t^T I = \mathbf{r}_t \tag{2.11}$$

This approach can not be thought as a pairs trading strategy and is instead used to assess whether more sophisticated approaches, like the ones presented above, do actually provide an edge that can not be explained by naive approaches.

Configurations Tested

Here we list all the configuration we tested for each family of approaches:

1. **Identity**;

2. **Corr-K** with $K \in \{1, 5, 20, 50\}$ in order to test all the approaches presented in the literature;

3. **SGS**, given L the solution to the optimization problem, we tested the following variations: i) L not adjusted, ii) L adjusted, iii) $L^{\frac{1}{2}}$ not adjusted, iv) $L^{\frac{1}{2}}$ adjusted;

4. **LGRMF**, given L the solution to the optimization problem, we tested the following variations: i) L not adjusted, ii) L adjusted, iii) $L^{\frac{1}{2}}$ not adjusted, iv) $L^{\frac{1}{2}}$ adjusted;

For notation purposes from now on we will refer to the output of the spreads evaluation step as $\mathbf{s}_t$ independently of the configuration chosen.

2.3 Trading

The trading step takes in input the spreads and produces the trading signals as output. In quasi-multivariate pairs trading, contrary to the univariate case where both the assets are traded, only the reference component is bought or sold and not the comoving portfolio, in order to avoid potentially high transaction costs. We identify two main approaches: i) **non-parametric**, ii) **quantiles based**.

Non-Parametric

Widely the most used in the univariate framework (GGR, Vidyamurthy 2004, Girma and Paulson 1999), this approach is based on thresholds estimated at pairs formation time, typically a pairs trade is opened when the price spread between the two securities exceeds the two historical standard deviations and is closed when the spread falls below such a threshold. In the quasi-multivariate case the undervalued (overvalued) asset is bought (sold) when the price spread with its comoving portfolio is below (above) the threshold $k_{\text{long}} = -2 \cdot std$ $(= -k_{\text{short}})$.

Unfortunately, since our spreads are estimated starting from the returns rather than the prices this framework can not be directly applied in our case. In order to embed past information in our approach we define the rolling spreads as the spreads rolling sum with window w:

$$\mathbf{s}_t^{(w)} = \sum_{\tau=t-w}^{t} \mathbf{s}_\tau \tag{2.12}$$

Since we used the log-returns to evaluate the spreads, if $w = t$ we recover the original price spreads normalized at 0 for $t = 0$.

We can now compute the rolling spreads standard deviation $s_{std,i}^{(w)}$ for each asset i during the pairs formation period and produce the associated trading signals:

- we **long** asset i if $s_{t,i}^{(w)} \leq -\gamma \cdot s_{std,i}^{(w)}$; while

- we **short** asset i if $s_{t,i}^{(w)} \geq \gamma \cdot s_{std,i}^{(w)}$.

If multiple stocks are longed (shorted) we weigh them evenly. In order to achieve zero exposition to the market the amount invested in the long leg is equal to the amount invested in the short leg.

Quantiles-Based

Proposed by C. W. Chen et al. 2017, this approach does not rely on threshold to compute the trading signals, instead, in the original version the authors first evaluate the return divergence as presented in 1.2 then sort the assets in descending order according to such divergence. The traded portfolio is thus constructed longing the stocks belonging to the top 10% and shorting the stocks belonging to the bottom 10%. Stocks with a low return divergence are assumed to be undervalued with respect to the comoving portfolios, on the contrary, stocks with a high return divergence are assumed to be overvalued. This trading method, unlike the threshold based, allows us to know how many pairs are going to be traded in each period and thus giving us information on the amount of capital that has to be allocated.

We generalized this method including past information using the rolling spreads instead of the naive return spreads in place of the return divergence and we also tested the effect of varying the amount of assets traded q on the performance.

Combining the two: Quantiles-Std

The trading signals, with the quantiles-based method, are triggered by divergences in the cross-section while in the non-parametric case the signals are triggered by divergences from the past. Due to their different nature one can think of combining the two trying to improve the performance. Therefore, we define a new method which we refer to as **Quantiles-Std**: a stock is bought (sold) if and only if both the non-parametric and the quantiles-based methods generate a long (short) signal.

We can also define the original non-parametric and quantiles-based methods as special cases of the Quantiles-Std method, i.e. Quantiles-Std with parameter $\gamma = None$ is equivalent to a quantiles-based method, similarly, Quantiles-Std with parameter $q = None$ is equivalent to a non-parametric method.

Configurations Tested

Now that we have defined all the methods we are going to use and how to recover the basic methods from Quantiles-Std, here we list all the configurations we tested. To run our tests we use only the Quantiles-Std method, as presented above this method requires the setting of three hyper-parameters, i.e. the moving window size w, the number of standard deviation γ for the non-parametric method and the share of securities traded q by the quantiles method. For each of the hyper-parameters above we define a set of possible values, the set of configurations we tested is simply the Cartesian product of all these sets:

A. $w \in \{1, 2, 4, 12, 24\}$;

B. $\gamma \in \{None, 2\}$;

C. $q \in \{None, 0.1, 0.2\}$.

Chapter 3

Learning Graph

3.1 Graph Algorithm

We have already shown how a, possibly sparse, undirected graphical model optimizing for the minimum Dirichlet energy is able to minimize the reconstruction error, it can be shown that the Dirichlet energy minimization naturally arises in the estimation of the inverse covariance matrix.

A basic model assumes the data coming from a multivariate Gaussian with mean vector μ and covariance matrix Σ. Variables i and j are assumed to be conditionally independent if and only if $\Sigma_{ij}^{-1} = 0$, hence it makes sense to model the partial dependence structure as a sparse inverse covariance matrix. Meinshausen and Bühlmann 2006 propose a simple algorithm relying on Lasso: they estimate a Lasso model to each variable using all the others as predictors. Friedman, Hastie, and Tibshirani 2008 explain why this approach can be seen as an approximation to the L1-penalized inverse covariance log-likelihood maximization problem and propose the Graphical Lasso algorithm that is both exact and remarkably fast. It became the benchmark for all future works. Following the work by Banerjee, Ghaoui, and Edu 2008, Friedman, Hastie, and Tibshirani 2008 define the penalized log-likelihood optimization problem as:

$$\underset{\Theta \succ 0}{\text{minimize}}\ tr(S\Theta) - \log\det(\Theta) + \lambda\|\Theta\|_1 \tag{3.1}$$

where $\Theta = \Sigma^{-1}$ is the non-negative inverse covariance matrix, $S \propto X^T X$ is the sample covariance matrix and $\|\Theta\|_1$ is the L_1 norm (sum of the absolute values) of Θ.

3.2 Laplacian constrained Gaussian Markov Random Field - LGMRF

More recent models assume the data to be coming from a zero-mean Improper Gaussian Markov Random Field (IGMRF) (Rue and Held 2005, Slawski and Hein 2015), whose rank-deficient inverse precision matrix $\Xi = \mathbb{E}\big[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T\big]^\dagger$, where $X^\dagger$ denotes the Moore-Penrose inverse of matrix X, is modeled by a graph Laplacian. The original Graphical Lasso problem can be thus generalized for the Laplacian constrained Gaussian Markov Random Field (LGMRF) model (Ying et al. 2020):

$$
\begin{aligned}
&\underset{L\succeq 0}{\text{minimize}} && tr(LS) - \log\det{}^*(L) + h_{\boldsymbol{\alpha}}(L) \\
&\text{subject to} && L\mathbf{1} = \mathbf{0},\ L_{ij} = L_{ji} \leq 0\ i \neq j
\end{aligned}
\tag{3.2}
$$

where $\det^*(L)$ is the pseudo determinant of L, i.e. the product of its positive eigenvalues, and $h_{\boldsymbol{\alpha}}(\cdot)$ is a regularization function with hyperparameter vector $\boldsymbol{\alpha}$ to impose properties such as sparsity. Thanks to the cyclic property of the trace, i.e. the trace is invariant under cyclic permutations, it is trivial to show that we are indeed minimizing the Dirichlet energy: $tr(XLX^T) = tr(LX^TX) = tr(LS)$.

This problem although convex, assuming a proper choice of $h_{\boldsymbol{\alpha}}$, cannot be easily solved due to scalability issues related to the computation of the $\log\det^*(L)$ term. In order to overcome that many scalable, iterative algorithms have been proposed. Among the others, Miranda Cardoso, Ying, and Daniel Perez Palomar 2020 propose several algorithms, based on Alternating Direction Method of Multipliers (ADMM) (Boyd et al. 2010) and Majorization-Minimization (MM) (Sun, Babu, and Daniel P. Palomar 2017), suited to various graph topologies (connected or k-components), and various types of distribution (normal or heavy-tailed). We tested the algorithm for connected normal graph.

The specialized version of problem 3.2 for connected graph makes use of several adjustment: i) in order to make optimization easier and allow for additional constraints onto the structure of the Laplacian, i.e. node degree constraints, they introduce the transformation $L = \mathcal{L}\mathbf{w}$, where $\mathbf{w} \in \mathbb{R}_+^{N(N-1)/2}$ contains the non-negative weights of the adjacency matrix and $\mathcal{L} : \mathbb{R}_+^{N(N-1)/2} \to \mathbb{R}^{N\times N}$ is the Laplacian operator; ii) similarly to the Laplacian operator, they define the degree operator $\mathcal{D} : \mathbb{R}^{N(N-1)/2} \to \mathbb{R}^N$, $\mathcal{D}\mathbf{w} = diag(L) = diag(D)$, which takes in input the weights vector $\mathbf{w}$ and outputs the diagonal of the Degree matrix; iii) they also rely on the equivalence $det^*(\Theta) = det(\theta + J)$, where $J = \frac{1}{N}\mathbf{1}\mathbf{1}^T$, to replace the pseudo-determinant term in the optimization problem. The connected graph optimization problem can then be formulated as:

$$
\begin{aligned}
&\underset{\mathbf{w}\geq\mathbf{0},\Theta\succeq 0}{\text{minimize}} && tr(S\mathcal{L}\mathbf{w}) - \log\det(\Theta + J) \\
&\text{subject to} && \Theta = \mathcal{L}\mathbf{w},\ \mathcal{D}\mathbf{w} = \mathbf{d}
\end{aligned}
\tag{3.3}
$$

where $\mathbf{d}$ is the desired degree vector, in our case $\mathbf{d} = \mathbf{1}$.

In order to solve the former optimization problem they define $\mathbf{Y}$ and $\mathbf{y}$ the dual variables associated with the constraints $\Theta = \mathcal{L}\mathbf{w}$ and $\mathcal{D}\mathbf{w} = \mathbf{d}$ respectively and they also define the transformation $\Omega^{l+1} = \Theta^{l+1} + J$. For each variable the following update rules are presented:

$$\begin{aligned} \Theta^{l+1} &= \Omega^{l+1} - J \\ \Omega^{l+1} &= \frac{1}{2\rho} U\big(\Gamma + \sqrt{\Gamma^2 + 4\rho I}\big) U^T \end{aligned} \tag{3.4}$$

where $U\Gamma U^T$ is the eigenvalue decomposition of $\rho(\mathcal{L}\mathbf{w}^l + J) - Y^l$ and $\rho > 0$ is the auxiliary ADMM parameter;

$$\begin{aligned} \mathbf{w}^{i+1} &= \left(\mathbf{w}^i - \frac{\mathbf{a}^i - \mathbf{b}^i}{2\rho(2N-1)}\right)^+ \\ \mathbf{a}^i &= \mathcal{L}^*\big(S - Y^l - \rho(\Theta^{l+1} - \mathcal{L}\mathbf{w}^i)\big) \\ \mathbf{b}^i &= \mathcal{D}^*\big(y^l - \rho(\mathbf{d} - \mathcal{D}\mathbf{w}^i)\big) \end{aligned} \tag{3.5}$$

where $\mathcal{L}^*$ and $\mathcal{D}^*$ are the adjoint of Laplacian and adjoint of degree operators respectively and $(\cdot)^+$ represent the elementwise ReLU function, i.e. $ReLU(x) = \max(0, x)$;

$$\begin{aligned} \mathbf{Y}^{l+1} &= \mathbf{Y}^l + \rho(\Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1}) \\ \mathbf{y}^{l+1} &= \mathbf{y}^l + \rho(\mathcal{D}\mathbf{w}^{l+1} + \mathbf{d}) \end{aligned} \tag{3.6}$$

The final algorithm can be written as follows.

Algorithm 1 LGMRF Algorithm

Require: $S, \mathbf{w}^0, \mathbf{d},$ penalty parameter $\rho > 0,$ tolerance $\epsilon > 0$
Ensure: Laplacian estimation $\mathcal{L}\mathbf{w}^*$
  initialize $\mathbf{Y} = \mathbf{0}, \mathbf{y} = \mathbf{0}, l \leftarrow 0$
  while $\max(|r^l|) > \epsilon$ *or* $\max(|s^l|) > \epsilon$ **do**
    update Ω^{l+1} according to 3.4
    update $\mathbf{w}^{l+1}$ iterating 3.5 until convergence ▷ ≈ 5 iterations are sufficient
    update $\mathbf{Y}^{l+1}, \mathbf{y}^{l+1}$ according to 3.6
    $\mathbf{r}^{l+1} = \Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1}$ ▷ compute residuals
    $\mathbf{s}^{l+1} = \mathcal{D}\mathbf{w}^{l+1} - \mathbf{d}$ ▷ compute residuals
    $l \leftarrow l + 1$
  end while

3.3 Smooth Graph Signal Representation - SGS

On the other hand, given the problems it carries some researcher abandoned the LGMRF model and focused on the assumption that underling graph signal is smooth.

Of particular interest is the work done by Kalofolias 2016, in which the author extend the original smooth graph signal representation proposed by Dong et al. 2016 providing the following convex formulation:

$$\begin{aligned} &\underset{A}{\text{minimize}} && \|A \circ Z\|_1 - \alpha \mathbf{1}^T \log(A\mathbf{1}) + \beta \|A\|_F^2 \\ &\text{subject to} && A_{ij} = A_{ji} \geq 0 \ i \neq j, \ \text{diag}(A) = \mathbf{0} \end{aligned} \tag{3.7}$$

where Z is the pairwise distance matrix, $Z_{ij} = ||x_i - x_j||_2^2$, and the original log determinant is replaced with its approximation $\log(A\mathbf{1})$ in order speed up the computation and avoid isolated nodes. Kalofolias 2016 also proved the equivalence between the weighted sum of the pairwise distances and the Dirichlet energy, i.e. $\|A \circ Z\|_1 = 2tr(XLX^T)$

Also in this case the author makes use of the vector form of A, i.e. $\mathbf{w} \in \mathbb{R}^{N(N-1)/2}$. He defines $d = \mathcal{D}\mathbf{w}$ and $\mathbf{y}, \bar{\mathbf{y}} \in \mathbb{R}^{N(N-1)/2}$ the dual variables associated with the optimization of the Dirichlet energy and the node degree penalty respectively.

Algorithm 2 SGS Algorithm

Require: $\mathbf{z}, \mathbf{w}^0, \mathbf{d}^0, \text{stepsize } \gamma, \text{tolerance } \epsilon > 0$
Ensure: Laplacian estimation $\mathcal{L}\mathbf{w}^*$
  $l \leftarrow 0$
  while $l < l_{max}$ **do**
    $\mathbf{y}^{l+1} = \mathbf{w}^l - \gamma(2\beta\mathbf{w}^l + \mathcal{D}^*\mathbf{d}^l)$
    $\mathbf{y}^{l+1} = \mathbf{d}^l - \gamma\mathcal{D}\mathbf{d}^l$
    $\mathbf{p}^{l+1} = \max(\mathbf{0}, \mathbf{y}^l - 2\gamma\mathbf{z})$ ▷ elementwise
    $\bar{\mathbf{p}}^{l+1} = (\bar{\mathbf{y}}^l - \sqrt{(\bar{\mathbf{y}}^l)^2 + 4\alpha\gamma})/2$ ▷ elementwise
    $\mathbf{q}^{l+1} = \mathbf{p}^l - \gamma(2\beta\mathbf{p} +^l + \mathcal{D}^*\mathbf{p}^l)$
    $\bar{\mathbf{q}}^{l+1} = \bar{\mathbf{p}}^l + \gamma(\mathcal{D}\mathbf{p}^l)$
    $\mathbf{w}^{l+1} = \mathbf{w}^l - \mathbf{y}^{l+1} + \mathbf{p}^{l+1}$
    $\mathbf{d}^{l+1} = \mathbf{d}^l - \bar{\mathbf{y}}^{l+1} + \bar{\mathbf{q}}^{l+1}$
    $l \leftarrow l + 1$
    if $\|\mathbf{w}^l - \mathbf{w}^{l1}\|/\|\mathbf{w}^{l1}\| < \epsilon$ **and** $\|\mathbf{d}^l - \mathbf{d}^{l1}\|/\|\mathbf{d}^{l1}\| < \epsilon$ **then**
      break
    end if
  end while

Where $\mathbf{z}$ is the vector form of the pairwise distance matrix Z. In this case the resulting Laplacian has not an even node degree therefore we compute the normalized Laplacian as $L^* = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ so as to fit our framework.

3.4 Visualize Graph

In order to understand where the Laplacian estimation methods differ here we show the networks estimated by the various method. The data used are the 5-factors residuals of the 50 stocks with the biggest market capitalization in the S&P500. The residuals are also normalized by dividing them by each stock's standard deviation. In the case of Corr-K a value of $K = 10$ is chosen.

The stock's GICS sector is represented by the node color [1], therefore, we can see that all the methods proposed, besides Identity, tend to bring closer stocks within the same sector.

It is worth of notice that, differently from Corr-K and LGMRF, the SGS method tends to prefer fewer but stronger connections, i.e. each stock movement is mostly explained by a handful of securities, rather than many weaker ones. If required this behavior can be changed by increasing the parameter α.

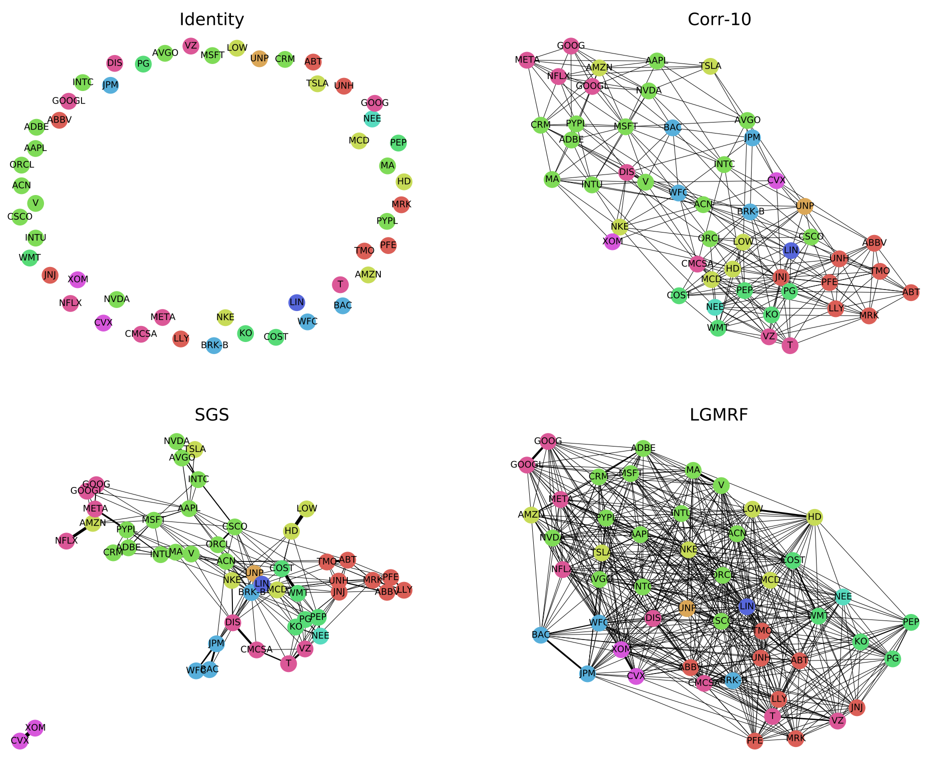

Figure 3.1: Graphs of the 50 U.S. largest stocks estimated with different learning methods.

[1] GICS Sector: color. Communication Services: pink. Consumer Cyclical: chartreuse green. Technology: green. Financial Services: light blue. Healthcare: red. Energy: magenta. Utilities: turquoise. Industrials: orange. Basic Materials: violet.

Chapter 4

Empirical Results

In this chapter, we are going to present the empirical results we obtained together with the description of the datasets used.

We utilized two different datasets: the first, smaller, consists of the monthly prices of 427 stocks currently listed in the S&P500 index from 1995-02 to 2022-09; the latter, larger, made up of the monthly prices of 3,218 carefully selected stocks from the CRSP (Center for Research in Security Prices) U.S. Stock Databases from 1962-06 to 2022-10.

In order to perform the trading simulation the graph Laplacians are computed once a year using the last 5 years' worth of data, i.e. 60 samples, and are used to evaluate the spreads in the following 12 months. We opted for a slow update frequency of the Laplacian in order to reduce the computational cost, however, we have not detected a degradation in performance in the months following the estimation, this proves the robustness and the long-lasting property of the retrieved relationships.

As anticipated, in order to understand which approach, or combination of approaches, is preferable and leads to higher, less volatile, returns we tested a total of 780 different configurations. Since a complete grid search on both datasets is infeasible we test the whole set of configurations solely on the S&P500 dataset. Then, using the information obtained, a subset of the configurations is extracted and tested on both the S&P500 and the U.S. CRSP datasets.

CRSP filtering criteria

Although the CRSP U.S. Stock Databases contains $\sim$ 27,000 stocks we kept just the firms who met the following criteria:

1. there must have been at least 10 years worth of data, i.e. 120 samples;

2. in order to remove the penny stocks all the securities whose minimum price fell below 5$ are dropped;

3. in order to remove illiquid stocks first we define the discounted log-volumes, we used $\log_{10}$, of stock i as $\mathbf{v}_i^* = \mathbf{v}_i - \mathbf{v}_t + v_t$ where $\mathbf{v}_i$ is the i-th stock log-volumes, $\mathbf{v}_t$ contains the average cross-sectional log-volumes and v_t is the average log-volumes at

the end of the period; then we remove the stocks whose average discounted log-volumes is smaller than 6; this can be thought as dropping the securities with less than a million shares traded every month on average, assuming the average constant over time.

Metrics and estimation procedure

To assess the performance the following metrics have been chosen: i) **monthly expected returns**; ii) **Sharpe ratio**, defined as $S_hR = \mathbb{E}[R - R_f]/\sqrt{\mathbb{V}ar[R - R_f]}$, where R is the portfolio return and R_f is the risk-free return, in our case $R_f = 0$; and iii) **Sortino ratio**, defined as $S_oR = \mathbb{E}[R]/DR$, where DR is the downside risk $DR = \sqrt{\int_{-\infty}^{0} r^2 f(r)dr}$ and $f(r)$ is the distribution of monthly returns.

The quantity of interest empirical distributions, together with their 95% confidence intervals, are computed using nonparametric bootstrap [1].

Let $\mathcal{C}$ bet the set of all the tested configurations and $R_{c,t}$ the return produced by the configuration c at time t, in order to compare different configurations we split the original configuration-return dataset $\mathcal{R} = \{\langle R_{c,t}, c\rangle, \forall\ c \in \mathcal{C}, t \in \mathcal{T}\}$ into n partitions $\mathcal{R}_p = \{\langle R_{c,t}, c\rangle, \forall\ c \in \mathcal{C}_p, t \in \mathcal{T}\}$, $\mathcal{C} = \bigcup_{p\in\mathcal{P}} \mathcal{C}_p$, and we evaluate the former metrics on these datasets independently; finally we analyze if any difference can be observed.

4.1 Results - Sp500

Due to the large number of configurations tested reporting the maximum performance achieved by one configuration will likely result in misleading claims since we would encounter overfitting. To prevent that while seeking for the optimal configuration we assume the 3 phases to be independent of each other thus we optimize each step individually, i.e. for each method we tested in each step we evaluate the empirical distribution of the metrics of interest together with some point estimates.

4.1.1 Data Preprocessing

The first step is the data preprocessing, here we tested two approaches: **residuals** and **returns**.

As we can see from Table 4.1 the approaches based on returns generate slightly higher excess returns with respect to the ones based on residuals. However, this difference is partially negligible due to the large overlapping of the distributions, Figure A.1; on the contrary, if we analyze risk-adjusted metrics, like Sharpe and Sortino ratios, we realize that, although lower, the excess returns generated using residuals are strongly less volatile hence less risky.

[1] In the nonparametric bootstrap a sample, also called bootstrap sample, of the same size as the data is obtained sampling the data with replacement. The empirical distribution of the metric of interest is then computed on the bootstrap samples.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
residuals	0.82%	**0.47**	**0.76**
	(0.78%-0.86%)	(0.45-0.49)	(0.72-0.81)
returns	**0.86%**	0.36	0.58
	(0.8%-0.91%)	(0.34-0.39)	(0.54-0.62)

Table 4.1: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for different data preprocessing methods on the S&P500 dataset. Bold values represent the best results.

For these reasons we believe that the approaches based on residuals should be preferred.

4.1.2 Spreads

In order to understand which spreads computation method is able to achieve the best performance we first analyze each class of approaches individually, i.e. correlation based, sparse inverse covariance, then we will compare the best of each class with the benchmark.

Corr-K

For what concerns the **Corr-K** method it is trivial to see, from Table 4.2, that an increase in the number of comoving assets is associated to an improvement in performance. This confirms what C. W. Chen et al. 2017 and M. Perlin 2007 noticed in their works. It is also worth noting that the improvement is not symmetric in the long and the short leg: Figure C.2 shows that increasing K the expected return of the long leg remains unchanged while the expected return of the short leg decreases. This suggests that movements of the long and the short leg might have different causes hence different methods should be employed to detect and predict them.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
Corr-1	0.57%	0.44	0.65
	(0.5%-0.65%)	(0.38-0.5)	(0.54-0.75)
Corr-5	0.69%	0.44	0.72
	(0.6%-0.78%)	(0.38-0.5)	(0.61-0.84)
Corr-20	0.77%	0.44	0.74
	(0.67%-0.87%)	(0.39-0.5)	(0.63-0.86)
Corr-50	**0.87%**	**0.48**	**0.79**
	(0.77%-0.98%)	(0.42-0.53)	(0.68-0.91)

Table 4.2: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for the Corr-K method with different values of K on the S&P500 dataset. Bold values represent the best results.

Sparse Inverse Covariance

In the sparse inverse covariance class we have the following two methods: **SGS** and **LGMRF**. For each of these methods we tested several configurations, unfortunately due to the extensive overlap of the distribution, see Figure A.3, it is impossible to say which configuration is better that the other definitively. In general, the usage of the Laplacian root seems beneficial for both methods whilst we would suggest to adjust the Laplacian solely when using the SGS method.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
LGMRF no-root, no-adjust	0.79%	0.44	0.7
	(0.68%-0.89%)	(0.38-0.49)	(0.59-0.81)
LGMRF no-root, adjust	**0.89%**	0.42	0.69
	(0.77%-1.02%)	(0.37-0.48)	(0.58-0.79)
LGMRF root, no-adjust	0.84%	**0.45**	**0.73**
	(0.73%-0.95%)	(0.39-0.5)	(0.62-0.84)
LGMRF root, adjust	0.87%	0.4	0.62
	(0.75%-1.0%)	(0.34-0.45)	(0.52-0.73)
SGS no-root, no-adjust	0.88%	**0.4**	**0.61**
	(0.76%-1.01%)	(0.34-0.46)	(0.51-0.71)
SGS no-root, adjust	**0.91%**	0.38	0.6
	(0.78%-1.05%)	(0.32-0.44)	(0.5-0.71)
SGS root, no-adjust	0.89%	**0.4**	**0.61**
	(0.75%-1.01%)	(0.34-0.46)	(0.5-0.71)
SGS root, adjust	**0.91%**	0.38	**0.61**
	(0.77%-1.06%)	(0.32-0.44)	(0.5-0.71)

Table 4.3: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for the sparse inverse covariance methods with different configurations on the S&P500 dataset. Bold values represent the best results for each method.

Spreads Methods

Here we compare the best configuration of each method (**Corr-50**; **LGMRF** root, no-adjust; **SGS** root, adjust) with the benchmark method (**Identity**). From Table 4.4 we can see that not only the excess returns generated by the benchmark are statistically significant but they also exceed the excess returns generated by more sophisticated approaches. It is worth to notice that the benchmark excess returns can not be explained as the premium for the reconstruction of some hidden relationships between couples of assets since it completely neglects such dependencies in the spreads computation. Instead, we argue that they should be interpreted as the compensation for closing the gap between a stock's returns and the returns of its factors: since trades are mainly opened on diversion from the average cross-section return, i.e. the market factor, (quantiles-based, quantiles-std trading rules), the robust and statistically significant excess return proves the reversion to the market property, i.e. stocks that under-performed the market will experience higher returns in the following month and vice versa.

In this scenario approaches like Corr-K or LGMRF, at the cost of diminished excess returns, can considerably reduce the risk associated with the strategy, as shown by the Sharpe and Sortino ratios of these methods. We believe that a comoving portfolio so created can be interpreted as a good approximation for a stock's linear factor model hence deviations from those although smaller show a higher convergence probability and a lower risk.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
Corr-50	0.87%	**0.47**	**0.79**
	(0.76%-0.99%)	(0.42-0.53)	(0.68-0.91)
Identity	**0.99%**	0.35	0.65
	(0.83%-1.17%)	(0.31-0.4)	(0.53-0.77)
LGMRF	0.84%	0.45	0.73
	(0.73%-0.95%)	(0.39-0.5)	(0.62-0.84)
SGS	0.91%	0.38	0.6
	(0.77%-1.06%)	(0.33-0.44)	(0.5-0.71)

Table 4.4: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for different spreads computation methods on the S&P500 dataset. Bold values represent the best results.

4.1.3 Trading Rules

As the last step of the chain we analyze the effect of the trading rule on the performance. In Table 4.5 we report the top 3 configurations that maximize each of the quantities of interest. As we can see they all share a short rolling window $w = 1, 2$ and indeed from

Configuration	Excess Returns	Sharpe Ratio	Sortino Ratio
w=1 - γ=None - q=0.1	0.96% (0.85%-1.09%)	0.65 (0.58-0.73)	1.0 (0.85-1.17)
w=1 - γ=None - q=0.2	0.8% (0.71%-0.89%)	**0.72** (0.65-0.8)	1.17 (1.02-1.34)
w=2 - γ=2 - q=0.1	**1.43%** (1.25%-1.61%)	0.64 (0.56-0.73)	**1.18** (1.01-1.37)
w=2 - γ=2 - q=0.2	1.3% (1.13%-1.47%)	0.63 (0.54-0.73)	1.14 (0.97-1.33)
w=2 - γ=2 - q=None	1.25% (1.09%-1.42%)	0.63 (0.54-0.72)	1.12 (0.95-1.3)

Table 4.5: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for different configurations of the trading rule on the S&P500 dataset. Bold values represent the best results. Here: i) w is the window size employed in the rolling spreads computation and it represents the number of months taken into consideration at trading time; ii) γ is the threshold utilized by the non-parametric trading rule, i.e. a stock is bought (sold) if its rolling spread exceeds γ ($-\gamma$) historical standard deviations; and iii) q is the amount of assets traded by the quantiles-based trading rule.

Figure A.5 we notice that an increase of the rolling window above $w = 2$ is associated to a worsening of the performance.

From Table 4.5 we can see that the use of a $\gamma \neq None$ is beneficial because it helps reducing the false positive rate and hence increasing the excess returns while keeping the risk under control.

We choose as the best configurations the following: $\{w = 1, \gamma = None, q = 0.2\}$ and $\{w = 2, \gamma = 2, q = 0.1\}$.

4.1.4 Performance Over Time

In this and the following section we analyze the performance over time and the 5-factor alpha produced by a small subset of the tested configurations. Specifically we select the best configurations for each step:

1. Data preprocessing: **Residuals**;
2. Spreads methods: **Identity**; **Corr-50**; **LGMRF** root, no-adjust; **SGS** root, adjust;
3. Trading rule: **Quantiles-std** with parameters $\{w = 1, \gamma = None, q = 0.2\}$ and $\{w = 2, \gamma = 2, q = 0.1\}$.

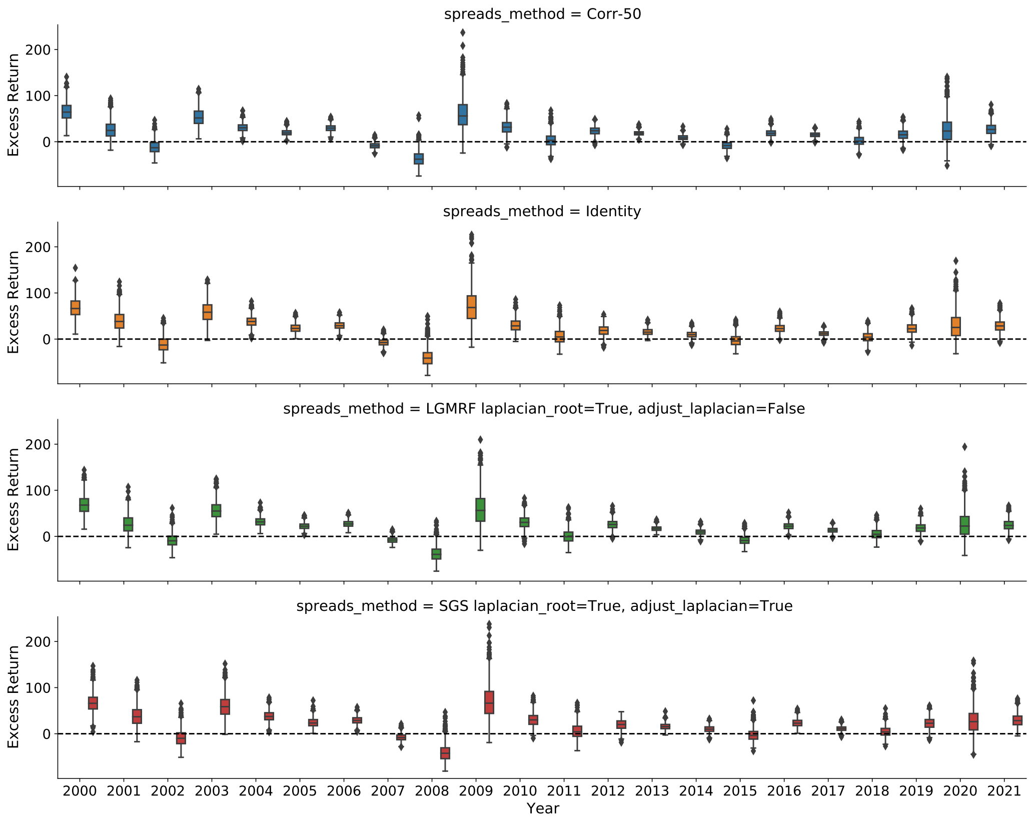

Figure 4.1: Evolution of the excess return distribution from 2000 to 2021, grouped by the spreads method, evaluated on the S&P500 dataset.

As many author reported and as we can see from Figures 4.1, B.1, we notice a decline in profitability after the early 2000's. Thus we decided to estimate the empirical distribution of the excess return in each year: we found that starting from 2007, with the only exception of the year 2009, positive excess returns on the S&P500 are no longer statistically significant (see Figures 4.1, B.1).

4.1.5 5-Factors α

Here we analyze the 5-Factors α which is defined as the intercept of the linear factor model:

$$R_j(t) = \alpha + \sum_i \beta_i f_i(t) + \epsilon,\ \ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{4.1}$$

where $R_j(t)$ is the return of the strategy j at time t. This quantity is crucial since it represents the portion of the returns that can not be explained by standard market factors nor by random movement.

	Spread	w	q	γ	Mkt-RF	SMB	HML	RMW	CMA	RF	α
0	Corr-50	1	0.2	None	0.119	-0.123	-0.014	-0.222	0.357	0.951	1.098%**
					(-0.04, 0.277)	(-0.36, 0.111)	(-0.26, 0.233)	(-0.5, 0.059)	(-0.04, 0.75)	(-3.27, 5.175)	(0.27, 1.93)
1	Corr-50	2	0.1	2	0.126	-0.131	-0.036	-0.263	0.322	1.038	1.264%**
					(-0.05, 0.304)	(-0.4, 0.133)	(-0.32, 0.243)	(-0.58, 0.054)	(-0.12, 0.765)	(-3.72, 5.801)	(0.33, 2.202)
2	Identity	1	0.2	None	0.083	-0.084	-0.035	-0.213	0.287	0.405	1.23%**
					(-0.08, 0.247)	(-0.33, 0.16)	(-0.29, 0.222)	(-0.5, 0.08)	(-0.12, 0.696)	(-3.99, 4.805)	(0.36, 2.096)
3	Identity	2	0.1	2	0.09	-0.066	-0.043	-0.186	0.176	2.194	1.444%**
					(-0.11, 0.294)	(-0.37, 0.236)	(-0.36, 0.276)	(-0.55, 0.177)	(-0.33, 0.684)	(-3.26, 7.648)	(0.37, 2.517)
4	LGMRF	1	0.2	None	0.105	-0.141	0.01	-0.26	0.319	0.735	1.135%**
					(-0.05, 0.263)	(-0.38, 0.093)	(-0.24, 0.258)	(-0.54, 0.021)	(-0.08, 0.712)	(-3.49, 4.963)	(0.3, 1.967)
5	LGMRF	2	0.1	2	0.126	-0.134	0.017	-0.3	0.235	1.808	1.317%**
					(-0.06, 0.311)	(-0.41, 0.14)	(-0.27, 0.306)	(-0.63, 0.029)	(-0.22, 0.695)	(-3.13, 6.749)	(0.34, 2.29)
6	SGS	1	0.2	None	0.083	-0.084	-0.035	-0.213	0.287	0.405	1.23%**
					(-0.08, 0.247)	(-0.33, 0.16)	(-0.29, 0.222)	(-0.5, 0.08)	(-0.12, 0.696)	(-3.99, 4.805)	(0.36, 2.096)
7	**SGS**	**2**	**0.1**	**2**	0.088	-0.07	-0.043	-0.191	0.177	2.15	**1.466**%**
					(-0.12, 0.292)	(-0.37, 0.232)	(-0.36, 0.277)	(-0.55, 0.172)	(-0.33, 0.685)	(-3.31, 7.607)	(0.39, 2.54)

Table 4.6: This table reports the factor loadings and the strategies monthly α for the top-8 configurations tested on the S&P500 dataset. Bold values represent the best α and the best configuration. Statistical significance at the 5%, 1% and 0.1% levels are indicated by *, ** and *** respectively.

In Table 4.6 we report the estimated β_i and α for each of the chosen configurations. As we can see most of the βs are close to 0 and not statistically significant, this shows that statistical arbitrage strategies do not indeed rely on common risk factors as source of profitability and can be thought independent from those.

Unfortunately, as we have seen above, excess returns tend to decline over time therefore a single point estimate might miss relevant information about the evolution of α through the years. To the purpose of recovering this information, using a rolling OLS with a moving window of 60 samples (5 years), we plot below, in Figure 4.2, the values of α over time together with its p-value and the S&P500 volatility estimated through a GARCH model. There we can clearly see the α values declining in the decade 2010-2020. We can also see that the significance of those values is often inversely related to the volatility of the market and in general well above the canonical 5% threshold.

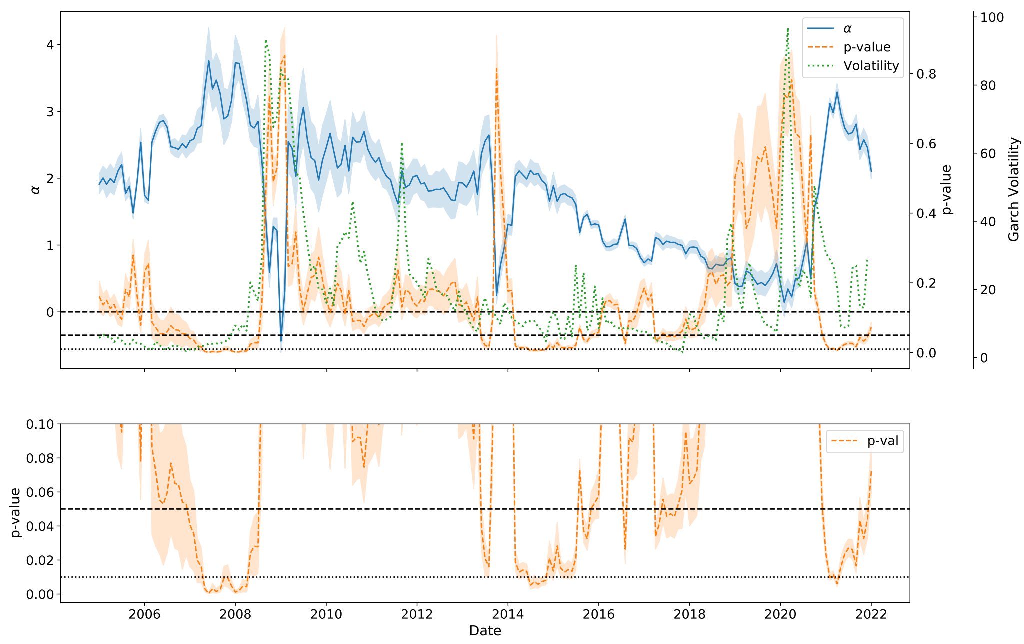

Figure 4.2: *Top*: Rolling α, α p-value and volatility estimated using a GARCH model from 2005 to end 2022 evaluated on the S&P500 dataset. *Bottom*: Zoom of the p-value time series in the 0-0.1 area.

4.2 Results - CRSP

In this section we analyze the results produced by the best configurations identified above on the U.S. CRSP dataset. The original CRSP U.S. Stock Databases include all the stocks listed on the NYSE, NYSE American, NASDAQ, NYSE Arca and Bats exchanges. The U.S. CRSP dataset contains a subset of the stocks included in the CRSP U.S. Stock Databases filtered to meet the criteria specified above.

4.2.1 Spreads Methods

The results split by spreads method confirms what we found above: although all the methods proposed are able to achieve similar excess returns, see Figures A.6, C.5, techniques such as Corr-50 and LGMRF show considerably lower risk thanks to their capability in reducing unsuccessful trades.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
Corr-50	2.07%	**2.42**	**5.15**
	(1.9%-2.23%)	(2.22-2.62)	(4.15-6.33)
Identity	**2.1%**	1.86	3.31
	(1.88%-2.31%)	(1.67-2.06)	(2.63-4.19)
LGMRF	**2.1%**	2.31	4.75
	(1.92%-2.27%)	(2.11-2.51)	(3.84-5.84)
SGS	**2.1%**	1.87	3.36
	(1.89%-2.31%)	(1.67-2.07)	(2.67-4.24)

Table 4.7: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for different spreads computation methods on the U.S. CRSP dataset. Bold values represent the best results.

4.2.2 Trading Rules

For what concerns the trading rules, as we can see from Table 4.8 it is not possible to define a configuration definitively better than the other: Quantiles-std with parameters $\{w = 1, \gamma = None, q = 0.2\}$ shows lower excess returns but higher risk-adjusted metrics, on the contrary, Quantiles-std with parameters $\{w = 2, \gamma = 2, q = 0.1\}$ has considerably higher returns but so is the associated risk (see Table 4.8). Therefore, depending on the quantity we aim to maximize one or the other configuration should be chosen.

Configuration	Excess Returns	Sharpe Ratio	Sortino Ratio
w=1 - γ=None - q=0.2	1.83%	**2.29**	**3.99**
	(1.72%-1.93%)	(2.14-2.44)	(3.32-4.83)
w=2 - γ=2 - q=0.1	**2.35%**	1.99	3.82
	(2.19%-2.51%)	(1.85-2.13)	(3.26-4.48)

Table 4.8: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for different configurations of the trading rule on the U.S. CRSP dataset. Bold values represent the best results. Here: i) w is the window size employed in the rolling spreads computation and it represents the number of months taken into consideration at trading time; ii) γ is the threshold utilized by the non-parametric trading rule, i.e. a stock is bought (sold) if its rolling spread exceeds γ ($-\gamma$) historical standard deviations; and iii) q is the share of assets traded by the quantiles-based trading rule.

4.2.3 Performance Over Time

As before, here we show the excess return over time split by spreads method, Figure 4.3. Looking at a longer time period we can confirm that the excess returns have been largely positive and robust until 2002 when they started declining and becoming not statistically significant. Such decline can also be observed when we split according to the trading method, see Figure B.2.

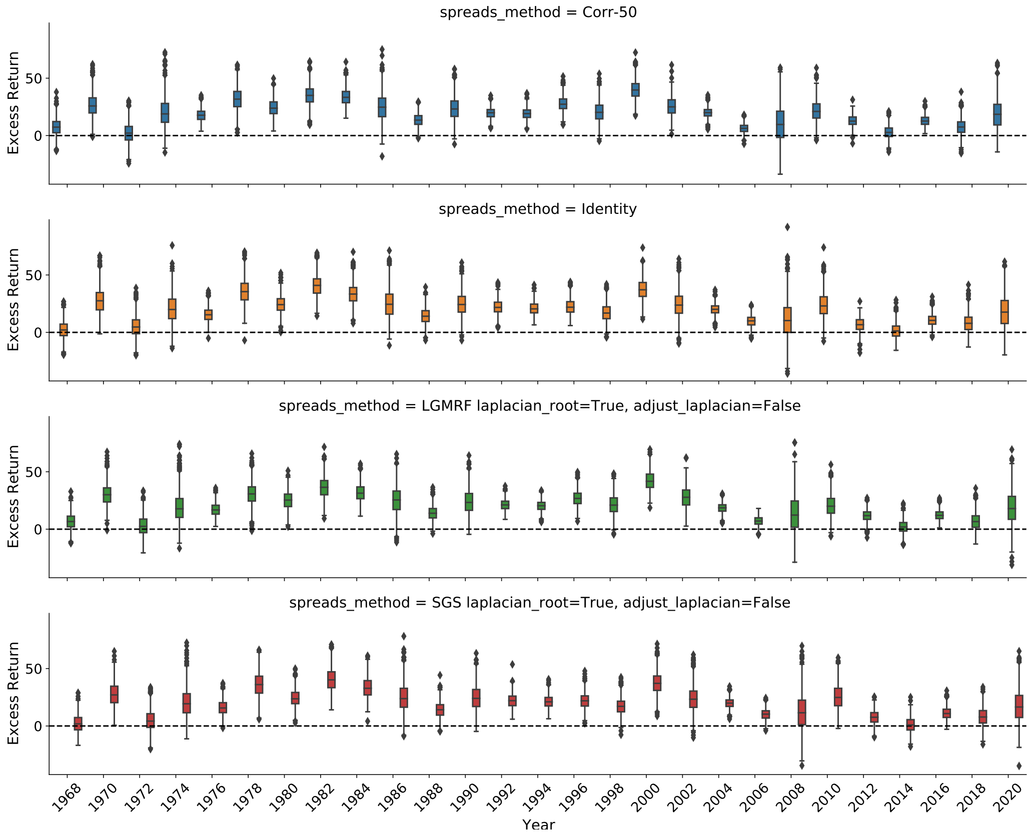

Figure 4.3: Evolution of the annualized excess return distribution from 2000 to 2021, grouped by the spreads method, evaluated on the U.S. CRSP dataset.

4.2.4 5-Factors α

The point estimates of the β_is and α for the tested configurations confirm what we found using the S&P500 dataset: statistical arbitrage strategies tend to be uncorrelated to common market factors and show instead a strongly positive and statistically significant α (as it can be seen in Table 4.9).

	Spread	w	q	γ	Mkt-RF	SMB	HML	RMW	CMA	RF	α
0	Corr-50	1	0.2	None	0.033	0.045	0.019	0.027	-0.046	0.805	1.103%***
					(-0.04, 0.111)	(-0.07, 0.159)	(-0.12, 0.162)	(-0.12, 0.179)	(-0.27, 0.182)	(-0.32, 1.927)	(0.57, 1.632)
1	Corr-50	2	0.1	2	0.048	0.017	-0.033	0.011	-0.013	0.628	1.387%***
					(-0.04, 0.132)	(-0.11, 0.141)	(-0.19, 0.123)	(-0.15, 0.177)	(-0.26, 0.235)	(-0.59, 1.847)	(0.81, 1.961)
2	Identity	1	0.2	None	0.012	0.061	0.06	0.04	-0.103	0.866	0.984%***
					(-0.07, 0.094)	(-0.06, 0.18)	(-0.09, 0.21)	(-0.12, 0.2)	(-0.34, 0.136)	(-0.31, 2.045)	(0.43, 1.539)
3	Identity	2	0.1	2	0.033	0.024	-0.013	0.08	-0.052	1.012	1.348%***
					(-0.06, 0.127)	(-0.11, 0.162)	(-0.19, 0.16)	(-0.1, 0.264)	(-0.33, 0.222)	(-0.34, 2.364)	(0.71, 1.985)
4	LGMRF	1	0.2	None	0.031	0.045	0.015	0.025	-0.028	0.853	1.074%***
					(-0.05, 0.109)	(-0.07, 0.161)	(-0.13, 0.16)	(-0.13, 0.178)	(-0.26, 0.203)	(-0.28, 1.987)	(0.54, 1.608)
5	**LGMRF**	**2**	**0.1**	**2**	0.043	0.014	-0.026	0.001	-0.016	0.636	**1.433**%***
					(-0.04, 0.128)	(-0.11, 0.14)	(-0.18, 0.132)	(-0.17, 0.169)	(-0.27, 0.236)	(-0.6, 1.874)	(0.85, 2.017)
6	SGS	1	0.2	None	0.012	0.061	0.06	0.04	-0.103	0.866	0.984%***
					(-0.07, 0.094)	(-0.06, 0.18)	(-0.09, 0.21)	(-0.12, 0.2)	(-0.34, 0.136)	(-0.31, 2.045)	(0.43, 1.539)
7	SGS	2	0.1	2	0.036	0.021	-0.015	0.077	-0.045	0.966	1.37%***
					(-0.06, 0.13)	(-0.12, 0.158)	(-0.19, 0.158)	(-0.11, 0.26)	(-0.32, 0.229)	(-0.38, 2.316)	(0.73, 2.006)

Table 4.9: This table reports the factor loadings and the strategies monthly α for the 8 configurations tested on the U.S. CRSP dataset. Bold values represent the best α and the best configuration. Statistical significance at the 5%, 1% and 0.1% levels are indicated by *, ** and *** respectively.

5

The evolution of α over time, shown in 4.4, reveals how such strategies perform poorly both during periods of high volatility and during crises, i.e. before 1980, dot-com bubble (1995-2005), financial crisis (2008-2010) and finally COVID-19 crisis (2020).

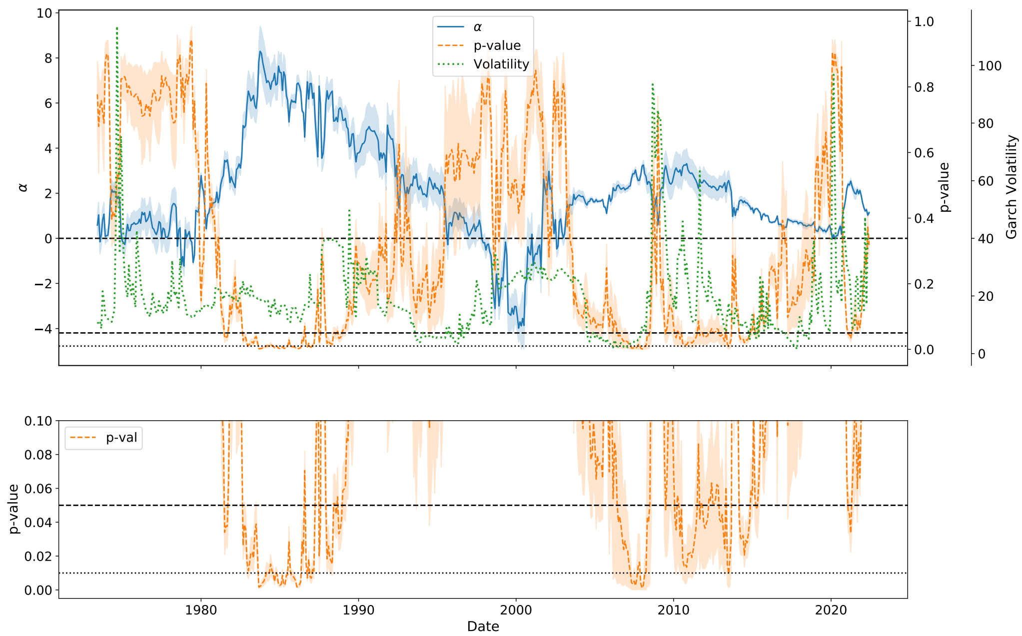

Figure 4.4: *Top*: Rolling α and α p-value evaluated on the U.S. CRSP dataset; volatility estimated for the S&P500 index using a GARCH model from 1973 to end 2022. *Bottom*: Zoom of the p-value time series in the 0-0.1 area.

4.3 Market Capitalization and Liquidity

Prior works have linked pairs trading profitability to poor information environment (C. W. Chen et al. 2017) and illiquidity (Engelberg, Gao, and Jagannathan 2009).

Here we used the monthly average volumes as a proxy measures for market capitalization and liquidity. In order to assess the effects that those might have on pairs trading we split the U.S. CRSP trading universe into 3 tertiles based on the monthly average volumes. We name the partitions **Low**, **Mid**, **High** containing stocks with low, medium and high liquidity, respectively. We evaluate the spreads using the entire stock universe whereas we compute the trading signals individually for each group.

To evaluate the performance of the three partitions we use the following configuration:

1. **Data preprocessing**: residuals;
2. **Spreads method**: Corr-50;
3. **Trading Rule**: Quantiles-std with $w = 2$, $\gamma = 2$, $q = 0.1$.

In Table 4.10, below we can see that there are indeed differences between the partitions: i) **Low** liquidity firm such the highest excess returns; while **Mid** achieve the best performance in the risk-adjusted metrics.

Liquidity	Excess Returns	Sharpe Ratio	Sortino Ratio
High	1.22%	1.55	2.93
	(1.01%-1.43%)	(1.29-1.82)	(2.24-3.73)
Mid	1.58%	**2.51**	**5.32**
	(1.42%-1.75%)	(2.24-2.79)	(4.2-6.8)
Low	**1.69%**	2.39	4.12
	(1.51%-1.88%)	(2.07-2.69)	(3.07-5.54)

Table 4.10: This table reports monthly excess returns, Sharpe ratio, Sortino ratio for different liquidity levels on the U.S. CRSP dataset. Bold values represent the best results.

4.3.1 Performance Over Time

Evaluating the excess returns over time it can be noticed, from Figure 4.5, how the high liquidity partition shows a decline in profitability, becoming largely unprofitable between 2012 and 2018. On the other hand, low liquidity partition shows robust excess returns throughout the trading period. This confirms the idea that market frictions are beneficial to pairs trading.

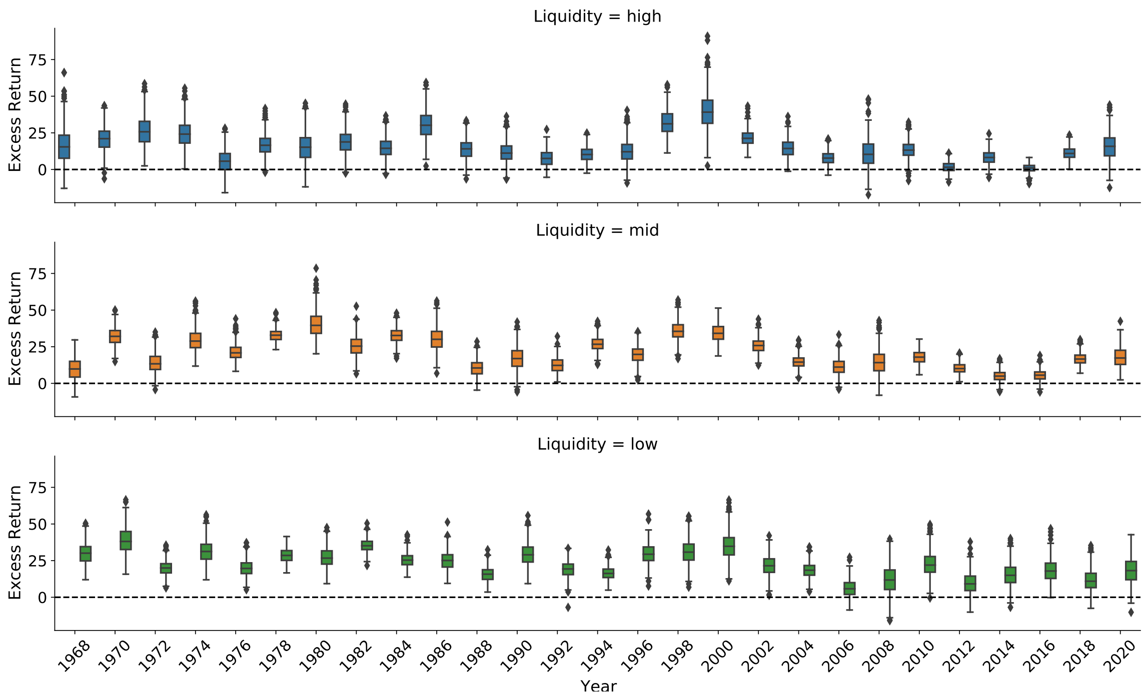

Figure 4.5: Evolution of the annualized excess return distribution from 2000 to 2021, grouped by the liquidity level, evaluated on the U.S. CRSP dataset.

4.3.2 5-Factors α

Finally, we evaluate the 5-Factors α: as expected the α generated by the low liquidity split is the highest of the three (see Table 4.11). We also found that the coefficients associated to the risk-free factor (RF) in the case of Mid and Low liquidity are positive at 0.001 significance level. This suggests that pairs trading profitability, and other market inefficiencies, might indeed be related to the investors' uncertainty. To the best of our knowledge, this relation has not been reported in the literature yet.

Liquidity	Mkt-RF	SMB	HML	RMW	CMA	RF	α
High	0.006	−0.013	−0.017	−0.015	0.07	0.55	1.001%***
	(-0.05, 0.058)	(-0.09, 0.064)	(-0.11, 0.08)	(-0.12, 0.088)	(-0.08, 0.223)	(-0.21, 1.306)	(0.64, 1.357)
Mid	0.009	−0.015	−0.038	0.036	0.064	1.326***	1.071%***
	(-0.03, 0.051)	(-0.08, 0.046)	(-0.12, 0.038)	(-0.04, 0.118)	(-0.06, 0.186)	(0.73, 1.926)	(0.79, 1.354)
Low	−0.053*	−0.025	−0.04	0.025	0.015	0.996***	**1.36%*** **
	(-0.1, -0.006)	(-0.09, 0.043)	(-0.13, 0.046)	(-0.07, 0.117)	(-0.12, 0.153)	(0.32, 1.674)	(1.04, 1.68)

Table 4.11: This table reports the factor loadings and the strategies α for different liquidity levels on the U.S. CRSP dataset. The bold value represents the best α. Statistical significance at the 5%, 1% and 0.1% levels are indicated by *, ** and *** respectively.

From the evolution of the α's over time in Figure 4.6 we can notice that not only α-low tends to be higher in value but it is more statistically significant too. In

Table 4.12 we report the time spent by the splits inside various statistically significant regions.

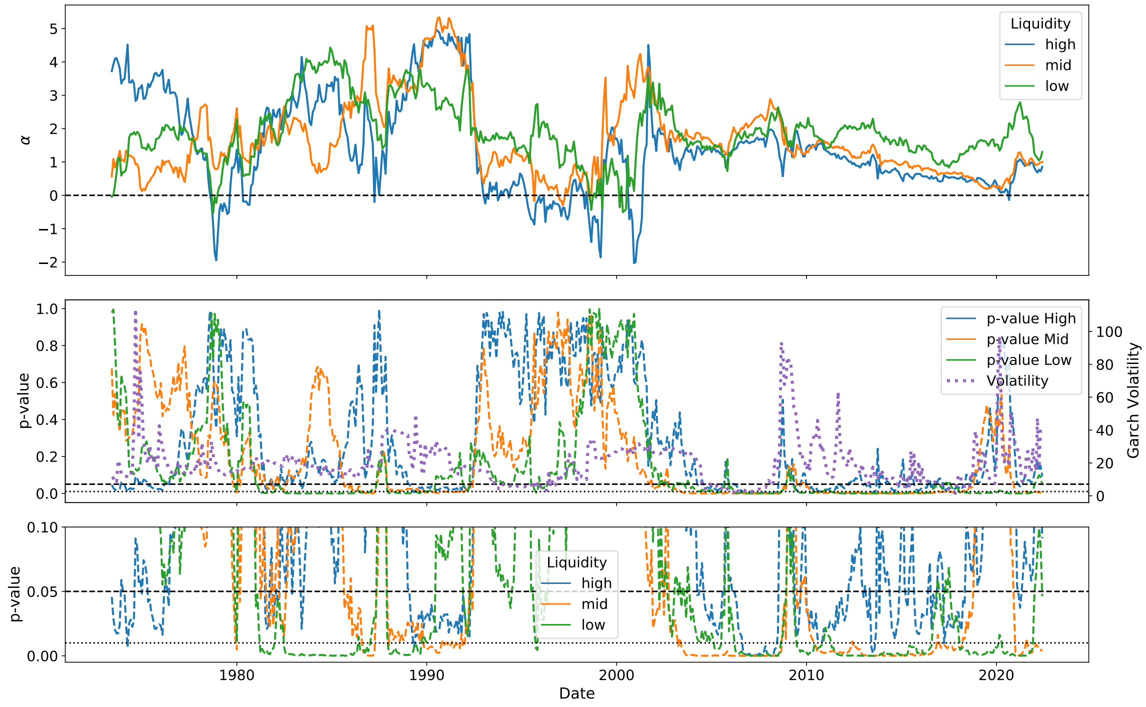

Figure 4.6: *Top*: Rolling αs grouped by liquidity level. *Center*: α p-values grouped by liquidity level and volatility estimated for the S&P500 index using a GARCH model. *Bottom*: Zoom of the p-value time series in the 0-0.1 area. The time series are evaluated on the U.S. CRSP dataset from 1973 to end 2022.

Liquidity	Time $pval < 0.05$	Time $pval < 0.01$	Time $pval < 0.001$
High	32.9%	4.8%	1.5%
Mid	50.3%	35.1%	17.1%
Low	**58.6%**	**41.8%**	**19.0%**

Table 4.12: This table reports the percentage of the time spent, by the rolling α's p-value, below the 5%, 1% and 0.1% significance level respectively for different liquidity levels on the U.S. CRSP dataset. Bold values represent the best results.

Conclusion

In our thesis we examined the pairs trading problem, we presented a unified framework based on graph Laplacians that fits all the quasi-multivariate approaches proposed in the literature and extends them through the use of factor residuals instead of raw returns, which we have shown to achieve similar performance with a considerably lower risk.

For what concerned the spreads, we proposed two novel methods for extracting the idiosyncratic component of the stocks based on graphical models under Laplacian structural constraints however they did not outperform previous correlation-based methods. Interestingly, we also found that those advanced pairs trading methods do not generate statistically higher α's nor higher excess returns with respect to simpler statistical arbitrage strategies which do not rely on explicit relationships between stocks. Given that the statistical arbitrage strategies employed make a profit through the mean- reversion of the residuals this led us to believe that so do the pairs trading strategies we tested. In this perspective, the comoving portfolios can be thought of as company-specific factors. Further research should investigate the relationship between the two and assess whether pairs trading is still profitable after accounting for those statistical arbitrage strategies.

About the trading signals: we tested the two trading rules most commonly used in the literature and we also defined a new trading rule combining the former. We concluded that the combination of the two is in general preferable since it leads to higher excess returns.

We extended the sample period up to recent days proving that the profitability decline has persisted until the 2020 COVID-19 crisis. We also have shown how a low liquidity environment is advantageous: it increases and stabilizes excess returns while reducing risks; this confirms the hypothesis that pairs trading profitability is partially driven by delays in information absorption.

Finally, our analysis over time reveals a plunge in the significance of the α's during crises and high volatility periods.

Bibliography

Avellaneda, Marco and Jeong Hyun Lee (2010). "Statistical arbitrage in the US equities market". In: *Quantitative Finance* 10 (7), pp. 761–782. ISSN: 14697688. DOI: 10.1080/14697680903124632.

Avellaneda, Marco and Jeong-Hyun Lee (2008). *Statistical Arbitrage in the U.S. Equities Market*.

Banerjee, Onureena, Laurent El Ghaoui, and Aspremon@princeton Edu (2008). "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data Alexandre d'Aspremont". In: *Journal of Machine Learning Research* 9, pp. 485–516.

Boyd, Stephen et al. (2010). "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends in Machine Learning* 3 (1), pp. 1–122. ISSN: 19358237. DOI: 10.1561/2200000016.

Caldeira, João Frois and Guilherme Valle Moura (2013). "Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy". In: *Brazilian Review of Finance* 11 (1), pp. 49–90.

Cardoso, Jose Vinicius De Miranda and Daniel P. Palomar (Nov. 2020). "Learning Undirected Graphs in Financial Markets". In: vol. 2020-November. IEEE Computer Society, pp. 741–745. ISBN: 9780738131269. DOI: 10.1109/IEEECONF51394.2020.9443573.

Chen, Cathy W.S. et al. (Jan. 2017). "Pair trading based on quantile forecasting of smooth transition GARCH models". In: *North American Journal of Economics and Finance* 39, pp. 38–55. ISSN: 10629408. DOI: 10.1016/j.najef.2016.10.015.

Chen, Huafeng et al. (2017). *Empirical Investigation of an Equity Pairs Trading Strategy*.

Cheng, Xixin, Philip L.H. Yu, and W. K. Li (Sept. 2011). "Basket trading under co-integration with the logistic mixture autoregressive model". In: *Quantitative Finance* 11 (9), pp. 1407–1419. ISSN: 14697688. DOI: 10.1080/14697688.2010.506445.

Cummins, Mark and Andrea Bucca (Dec. 2012). "Quantitative spread trading on crude oil and refined products markets". In: *Quantitative Finance* 12 (12), pp. 1857–1875. ISSN: 14697688. DOI: 10.1080/14697688.2012.715749.

d'Aspremont, Alexandre (Aug. 2007). "Identifying Small Mean Reverting Portfolios". In: URL: http://arxiv.org/abs/0708.3048.

Danaher, Patrick, Pei Wang, and Daniela M Witten (2014). "The joint graphical lasso for inverse covariance estimation across multiple classes". In: *Source: Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 76 (2), pp. 373–397. URL: https://www.jstor.org/stable/24772461.

Do, Binh and Robert Faff (2010). *Does Simple Pairs Trading Still Work?*, pp. 83–95.

— (June 2012). "Are pairs trading profits robust to trading costs?" In: *Journal of Financial Research* 35 (2), pp. 261–287. ISSN: 02702592. DOI: 10.1111/j.1475-6803.2012.01317.x.

Do, Binh, Robert Faff, and Kais Hamza (2006). *A New Approach to Modeling and Estimation for Pairs Trading.* URL: https://www.researchgate.net/publication/267721035.

Dong, Xiaowen et al. (June 2014). "Learning Laplacian Matrix in Smooth Graph Signal Representations". In: URL: http://arxiv.org/abs/1406.7842.

— (Dec. 2016). "Learning Laplacian Matrix in Smooth Graph Signal Representations". In: *IEEE Transactions on Signal Processing* 64 (23), pp. 6160–6173. ISSN: 1053587X. DOI: 10.1109/TSP.2016.2602809.

Dunis, Christian L et al. (2010). "Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities". In.

Elliott, Robert J., John Van Der Hoek, and William P. Malcolm (June 2005). "Pairs trading". In: vol. 5, pp. 271–276. DOI: 10.1080/14697680500149370.

Engelberg, Joseph, Pengjie Gao, and Ravi Jagannathan (2009). "An Anatomy of Pairs Trading: the Role of Idiosyncratic News, Common Information and Liquidity". In: *Third Singapore International Conference on Finance.* URL: http://ssrn.com/abstract=1330689Electroniccopyavailableat:https://ssrn.com/abstract=1330689Electroniccopyavailableat:http://ssrn.com/abstract=1330689.

Engle, Robert F and C W J Granger (1987). *Co-Integration and Error Correction: Representation, Estimation, and Testing*, pp. 251–276. URL: https://about.jstor.org/terms.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds*". In: *Journal of Financial Economics* 33, pp. 3–56.

— (2014). *A Five-Factor Asset Pricing Model.*

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (July 2008). "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9 (3), pp. 432–441. ISSN: 1465-4644. DOI: 10.1093/BIOSTATISTICS/KXM045. URL: https://academic.oup.com/biostatistics/article/9/3/432/224260.

Gatarek, Lukasz, Lennart F Hoogerheide, and Herman K Van Dijk (2014). *Return and Risk of Pairs Trading using a Simulation-based Bayesian Procedure for Predicting Stable Ratios of Stock Prices.* Tinbergen Institute. URL: http://www.tinbergen.nl.

Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst (Sept. 2006). “Pairs trading: Performance of a relative-value arbitrage rule”. In: *Review of Financial Studies* 19 (3), pp. 797–827. ISSN: 08939454. DOI: 10.1093/rfs/hhj020.

Girma, Paul Berhanu and Albert S. Paulson (1999). “Risk arbitrage opportunities in petroleum futures spreads”. In: *Journal of Futures Markets* 19 (8), pp. 931–955. ISSN: 02707314. DOI: 10.1002/(SICI)1096-9934(199912)19:8<931::AID-FUT5>3.0.CO;2-L.

Hou, Kewei, Chen Xue, and Lu Zhang (May 2020). “Replicating Anomalies”. In: *Review of Financial Studies* 33 (5), pp. 2019–2133. ISSN: 14657368. DOI: 10.1093/rfs/hhy131.

Huang, Chien Feng et al. (2015). “An intelligent model for pairs trading using genetic algorithms”. In: *Computational Intelligence and Neuroscience* 2015. ISSN: 16875273. DOI: 10.1155/2015/939606.

Jacobs, Heiko (Aug. 2015). “What explains the dynamics of 100 anomalies?” In: *Journal of Banking and Finance* 57, pp. 65–85. ISSN: 03784266. DOI: 10.1016/j.jbankfin.2015.03.006.

Jacobs, Heiko and Martin Weber (2011). *Losing sight of the trees for the forest? Pairs trading and attention shifts.*

Jegadeesh, Narasimhan and Sheridan Titman (2001). “Profitability of Momentum Strategies: An Evaluation of Alternative Explanations”. In: *Source: The Journal of Finance* 56 (2), pp. 699–720.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen (n.d.). *Is There a Replication Crisis in Finance?* URL: www.bryankellyacademic.org..

Kalofolias, Vassilis (Jan. 2016). “How to learn a graph from smooth signals”. In: URL: http://arxiv.org/abs/1601.02513.

Karakas, Oguzhan (2008). “Mean Reversion Between Different Classes of Shares in Dual-Class Firms: Evidence and Implications”. In: ISBN: 44(0)2070007. URL: http://ssrn.com/abstract=1343668at:https://ssrn.com/abstract=1343668Electroniccopyavai http://ssrn.com/abstract=1343668.

Keuzenkamp, Jelmer and Chen Zhou (2018). *Multivariate Pairs Trading Using Temporal Dependence Structures.*

Komodakis, Nikos and Jean-Christophe Pesquet (June 2014). “Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems”. In: URL: http://arxiv.org/abs/1406.5429.

Krauss, Christopher (2015). “Statistical Arbitrage Pairs Trading Strategies: Review and Outlook”. In: ISSN: 1867-6707.

Lake, Brenden M and Joshua B Tenenbaum (2010). “Discovering Structure by Learning Sparse Graphs”. In.

Mantegna, R N (1999). *Hierarchical structure in financial markets*, pp. 193–197. URL: www.forbes.com..

Mantegna, Rosario N (1998). “Hierarchical Structure in Financial Markets”. In.

Marti, Gautier et al. (Mar. 2017). “A review of two decades of correlations, hierarchies, networks and clustering in financial markets”. In: DOI: 10.1007/978-3-030-65459-7. URL: http://arxiv.org/abs/1703.00485%20http://dx.doi.org/10.1007/978-3-030-65459-7.

Meinshausen, Nicolai and Peter Bühlmann (June 2006). “High-dimensional graphs and variable selection with the Lasso”. In: *Annals of Statistics* 34 (3), pp. 1436–1462. ISSN: 00905364. DOI: 10.1214/009053606000000281.

Miranda Cardoso, José Vinícius de, Jiaxi Ying, and Daniel Perez Palomar (Dec. 2020). “Algorithms for Learning Graphs in Financial Markets”. In: URL: http://arxiv.org/abs/2012.15410.

Pal, Soumyasundar, Florence Regol, and Mark Coates (Oct. 2019). “Bayesian Graph Convolutional Neural Networks Using Non-Parametric Graph Learning”. In: URL: http://arxiv.org/abs/1910.12132.

Perlin, Marcelo (2007). *M of a kind: A Multivariate Approach at Pairs Trading.* URL: https://ssrn.com/abstract=952782.

Perlin, Marcelo Scherer (Aug. 2009). “Evaluation of pairs-trading strategy at the brazilian financial market”. In: *Journal of Derivatives and Hedge Funds* 15 (2), pp. 122–136. ISSN: 17539641. DOI: 10.1057/jdhf.2009.4.

Peters, Gareth W. et al. (2011). “Bayesian Cointegrated Vector Autoregression Models Incorporating ®-stable Noise for Inter-day Price Movements Via Approximate Bayesian Computation”. In: *Bayesian Analysis* 6 (4), pp. 755–792. ISSN: 19360975. DOI: 10.1214/11-BA628.

Ross, Stephen A (1976). *The Arbitrage Theory of Capital Asset Pricing.*

Rue, Havard and Leonhard Held (Feb. 2005). *Gaussian Markov Random Fields.* Chapman and Hall/CRC. ISBN: 9780203492024. DOI: 10.1201/9780203492024.

Slawski, Martin and Matthias Hein (May 2015). “Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields”. In: *Linear Algebra and Its Applications* 473, pp. 145–179. ISSN: 00243795. DOI: 10.1016/j.laa.2014.04.020.

Sun, Ying, Prabhu Babu, and Daniel P. Palomar (Feb. 2017). “Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning”. In: *IEEE Transactions on Signal Processing* 65 (3), pp. 794–816. ISSN: 1053587X. DOI: 10.1109/TSP.2016.2601299.

Tsay, Ruey S. (2010). *Analysis of financial time series.* Wiley, p. 677. ISBN: 9780470414354.

Vidyamurthy, Ganapathy (2004). “Pairs Trading: Quantitative Methods and Analysis”. In.

Wu, Zonghan et al. (Jan. 2019). “A Comprehensive Survey on Graph Neural Networks”. In: DOI: 10.1109/TNNLS.2020.2978386. URL: http://arxiv.org/abs/1901.00596%20http://dx.doi.org/10.1109/TNNLS.2020.2978386.

Yang, Jakub W Jurek Halla et al. (2007). *Dynamic Portfolio Selection in Arbitrage **. URL: http://ssrn.com/abstract=882536https://ssrn.com/abstract=882536.

Ying, Jiaxi et al. (2020). "Nonconvex Sparse Graph Learning under Laplacian Constrained Graphical Model". In: *Advances in Neural Information Processing Systems (NeurIPS)*. URL: https://github.com/mirca/sparseGraph..

Zou, Hui and Trevor Hastie (2005). "Regularization and variable selection via the elastic net". In: *J. R. Statist. Soc. B* 67 (2), pp. 301–320.

Appendix A

Empirical Distributions

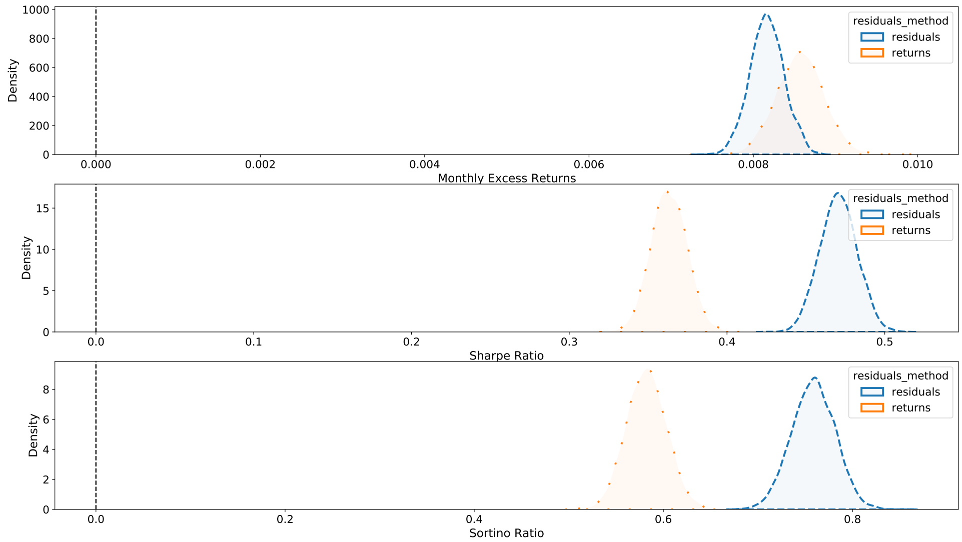

Figure A.1: Empirical distribution functions of the metrics of interest (from top to bottom: monthly excess returns, Sharpe ratio, Sortino ratio) grouped by the data preprocessing method, evaluated on the S&P500 dataset.

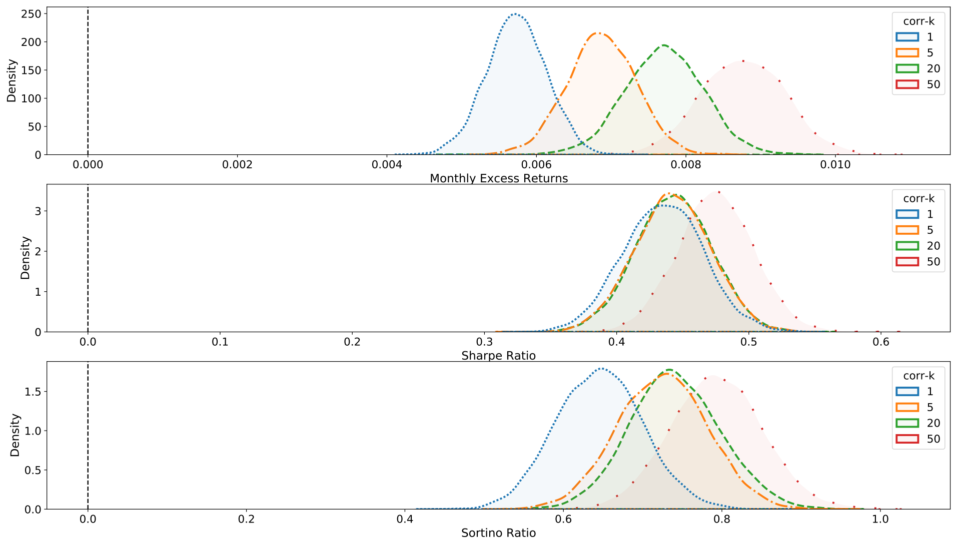

Figure A.2: Empirical distribution of the metrics of interest (from top to bottom: monthly excess returns, Sharpe ratio, Sortino ratio) for the Corr-K method grouped by the comoving portfolio size K, evaluated on the S&P500 dataset.

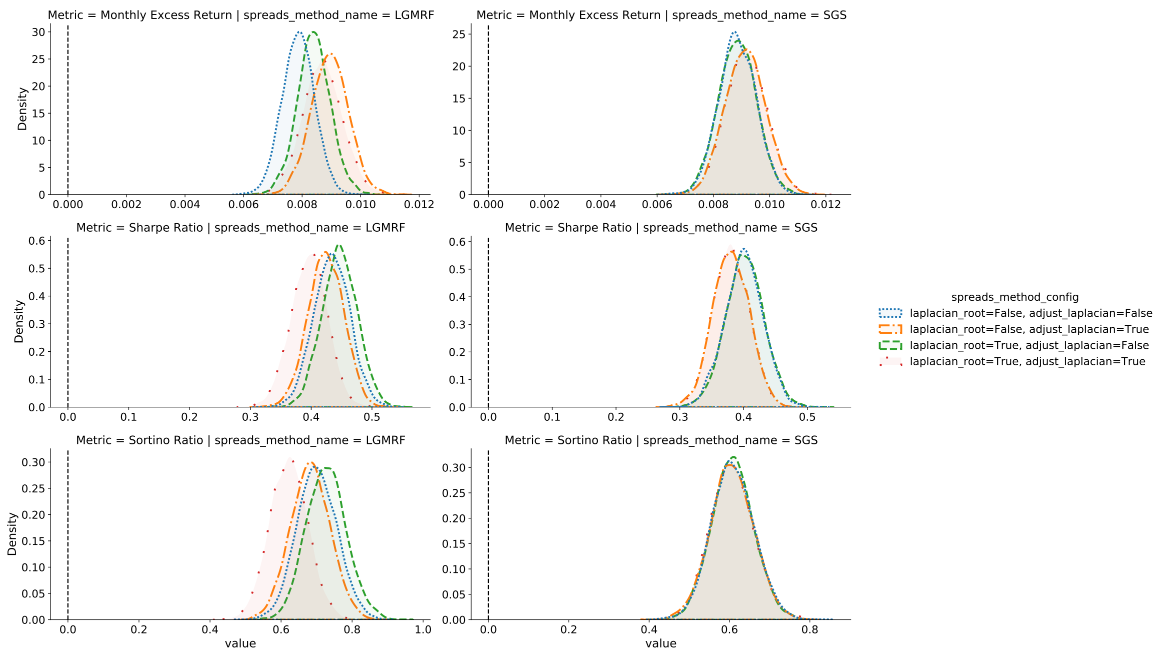

Figure A.3: Empirical distribution of the metrics of interest (from top to bottom: monthly excess returns, Sharpe ratio, Sortino ratio) for the sparse inverse covariance methods grouped by the configuration used, evaluated on the S&P500 dataset.

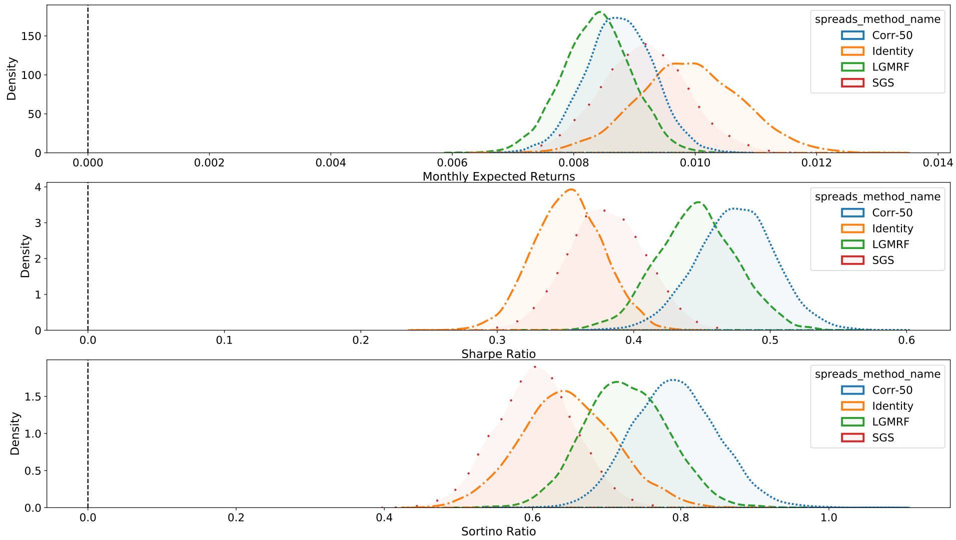

Figure A.4: Empirical distribution of the metrics of interest (from top to bottom: monthly excess returns, Sharpe ratio, Sortino ratio) grouped by the spreads method, evaluated on the S&P500 dataset. The configuration chosen for LGMRF is Laplacian root=True, adjust Laplacian=False. The configuration chosen for SGS is Laplacian root=True, adjust Laplacian=True.

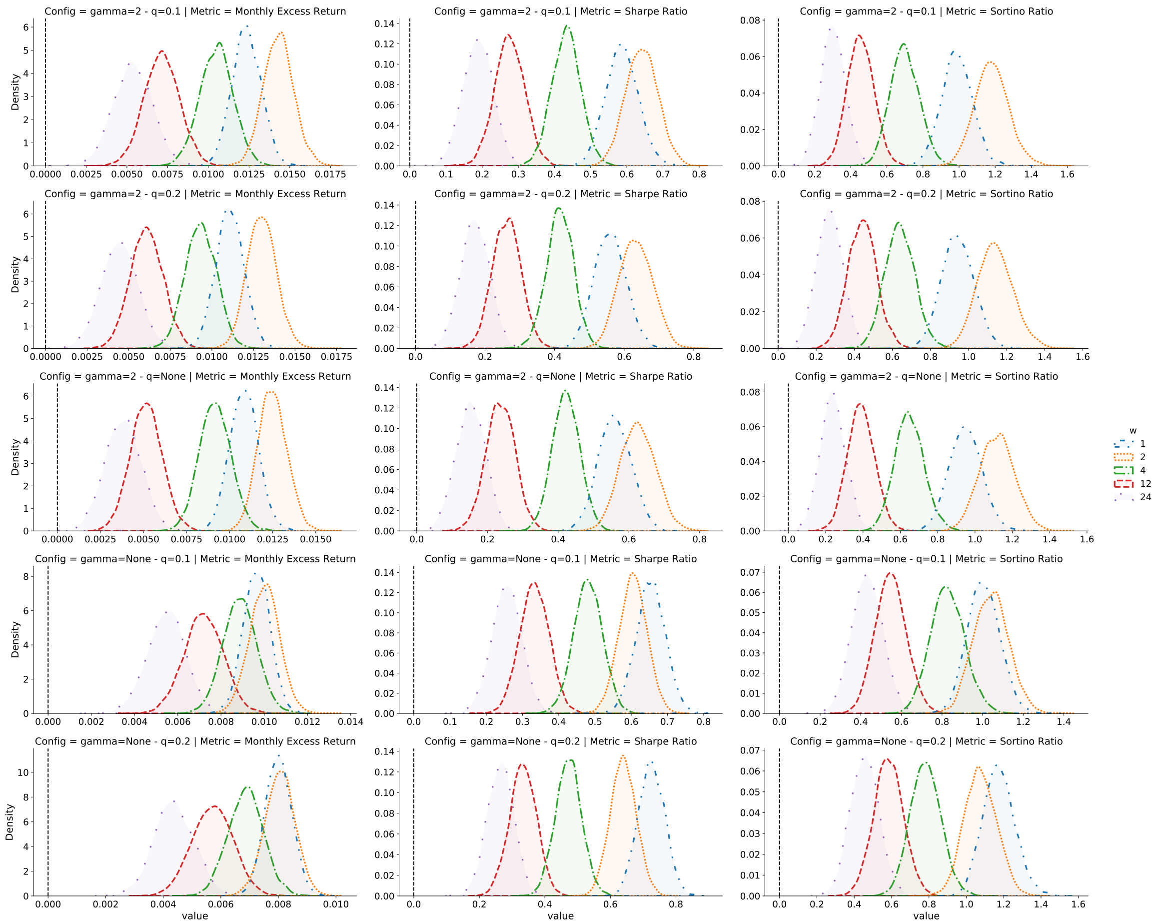

Figure A.5: Empirical distribution of the metrics of interest (from left to right: monthly excess returns, Sharpe ratio, Sortino ratio) grouped by the moving window w (hues and line styles) and the Quantiles-std hyper-parameters q, γ (rows), evaluated on the S&P500 dataset.

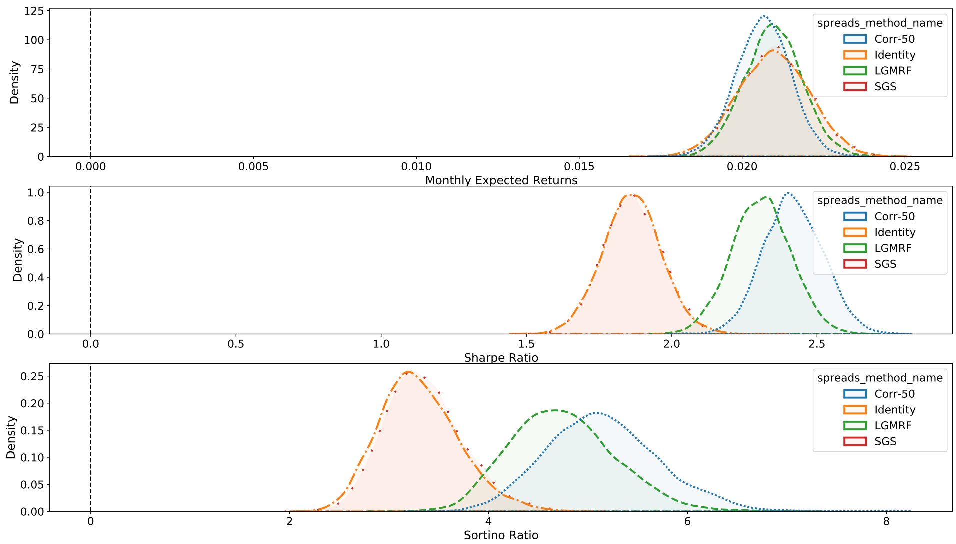

Figure A.6: Empirical distribution of the metrics of interest (from top to bottom: monthly excess returns, Sharpe ratio, Sortino ratio) grouped by the spreads method, evaluated on the CRSP dataset. The configuration chosen for LGMRF is Laplacian root=True, adjust Laplacian=False. The configuration chosen for SGS is Laplacian root=True, adjust Laplacian=True.

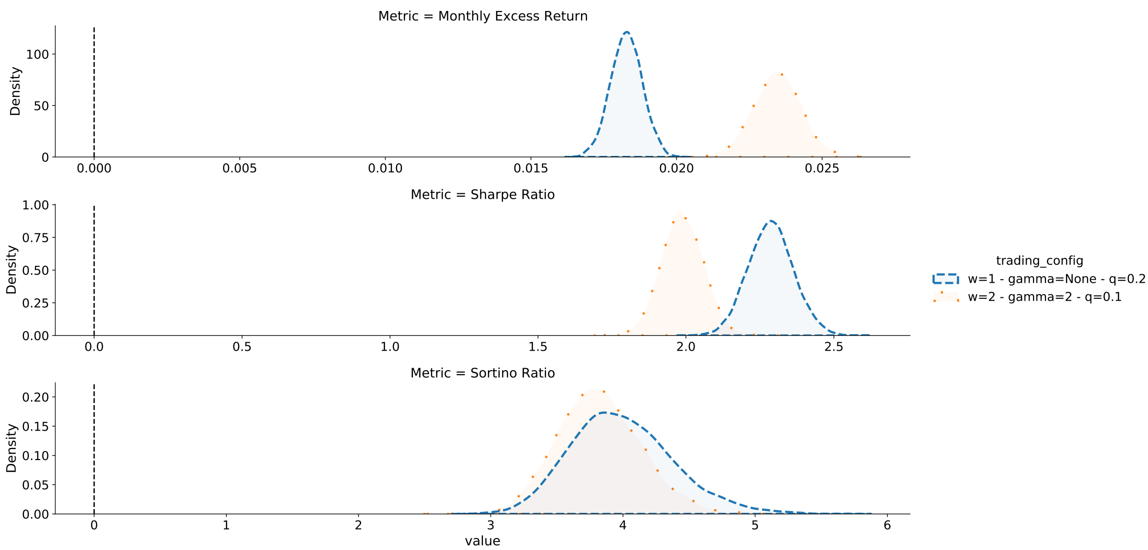

Figure A.7: Empirical distribution of the metrics of interest (from left to right: monthly excess returns, Sharpe ratio, Sortino ratio) grouped by the trading configuration, evaluated on the CRSP dataset.

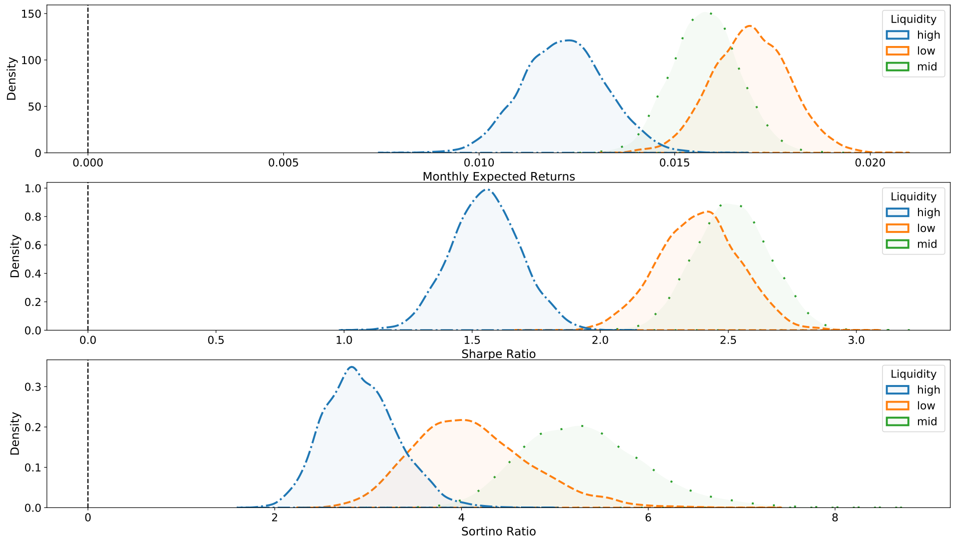

Figure A.8: Empirical distribution of the metrics of interest (from top to bottom: monthly excess returns, Sharpe ratio, Sortino ratio) grouped by liquidity level.

Appendix B

Performance Over Time

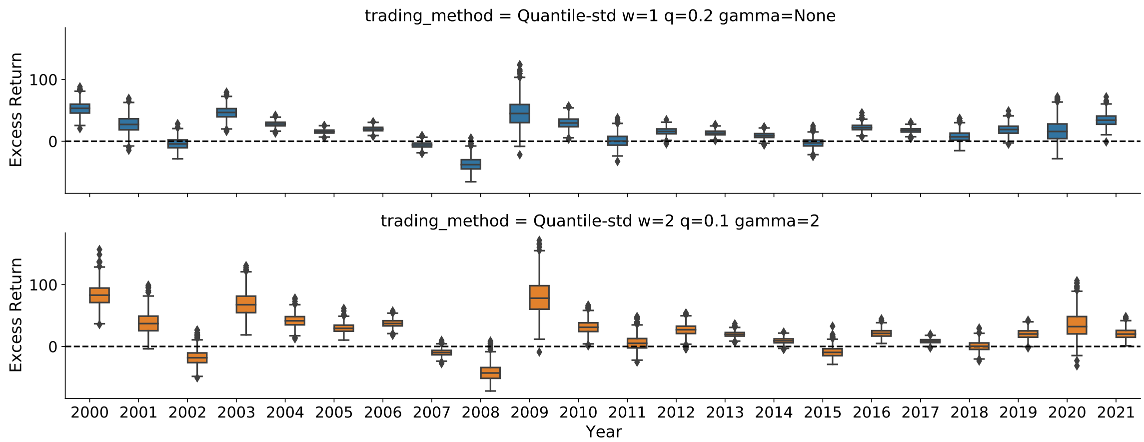

Figure B.1: Evolution of the excess return distribution from 2000 to 2021, grouped by trading rule, evaluated on the S&P500 dataset.

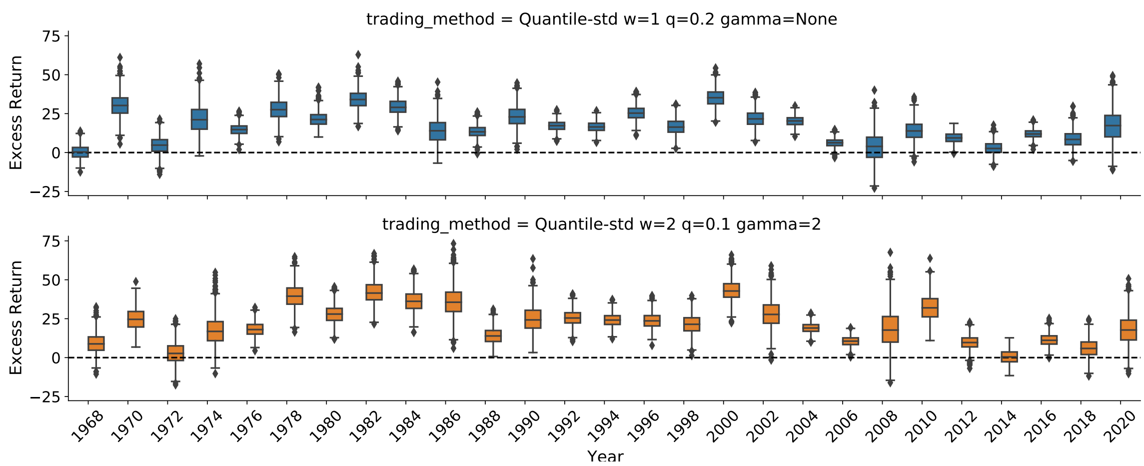

Figure B.2: Evolution of the excess return distribution from 2000 to 2021, grouped by trading rule, evaluated on the CRSP dataset.

Appendix C

Long vs Short Performance

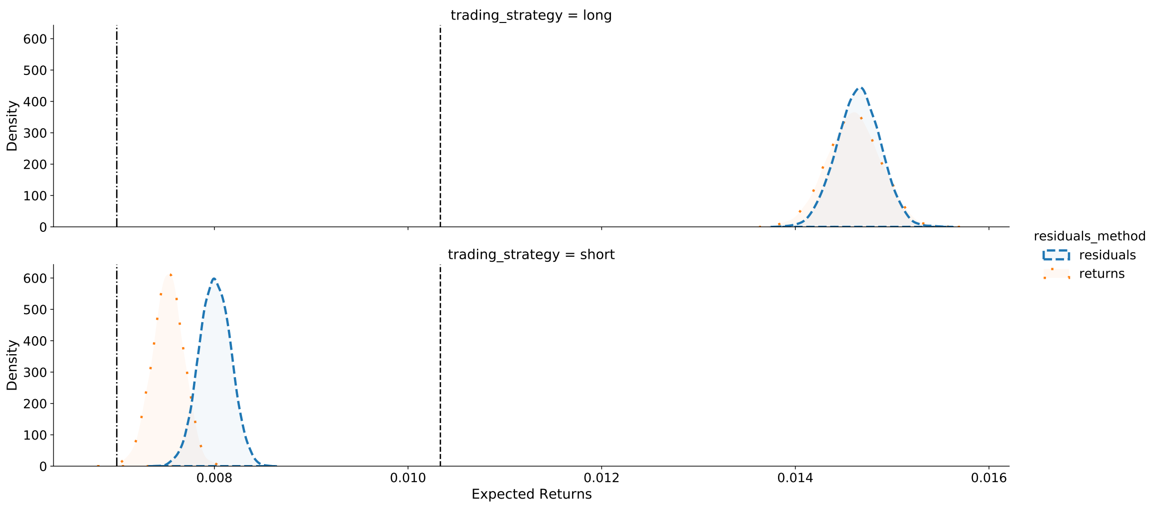

Figure C.1: Comparison between the expected return distribution of the long and the short leg, grouped by the data preprocessing method, evaluated on the S&P500 dataset.

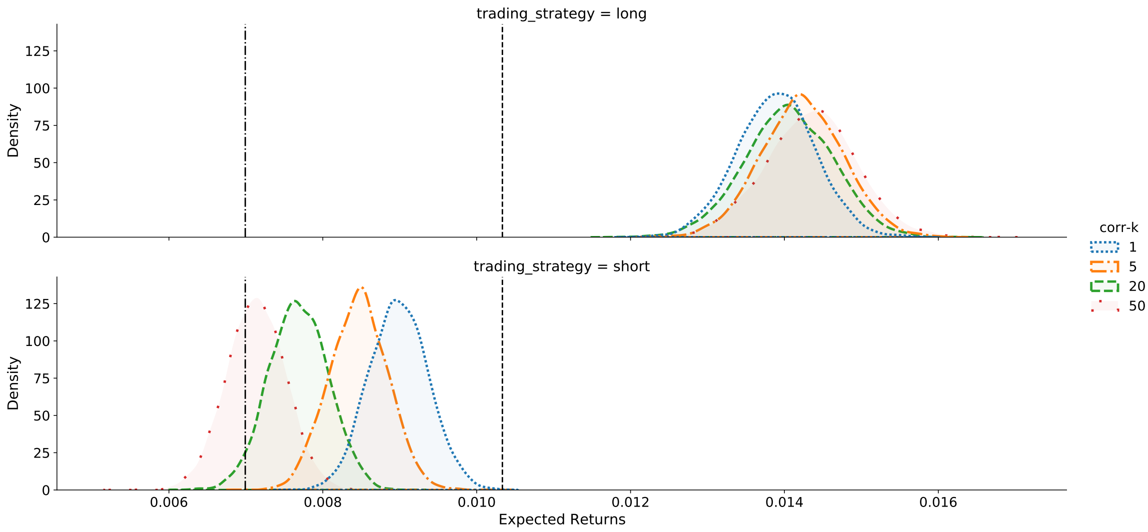

Figure C.2: Comparison between the expected return distribution of the long and the short leg, for the Corr-K method grouped by the comoving portfolio size K, evaluated on the S&P500 dataset

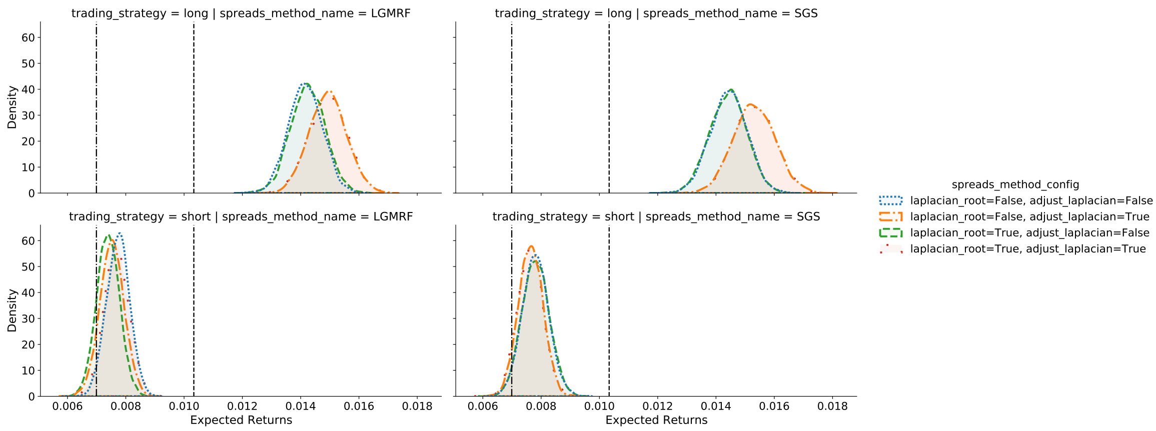

Figure C.3: Comparison between the expected return distribution of the long and the short leg, for the sparse inverse covariance methods grouped by the configuration used, evaluated on the S&P500 dataset.

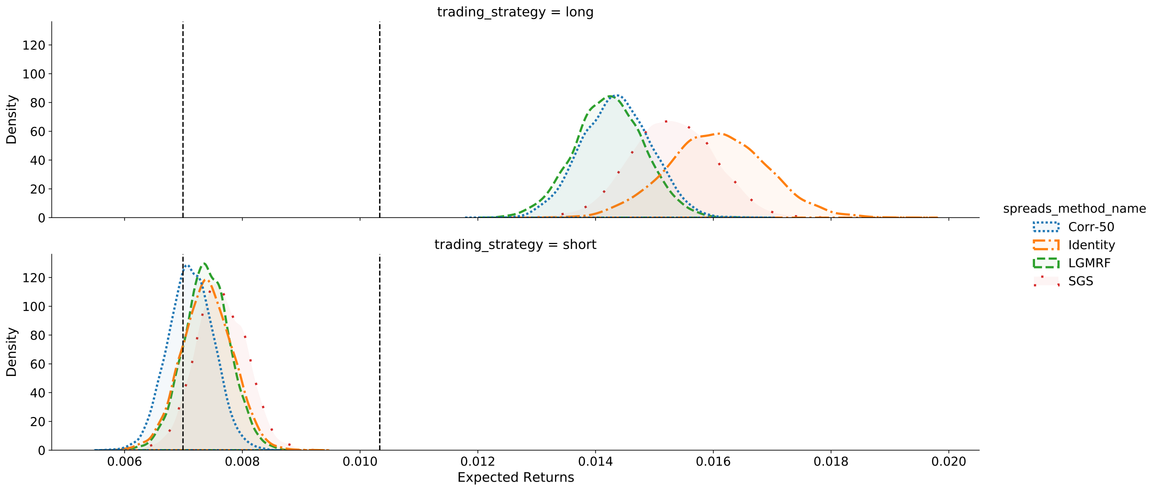

Figure C.4: Comparison between the expected return distribution of the long and the short leg, grouped by the spreads method, evaluated on the S&P500 dataset. The configuration chosen for LGMRF is Laplacian root=True, adjust Laplacian=False. The configuration chosen for SGS is Laplacian root=True, adjust Laplacian=True.

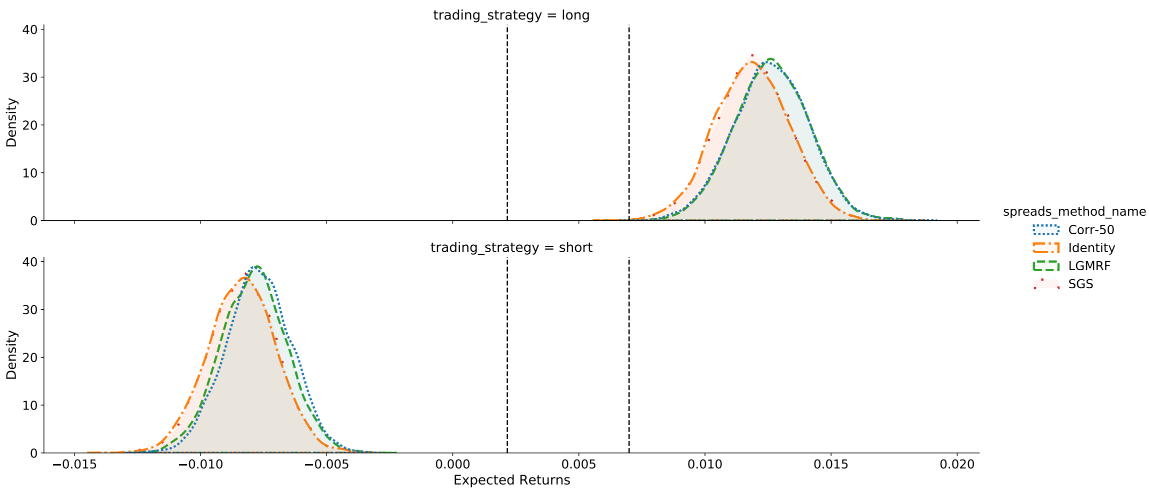

Figure C.5: Comparison between the expected return distribution of the long and the short leg, grouped by the spreads method, evaluated on the CRSP dataset. The configuration chosen for LGMRF is Laplacian root=True, adjust Laplacian=False. The configuration chosen for SGS is Laplacian root=True, adjust Laplacian=True.