# ADVANCES IN COINTEGRATION AND SUBSET CORRELATION HEDGING METHODS

**Marcos López de Prado**
Head of Global Quantitative Research, Tudor Investment Corporation
Research Affiliate, Lawrence Berkeley National Laboratory
lopezdeprado@lbl.gov

**David Leinweber**
Head of CIFT, Lawrence Berkeley National Laboratory
dleinweber@lbl.gov

First draft: May 2008
Last update[1]: January 2012

***Journal of Investment Strategies, Vol. 1, No. 2, pp. 67-115, Spring 2012***

[1] We thank the editor, Arthur Berd, as well as two anonymous referees for useful comments. We also thank Tudor Investment Corp., David Bailey, John Wu and our CIFT colleagues (Lawrence Berkeley National Laboratory), Maureen O'Hara, David Easley and Robert Jarrow (Cornell University), Paul Glasserman (Columbia University), Luis Viceira (Harvard Business School), Eva del Pozo (Complutense University), José Blanco (UBS, Univ. Zurich), Attilio Meucci (Kepos Capital, Baruch College), Riccardo Rebonato (PIMCO, University of Oxford), Alexandre d'Aspremont (Princeton University) and Yin Luo (Deutsche Bank) for their help and encouragement.

# ADVANCES IN COINTEGRATION AND SUBSET CORRELATION HEDGING METHODS


## ABSTRACT

We divide hedging methods between single-period and multi-period. After reviewing some well-known hedging algorithms, two new procedures are introduced, called *Dickey-Fuller Optimal* (DFO), *Mini-Max Subset Correlation* (MMSC). The former is a multi-period, cointegration-based hedging method that estimates the holdings that are most likely to deliver a hedging error absent of unit root. The latter is a single-period method that studies the geometry of the hedging errors and estimates a hedging vector such that subsets of its components are as orthogonal as possible to the error. We test each method for stability and robustness of the derived hedged portfolio. Results indicate that DFO produces estimates similar to the Error Correction Method, but more stable. Likewise, MMSC estimates are similar to Principal Component Analysis but more stable. Finally, a generalized Box-Tiao Canonical Decomposition (BTCD) method is proposed, which is of the multi-period class. BTCD estimates are also very stable, and cannot be related to any of the aforementioned methodologies. Finally, we find that all three advanced hedging methods (MMSC, BTCD, DFO) perform well.



**Keywords**: Hedging portfolios, robustness, portfolio theory, stationarity, Maeloc spread, ECM, DFO, PCA, BTCD, MMSC.

**JEL codes**: C01, C02, C61, D53, G11.

## 1.- INTRODUCTION

Hedging is an important risk management technique that is widely used in virtually all capital markets activities, from investment management and trading to market making and derivatives structuring. In all of these situations, an investor/trader is holding a portfolio of securities whose risk, defined as fluctuations of its market value, must be maintained within acceptable limits.

The purpose of hedging is to reduce a portfolio's exposure to a certain source of risk. Closing the positions responsible for that risk source is not always possible, either due to liquidity constraints or because that would impact the portfolio's exposure to other desired risk sources.

Hedging is also an inextricable part of an alpha generating strategy, such as pairs trading, equity market neutral, long-short and most "hedge" fund trading styles (López de Prado and Rodrigo (2004)). In that context, hedging involves removing the exposures on which we have no forecasting power, while leveraging our capital on those exposures over which we have a skill.

Portfolio replication may also be viewed as an application of hedging methods, for its goal is to define a mimicking portfolio with minimum tracking error. The methods discussed henceforth can be used to *"summarize"* or reduce the dimension of a portfolio into its core components (even if those are unknown ex-ante).

After reviewing a few well-known methods, we take the opportunity to extend or generalize some of them. Specifically, we introduce a generalized PCA method, applicable to any dimension or asset class. A generalized Box-Tiao Canonical Decomposition (BTCD) procedure is presented, which accepts any number of lags, regressors and forecasting horizons in the specification of its VAR system.

Although unit root tests have been used to assess the quality of a hedge (like in Vidyamurthy (2004)), we believe that this study is the first to propose a procedure for computing a Dickey-Fuller (DF) optimal hedging strategy, whereby the unit root test statistic is the objective function. Based on our direct estimation of the DF statistic, we develop the DFO hedging methodology.

It seems intriguing that some of the most applied hedging methods happen to be among the most unreliable. Regression approaches in particular are known to deliver unstable results, and yet they are ubiquitous (e.g., CAPM, APT and stocks' betas). Besides their simplicity, a possible explanation may be that they search for a "concrete" solution (as opposed to "hidden" factors analysis such as PCA or BTCD). A good compromise would consist on developing a regression-like analysis that imposes a strong structure with the aim of improving the hedge's stability. This goal motivates our new Mini-Max Subset Correlation (MMSC) model, as well as the concept of *Maeloc* spread.

Robustness is a key characteristic of a good hedging procedure. Its absence indicates that the solution is either unstable or arbitrary. An unstable method delivers significant hedging errors and substantial transaction costs associated with its rebalance. In the empirical part of this study we will analyze how robust each procedure is, outlining which method should be preferred among those comparable.

Finally, among those methods that are distinct and stable, we will indicate which ones perform best in different hedging horizons.

This paper is organized as follows. Section 2 formalizes the hedging problem. Section 3 proposes a taxonomy of hedging methods. Section 4 reviews some of the most applied hedging methods, pointing out their virtues and pitfalls. Section 5 describes advanced hedging methods, developed to address the caveats of the traditional approaches. Section 6 applies the hedging methodologies to pairs of index futures, indicating which should be preferred in terms of stability and performance over different horizons. Section 7 outlines the conclusions. The Appendices complement the mathematical apparatus involved in these methods.

## 2.- THE HEDGING PROBLEM

From a portfolio management perspective, the hedging problem is posed in the following terms[2]. Let $P_{1,t}$ represent the market value at observation *t* of a portfolio we wish to hedge, with *t=1,…,T*. $\Delta P_{1,t}$ is the change in market value between observation *t-1* and *t*, induced by the risk drivers we intend to hedge against. Provided a set of *n=2,…,N* variables (instruments or portfolios) available for building a hedge, the hedging problem consists in computing the vector of holdings $\boldsymbol{\omega}$ that is optimal according to a particular method. The market value of the combined position of portfolio plus hedge,

$$S_t = P_{1,t} + \sum_{n=2}^{N} \omega_n P_{n,t} \tag{1}$$

is denoted *spread*, and

$$e(h) = S_{T+h} - S_T \tag{2}$$

is the hedging error after *h* observations. Generally speaking, portfolio managers fear the case that $e(h)$ is *non-stationary* in variance, because in that scenario the hedging error is unbounded. A special case of non-stationarity occurs when $e(h)$ has a unit root, in which situation the hedging errors follow what is commonly known as a *random walk*. This can be dealt with if we are able to find a hedging portfolio that, combined with $P_1$, makes $e(h)$ stationary. Then that hedging portfolio is said to be *cointegrated* to $P_1$, and finding it is the goal of several dynamic methods that we will study.

We will assume that the set of instruments to form the hedge is predefined. If that is not the situation faced by the modeler, the selection could be done applying a standard factor selection algorithm[3] on the hedging procedures presented.

---

[2] Other possible standpoints, such as credit risk, liquidity risk or inventory management, are beyond the scope of this paper.

[3] For instance, a forward algorithm will simply require an evaluation criterion, such as the $R^2$ in the regression case, minimum variance in the minimum risk case, residual unexplained variance in the PCA or BTCD cases, or DF stat in that analysis.

## 3.- A TAXONOMY OF HEDGING METHODOLOGIES

We can divide hedging methods in two classes:

1. Single-period, distributional-based or static methods: They make certain assumptions about the distribution of the portfolio and security returns. Their objective functions are formulated in terms of minimizing or controlling certain distributional parameters. In particular, they assume that random perturbations are IID. Such assumption is very convenient and is ubiquitous in the Financial literature about portfolio theory. However, there is overwhelming evidence that returns are serially-dependent, particularly in the high frequency domain (see Easley, López de Prado and O'Hara (2012) for a discussion).
2. Multi-period, equilibrium or dynamic methods: Not assuming IID random perturbations requires dealing with the cumulative hedging error. This problem is addressed by multi-period hedging methods, which in turn require making certain assumptions regarding the serial dependence of the returns. Their objective functions are defined in terms of minimizing the cumulative hedging error that results from such dynamics.

One type of hedging method is not necessarily superior to the other. The choice for a class and particular method will largely depend on the hedging horizon, the dimension of the portfolio and the statistical properties of the instruments' returns, among other factors.

The approaches discussed in this paper are numerous, and for convenience we must define a few acronyms:

1. Single-period, distributional-based or static hedging methods:
   - OLSD: Ordinary Least Squares in Differences.
   - MVP: Minimum Variance Portfolio.
   - PCA: Principal Components Analysis.
   - ERC: Equal Risk Contribution.
   - MDR: Maximum Diversification Ratio.
   - MMSC: Mini-Max Subset Correlation.
2. Multi-period, equilibrium or dynamic hedging methods:
   - OLSL: Ordinary Least Squares in Levels.
   - ECM: Error Correction Model.
   - DFO: Dickey-Fuller Optimal.
   - BTCD: Box-Tiao Canonical Decomposition.

## 4.- A REVIEW OF EXISTING HEDGING ALGORITHMS

We will start by reviewing some of the best known hedging methodologies. They incorporate multiple concepts from APT, portfolio replication, time series analysis, "modern" portfolio theory, spectral theory and canonical analysis among other fields.

### 4.1.- SINGLE-PERIOD METHODS

#### 4.1.1.- OLS IN DIFFERENCES (OLSD)

Despite its limitations, this is one of the most widely used methods (Moulton and Seydoux (1998)), perhaps because of its simplicity.

The regression is specified as $\Delta P_{1,t} = \alpha + \sum_{n=2}^{N} \beta_n \Delta P_{n,t} + \varepsilon_t$, where $P_n$, $n=1,\ldots, N$ are market values of the $n$-th position and $n=1$ corresponds to the portfolio we want to hedge. A necessary condition for the hedge to be effective is that the drift ($\alpha$) is statistically insignificant. The goodness of the fit can be evaluated through the adjusted $R^2$, and the solution is $\omega_n = -\beta_n$, $n=2,\ldots,N$.

In summary, this approach may be applied under the conditions that $\alpha \approx 0$ and $\varepsilon$ is *IID*, with $\varepsilon \rightarrow N(0, \sigma_\varepsilon^2)$, $E\varepsilon_t\varepsilon_s = 0, \forall t \neq s$. This is extremely restrictive, and a pitfall common to single-period methods, as they assume that any change in the $P_{1,t}$ portfolio must be *synchronously* offset by the hedging portfolio, $\sum_{n=2}^{N} \omega_n P_{n,t}$. It will not suffice to establish the stationarity of $\Delta P_{1,t} - \sum_{n=2}^{N} \beta_n \Delta P_{n,t}$, for that would not prevent $S_t = P_{1,t} + \sum_{n=2}^{N} \omega_n P_{n,t}$ from following a random walk[4]. In other words, this model fails to impose any condition on the behavior of the cumulative hedging errors, $e(h) = S_{T+h} - S_T$, implying that hedging errors may not be corrected over time. This is a direct consequence of the specification in differences, which has removed all memory of the process. This approach is also somewhat arbitrary, as switching places between the portfolio and one of the hedging constituents may lead to vectors $\boldsymbol{\omega}$ in different directions.

Among other reasons, these three critiques (restrictiveness, absence of error correction, arbitrarity) make the regression of differences an undesirable hedging method.

**4.1.2.- MINIMUM VARIANCE PORTFOLIO (MVP)**

First introduced by Markowitz (1952), it consists in solving the basic quadratic optimization problem, with a single linear constraint in equality[5]. Its popularity has grown ever since, with studies as recent as Clarke, de Silva and Thorley (2011) or Scherer (2010).

$\Delta P$ observations are assumed to be *IID* Normal. Let $V$ be the covariance of matrix $\Delta P$, where the first column represents the covariances against the portfolio to be hedged. $V$ must be invertible, thus steps should be taken to prevent singularity (Stevens, 1998).

$$\begin{aligned} &\underset{\beta}{Min} \quad \beta' V \beta \\ &s.t. \quad \beta' a = 1 \end{aligned} \tag{3}$$

This program can be solved through the lagrangian $L(\beta, \lambda) = \frac{1}{2}\beta' V \beta - \lambda(\beta' a - 1)$, with first order conditions

$$\begin{aligned} F1 &: \frac{\partial L(\beta,\lambda)}{\partial \beta} = V\beta - \lambda a = 0 \\ F2 &: \frac{\partial L(\beta,\lambda)}{\partial \lambda} = \beta' a - 1 = 0 \end{aligned} \tag{4}$$

[4] $S_t$ could be I(1).

[5] A constraint is needed to exclude the zero-holding solution.

Operating, $F1 \rightarrow \beta = \lambda V^{-1}a$ and $F2 \rightarrow \beta' a = a' \beta = 1; \lambda a' V^{-1} a = 1 \Rightarrow \lambda = \frac{1}{a' V^{-1} a}$.

Thus, $\beta = \frac{V^{-1}a}{a' V^{-1} a}$, and $\omega_j = \frac{\beta_j}{\beta_1}$, *j=1,…,N*, to meet the constraint of unit holding of the hedged portfolio (first column of the covariance matrix *V*).

We can verify that we have indeed computed the minimum through the second order condition.

$$\begin{vmatrix} \frac{\partial^2 L(\beta,\lambda)}{\partial \beta^2} & \frac{\partial^2 L(\beta,\lambda)}{\partial \beta \partial \lambda} \\ \frac{\partial^2 L(\beta,\lambda)}{\partial \lambda \partial \beta} & \frac{\partial^2 L(\beta,\lambda)}{\partial \lambda^2} \end{vmatrix} = \begin{vmatrix} V' & -a' \\ a & 0 \end{vmatrix} = a' a \geq 0 \quad (5)$$

This is the general convex minimization program for computing characteristic portfolios, of which MVP is the class that results from setting *a* equal to a vector of 1s (Grinold and Kahn, 1999).[6] The solution corresponds to the portfolio on the left-most point of the efficient frontier. An empirical study of the performance of MVPs on stocks can be found in Luo et al. (2011). This approach presents similar caveats as the regression of differences (OLSD).

It is worth noting that the MVP method delivers the minimum risk solution under the assumption of Normality, but beyond that assumption a number of alternative objective functions could be chosen. This would lead to CVaR and Cornish-Fisher related methods, to name only a couple.

### 4.1.3.- PRINCIPAL COMPONENTS ANALYSIS (PCA)

Steely (1990) and Litterman and Sheinkman (1991) were among the first to see the potential applications of the eigendecomposition of variance to hedging. Their analysis focused on explaining how common factors affect bond returns, which in the case of the term structure of interest rates they identified as parallel shift, slope and convexity (Lord and Pelsser, 2007). This was later applied by Moulton and Seydoux (1998) to construct portfolios of 3 bonds hedged against the first two principal components (parallel shifts and slope changes). In this paper we will generalize that analysis to portfolios of any size, without restricting its use to the term structure of interest rates.

$\Delta P$ observations are assumed to be *IID* Normal. Let *V* be the *NxN* covariance of matrix $\Delta P$, where the first column represents the covariances against the portfolio to be hedged. The target is to compute the vector of weightings $\boldsymbol{\beta}$ such that $\Delta P \beta$ is hedged against moves of the *m* largest principal components (typically, *m=N-1*), leaving the combined position solely exposed to moves of the *N-m* components with lowest variances (eigenvalues). In other words, we wish to compute a *N*-vector $\boldsymbol{\beta}$ such that $W^{*\prime}\beta = 0_m$, where $W^{*\prime}$ is the transposed eigenvector matrix after having

[6] When working with returns, this is the 'fully invested portfolio constraint'. It is somewhat arbitrary to choose a vector of 1s in the context of optimal holdings, but any other non-null number will simply re-scale the solution. An equivalent approach would be fixing the weight of a portfolio constituent (numeraire) and minimizing the overall risk without imposing a constraint of the sum of weights.

removed the columns associated with the unhedged eigenvectors, and $\beta_i = 1, \forall i > m$. In order to explain why, we must describe a few matrix operations.

$W^{*\prime}\beta = 0_m$ is an homogeneous system with infinite non-trivial solutions, because $Rank[W^{*\prime}] = m < N$. In order to find a single solution, we impose $\beta_i = 1$ on the $N - m$ last columns. Let $W^{*'*}$ be the *mxm* matrix which results from moving the last $N - m$ columns (numeraires) from $W^{*\prime}$ to the right side of the equation, which we denote $W^{*'-*}$.

$\beta^*$ is the submatrix of $\boldsymbol{\beta}$ that excludes those $\beta_i = 1, \forall i > m$. This leads us to express the problem as $W^{*'*}\beta^* = -W^{*'-*}I_{N-m}$. For $-W^{*'-*}I_{N-m} \neq 0_m$, the solution is unique and non-trivial, which can be computed as $\beta^* = -\left[W^{*'*}\right]^{-1}W^{*'-*}I_{N-m}$. Finally, the holdings are obtained as $\omega_j = \frac{\beta_j}{\beta_1}$, *j=1,…,N.*

This approach presents the advantage of searching for a solution which hedges against the principal sources of risk. Like the prior two methods, it doesn't guarantee that the source of risk we remain exposed to is stationary[7]. It could be argued however that, having the smallest variance (in differences), the stationarity of the eigenvectors with smallest eigenvalues is a minor concern[8]. This makes of PCA a valid, consistent method of hedging.

## 4.2.- MULTI-PERIOD METHODS

### 4.2.1.- OLS IN LEVELS (OLSL)

The target is to solve $P_{1,t} = \sum_{n=2}^{N}\beta_n P_{n,t} + S_t$, with $\omega_n = -\beta_n$, *n=2,…,N*. The hedge is effective as long as *S* is stationary in mean and variance, which can be tested through KPSS or unit root tests (ADF, PP).

OLSL may not be considered a hedging procedure by itself, but a methodology that assesses whether the results from a regression of levels can be applied as a hedge. The reason is, the outcome of the ADF or KPSS style-test is not used to determine the vector $\boldsymbol{\omega}$, but rather to determine with what confidence we may assume that the hedging errors are stationary. This method has the additional disadvantage that, because the error correction component is not separated from the observed levels in the equation, the $\boldsymbol{\beta}$ may not accurately capture the equilibrium relationship. That inconvenience is formally addressed by the ECM.

### 4.2.2.- ERROR CORRECTION MODEL (ECM)

Engle and Granger (1987) show that if two series are cointegrated, there must exist an error correction representation, and conversely, if an error correction representation is verified, the two series are cointegrated. Following Gosh (1993) among others, the procedure consists on solving a dynamic equilibrium system between the portfolio that we wish to hedge and a hedging portfolio, estimated through a regression

---

[7] The components we remain exposed to have the smallest variance *in differences*. Once again, this doesn't imply that the components are stationary in levels, as they could be I(1).

[8] Alternatively, $V$ could have been estimated on $P$ rather than $\Delta P$, provided that the elements of $P$ are stationary, which generally is not the case.

$$\Delta p_{1,t} = \beta_0 + \beta_1 \Delta p_{2,t} + \gamma(p_{2,t-1} - p_{1,t-1}) + \varepsilon_t \tag{6}$$

where $p_1$, $p_2$ are the natural logarithms of market values $P_1$, $P_2$, and $\gamma$ must be tested to be positive ($H_0\colon \gamma \leq 0$). The spread is characterized by the holdings $(\omega_1, \omega_2) = (1, -K)$, where $K = e^{\frac{\beta_0}{\gamma}}$ (see Appendix 1 for its derivation). As originally formulated, the approach is limited to only two variables, although an extension could be built upon Johansen (1991). We do not see a need for that, as that approach would not be in practice substantially dissimilar from the BTCD and DFO methods, discussed later.

## 5.- ADVANCED HEDGING METHODS

We are now in a position to discuss several approaches that overcome some of the limitations listed earlier.

### 5.1.- MULTI-PERIOD METHODS

#### 5.1.1.- BOX-TIAO CANONICAL DECOMPOSITION (BTCD)

Box and Tiao (1977) introduced a canonical transformation of a *N*-dimensional stationary autoregressive process. The components of the transformed process can then be ordered from least to most predictable. The authors' original intent was not to produce a new hedging method, however their discovery can be adapted to this purpose. In short, the objective is to come up with the matrix of coefficients that deliver a vector of forecasts with the most predictive power over the next observation. To understand this procedure, it is best to start with a single-equation, two-dimensional example, i.e. *AR(1)* specification, and then move up to a multi-equation, multi-dimensional (or *VAR(L)*, where *L* is the number of lags) specification.

In the AR(1) case, $P_t = \beta P_{t-1} + \varepsilon_t$, and $E[P_t^2] = E[(\beta P_{t-1})^2] + E[\varepsilon_t^2]$. Box-Tiao defined a measure of *predictability*, $\lambda = \frac{E\left[(\beta P_{t-1})^2\right]}{E\left[P_t^2\right]} = 1 - \frac{E\left[\varepsilon_t^2\right]}{E\left[P_t^2\right]}$, as a proxy for the mean reversion parameter of the Orstein-Uhlenbeck (O-U) stochastic process. When $\lambda$ is small, $E\left[\varepsilon_t^2\right]$ dominates $E_{t-1}\left[P_t^2\right]$ and $P_t$ is almost pure noise. When $\lambda$ is large, $E_{t-1}\left[P_t^2\right]$ dominates $E\left[\varepsilon_t^2\right]$ and $P_t$ is almost perfectly predictable. This makes the connection between ECM and BTCD in the two-dimensional case evident, as O-U processes are a continuous time representation of discrete-time mean-reverting processes.

Let's move now to the VAR(1) specification. This is an system of AR(1) equations on each time series of a set of variables, *n=1,…,N*, where *n=1* corresponds to the portfolio to be hedged.

$$P_{t,n} = \sum_{i=1}^{N} \beta_{i,n} P_{t-1,i} + \varepsilon_{t,n} \tag{7}$$

Because the explanatory variables are the same in each equation, the Multi-equation Least Square is equivalent to the Ordinary least squares (OLS) estimator applied to each equation separately, as shown by Zellner (1962). We can fit the model[9] for the entire set: $\hat{\beta} = \left(P_{t-1}' P_{t-1}\right)^{-1} P_{t-1}' P_t$ . We can derive a similar measure of predictability for

[9] Select only those statistically significant regressors, following a *stepwise* algorithm.

linear combinations of $P_t$. Rewriting, $P_t\Omega = P_{t-1}\beta\,\Omega + \varepsilon_t\Omega$. The series' predictability is then characterized as $\lambda_\Omega = \frac{\Omega'\beta'\Gamma\beta\Omega}{\Omega'\Gamma\Omega}$,[10] where $\Gamma = P_t'P_t$.[11]

We would like to compute a *Nx1* vector $\Omega$ such that $\lambda_\Omega$ is minimized, i.e. $\underset{\Omega}{Min}\ \frac{\Omega'\beta'\Gamma\beta\Omega}{\Omega'\Gamma\Omega}$. This is equivalent to solving the *generalized eigenvalue problem* in $\lambda_\Omega \in \Re$ characterized by $\det(\lambda_\Omega\Gamma - \beta'\Gamma\beta) = 0$.[12]

A closer examination of the ratio $\frac{\Omega'\beta'\Gamma\beta\Omega}{\Omega'\Gamma\Omega}$ leads us to treat it as a *generalized Rayleigh quotient* of the form $R(A,B;x) := \frac{x'Ax}{x'Bx}$, where $A = \beta'\Gamma\beta$ and $B = \Gamma$ are real symmetric positive-definite matrices and *x* is a given non-zero vector. We can reduce it to the standard *Rayleigh quotient* through the change of variables $z = Cx$ and $D = (C^{-1})'AC^{-1}$, where *C* is the *Cholesky decomposition* of matrix *B*. This approach is useful, because we know that a standard *Rayleigh quotient* such as $R(D,z) := \frac{z'Dz}{z'z}$ reaches its minimum value (the smallest eigenvalue) when *z* equals the eigenvector corresponding to the smallest eigenvalue of *D*. For a succinct proof of this, consider $\begin{matrix} \underset{x}{Max} & x'Ax \\ s.t. & x'x = 1 \end{matrix}$, where *A* is symmetric. Take derivatives on its Lagrangian $L(x) = x'Ax + \tilde{\lambda}(x'x - 1)$. The first order necessary condition, $\frac{\partial L(x)}{\partial x} = x'(A + A') + 2\tilde{\lambda}x' = 0 \Rightarrow Ax = \tilde{\lambda}x$, and thus the Lagrange multiplier is an eigenvalue. $\frac{\partial L(x)}{\partial \lambda} = 0 \Rightarrow x'x = 1$. Furthermore, $x'Ax = x'\lambda x = \lambda$, thus all *critical points* (and *extreme values* in particular) are derived from computing the eigenvectors of *A*, and the *stationary values* from the respective eigenvalues. The same argument can be used to find the maximum value of $R(D,z)$.

Assuming that $\Gamma$ is positive definite, the solution is $\Omega^* = \Gamma^{-1/2}z$, where *z* is the eigenvector corresponding to the smallest eigenvalue of the matrix $\Gamma^{-1/2}\beta'\Gamma\beta\Gamma^{-1/2}$.[13]

---

[10] This can also be interpreted as a *mean reversion* coefficient. The smaller, the stronger the trend (and more predictable). The larger, the noisier (and more unpredictable).

[11] Alternatively, $\Gamma$ can be defined as a covariance matrix of $P_t$.

[12] This is derived from rearranging $\Omega'[\lambda_\Omega\Gamma - \beta'\Gamma\beta]\Omega = 0$. Since $\Omega \neq 0$, it must occur that $\det(\lambda_\Omega\Gamma - \beta'\Gamma\beta) = 0$.

[13] Note that the matrix $C^{-1} = \Gamma^{-1/2}$ is symmetric in the $\Re$ domain.

Once $\Omega^*$ is known, there is no need to compute $\lambda_{\Omega^*} = \frac{\Omega^{*'}\beta'\Gamma\beta\Omega^*}{\Omega^{*'}\Gamma\Omega^*}$, because its value is precisely the eigenvalue that corresponds to the eigenvector $z$.

So much for the VAR(1) case. Now we would like to outline the solution for the case where $L$ lags are used on each forecasting variables, and additional exogenous variables are admissible (including the possibility of an intercept):

1. Fit $\hat{\beta}$ on the forecasting equation, which is now of the general form $P_{t,n} = \sum_{l=1}^{L}\sum_{i=1}^{N}\beta_{i,l,n}P_{t-l,i} + \beta_{n,0}X_{t-1,n} + \varepsilon_{t,n}$.[14]
2. Estimate $\hat{P}_t$ applying $\hat{\beta}$.
3. Compute $\left(\hat{P}_t'\hat{P}_t\right)$. This is the matrix $A$ of the generalized Rayleigh quotient.
4. Compute the spectral decomposition of $\left(P_t'P_t\right) = W\Lambda W'$. This is the matrix $B$ of the generalized Rayleigh quotient.
5. Compute $\left(P_t'P_t\right)^{-1/2} = W\Lambda^{-1/2}W'$.
6. Compute a PCA on $\left(P_t'P_t\right)^{-1/2}\left(\hat{P}_t'\hat{P}_t\right)\left(P_t'P_t\right)^{-1/2}$, which is the matrix $D$ of the standard Rayleigh quotient.
7. Determine $\Omega^* = \left(P_t'P_t\right)^{-1/2}z$, where $z$ is the eigenvector associated to the smallest eigenvalue ($\lambda_\Omega$).
8. As a verification, we can check that the ratio $\frac{\Omega^{*'}\left(\hat{P}_t'\hat{P}_t\right)\Omega^*}{\Omega^{*'}\left(P_t'P_t\right)\Omega^*}$ merely recovers the previously selected eigenvalue $\lambda_\Omega$.
9. Set a unit position on the portfolio to be hedged (*i=1*): $\omega = \Omega^*\frac{1}{\Omega_1^*}$.

Although computing trending portfolios is not relevant in the context of hedging, this procedure can also be applied to determine them. In order to deliver the most trending portfolio, it suffices to select $z$ to be the eigenvector associated to the largest eigenvalue in Step 7.

A caveat of this approach is that estimates of $\Gamma$ and $\beta$ usually are quite unstable, particularly as the number of variables increases. A classic remedy is to penalize the covariance estimation using, for example, a multiple of the norm of $\Gamma$,[15] though not satisfactory solution seems available at the moment.

### 5.1.2.- DICKEY-FULLER OPTIMAL (DFO)

We have seen that ECM is a dynamic model limited to two dimensions, and that this limitation could be circumvented through a canonical transformation of a multivariate, multi-equation specification, like in BTCD. That approach introduced a strong structure through a system of equations, each imposing an *individual* autoregressive equilibrium condition. It may however be more convenient to search

[14] This allows adding an intercept and additional lags to our specification.
[15] See d'Aspremont (2008) for an in-depth discussion, in the context of small mean-reverting portfolios.

for a *joint* equilibrium in cointegrated form, removing the "$\beta$" part of the structure and estimating $\Gamma$ directly. Next, we present such alternative approach.

The target is to find a vector of holdings $\boldsymbol{\omega}$ for $S_t = P_{1,t} + \sum_{n=2}^{N} \omega_n P_{n,t}$ such that the probability of having a unit root *in the spread* is minimized. Dickey and Fuller (1979) test whether a unit root is present in an autoregressive model, which would be the case should $\beta = 1$ in $S_t = \alpha + \beta S_{t-1} + \varepsilon_t$. A unit root means that $S_t$ follows a random walk, which makes its outcome unpredictable (a particular case of martingale). $\beta > 1$ is a sufficient condition for $S$ not being stationary. Thus, our best hope is for $\beta < 1$. Since the null hypothesis is $H_0: \beta = 1$, we are more confidence in the hedge the more negative the Dickey-Fuller test statistic is, $DF = \frac{\hat{\beta}-1}{\sigma_{\hat{\beta}}} \ll 0$.

Said and Dickey (1984) "augmented" the test to encompass a more complicated set of time series models. Similar tests include Phillips and Perron (1988) and Elliot, Rothenberg and Stock (1996).

**5.1.2.1.- DIRECT ESTIMATION OF THE DF STAT**

Consider the standard autoregressive specification $S_t = \alpha + \beta S_{t-1} + \varepsilon_t$, which can be rewritten as $\Delta S_t = \alpha + (\beta - 1)\, S_{t-1} + \varepsilon_t$. Rather than having to estimate the DF statistic based upon the statistical significance of $\beta$, $DF = \frac{\hat{\beta}-1}{\sigma_{\hat{\beta}}}$, through OLS, we would like to devise a direct estimation that does not require matrix inversions, multiplications, and other computationally inefficient calculations.

In matrix form, $\Delta S = X\delta + \varepsilon$, where $\delta = \begin{bmatrix} \alpha \\ \beta - 1 \end{bmatrix}$ and $X = [1_T \quad L(S)]$, where $1_T$ is a column-vector of 1s of $T$ elements and $L$ is the lag operator. Then, $\delta = [X'X]^{-1}X'\Delta S$, with $X'X = \begin{bmatrix} T & \sum_{t=1}^{T} S_{t-1} \\ \sum_{t=1}^{T} S_{t-1} & \sum_{t=1}^{T} S_{t-1}^2 \end{bmatrix}$ and $X'\Delta S = \begin{bmatrix} S_T - S_1 \\ \sum_{t=1}^{T} S_t S_{t-1} - \sum_{t=1}^{T} S_{t-1}^2 \end{bmatrix}$. This can be solved in terms of $\hat{\beta} - 1 = \frac{\sigma_{\Delta S, L(S)}}{\sigma_{L(S)}^2} = \frac{\sigma_{\Delta S}\rho_{\Delta S, L(S)}}{\sigma_{L(S)}}$, with $\sigma_{\hat{\beta}}^2 = \frac{\sigma_\varepsilon^2}{T\sigma_{L(S)}^2}$ and $\sigma_\varepsilon^2 = \sigma_{\Delta S}^2 - \frac{\sigma_{\Delta S, L(S)}^2}{\sigma_{L(S)}^2}$. Because $\hat{\sigma}_\varepsilon^2 = \frac{T}{T-2}\left(\sigma_{\Delta S}^2 - \frac{\sigma_{\Delta S, L(S)}^2}{\sigma_{L(S)}^2}\right)$, further operations lead to $\hat{\sigma}_{\hat{\beta}} = \frac{\sigma_{\Delta S}}{\sigma_{L(S)}}\sqrt{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}}$. Finally,

$$\widehat{DF} = \frac{\frac{\sigma_{\Delta S}\rho_{\Delta S, L(S)}}{\sigma_{L(S)}}}{\frac{\sigma_{\Delta S}}{\sigma_{L(S)}}\sqrt{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}}} = \frac{\rho_{\Delta S, L(S)}}{\sqrt{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}}} \tag{8}$$

which can be computed directly without having to make the intermediate calculations of $[X'X]^{-1}X'\Delta S$, etc.

Furthermore, $\rho_{\Delta S, L(S)}$ can be easily updated for each new observation without having to re-use the whole sample, thus allowing a continuous estimation of $\widehat{DF}$ after a few basic arithmetic operations.

#### 5.1.2.2.- DF STAT MINIMIZATION

The previous epigraph has shown how to estimate the DF stat directly, in one step. Considering that $\boldsymbol{S}$ is a linear function of $\boldsymbol{\omega}$, and DF a function of $\boldsymbol{S}$, we can compute the partial derivatives $\frac{\partial DF(S(\omega))}{\partial \omega}$, $\frac{\partial^2 DF(S(\omega))}{\partial \omega^2}$. These in turn can be used to compute the vector $\boldsymbol{\omega}$ that delivers a hedge with minimum DF:

$$\underset{\substack{\omega \\ \omega_0 = 1}}{Min} \frac{\rho_{\Delta S(\omega), L(S(\omega))}}{\sqrt{\frac{1-\rho^2_{\Delta S(\omega), L(S(\omega))}}{T-2}}} \tag{9}$$

Appendix 2 obtains the first and second analytical derivatives of our objective function, which can be applied on standard gradient-search algorithms. This hedge optimization procedure addresses the three critiques discussed in Section 4.1.1. For $N$ sufficiently small, $\boldsymbol{\omega}$ can be reliably computed through a brute force, grid search algorithm. A similar procedure could be devised on the KPSS test for stationarity.

## 5.2.- ADVANCED SINGLE-PERIOD METHOD

### 5.2.1.- MINI-MAX SUBSET CORRELATION (MMSC)

Of the approaches discussed in the previous Sections, three seemed particularly interesting. DFO searched for the linear combination of positions that minimized the probability that the hedging error contains a unit root. PCA and BTCD looked deep into the geometry of the hedging set, and identified uncorrelated sources of variability responsible for most of the risk (Principal or Canonical Components). On the negative side, none of these approaches impose a balanced structure on the combined position (*spread*). For example, the DFO solution may be exceedingly biased towards a particular instrument with strong mean reversion, but that otherwise provides little hedge to the original portfolio. Detecting *irrelevant* hedging positions is even harder in the case of PCA and BTCD, since all instruments participate in the definition of each principal component.

In this Section we will introduce a new approach, called MMSC, which imposes a strong balancing structure on the hedging portfolio. The mathematics of the solution may appear complex, but the intuition is simple: Hedging errors move the spread (i.e., combined portfolio + hedge positions) away from its equilibrium level. Spread changes should not be highly correlated to any individual position or subset of positions. If one particular "leg" or subset of legs is highly correlated to the spread, the spread is imbalanced, meaning that it is dominated by that leg or subset. Ideally, we should find a vector of holdings such that the maximum correlation of any leg *or subset of legs* to the spread is minimal (thus the name *Mini-Max Subset Correlation*).

#### 5.2.1.1.- MOTIVATION

More formally, suppose a *n*-legged spread, characterized by its holdings, $\boldsymbol{\omega}$, and the covariance matrix of value changes, $V$. The spread's risk can be decomposed in terms of its legs' contributions as

$$\sigma_{\Delta S}^2 = \omega' V \omega = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i \ \sigma_{i,j} \omega_j \ = \sum_{i=1}^{n} \omega_i \ \sigma_{i,\Delta S} \\ = \sigma_{\Delta S} \sum_{i=1}^{n} \omega_i \ \sigma_i \ \rho_{i,\Delta S} \tag{10}$$

Therefore, the spread's risk is a weighted average of the instruments' standard deviations, where the weightings are the product of the instrument's holdings and their correlations to the spread.

$$\sigma_{\Delta S} = \sum_{i=1}^{n} \omega_i \ \rho_{i,\Delta S} \sigma_i \tag{11}$$

One approach to risk diversification would consist in computing the Equally-weighted Risk Contribution (ERC) spread (see Maillard, Roncalli and Teiletche (2009) for a thorough study), such that

$$\omega_i \ \sigma_i \ \rho_{i,\Delta S} = \frac{\sigma_{\Delta S}}{n}, \forall i \tag{12}$$

ERC provides better diversification than equal weights (also called "1/n") solutions, but still it is under general circumstances objectionable. For example, in a portfolio of three assets, two of which are highly correlated, 2/3 of the risk would be allocated to the same exposure.

A second approach is proposed by Choueifaty and Coignard (2008). These authors compute the vector of holdings that Maximize the *Diversification Ratio* (MDR), as defined by

$$\max_{\omega} \quad \frac{\sum_{i=1}^{n} \omega_i \ \sigma_i}{\sigma_{\Delta S}} \tag{13}$$

This diversification ratio is the ratio of weighted volatilities divided by the portfolio volatility. MDR is an intuitive method that penalizes the risk associated with cross-correlations, as they are accounted by the denominator but absent in the numerator of the maximized ratio. Still, MDR only takes into account correlation of every holding to the overall spread, ignoring the possibility of exposure imbalances from subsets of holdings. MDR does not prevent that subsets of holdings may dominate the overall risk, because Eq. (13) uses $\rho_{i,\Delta S}$ (correlation of an individual holding to the spread) as the only balancing argument. For example, in a three asset portfolio, even though the correlation of every holding to the overall portfolio may be the same, the correlation of the subsets of 2 holdings to the overall portfolio can indeed be very imbalanced. This makes the solution somewhat arbitrary, as different clustering criteria (by country, asset class, currency, liquidity, capitalization, etc.) will yield different hedging vectors. The authors also acknowledge that the solution may not be unique or robust, particularly with ill-conditioned covariance matrices. Adding some structure to the optimization program would alleviate these problems.

The natural question becomes, for what ω occurs that a portfolio is balanced, in the sense that the correlation of each constituent *or subset of constituents* to the spread is overall minimized? Before providing the mathematical solution to this highly-dimensional problem, we will have to introduce a few new concepts.

**5.2.1.2.- SUBSET MATRIX (D)**

Consider a set ***X*** of *n* instruments. Let be Φ(***X***)-Ø the *σ-algebra* formed by ***X***'s power set Φ(***X***), from which we exclude the empty set. (***X***, Φ(***X***)-Ø) constitutes our *σ-field* or *measurable space*. ***D*** represents our *σ-algebra* Φ(***X***)-Ø as a binary *(nxN)* matrix, $N=\sum_{i=1}^{n}\binom{n}{i}=2^{n}-1$, where $D_{i,p}=1$ if subset *p* contains instrument *i*, *p=1, …, N*, *i=1, …, n*, and $D_{i,p}=0$ otherwise. $D_N$, the last column of matrix ***D***, is an identity matrix, i.e. *the last subset is the spread itself*.

Denote $P_{i,t}$ the market value associated with variable *i* at observation *t*, *i=1, …, n*, *t=1,…,T*. *i=1* corresponds to the portfolio we wish to hedge. A vector (*nx1*) of holdings ***ω*** allows to define a *n*-legged spread with market value $S_t=\sum_{i=1}^{n}\omega_i P_{i,t}$.

Additionally, we define ***D**** as the result of removing from matrix ***D*** any column $i\middle| k<D_i^{'}D_i<n$, where $1\leq k<n$[16]. Likewise, $N^{*}=\sum_{i=1}^{k<n}\binom{n}{i}+1\leq N$.

**5.2.1.3.- SUBSET COVARIANCE MATRIX (B)**

Let ***B*** be a *(N*xN*)* matrix, $B_{p,q}=\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_i D_{i,p}^{*}\sigma_{i,j}D_{j,q}^{*}\omega_j=\omega^{'}D^{*'}VD^{*}\omega$, *V* is the covariance matrix of Δ*P* (which are assumed *IID* Normal), and $\sigma_{i,j}$ represents the covariance of changes between instruments *i* and *j*, *p=1,…,N**, *q=1,…,N**.

**5.2.1.4.- SUBSET CORRELATION MATRIX (C)**

Let ***C*** be a *(N*xN*)* matrix, defined as the correlation matrix implied by ***B***. $C_{p,q}=B_{p,q}\left(B_{p,p}B_{q,q}\right)^{-\frac{1}{2}}$.

**5.2.1.5.- MAXIMUM SUBSET CORRELATION (MSC)**

The last column of matrix ***C*** has special significance. It represents the correlation of each subset to the spread. $MSC=Max\left\{\left|C_{p,N^{*}}\right|\right\}$, $p=1,\ldots,N^{*}\text{-}1$. Note that, like any diagonal element of a correlation matrix, $C_{N^{*},N^{*}}=1$.

**5.2.1.6.- MAELOC SPREAD**

Given a set of variables *n*, let's designate as *Maeloc* the spread characterized by a vector (*nx1*) of holdings ***ω*** such that solves the following non-linear program:

---

[16] Note that the last column of $D_N^{*}$ will still be a vector of 1s.

$$\begin{array}{lll} \underset{\omega}{Min} & Max\left\{\left|C_{p,N^*}\right|\right\} & p = 1,...,N^*\text{-}1 \\ s.t. & \omega_1 = 1 & \end{array} \quad (14)$$

The spread's first leg corresponds to the portfolio we wish to hedge, and its holding is set to $\omega_1 = 1$. A solution to a *Maeloc* spread always exists and it is obviously unique[17].

As $n$ increases, the value of $N^*$ explodes, which makes this non-linear problem highly dimensional. To make matters worse, the objective function is by no means continuous nor differentiable. Traditional optimization approaches may not offer a viable solution to this problem. An optimization algorithm specially designed to solve this program is presented in Appendix 3.

A *Maeloc* spread has the following properties:

1. **Balanced**: The exposure to any *k-subset* is minimized. No one instrument or set of instruments dominates the spread.
2. **Economic**: Unnecessary legs are removed, as any subset including them would exhibit a high correlation to the spread. See the next section for an explanation of how to eliminate unnecessary legs.
3. **Customizable**: The algorithm converges in presence of any number of constrained holdings.
4. **Control over lead-lag effects**: The interval used to compute changes, on which the covariance matrix is estimated, can be interpreted as the horizon beyond which lead-lag effects should be penalized. Asynchronous co-movements occurred within that interval are indistinguishable from synchronous, and therefore do not increase the correlation between the spread and the leader. Otherwise, they will increase the correlation between the spread and the leader, which will impact the holdings of the *Maeloc*-spread in order to provide a hedge.
5. **Robust**: Similar to robust asset allocation methods (Meucci (2005)), the *Maeloc*-spread is obtained by minimizing the impact of the worst case scenario (a risk driver affecting a subset highly correlated to the spread). There may be other vectors of holdings that could hedge better against particular risk drivers, but this is the solution that provides the best overall hedge (including the worst scenarios).

MMSC's control over lead-lag effects is an interesting feature. Lead-lag effects increase the correlation between the spread and the leader, which the *Maeloc* spread subdues as it balances the spread across all constituents. Although the model itself makes no assumption regarding the dynamics of the spread, the weights are impacted by the spread's multi-period behaviour. We have classified MMSC as a single-period method, but in fact it can be argued to be a hybrid.

#### 5.2.1.7.- MINIMUM LEG CORRELATION (MLC)

We define Minimum Leg Correlation (*MLC*) as the minimum correlation among any leg or subset of legs (excluding the entire spread) of the *Maeloc*-spread. More formally,

---

[17] $n \le N^*, \forall k$, and in particular $n < N^*, \forall k > 1$.

$$MLC = Min\left\{\left|C_{p,q}\right|\right\}, p = 1,...,N^*\text{-}1, q = 1,...,N^*\text{-}1 \quad (15)$$

We must stress that MLC is computed after the *Maeloc*-spread has been determined. MLC's role is to determine whether there are expendable legs. Consider a ***C*** matrix of a spread with 3 legs. Figure 1 illustrates how the ***C*** matrix is divided into two regions where MSC and MLC are to be found. A low *MSC* (maximum of the outer area) indicates that the spread is *well-balanced*, because no leg or subset of legs dominates the spread. However, that the spread is well-balanced is not a sufficient condition for being *meaningful*. As Meucci (2010) shows, the potential for improving a portfolio's diversification is a function of the system's correlation. A necessary condition must therefore be imposed, namely that the legs and subset of legs are highly correlated with each other, i.e. a high *MLC* (minimum of the inner area).

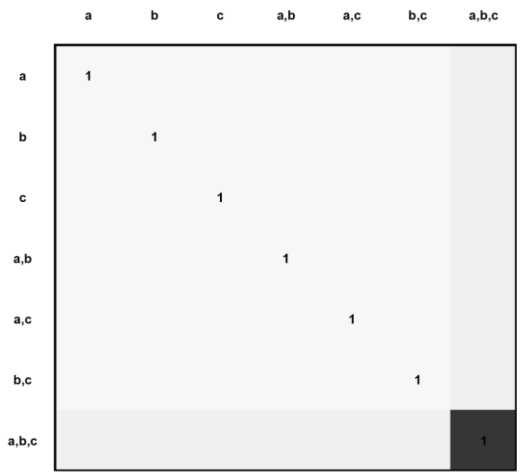


*Figure 1 – The MSC and MLC regions of the C matrix*

Whereas the MSC computed on the *Maeloc*-spread points to the areas of the spread that are imbalanced, the MLC computed on the *Maeloc*-spread indicates which constituents are not playing a relevant role in terms of adding diversification.

Should *MLC* be low, it will be easy to form a spread with low *MSC* (e.g., equal holdings of alternating sign), however meaningless it may be. In that case, the unnecessary instrument, responsible for reducing the value of *MLC*, can be easily identified and removed. This sequential two-step process of MSC minimization (*Maeloc*-spread determination) and MLC evaluation delivers a spread that is both, meaningful and well-balanced.

## 6.- EMPIRICAL RESULTS

### 6.1.- THE DATA

We will discuss in this section the results of estimating the previous hedging procedures. The investment universe is comprised of the 11 most liquid index futures worldwide, converted into USD: ES1 Index (CME E-Mini S&P500), DM1 Index (CBOT Mini Dow Jones), NQ1 Index (CME Nasdaq 100), VG1 Index (EUREX Eurostoxx 50), GX1 Index (EUREX DAX), CF1 Index (Euronext LIFFE CAC), Z 1 Index (Euronext LIFFE FTSE), EO1 Index (Euronext LIFFE Amsterdam), RTA1 Index (ICE Mini Russell 2000), NX1 Index (CME Nikkei 225 Dollar) and FA1 Index (CME Mini S&P MID 400). The data source is Bloomberg's 1-minute bar history from December 31st 2007 to February 4th 2011. Contracts are rolled forwards with the transfer of volume from the front contract to the next.

Before running a procedure over a particular combination of securities, the relevant data is preprocessed as follows:

1. Alignment: Minute bars on which one of the securities did not trade are eliminated.
2. Observation weight: Units traded for different securities represent different bet sizes. In order to assign an observation weight to each aligned 1-minute bar, we must make the different volumes traded of each security comparable. To this purpose, we multiply the units traded of each security by that security's risk. The sum of these products for each time bar is that observation's weight.
3. Sample: At the beginning of each session, 1-minute bars are gathered for the previous 5 sessions. The cumulative observation weights (as derived from the previous point) are divided into 250 buckets, equivalent to 50 buckets per session. A price time series is formed by taking the price of the last transaction from each bucket.

The result is a time series of aligned prices sampled by equidistant observation weights, in excess of 40,000 datapoints and 810 rebalances per combination of securities.

**6.2.- TESTING FOR STABILITY**

Each procedure is optimal under its own (utility or satisfaction) criterion, and in that respect we cannot prefer one over the other in-sample. For example, a measure of diversification based on PCA-risk decomposition will prefer PCA over the other risk criteria, however subjective that choice is. An objective criterion for assessing the quality of a hedge is its stability. Thus, we will evaluate the alternative methods, preferring the most stable among the similar ones. For each combination of instruments, we measure stability in two different ways: Dispersion of the hedging ratios (s*tatic stability*) and dispersion of the change in the hedging ratio between consecutive rebalances (*dynamic stability*). Static instability evidences lack of robustness, because the solutions exhibit a greater dispersion over time. Dynamic instability makes a solution impracticable, due to the need for frequent and costly rebalances.

We have computed the number of E-mini S&P500 futures to be sold as a hedge against one contract owned of DM1 Index, etc. Hedging ratios are estimated over the entire sample (LR) and sequentially re-estimated every session based on the prior 5 sessions (SR). *LR w2* is the hedge estimated over the entire sample. *Avg (SR w2)* is the average value of the hedging ratios as estimated every day using data from the previous 5 sessions. *StDev (SR w2)* is the standard deviation of the same. That gives us the static stability, as we are computing the dispersion of the hedging ratios against the overall mean that *Avg (SR w2)* represents. *t-Stat* is the ratio of the prior two.

Dynamic stability is assessed in the following terms: If $\omega_{2,t}$ is the hedging ratio of a particular pair at session $t$ and $\omega_{2,t-1}$ is the hedging ratio of that same pair as of the previous session, *StDev (d1 SR w2)* is the standard deviation of the change ($\omega_{2,t} - \omega_{2,t-1}$) from between two sessions. *StDev (d2 SR w2)* is the standard deviation of the change over 2 sessions, etc. *StDev (d5 SR w2)* is the standard deviation of the changes from each day to the week after.

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.87 | -0.84 | 0.05 | -15.66 | 0.05 | 0.06 | 0.06 | 0.06 | 0.06 |
| NQ1 Index | ES1 Index | 1 | -0.58 | -0.71 | 0.17 | -4.21 | 0.17 | 0.18 | 0.19 | 0.20 | 0.20 |
| VG1 Index | ES1 Index | 1 | -1.00 | -0.94 | 0.37 | -2.56 | 0.34 | 0.37 | 0.43 | 0.45 | 0.48 |
| GX1 Index | ES1 Index | 1 | -3.61 | -4.48 | 2.35 | -1.91 | 1.46 | 2.03 | 2.49 | 2.73 | 2.94 |
| CF1 Index | ES1 Index | 1 | -1.30 | -1.20 | 0.55 | -2.18 | 0.71 | 0.54 | 0.73 | 0.74 | 0.74 |
| Z 1 Index | ES1 Index | 1 | -1.38 | -1.72 | 0.61 | -2.82 | 0.56 | 0.73 | 0.78 | 0.73 | 0.74 |
| EO1 Index | ES1 Index | 1 | -1.67 | -1.88 | 0.50 | -3.73 | 0.55 | 0.59 | 0.62 | 0.64 | 0.66 |
| RTA1 Index | ES1 Index | 1 | -1.27 | -1.54 | 0.34 | -4.55 | 0.34 | 0.35 | 0.33 | 0.37 | 0.36 |
| NX1 Index | ES1 Index | 1 | -0.97 | -1.30 | 0.65 | -2.01 | 0.72 | 0.81 | 0.83 | 0.88 | 0.90 |
| FA1 Index | ES1 Index | 1 | -1.32 | -1.49 | 0.24 | -6.34 | 0.22 | 0.24 | 0.23 | 0.25 | 0.24 |

*Table 1 – OLSD hedge stability*

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.95 | -0.68 | 0.42 | -1.63 | 0.45 | 0.54 | 0.55 | 0.57 | 0.58 |
| NQ1 Index | ES1 Index | 0 | 0.00 | -0.51 | 0.24 | -2.17 | 0.28 | 0.32 | 0.33 | 0.33 | 0.33 |
| VG1 Index | ES1 Index | 1 | -0.65 | -0.58 | 0.24 | -2.42 | 0.29 | 0.34 | 0.34 | 0.32 | 0.33 |
| GX1 Index | ES1 Index | 1 | -3.54 | -3.11 | 1.18 | -2.63 | 1.30 | 1.60 | 1.70 | 1.67 | 1.69 |
| CF1 Index | ES1 Index | 1 | -0.86 | -0.76 | 0.33 | -2.30 | 0.37 | 0.45 | 0.46 | 0.46 | 0.46 |
| Z 1 Index | ES1 Index | 1 | -1.53 | -1.32 | 0.52 | -2.55 | 0.64 | 0.75 | 0.77 | 0.76 | 0.72 |
| EO1 Index | ES1 Index | 1 | -1.57 | -1.33 | 0.57 | -2.32 | 0.68 | 0.78 | 0.80 | 0.77 | 0.79 |
| RTA1 Index | ES1 Index | 1 | -1.14 | -0.86 | 0.47 | -1.84 | 0.52 | 0.65 | 0.69 | 0.68 | 0.66 |
| NX1 Index | ES1 Index | 1 | -0.79 | -0.72 | 0.35 | -2.04 | 0.40 | 0.45 | 0.46 | 0.46 | 0.47 |
| FA1 Index | ES1 Index | 1 | -1.35 | -1.01 | 0.52 | -1.95 | 0.61 | 0.69 | 0.73 | 0.74 | 0.73 |

*Table 2 – ECM hedge stability*

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.69 | -0.55 | 0.16 | -3.39 | 0.09 | 0.10 | 0.12 | 0.13 | 0.14 |
| NQ1 Index | ES1 Index | 1 | -0.43 | -0.32 | 0.25 | -1.30 | 0.21 | 0.22 | 0.27 | 0.28 | 0.30 |
| VG1 Index | ES1 Index | 1 | 0.02 | 2.30 | 34.52 | 0.07 | 52.99 | 35.05 | 52.36 | 49.17 | 49.38 |
| GX1 Index | ES1 Index | 1 | -5.03 | -5.94 | 2.51 | -2.36 | 1.31 | 1.95 | 2.46 | 2.78 | 3.00 |
| CF1 Index | ES1 Index | 1 | -5.44 | 3.34 | 87.73 | 0.04 | 124.56 | 123.16 | 127.58 | 124.42 | 124.04 |
| Z 1 Index | ES1 Index | 1 | -2.58 | -4.04 | 8.72 | -0.46 | 11.40 | 12.04 | 12.07 | 12.50 | 12.50 |
| EO1 Index | ES1 Index | 1 | -2.82 | -3.93 | 12.24 | -0.32 | 19.83 | 17.62 | 17.46 | 17.37 | 17.57 |
| RTA1 Index | ES1 Index | 1 | -2.15 | -1.91 | 35.06 | -0.05 | 49.42 | 50.76 | 51.23 | 50.11 | 49.56 |
| NX1 Index | ES1 Index | 1 | 11.07 | 8.00 | 178.92 | 0.04 | 271.81 | 255.03 | 248.89 | 256.51 | 255.36 |
| FA1 Index | ES1 Index | 1 | -1.87 | -2.68 | 10.39 | -0.26 | 15.04 | 14.75 | 14.70 | 14.74 | 14.72 |

*Table 3 – MVP hedge stability*

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.85 | -0.84 | 0.04 | -19.25 | 0.03 | 0.03 | 0.04 | 0.04 | 0.04 |
| NQ1 Index | ES1 Index | 1 | -0.60 | -0.67 | 0.11 | -5.98 | 0.08 | 0.09 | 0.10 | 0.10 | 0.10 |
| VG1 Index | ES1 Index | 1 | -0.89 | -0.90 | 0.29 | -3.15 | 0.17 | 0.23 | 0.30 | 0.34 | 0.38 |
| GX1 Index | ES1 Index | 1 | -4.66 | -5.17 | 2.14 | -2.41 | 1.08 | 1.68 | 2.12 | 2.42 | 2.64 |
| CF1 Index | ES1 Index | 1 | -1.18 | -1.20 | 0.49 | -2.43 | 0.54 | 0.43 | 0.60 | 0.62 | 0.65 |
| Z 1 Index | ES1 Index | 1 | -1.69 | -1.82 | 0.46 | -3.95 | 0.26 | 0.38 | 0.44 | 0.48 | 0.49 |
| EO1 Index | ES1 Index | 1 | -1.96 | -2.05 | 0.36 | -5.71 | 0.22 | 0.27 | 0.31 | 0.37 | 0.39 |
| RTA1 Index | ES1 Index | 1 | -1.39 | -1.61 | 0.29 | -5.48 | 0.10 | 0.13 | 0.15 | 0.16 | 0.18 |
| NX1 Index | ES1 Index | 1 | -1.19 | -1.30 | 0.49 | -2.64 | 0.38 | 0.47 | 0.57 | 0.61 | 0.62 |
| FA1 Index | ES1 Index | 1 | -1.42 | -1.54 | 0.22 | -7.00 | 0.09 | 0.11 | 0.12 | 0.14 | 0.15 |

*Table 4 – PCA hedge stability*

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.94 | -0.94 | 0.03 | -36.67 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| NQ1 Index | ES1 Index | 1 | -0.66 | -0.62 | 0.17 | -3.74 | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 |
| VG1 Index | ES1 Index | 1 | -0.59 | -0.68 | 0.04 | -18.67 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| GX1 Index | ES1 Index | 1 | -3.67 | -3.55 | 0.12 | -30.83 | 0.04 | 0.04 | 0.05 | 0.05 | 0.05 |
| CF1 Index | ES1 Index | 1 | -0.80 | -0.90 | 0.04 | -22.03 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| Z 1 Index | ES1 Index | 1 | -1.52 | -1.54 | 0.38 | -4.02 | 0.54 | 0.54 | 0.55 | 0.55 | 0.55 |
| EO1 Index | ES1 Index | 1 | -1.56 | -1.57 | 0.05 | -29.22 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| RTA1 Index | ES1 Index | 1 | -1.18 | -1.11 | 0.07 | -16.60 | 0.04 | 0.05 | 0.05 | 0.05 | 0.05 |
| NX1 Index | ES1 Index | 1 | -0.79 | -0.86 | 0.43 | -1.99 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 |
| FA1 Index | ES1 Index | 1 | -1.36 | -1.27 | 0.08 | -16.11 | 0.03 | 0.04 | 0.04 | 0.04 | 0.04 |

*Table 5 – BTCD hedge stability*

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.94 | -0.93 | 0.23 | -4.08 | 0.30 | 0.31 | 0.32 | 0.32 | 0.32 |
| NQ1 Index | ES1 Index | 1 | -1.18 | -0.64 | 0.30 | -2.15 | 0.40 | 0.40 | 0.40 | 0.40 | 0.38 |
| VG1 Index | ES1 Index | 1 | -0.34 | -0.68 | 0.20 | -3.33 | 0.28 | 0.29 | 0.29 | 0.29 | 0.29 |
| GX1 Index | ES1 Index | 1 | -4.40 | -3.64 | 1.18 | -3.08 | 1.34 | 1.70 | 1.71 | 1.71 | 1.71 |
| CF1 Index | ES1 Index | 1 | -0.51 | -0.91 | 0.16 | -5.80 | 0.19 | 0.22 | 0.22 | 0.22 | 0.22 |
| Z 1 Index | ES1 Index | 1 | -1.31 | -1.51 | 0.36 | -4.18 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 |
| EO1 Index | ES1 Index | 1 | -1.09 | -1.58 | 0.22 | -7.32 | 0.28 | 0.30 | 0.30 | 0.30 | 0.30 |
| RTA1 Index | ES1 Index | 1 | -1.45 | -1.13 | 0.42 | -2.68 | 0.56 | 0.60 | 0.60 | 0.60 | 0.60 |
| NX1 Index | ES1 Index | 1 | -0.03 | -0.86 | 0.31 | -2.78 | 0.41 | 0.41 | 0.41 | 0.42 | 0.41 |
| FA1 Index | ES1 Index | 1 | -1.90 | -1.30 | 0.43 | -3.05 | 0.46 | 0.57 | 0.61 | 0.62 | 0.62 |

*Table 6 – DFO hedge stability*

| Portf. 1 | Portf. 2 | LR w1 | LR w2 | Avg (SR w2) | StDev (SR w2) | t-Stat (SR w2) | StDev (d1 SR w2) | StDev (d2 SR w2) | StDev (d3 SR w2) | StDev (d4 SR w2) | StDev (d5 SR w2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 1 | -0.86 | -0.84 | 0.04 | -20.77 | 0.02 | 0.03 | 0.03 | 0.03 | 0.04 |
| NQ1 Index | ES1 Index | 1 | -0.63 | -0.69 | 0.10 | -7.28 | 0.04 | 0.05 | 0.06 | 0.06 | 0.07 |
| VG1 Index | ES1 Index | 1 | -0.90 | -0.90 | 0.15 | -5.86 | 0.09 | 0.11 | 0.15 | 0.17 | 0.19 |
| GX1 Index | ES1 Index | 1 | -4.12 | -4.27 | 0.98 | -4.34 | 0.48 | 0.73 | 0.91 | 1.07 | 1.19 |
| CF1 Index | ES1 Index | 1 | -1.16 | -1.15 | 0.23 | -5.08 | 0.19 | 0.17 | 0.24 | 0.26 | 0.28 |
| Z 1 Index | ES1 Index | 1 | -1.62 | -1.63 | 0.21 | -7.70 | 0.12 | 0.16 | 0.18 | 0.20 | 0.21 |
| EO1 Index | ES1 Index | 1 | -1.84 | -1.82 | 0.22 | -8.37 | 0.13 | 0.18 | 0.21 | 0.24 | 0.26 |
| RTA1 Index | ES1 Index | 1 | -1.36 | -1.52 | 0.22 | -6.85 | 0.08 | 0.10 | 0.11 | 0.13 | 0.14 |
| NX1 Index | ES1 Index | 1 | -1.16 | -1.18 | 0.23 | -5.15 | 0.15 | 0.20 | 0.24 | 0.27 | 0.28 |
| FA1 Index | ES1 Index | 1 | -1.40 | -1.48 | 0.18 | -8.13 | 0.07 | 0.08 | 0.09 | 0.10 | 0.11 |

*Table 7 – MMSC hedge stability*

Short run Avg DFO and Avg ECM vectors are very close, although DFO seems to provide more robust results (an average *StDev(SR w2)* of 0.38 for DFO compared to a 0.48 for ECM). Likewise, short run Avg MMSC and Avg PCA vectors are very similar, with MMSC delivering more robust estimates (an average *StDev(SR w2)* 0.26 for MMSC compared to a 0.49 for PCA).[18] These similarities are not surprising and are consistent with the theory outlined earlier. DFO and ECM approach the hedging problem from a cointegration perspective. The difference is, ECM's solution tries to maximize the $R^2$ of portfolio changes, while DFO focuses on minimizing the probability that the cumulative hedging errors incorporates a unit root (perhaps a more critical question for the purpose of hedging). Like PCA, MMSC also looks into the geometry of the hedging problem, deriving the holdings that are most orthogonal to the hedging error (in terms of legs and subsets of legs).

Using *StDev(SR w2)* from Tables 1-7, Table 8 provides the rank for each hedging method and pair in terms of static stability. The greatest average rank is obtained by BTCD and MMSC, followed by DFO.

[18] These discrepancies in stability are statistically significant beyond a 99% confidence level for an F-test of homogeneous variances.

| Portf. 1 | Portf. 2 | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 4 | 7 | 5 | 3 | 1 | 6 | 2 |
| NQ1 Index | ES1 Index | 4 | 5 | 6 | 2 | 3 | 7 | 1 |
| VG1 Index | ES1 Index | 6 | 4 | 7 | 5 | 1 | 3 | 2 |
| GX1 Index | ES1 Index | 6 | 4 | 7 | 5 | 1 | 3 | 2 |
| CF1 Index | ES1 Index | 6 | 4 | 7 | 5 | 1 | 2 | 3 |
| Z 1 Index | ES1 Index | 6 | 5 | 7 | 4 | 3 | 2 | 1 |
| EO1 Index | ES1 Index | 5 | 6 | 7 | 4 | 1 | 2 | 3 |
| RTA1 Index | ES1 Index | 4 | 6 | 7 | 3 | 1 | 5 | 2 |
| NX1 Index | ES1 Index | 6 | 3 | 7 | 5 | 4 | 2 | 1 |
| FA1 Index | ES1 Index | 4 | 6 | 7 | 3 | 1 | 5 | 2 |
| **Average rank** | | 5.10 | 5.00 | 6.70 | 3.90 | 1.70 | 3.70 | 1.90 |

*Table 8 – Static stability ranking*

Dynamic stability results are broadly consistent with their static counterpart. For example, average *StDev(d1 SR w2)* for DFO is 0.47, compared to a 0.55 for ECM. Likewise, MMSC is only 0.14, compared to 0.29 of PCA. These discrepancies in stability are again statistically significant beyond the 99% confidence level for an F-test of variance homogeneity. MMSC's stability, both static and dynamic, makes it a very strong candidate in those situations where rebalancing is costly, like in the case of illiquid instruments. It also indicates that MMSC should be preferred to PCA when the sample length is limited. These advantages of MMSC are a consequence of its strong structure.

Table 9 delivers the rankings in terms of dynamic stability per method and pair, based on *StDev(d1 SR w2)*. Table 10 offers the equivalent rank, but based on *StDev(d5 SR w2)*. It is interesting to note that this ranking for multi-period methods tends to improve as the stability horizon increases, while the same ranking worsens for single-period methods. This result is consistent with the theory outlined earlier.

| Portf. 1 | Portf. 2 | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 4 | 7 | 5 | 3 | 1 | 6 | 2 |
| NQ1 Index | ES1 Index | 3 | 6 | 4 | 2 | 5 | 7 | 1 |
| VG1 Index | ES1 Index | 6 | 5 | 7 | 3 | 1 | 4 | 2 |
| GX1 Index | ES1 Index | 7 | 4 | 5 | 3 | 1 | 6 | 2 |
| CF1 Index | ES1 Index | 6 | 4 | 7 | 5 | 1 | 3 | 2 |
| Z 1 Index | ES1 Index | 5 | 6 | 7 | 2 | 4 | 3 | 1 |
| EO1 Index | ES1 Index | 5 | 6 | 7 | 3 | 1 | 4 | 2 |
| RTA1 Index | ES1 Index | 4 | 5 | 7 | 3 | 1 | 6 | 2 |
| NX1 Index | ES1 Index | 6 | 3 | 7 | 2 | 5 | 4 | 1 |
| FA1 Index | ES1 Index | 4 | 6 | 7 | 3 | 1 | 5 | 2 |
| **Average rank** | | 5.00 | 5.20 | 6.30 | 2.90 | 2.10 | 4.80 | 1.70 |

*Table 9 – Dynamic stability ranking over one session*

| Portf. 1 | Portf. 2 | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 4 | 7 | 5 | 3 | 1 | 6 | 2 |
| NQ1 Index | ES1 Index | 3 | 6 | 5 | 2 | 4 | 7 | 1 |
| VG1 Index | ES1 Index | 6 | 4 | 7 | 5 | 1 | 3 | 2 |
| GX1 Index | ES1 Index | 6 | 3 | 7 | 5 | 1 | 4 | 2 |
| CF1 Index | ES1 Index | 6 | 4 | 7 | 5 | 1 | 2 | 3 |
| Z 1 Index | ES1 Index | 6 | 5 | 7 | 2 | 4 | 3 | 1 |
| EO1 Index | ES1 Index | 5 | 6 | 7 | 4 | 1 | 3 | 2 |
| RTA1 Index | ES1 Index | 4 | 6 | 7 | 3 | 1 | 5 | 2 |
| NX1 Index | ES1 Index | 6 | 3 | 7 | 5 | 4 | 2 | 1 |
| FA1 Index | ES1 Index | 4 | 6 | 7 | 3 | 1 | 5 | 2 |
| **Average rank** | | 5.00 | 5.00 | 6.60 | 3.70 | 1.90 | 4.00 | 1.80 |

*Table 10 – Dynamic stability ranking over five sessions (one week)*

Next, we would like to determine which stable methods should be chosen among those similar. To that purpose, we have computed the correlation between procedures on the daily re-estimated hedging vectors. We should prefer DFO over ECM and MMSC over PCA, as they yield highly correlated results (see Table 11) with the first of each couple delivering more stable estimates (see Tables 1-8).[19] MVP is the most unstable of all procedures, with an average *StDev(SR w2)* of 37.05 and an average *StDev (d1 SR w2)* of 54.66. OLSD's theoretical inconsistencies make it an unreliable choice. BTCD is not highly correlated with DFO or MMSC, while also delivering stable results (average *StDev(SR w2)* of 0.14 and an average *StDev (d1 SR w2)* of 0.16). Hence, we advocate for BTCD, DFO and MMSC as stable, mutually different hedging procedures.

| **DM_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.878 | 0.487 | 0.992 | 0.877 | 0.686 | 0.992 |
| ECM | 0.878 | 1 | 0.158 | 0.876 | 1.000 | 0.996 | 0.884 |
| MVP | 0.487 | 0.158 | 1 | 0.496 | 0.163 | 0.146 | 0.481 |
| PCA | 0.992 | 0.876 | 0.496 | 1 | 0.878 | 0.691 | 1.000 |
| BTCD | 0.877 | 1.000 | 0.163 | 0.878 | 1 | 0.768 | 0.886 |
| DFO | 0.686 | 0.996 | 0.146 | 0.691 | 0.768 | 1 | 0.697 |
| MMSC | 0.992 | 0.884 | 0.481 | 1.000 | 0.886 | 0.697 | 1 |

| **NQ_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.911 | 0.512 | 0.961 | 0.912 | 0.746 | 0.969 |
| ECM | 0.911 | 1 | 0.751 | 0.962 | 1.000 | 0.999 | 0.949 |
| MVP | 0.512 | 0.751 | 1 | 0.619 | 0.752 | 0.623 | 0.558 |
| PCA | 0.961 | 0.962 | 0.619 | 1 | 0.963 | 0.784 | 0.995 |
| BTCD | 0.912 | 1.000 | 0.752 | 0.963 | 1 | 0.822 | 0.950 |
| DFO | 0.746 | 0.999 | 0.623 | 0.784 | 0.822 | 1 | 0.774 |
| MMSC | 0.969 | 0.949 | 0.558 | 0.995 | 0.950 | 0.774 | 1 |

| **VG_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.712 | -0.111 | 0.964 | 0.730 | 0.704 | 0.943 |
| ECM | 0.712 | 1 | 0.047 | 0.797 | 1.000 | 1.000 | 0.840 |
| MVP | -0.111 | 0.047 | 1 | -0.082 | 0.051 | 0.049 | -0.048 |
| PCA | 0.964 | 0.797 | -0.082 | 1 | 0.810 | 0.774 | 0.986 |
| BTCD | 0.730 | 1.000 | 0.051 | 0.810 | 1 | 0.951 | 0.846 |
| DFO | 0.704 | 1.000 | 0.049 | 0.774 | 0.951 | 1 | 0.808 |
| MMSC | 0.943 | 0.840 | -0.048 | 0.986 | 0.846 | 0.808 | 1 |

| **GX_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.821 | 0.890 | 0.942 | 0.830 | 0.692 | 0.943 |
| ECM | 0.821 | 1 | 0.693 | 0.793 | 1.000 | 1.000 | 0.933 |
| MVP | 0.890 | 0.693 | 1 | 0.983 | 0.705 | 0.587 | 0.891 |
| PCA | 0.942 | 0.793 | 0.983 | 1 | 0.803 | 0.671 | 0.954 |
| BTCD | 0.830 | 1.000 | 0.705 | 0.803 | 1 | 0.832 | 0.936 |
| DFO | 0.692 | 1.000 | 0.587 | 0.671 | 0.832 | 1 | 0.785 |
| MMSC | 0.943 | 0.933 | 0.891 | 0.954 | 0.936 | 0.785 | 1 |

| **CF_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.811 | 0.020 | 0.958 | 0.810 | 0.773 | 0.955 |
| ECM | 0.811 | 1 | -0.007 | 0.845 | 1.000 | 1.000 | 0.883 |
| MVP | 0.020 | -0.007 | 1 | 0.000 | -0.004 | 0.003 | -0.003 |
| PCA | 0.958 | 0.845 | 0.000 | 1 | 0.846 | 0.805 | 0.996 |
| BTCD | 0.810 | 1.000 | -0.004 | 0.846 | 1 | 0.944 | 0.882 |
| DFO | 0.773 | 1.000 | 0.003 | 0.805 | 0.944 | 1 | 0.838 |
| MMSC | 0.955 | 0.883 | -0.003 | 0.996 | 0.882 | 0.838 | 1 |

| **Z_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.927 | 0.120 | 0.940 | 0.921 | 0.914 | 0.950 |
| ECM | 0.927 | 1 | 0.120 | 0.936 | 1.000 | 1.000 | 0.978 |
| MVP | 0.120 | 0.120 | 1 | 0.147 | 0.127 | 0.126 | 0.139 |
| PCA | 0.940 | 0.936 | 0.147 | 1 | 0.934 | 0.924 | 0.983 |
| BTCD | 0.921 | 1.000 | 0.127 | 0.934 | 1 | 0.989 | 0.977 |
| DFO | 0.914 | 1.000 | 0.126 | 0.924 | 0.989 | 1 | 0.966 |
| MMSC | 0.950 | 0.978 | 0.139 | 0.983 | 0.977 | 0.966 | 1 |

| **EO_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.883 | 0.302 | 0.944 | 0.887 | 0.874 | 0.952 |
| ECM | 0.883 | 1 | 0.253 | 0.882 | 1.000 | 1.000 | 0.948 |
| MVP | 0.302 | 0.253 | 1 | 0.396 | 0.256 | 0.255 | 0.343 |
| PCA | 0.944 | 0.882 | 0.396 | 1 | 0.887 | 0.877 | 0.983 |
| BTCD | 0.887 | 1.000 | 0.256 | 0.887 | 1 | 0.980 | 0.951 |
| DFO | 0.874 | 1.000 | 0.255 | 0.877 | 0.980 | 1 | 0.937 |
| MMSC | 0.952 | 0.948 | 0.343 | 0.983 | 0.951 | 0.937 | 1 |

| **RTA_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.846 | 0.035 | 0.958 | 0.837 | 0.740 | 0.963 |
| ECM | 0.846 | 1 | -0.005 | 0.832 | 1.000 | 0.993 | 0.883 |
| MVP | 0.035 | -0.005 | 1 | 0.035 | 0.000 | 0.001 | 0.028 |
| PCA | 0.958 | 0.832 | 0.035 | 1 | 0.825 | 0.716 | 0.993 |
| BTCD | 0.837 | 1.000 | 0.000 | 0.825 | 1 | 0.847 | 0.878 |
| DFO | 0.740 | 0.993 | 0.001 | 0.716 | 0.847 | 1 | 0.759 |
| MMSC | 0.963 | 0.883 | 0.028 | 0.993 | 0.878 | 0.759 | 1 |

| **NX_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.838 | 0.018 | 0.927 | 0.833 | 0.728 | 0.925 |
| ECM | 0.838 | 1 | -0.029 | 0.879 | 1.000 | 0.956 | 0.926 |
| MVP | 0.018 | -0.029 | 1 | -0.006 | -0.027 | -0.031 | -0.010 |
| PCA | 0.927 | 0.879 | -0.006 | 1 | 0.878 | 0.745 | 0.990 |
| BTCD | 0.833 | 1.000 | -0.027 | 0.878 | 1 | 0.875 | 0.926 |
| DFO | 0.728 | 0.956 | -0.031 | 0.745 | 0.875 | 1 | 0.791 |
| MMSC | 0.925 | 0.926 | -0.010 | 0.990 | 0.926 | 0.791 | 1 |

| **FA_ES** | OLSD | ECM | MVP | PCA | BTCD | DFO | MMSC |
|---|---|---|---|---|---|---|---|
| OLSD | 1 | 0.918 | 0.010 | 0.974 | 0.917 | 0.798 | 0.974 |
| ECM | 0.918 | 1 | 0.033 | 0.925 | 1.000 | 0.995 | 0.947 |
| MVP | 0.010 | 0.033 | 1 | 0.036 | 0.034 | 0.045 | 0.029 |
| PCA | 0.974 | 0.925 | 0.036 | 1 | 0.923 | 0.807 | 0.997 |
| BTCD | 0.917 | 1.000 | 0.034 | 0.923 | 1 | 0.861 | 0.946 |
| DFO | 0.798 | 0.995 | 0.045 | 0.807 | 0.861 | 1 | 0.823 |
| MMSC | 0.974 | 0.947 | 0.029 | 0.997 | 0.946 | 0.823 | 1 |

*Table 11 – Correlation matrices of the time series of hedging vectors*

[19] Applying Fisher's transformation, a high correlation estimate such as 0.9 on 810 observations has a 99% confidence band between 0.881 and 0.916.

### 6.3.- TESTING FOR HEDGING ERRORS

Unstable methods are impracticable for operational (robustness) and economic reasons (rebalance cost) reasons. In the previous Section we have concluded that the three advanced methods (BTCD, DFO, MMSC) are the most stable among those similar.

The paper began by enunciating the hedging problem in terms of the minimizing the change in value of the *spread* over the hedged period (recall Eqs. (1)-(2)). We can finally turn our attention to answer that original question. Using the hedging ratios estimated in the previous Section, tables 12-14 assess the performance of the three distinct and stable methods in terms of the standard deviation of the hedging errors. *StDev(e1)x* is the standard deviation of the hedging error over one session, divided by the standard deviation of price changes for the first leg (which has a weight of 1) over one session.

$$StDev(e1)x = \frac{\sigma\big(e(1)\big)}{\sigma\big(P_{1,t} - P_{1,t-1}\big)} = \frac{\sigma\big(P_{1,t} - P_{1,t-1} + \sum_{n=2}^{N} \omega_n \big(P_{n,t} - P_{n,t-1}\big)\big)}{\sigma\big(P_{1,t} - P_{1,t-1}\big)} \tag{16}$$

Similarly, *StDev(e5)x* is the standard deviation of the hedging error over five sessions, divided by the standard deviation of price changes for the first leg over five sessions. In other words, we are measuring the dispersion of the hedging error relative to the exposure of remaining unhedged over the same horizon. The smallest these magnitudes, the better the hedge. A reading of zero would indicate a perfect hedge, and a reading of one would indicate that the hedge failed to deliver any protection.

$$StDev(e5)x = \frac{\sigma\big(e(5)\big)}{\sigma\big(P_{1,t} - P_{1,t-5}\big)} = \frac{\sigma\big(P_{1,t} - P_{1,t-5} + \sum_{n=2}^{N} \omega_n \big(P_{n,t} - P_{n,t-5}\big)\big)}{\sigma\big(P_{1,t} - P_{1,t-5}\big)} \tag{17}$$

| Portf. 1 | Portf. 2 | StDev (e1)x | StDev (e2)x | StDev (e3)x | StDev (e4)x | StDev (e5)x |
|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 0.28 | 0.26 | 0.26 | 0.26 | 0.26 |
| NQ1 Index | ES1 Index | 0.60 | 0.56 | 0.49 | 0.43 | 0.39 |
| VG1 Index | ES1 Index | 0.44 | 0.46 | 0.46 | 0.46 | 0.46 |
| GX1 Index | ES1 Index | 0.68 | 0.57 | 0.54 | 0.52 | 0.52 |
| CF1 Index | ES1 Index | 0.42 | 0.26 | 0.44 | 0.44 | 0.44 |
| Z 1 Index | ES1 Index | 0.42 | 0.40 | 0.42 | 0.46 | 0.41 |
| EO1 Index | ES1 Index | 0.41 | 0.42 | 0.42 | 0.42 | 0.42 |
| RTA1 Index | ES1 Index | 0.40 | 0.41 | 0.41 | 0.40 | 0.40 |
| NX1 Index | ES1 Index | 0.51 | 0.52 | 0.59 | 0.58 | 0.63 |
| FA1 Index | ES1 Index | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 |

*Table 12 – BTCD hedge performance*

| Portf. 1 | Portf. 2 | StDev (e1)x | StDev (e2)x | StDev (e3)x | StDev (e4)x | StDev (e5)x |
|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 0.33 | 0.32 | 0.31 | 0.30 | 0.34 |
| NQ1 Index | ES1 Index | 0.55 | 0.52 | 0.46 | 0.47 | 0.44 |
| VG1 Index | ES1 Index | 0.51 | 0.46 | 0.46 | 0.47 | 0.50 |
| GX1 Index | ES1 Index | 0.70 | 0.59 | 0.58 | 0.54 | 0.60 |
| CF1 Index | ES1 Index | 0.43 | 0.32 | 0.45 | 0.45 | 0.46 |
| Z 1 Index | ES1 Index | 0.46 | 0.45 | 0.44 | 0.40 | 0.41 |
| EO1 Index | ES1 Index | 0.42 | 0.43 | 0.43 | 0.43 | 0.43 |
| RTA1 Index | ES1 Index | 0.50 | 0.47 | 0.50 | 0.44 | 0.43 |
| NX1 Index | ES1 Index | 0.56 | 0.55 | 0.58 | 0.57 | 0.57 |
| FA1 Index | ES1 Index | 0.36 | 0.36 | 0.36 | 0.35 | 0.37 |

*Table 13 – DFO hedge performance*

| Portf. 1 | Portf. 2 | StDev (e1)x | StDev (e2)x | StDev (e3)x | StDev (e4)x | StDev (e5)x |
|---|---|---|---|---|---|---|
| DM1 Index | ES1 Index | 0.24 | 0.22 | 0.21 | 0.21 | 0.21 |
| NQ1 Index | ES1 Index | 0.40 | 0.41 | 0.41 | 0.41 | 0.41 |
| VG1 Index | ES1 Index | 0.40 | 0.44 | 0.45 | 0.47 | 0.48 |
| GX1 Index | ES1 Index | 0.73 | 0.62 | 0.60 | 0.57 | 0.56 |
| CF1 Index | ES1 Index | 0.40 | 0.22 | 0.43 | 0.43 | 0.45 |
| Z 1 Index | ES1 Index | 0.41 | 0.40 | 0.39 | 0.38 | 0.38 |
| EO1 Index | ES1 Index | 0.41 | 0.42 | 0.42 | 0.42 | 0.42 |
| RTA1 Index | ES1 Index | 0.39 | 0.39 | 0.39 | 0.38 | 0.38 |
| NX1 Index | ES1 Index | 0.54 | 0.55 | 0.54 | 0.54 | 0.54 |
| FA1 Index | ES1 Index | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 |

*Table 14 – MMSC hedge performance*

For example, consider the reading of *StDev(e1)x* for the first spread in Table 14. This means the MMSC's standard deviation of the hedging error over one session ($\sigma\left(e(1)\right)$) was 0.24 times the standard deviation of being unhedged ($\sigma\left(P_{1,t} - P_{1,t-1}\right)$). MMSC removed more than ¾ of the exposure we had to the first leg (DM1 Index).

Average *StDev(e1)x* are respectively 0.42, 0.45 and 0.48 for MMSC, BTDC and DFO. Although MMSC does perform better, the difference is not statistically significant at any reasonable confidence level.

Tables 15-16 provide the ranking of the three advanced hedging methods. MMSC delivers the best hedging performance over the horizon of one session, beating the two multi-period methods. However, as the horizon increases to five sessions (equivalent to one week), BTCD improves while MMSC worsens. This is again consistent with the theory presented in previous Sections. Because single-period methods do not take into account the serial conditionality of the hedging error, their performance is expected to worsen as the horizon increases.

| Portf. 1 | Portf. 2 | BTCD | DFO | MMSC |
|---|---|---|---|---|
| DM1 Index | ES1 Index | 2 | 3 | 1 |
| NQ1 Index | ES1 Index | 3 | 2 | 1 |
| VG1 Index | ES1 Index | 2 | 3 | 1 |
| GX1 Index | ES1 Index | 1 | 2 | 3 |
| CF1 Index | ES1 Index | 2 | 3 | 1 |
| Z 1 Index | ES1 Index | 2 | 3 | 1 |
| EO1 Index | ES1 Index | 2 | 3 | 1 |
| RTA1 Index | ES1 Index | 2 | 3 | 1 |
| NX1 Index | ES1 Index | 1 | 3 | 2 |
| FA1 Index | ES1 Index | 2 | 3 | 1 |
| **Average rank** | | 1.90 | 2.80 | 1.30 |

*Table 15 – Ranking advanced methods in terms of the standard deviation of the hedging error over one session*

| Portf. 1 | Portf. 2 | BTCD | DFO | MMSC |
|---|---|---|---|---|
| DM1 Index | ES1 Index | 2 | 3 | 1 |
| NQ1 Index | ES1 Index | 1 | 3 | 2 |
| VG1 Index | ES1 Index | 1 | 3 | 2 |
| GX1 Index | ES1 Index | 1 | 3 | 2 |
| CF1 Index | ES1 Index | 1 | 3 | 2 |
| Z 1 Index | ES1 Index | 3 | 2 | 1 |
| EO1 Index | ES1 Index | 2 | 3 | 1 |
| RTA1 Index | ES1 Index | 2 | 3 | 1 |
| NX1 Index | ES1 Index | 3 | 2 | 1 |
| FA1 Index | ES1 Index | 1 | 3 | 2 |
| **Average rank** | | 1.70 | 2.80 | 1.50 |

*Table 16 – Ranking advanced methods in terms of the standard deviation of the hedging error over five sessions (one week)*

As usual, we caution against overstretching the conclusions of this study. We have focused our attention on the 11 most liquid Index Futures. These methodologies may perform differently for other asset classes and sample lengths. In a coming paper we will expand our study to spreads with more than two legs.

## 7.- CONCLUSIONS

After characterizing the hedging problem, we have proposed a taxonomy of existing methodologies. Then, we have introduced two novel hedging procedures, Dickey-Fuller Optimal (DFO) and Mini-Max Subset Correlation (MMSC), and generalized an existing one, Box-Tiao Canonical Decomposition (BTCD). The first one estimates the vector of holdings that delivers a hedging error with the lowest probability of having a unit root. This is a useful property, as it limits the magnitude of the cumulative hedging errors. The second one computes a Maeloc spread, which is characterized by the holdings that generate the most orthogonal subset components. Thus, no particular holding or *subset* of holdings dominates the hedging error.

We evaluate traditional and advanced hedging methods in two stages: First, we wish to identify which ones are robust (static stability) and practicable (dynamic stability).

Second, among the robust and practicable, we would like to discern what methods deliver the lowest standard deviation of hedging error. Regarding the first stage, historical backtests show that DFO delivers estimates close to those derived by Error Correction Method (ECM), although the estimates from the first are more stable over time. For the same reason, we should prefer MMSC estimates over Principal Component Analysis' (PCA). DFO and MMSC yield distinct results, mutually and compared to BTCD. Of the seven hedging procedures discussed, we advocate for applying the last three (BTCD, DFO, MMSC) and disregard the other four (OLSD, ECM, MVP, PCA). Regarding the second stage, for the 11 most liquid Index Futures we find that all three advanced hedging methods perform well. Researchers may prefer one of the advanced methods over the others from a theoretical, technical or practical perspective, but as it relates to hedging performance, BTCD, DFO and MMSC deliver similar standard deviation of hedging errors.

## APPENDIX

### A.1.- SPECIFICATION OF THE SIMPLE ERROR CORRECTION MODEL

The starting point is a proportional, long-run equilibrium relationship between the market values of the portfolio we wish to hedge and the hedging portfolio.

$$P_{1,t} = KP_{2,t} \tag{18}$$

where K is the constant of proportionality. In log form, $p_{1,t} = k + p_{2,t}$, where the lower case indicates the natural logarithm of the variables in upper case. The dynamic relationship between $p_1$ and $p_2$ can be represented as:

$$p_{1,t} = \beta_0 + \beta_1 p_{2,t} + \beta_2 p_{2,t-1} + \alpha_1 p_{1,t-1} + \varepsilon_t \tag{19}$$

In order for this dynamic equation to converge to the long-run equilibrium $(p_1^*, p_2^*)$, it must occur that

$$p_1^* = \beta_0 + \beta_1 p_2^* + \beta_2 p_2^* + \alpha_1 p_1^* \tag{20}$$

which leads to

$$p_1^* = \frac{\beta_0}{1-\alpha_1} + \frac{\beta_1 + \beta_2}{1-\alpha_1} p_2^* \tag{21}$$

and sets the general equilibrium conditions as

$$k = \frac{\beta_0}{1-\alpha_1} \tag{22}$$
$$\beta_1 + \beta_2 = 1 - \alpha_1$$

Let's define $\gamma \equiv \beta_1 + \beta_2$. Under such equilibrium condition, this implies that $\beta_2 = \gamma - \beta_1$ and $\alpha_1 = 1 - \gamma$. Then, our general dynamic equation can be re-written as:

$$\Delta p_{1,t} = \beta_0 + \beta_1 \Delta p_{2,t} + \gamma\left(p_{2,t-1} - p_{1,t-1}\right) + \varepsilon_t \tag{23}$$

where $\gamma\left(p_{2,t-1} - p_{1,t}\right)$ is the "error correction" that over time corrects the cumulative hedging errors, hence ensuring the convergence of the spread towards the long run equilibrium.

$\gamma > 0$, because in absence of disturbances $(\Delta p_{2,t}, \varepsilon_t)$, $p_1$ should converge towards its equilibrium level. Let's see what occurs when we set $\Delta p_{2,t} = 0$, $\varepsilon_t = 0$. Then, $\Delta p_{1,t} = \beta_0 + \gamma\left(p_{2,t-1} - p_{1,t-1}\right)$. Applying the equilibrium conditions, this leaves us with $\Delta p_{1,t} = \gamma\left(k + p_{2,t-1} - p_{1,t-1}\right)$, where $k + p_{2,t-1}$ happens to be the equilibrium value of $p_1$ for observation *t-1*. If $k + p_{2,t-1} - p_{1,t-1} > 0$, then $p_1$ fell short of its equilibrium level in *t-1*, in which case the error correction should compensate for the difference (i.e., $\gamma$ ought to be positive). This has the important consequence that a test of significance on $\gamma$ should be one-tailed, with $H_0\colon \gamma \leq 0$.

How does this relate to the OLSD model? Consider the case that $k + p_{2,t-1} = p_{1,t-1}$, i.e. the model reached the equilibrium in observation *t-1*. In absence of disturbance, this implies that $\Delta p_{1,t} = \beta_0 + \beta_1 \Delta p_{2,t} + \gamma(p_{2,t-1} - k - p_{2,t-1})$, which after a few operations reduces to $\Delta p_{1,t} = \beta_1 \Delta p_{2,t}$. This illustrates the fact that an OLSD model incorporates the unlikely assumption that the spread is already in equilibrium and it won't be disturbed.

Finally, the hedge is characterized by the holdings $(\omega_1, \omega_2) = (1, -K)$, where $K = e^{\frac{\beta_0}{\gamma}}$.

## A.2.- DERIVATIVES OF THE DF STATISTIC

### A.2.1.- FIRST DERIVATIVE

$$\sigma_{\Delta S}^2 = \sum_{j=1}^{I}\sum_{k=1}^{I} \omega_j \ \omega_k \ \sigma_{\Delta \mathrm{P}_j, \Delta \mathrm{P}_k} = \omega_i^2 \sigma_i^2 + 2\omega_i \sum_{j \neq i}^{I} \omega_j \sigma_{\Delta \mathrm{P}_i, \Delta \mathrm{P}_j} + \sum_{j \neq i}^{I}\sum_{k \neq i}^{I} \omega_j \ \omega_k \ \sigma_{\Delta \mathrm{P}_j, \Delta \mathrm{P}_k} \tag{24}$$

$$\sigma_{L(S)}^2 = \omega_i^2 \sigma_i^2 + 2\omega_i \sum_{j \neq i}^{I} \omega_j \sigma_{L(\mathrm{P}_i), L(\mathrm{P}_j)} + \sum_{j \neq i}^{I}\sum_{k \neq i}^{I} \omega_j \ \omega_k \ \sigma_{L(\mathrm{P}_j), L(\mathrm{P}_k)} \tag{25}$$

$$\sigma_{\Delta S, L[S]} = \sum_{j=1}^{I}\sum_{k=1}^{I} \omega_j \ \omega_k \ \sigma_{\Delta \mathrm{P}_j, L(\mathrm{P}_k)} = \omega_i \sum_{j \neq i}^{I} \omega_j \ \left(\sigma_{\Delta \mathrm{P}_i, L(\mathrm{P}_j)} + \sigma_{\Delta \mathrm{P}_j, L(\mathrm{P}_i)}\right) + \sum_{j \neq i}^{I}\sum_{k \neq i}^{I} \omega_j \ \omega_k \ \sigma_{\Delta \mathrm{P}_j, L(\mathrm{P}_k)} \tag{26}$$

$$\frac{\partial \sigma_{\Delta S, L(S)}}{\partial \omega_i} = \sum_{j \neq i}^{I} \omega_j \ \left(\sigma_{\Delta \mathrm{P}_i, L(\mathrm{P}_j)} + \sigma_{\Delta \mathrm{P}_j, L(\mathrm{P}_i)}\right) \tag{27}$$

$$\frac{\partial \sigma_{\Delta S}}{\partial \omega_i} = \frac{1}{2}[\sigma_{\Delta S}^2]^{-\frac{1}{2}} \frac{\partial \sigma_{\Delta S}^2}{\partial \omega_i} \tag{28}$$

$$\frac{\partial \sigma_{\Delta S}^2}{\partial \omega_i} = 2\omega_i \ \sigma_{\Delta \mathrm{P}_i}^2 + 2\sum_{j \neq i}^{I} \omega_j \sigma_{\Delta \mathrm{P}_i, \Delta \mathrm{P}_j} \tag{29}$$

$$\frac{\partial \sigma_{L(S)}}{\partial \omega_i} = \frac{1}{2}\left[\sigma_{L(S)}^2\right]^{-\frac{1}{2}} \frac{\partial \sigma_{L(S)}^2}{\partial \omega_i} \tag{30}$$

$$\frac{\partial\sigma_{L(S)}^2}{\partial\omega_i} = 2\omega_i\ \sigma_{L(\mathrm{P}_i)}^2 + 2\sum_{j\neq i}^{I}\omega_j\sigma_{L(\mathrm{P}_i),L(\mathrm{P}_j)} \tag{31}$$

$$\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i} = \frac{\partial\sigma_{\Delta S}}{\partial\omega_i}\sigma_{L(S)} + \frac{\partial\sigma_{L(S)}}{\partial\omega_i}\sigma_{\Delta S} \tag{32}$$

$$\frac{\partial\rho_{\Delta S,L(S)}}{\partial\omega_i} = \frac{\frac{\partial\sigma_{\Delta S,L(S)}}{\partial\omega_i}\sigma_{\Delta S}\sigma_{L(S)} - \frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i}\rho_{\Delta S,L(S)}}{\sigma_{\Delta S}^2\sigma_{L(S)}^2} \tag{33}$$

$$\begin{aligned}\frac{\partial\widehat{DF}}{\partial\omega_i} &= \frac{\partial\rho_{\Delta S,L(S)}}{\partial\omega_i}\frac{\sqrt{\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}} + \frac{1}{(T-2)}\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{1}{2}}\rho_{\Delta S,L(S)}^2}{\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}} \\ &= \frac{\partial\rho_{\Delta S,L(S)}}{\partial\omega_i}\left(\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{1}{2}} \right. \\ &\quad \left. + \frac{1}{(T-2)}\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{3}{2}}\rho_{\Delta S,L(S)}^2\right)\end{aligned} \tag{34}$$

This gradient can be used to identify the set $\{\omega_i\}$ that delivers a $\{S_t\}$ with minimum DF Stat.

### A.2.2.- SECOND DERIVATIVE

$$\frac{\partial^2\sigma_{\Delta S,L(S)}}{\partial{\omega_i}^2} = 0 \tag{35}$$

$$\frac{\partial\left[\sigma_{\Delta S}^2\sigma_{L(S)}^2\right]}{\partial\omega_i} = 2\left[\sigma_{\Delta S}\sigma_{L(S)}\right]\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i} \tag{36}$$

$$\frac{\partial^2\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial{\omega_i}^2} = \frac{\partial^2\sigma_{\Delta S}}{\partial{\omega_i}^2}\sigma_{L(S)} + \frac{\partial^2\sigma_{L(S)}}{\partial{\omega_i}^2}\sigma_{\Delta S} + 2\frac{\partial\sigma_{L(S)}}{\partial\omega_i}\frac{\partial\sigma_{\Delta S}}{\partial\omega_i} \tag{37}$$

$$\frac{\partial^2\sigma_{\Delta S}}{\partial{\omega_i}^2} = -\frac{1}{8}[\sigma_{\Delta S}^2]^{-2}\left(\frac{\partial\sigma_{\Delta S}^2}{\partial\omega_i}\right)^2\frac{\partial^2\sigma_{\Delta S}^2}{\partial{\omega_i}^2} \tag{38}$$

$$\frac{\partial^2\sigma_{\Delta S}^2}{\partial{\omega_i}^2} = 2\sigma_{\Delta \mathrm{P}_i}^2 \tag{39}$$

$$\frac{\partial^2 \sigma_{L(S)}}{\partial \omega_i{}^2} = -\frac{1}{8}\left[\sigma_{L(S)}^2\right]^{-2}\left(\frac{\partial \sigma_{L(S)}^2}{\partial \omega_i}\right)^2 \frac{\partial^2 \sigma_{L(S)}^2}{\partial \omega_i{}^2} \tag{40}$$

$$\frac{\partial^2 \sigma_{L(S)}^2}{\partial \omega_i{}^2} = 2\sigma_{L(\mathrm{P}_i)}^2 \tag{41}$$

$$\begin{aligned}\frac{\partial^2 \rho_{\Delta S,L(S)}}{\partial \omega_i{}^2} = \left(\sigma_{\Delta S}\sigma_{L(S)}\right)^{-4}\Bigg[\Bigg(&\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i}\frac{\partial \sigma_{\Delta S,L(S)}}{\partial \omega_i} \\ &- \frac{\partial^2\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i{}^2}\rho_{\Delta S,L(S)} \\ &- \frac{\partial \rho_{\Delta S,L(S)}}{\partial \omega_i}\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i}\Bigg)\sigma_{\Delta S}^2\sigma_{L(S)}^2 \\ &- \frac{\partial\left[\sigma_{\Delta S}^2\sigma_{L(S)}^2\right]}{\partial \omega_i}\Bigg(\frac{\partial \sigma_{\Delta S,L(S)}}{\partial \omega_i}\sigma_{\Delta S}\sigma_{L(S)} \\ &- \frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i}\rho_{\Delta S,L(S)}\Bigg)\Bigg]\end{aligned} \tag{42}$$

$$\begin{aligned}\frac{\partial^2 \widehat{DF}}{\partial \omega_i{}^2} = \frac{\partial^2 \rho_{\Delta S,L(S)}}{\partial \omega_i{}^2}\Bigg(&\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{1}{2}} \\ &+ \frac{1}{(T-2)}\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{3}{2}}\rho_{\Delta S,L(S)}^2\Bigg) \\ &+ \frac{\partial \rho_{\Delta S,L(S)}}{\partial \omega_i}\Bigg(\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{3}{2}}\rho_{\Delta S,L(S)}\frac{\partial \rho_{\Delta S,L(S)}}{\partial \omega_i} \\ &+ \left(\frac{3}{(T-2)}\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{5}{2}} 2\rho_{\Delta S,L(S)}\frac{\partial \rho_{\Delta S,L(S)}}{\partial \omega_i}\right)\rho_{\Delta S,L(S)}^2 \\ &+ \frac{2}{T-2}\rho_{\Delta S,L(S)}\frac{\partial \rho_{\Delta S,L(S)}}{\partial \omega_i}\left(\frac{1-\rho_{\Delta S,L(S)}^2}{T-2}\right)^{-\frac{3}{2}}\Bigg)\end{aligned} \tag{43}$$

## A.3.- GRADIENT OPTIMIZATION OF MAELOC SPREADS

Let $S_t = \sum_{i=1}^{N} w_i \cdot P_{i,t}$. Taylor's expansion on $\rho_{\Delta S,w_j\Delta P_j}\left(w_i + \Delta w_i\right) = \sum_{n=0}^{\infty}\frac{\partial^n \rho_{\Delta S,w_j\Delta P_j}}{\partial w_i{}^n}\cdot\frac{\left(\Delta w_i\right)^n}{n!}$. Ignoring the residual beyond the second term, this reduces to

$$\rho_{\Delta S,w_j\Delta P_j}\left(w_i + \Delta w_i\right) = \rho_{\Delta S,w_j\Delta P_j}\left(w_i\right) + \frac{\partial \rho_{\Delta S,w_j\Delta P_j}}{\partial w_i}\Delta w_i + \frac{1}{2}\frac{\partial^2 \rho_{\Delta S,w_j\Delta P_j}}{\partial w_i{}^2}\Delta w_i{}^2 \tag{44}$$

We need to compute the first two partial derivatives.

**A.3.1.- FIRST DERIVATIVE**

We'll derive $\frac{\partial\rho(\Delta S, w_i\Delta P_i)}{\partial w_i}$ as follows:

1. $\sigma^2_{w_i\Delta P_i} = w_i^2\sigma_i^2$
2. $\sigma^2_{\Delta S} = \sum_{k=1}^{N}\sum_{j=1}^{N} w_k w_j \sigma_{kj} = w_i^2\sigma_i^2 + 2w_i\left[\sum_{j\neq i}^{N} w_j\sigma_{ij}\right] + \sum_{k\neq i}^{N}\sum_{j\neq i}^{N} w_k w_j \sigma_{kj}$
3. $\sigma_{\Delta S, w_i\Delta P_i} = w_i^2\sigma_i^2 + w_i\left[\sum_{j\neq i}^{N} w_j\sigma_{ij}\right] = w_i\left(w_i\,\sigma_i^2 + \sum_{j\neq i}^{N} w_j\sigma_{ij}\right) = \sigma_{\Delta S}\, w_i\sigma_i\rho_{\Delta S, w_i\Delta P_i}$
4. $\frac{\partial\sigma^2_{\Delta S}}{\partial w_i} = 2\left(w_i\,\sigma_i^2 + \sum_{j\neq i}^{N} w_j\sigma_{ij}\right) = \frac{2\sigma_{\Delta S, w_i\Delta P_i}}{w_i}$
5. $\frac{\partial\sigma_{\Delta S}}{\partial w_i} = \frac{1}{2}\left[\sigma^2_{\Delta S}\right]^{-\frac{1}{2}}\cdot\frac{\partial\sigma^2_{\Delta S}}{\partial w_i} = \frac{\sigma_{\Delta S, w_i\Delta P_i}}{w_i\cdot\sigma_{\Delta S}} = \rho_{\Delta S, w_i\Delta P_i}\cdot\sigma_i$
6. $\frac{\partial\sigma_{\Delta S, w_i\Delta P_i}}{\partial w_i} = 2w_i\,\sigma_i^2 + \sum_{j\neq i}^{N} w_j\sigma_{ij} = \frac{\sigma_{\Delta S, w_i\Delta P_i}}{w_i} + w_i\,\sigma_i^2$
7. $\frac{\partial\left[\sigma_{\Delta S}\cdot\sigma_{w_i\Delta P_i}\right]}{\partial w_i} = \frac{\partial\left[\sigma_{\Delta S}\cdot w_i\sigma_i\right]}{\partial w_i} = \sigma_i\left(\frac{\partial\sigma_{\Delta S}}{\partial w_i}\cdot w_i + \sigma_{\Delta S}\right) = \rho_{\Delta S, w_i\Delta P_i}\cdot w_i\sigma_i^2 + \sigma_S\sigma_i$
8. $\frac{\partial\rho_{\Delta S, w_i\Delta P_i}}{\partial w_i} = \frac{\partial\left[\frac{\sigma_{\Delta S, w_i\Delta P_i}}{\sigma_{\Delta S}\cdot\sigma_{w_i\Delta P_i}}\right]}{\partial w_i} = \frac{\frac{\partial\sigma_{\Delta S, w_i\Delta P_i}}{\partial w_i}\sigma_{\Delta S}\, w_i\sigma_i - \sigma_{\Delta S, w_i\Delta P_i}\cdot\left[\rho_{\Delta S, w_i\Delta P_i}\cdot w_i\sigma_i^2 + \sigma_{\Delta S}\sigma_i\right]}{\sigma^2_{\Delta S}w_i^2\sigma_i^2} =$

$$= \frac{\sigma_{\Delta S}w_i\,\sigma_i - \sigma_{\Delta S, w_i\Delta P_i}\cdot\rho_{\Delta S, w_i\Delta P_i}}{\sigma^2_{\Delta S}w_i} = \frac{\sigma_{\Delta S}w_i\,\sigma_i - \sigma_{\Delta S}w_i\sigma_i\rho^2_{\Delta S, w_i\Delta P_i}}{\sigma^2_{\Delta S}w_i} = \frac{\sigma_i}{\sigma_{\Delta S}}\left[1-\rho^2_{\Delta S, w_i\Delta P_i}\right]$$

And since $\frac{\partial\rho_{\Delta S, w_i\Delta P_i}}{\partial w_i} = \frac{\sigma_i}{\sigma_{\Delta S}}\left[1-\rho^2_{\Delta S, w_i\Delta P_i}\right]$, we can use $w_i$ to regulate $\partial\rho_{\Delta S, w_i\Delta P_i}$ by applying $\partial w_i = \partial\rho_{\Delta S, w_i\Delta P_i}\frac{\sigma_{\Delta S}}{\sigma_i\left[1-\rho^2_{\Delta S, w_i\Delta P_i}\right]}$. $w_i^* = w_i + \partial w_i$, where $w_i^*$ is the seed for the next iteration.

Note that the reason for $\frac{\partial\rho_{\Delta S, w_i\Delta P_i}}{\partial w_i} \geq 0$ is that, although $-1 \leq \rho_{\Delta S, \Delta P_i} \leq 1$, it must happen that $0 \leq \rho_{\Delta S, w_i\Delta P_i} \leq 1$ and $\rho_{\Delta S, w_i\Delta P_i} = \left|\rho_{\Delta S, \Delta P_i}\right|$.[20]

In order to control the cross effects on correlation, $\frac{\partial\rho_{\Delta S, w_j\Delta P_j}}{\partial w_i}$:

[20] This can be seen from the linear relation $\Delta S = \Delta P w$.

1. $\sigma_{\Delta S,w_j\Delta P_j} = w_j^2\sigma_j^2 + w_j\left[w_i\,\sigma_{ij} + \sum_{\substack{k\neq j\\k\neq i}}^{N} w_k\sigma_{kj}\right]$

2. $\frac{\partial\sigma_{\Delta S,w_j\Delta P_j}}{\partial w_i} = w_j\sigma_{ij}$

3. $\frac{\partial\left[\sigma_{\Delta S}\cdot\sigma_{w_j\Delta P_j}\right]}{\partial w_i} = \frac{\partial\left[\sigma_{\Delta S}\cdot w_j\sigma_j\right]}{\partial w_i} = \frac{\partial\sigma_{\Delta S}}{\partial w_i}w_j\sigma_j = w_j\rho_{\Delta S,w_i\Delta P_i}\sigma_i\sigma_j$

4. $\frac{\partial\rho_{\Delta S,w_j\Delta P_j}}{\partial w_i} = \frac{\partial\left[\frac{\sigma_{\Delta S,w_j\Delta P_j}}{\sigma_{\Delta S}\cdot\sigma_{w_j\Delta P_j}}\right]}{\partial w_i} = \frac{\frac{\partial\sigma_{\Delta S,w_j\Delta P_j}}{\partial w_i}\sigma_{\Delta S}w_j\sigma_j - \sigma_{\Delta S,w_j\Delta P_j}\cdot w_j\rho_{\Delta S,w_i\Delta P_i}\sigma_i\sigma_j}{\sigma_{\Delta S}^2 w_j^2\sigma_j^2} =$

$= \frac{\sigma_{\Delta S}w_j^2\sigma_j\sigma_{ij} - \sigma_{\Delta S}w_j\sigma_j\rho_{\Delta S,w_j\Delta P_j}w_j\rho_{\Delta S,w_i\Delta P_i}\sigma_i\sigma_j}{\sigma_{\Delta S}^2 w_j^2\sigma_j^2} = \frac{\sigma_i\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]}{\sigma_{\Delta S}}$

Therefore, $\partial w_i = \partial\rho_{\Delta S,w_j\Delta P_j}\frac{\sigma_{\Delta S}}{\sigma_i\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]}$. This makes possible to adjust the correlation between the spread and a leg *j* by changing any other leg *i*. This will prove useful in presence of constraints.

### A.3.2.- SECOND DERIVATIVE

$$\frac{\partial\left[\rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]}{\partial w_i} = \frac{\partial\rho_{\Delta S,w_j\Delta P_j}}{\partial w_i}\rho_{\Delta S,w_i\Delta P_i} + \frac{\partial\rho_{\Delta S,w_i\Delta P_i}}{\partial w_i}\rho_{\Delta S,w_j\Delta P_j} =$$

1. $= \frac{\sigma_i\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]}{\sigma_{\Delta S}}\rho_{\Delta S,w_i\Delta P_i} + \frac{\sigma_i}{\sigma_{\Delta S}}\left[1 - \rho_{\Delta S,w_i\Delta P_i}^2\right]\rho_{\Delta S,w_j\Delta P_j} =$

$$= \frac{\sigma_i}{\sigma_{\Delta S}}\left(\rho_{\Delta S,w_j\Delta P_j} + \rho_{\Delta S,w_i\Delta P_i}\left[\rho_{ij} - 2\rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]\right)$$

$$\frac{\partial^2 \rho_{\Delta S, w_j \Delta P_j}}{\partial w_i^2} = \frac{\partial \left[ \frac{\sigma_i \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}{\sigma_{\Delta S}} \right]}{\partial w_i} =$$

$$= \sigma_i \frac{\frac{\partial \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}{\partial w_i} \sigma_{\Delta S} - \frac{\partial \sigma_{\Delta S}}{\partial w_i} \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}{\sigma_{\Delta S}^2} =$$

2. $$= \sigma_i \frac{-\frac{\partial \left[\rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}{\partial w_i} \sigma_{\Delta S} - \rho_{\Delta S, w_i \Delta P_i} \sigma_i \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}{\sigma_{\Delta S}^2} =$$

$$= -\sigma_i^2 \frac{\rho_{\Delta S, w_j \Delta P_j} + \rho_{\Delta S, w_i \Delta P_i} \left[\rho_{ij} - 2\rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right] + \rho_{\Delta S, w_i \Delta P_i} \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}{\sigma_{\Delta S}^2} =$$

$$= -\frac{\sigma_i^2}{\sigma_{\Delta S}^2} \left[\rho_{\Delta S, w_j \Delta P_j} + \rho_{\Delta S, w_i \Delta P_i} \left(2\rho_{ij} - 3\rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right)\right]$$

3. $$\frac{\partial^2 \rho_{\Delta S, w_i \Delta P_i}}{\partial w_i^2} = -3 \frac{\sigma_i^2}{\sigma_{\Delta S}^2} \rho_{\Delta S, w_i \Delta P_i} \left[1 - \rho_{\Delta S, w_i \Delta P_i}^2\right]$$

### A.3.3.- TAYLOR'S EXPANSION

Let's denote $\Delta \rho_{\Delta S, w_j \Delta P_j} = \rho_{\Delta S, w_j \Delta P_j}(w_i + \Delta w_i) - \rho_{\Delta S, w_j \Delta P_j}(w_i)$. Substituting on Taylor's expansion,

$$\begin{aligned} &\Delta \rho_{\Delta S, w_j \Delta P_j} - \frac{\sigma_i}{\sigma_{\Delta S}} \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right] \Delta w_i + \\ &\frac{1}{2} \frac{\sigma_i^2}{\sigma_{\Delta S}^2} \left[\rho_{\Delta S, w_j \Delta P_j} + \rho_{\Delta S, w_i \Delta P_i} \left(2\rho_{ij} - 3\rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right)\right] \Delta w_i^2 = 0 \end{aligned} \quad (45)$$

which we solve as $\Delta w_i = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ for $a \neq 0$, where

$$\begin{aligned} a &= \frac{1}{2} \frac{\sigma_i^2}{\sigma_{\Delta S}^2} \left[\rho_{\Delta S, w_j \Delta P_j} + \rho_{\Delta S, w_i \Delta P_i} \left(2\rho_{ij} - 3\rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right)\right] \\ b &= -\frac{\sigma_i}{\sigma_{\Delta S}} \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right] \\ c &= \Delta \rho_{\Delta S, w_j \Delta P_j} \end{aligned} \quad (46)$$

The solution is two roots, $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, of which we use the one which produces the smallest $|\Delta w_i|$, i.e. $\Delta w_i \begin{cases} \frac{-b + \sqrt{b^2 - 4ac}}{2a} & if\, b \geq 0 \\ \frac{-b - \sqrt{b^2 - 4ac}}{2a} & if\, b < 0 \end{cases}$.[21]

And for *a=0*, the unique root is $\Delta w_i = -\frac{c}{b} = \Delta\rho_{\Delta S, w_j \Delta P_j} \frac{\sigma_{\Delta S}}{\sigma_i \left[\rho_{ij} - \rho_{\Delta S, w_j \Delta P_j} \rho_{\Delta S, w_i \Delta P_i}\right]}$.

**A.3.4.- BACKPROPAGATION FROM SUBSETS TO INSTRUMENTS**

Spreads can be thought as linear combinations of $N^*$ subsets of legs, rather than *n* instruments.

The series for subset *i* is $P_i^* = X\left(\Omega \bullet D_i^*\right)$, where $\Omega \bullet D_i^*$ is a *Hadamar* product between instruments' holdings $\boldsymbol{\Omega}$ and the *i*th-column of matrix subset definition $D^*$, $D_i^*$. $\boldsymbol{X}$ is the matrix of instruments' series. Let be $\boldsymbol{P^*}$ the matrix of $N^*$ subsets' series.

If we simply aggregate all subsets, we obtain $q \cdot S = P^* I_N$, where $q = \frac{1}{n} I_n' D^* I_{N^*}$. Denoting $w = I_{N^*}$, a ($N^*x1$) identity vector, then $q \cdot S = P^* w$.[22]

We have now defined the spread in terms of instruments, $S = X\Omega$, and subsets of legs, $q \cdot S = P^* w$. Expanding $\boldsymbol{P^*}$, $q \cdot S = \underbrace{X\left(\underbrace{\Omega \bullet D^*}_{nxN^*} \underbrace{w}_{N^*x1}\right)}_{nx1}$. This expression shows how to backpropagate changes in subsets' holdings into instruments' holdings.

$\Delta w_i$ is the change to subset *i*'s holdings that returns the desired $\rho_{\Delta S, w_j \Delta P_j}$. $w_i^* = w_i + \Delta w_i = 1 + \Delta w_i$, since each subset is set to a weight of 1. In order to backpropagate $\Delta w_i$ into instruments' holdings *Ω*, $\Omega \bullet D_i^* w_i^* = (1 + \Delta w_i)\Omega \bullet D_i^*$. In other words, we simply need to scale by $(1 + \Delta w_i)$ the holding on any instrument involved in subset *i*.[23]

---

[21] The reason is, since this is a Taylor expansion, we know the approximation error grows with $|\Delta w_i|$.

[22] Alternatively, $\Delta S = \Delta P^* w$ for a $w = \frac{1}{q} I_{N^*}$. Either definition will lead to identical results, since $\rho_{\Delta S, w_i \Delta P_i} = \rho_{q\Delta S, w_i \Delta P_i}$ for any *q>0*.

[23] Should all holdings be scaled, $\Omega^* = \Omega(1 + \Delta w_i)$, obviously nothing would change.

**A.3.5.- STEP SIZE**

At every iteration, we want to reduce the exposure to the subset that produces the $MSC = Max\left\{\left|\rho_{\Delta S,w_j\Delta P_j}\right|\right\}, j = 1,...,N^*\text{-}1$. Let's say that $j\left|MSC = \left|\rho_{\Delta S,w_j\Delta P_j}\right|\right.$. Any *i* subset containing no constrained instruments can be used to reduce *MSC*. *i* can be determined by rotation or searching for the unconstrained subset *i* with highest sensitivity to *j*.

Ideally, all *k-subset* correlations will converge to an average. This is guaranteed for *k=1*, but not for *k>1*[24]. We'll define $\overline{C} = \frac{1}{N^* - 1}\sum_{i=1}^{N^*-1}\rho_{\Delta S,w_i\Delta P_i}$ as the target for the next iteration, and $\Delta\rho_{\Delta S,w_j\Delta P_j} = \overline{C} - \rho_{\Delta S,w_j\Delta P_j}$.[25]

**A.3.6.- DEALING WITH CONSTRAINED INSTRUMENTS**

Any subset containing a constrained instrument shall not be iterated. Its exposure to the spread can be reduced by means of modifying another subset with no constrained instrument, using the cross-derivatives.

[24] Because $n = N^* - 1 = N - 1$.

[25] In practice, $\Delta\rho_{\Delta S,w_j\Delta P_j} = \left(\overline{C} - \rho_{\Delta S,w_j\Delta P_j}\right)\cdot\left(1 - \rho_{\Delta S,w_j\Delta P_j}{}^2\right)$ delivers a smoother convergence.

## REFERENCES


- Box, G.E. and G.C. Tiao (1977): *"A canonical analysis of multiple time series"*, Biometrika 64(2), 355.
- Clarke R., de Silva H., Thorley S. (2011): *"Minimum-Variance Portfolio Composition"*, The Journal of Portfolio Management, Winter.
- Choueifaty, Y. and Y. Coignard (2008): *"Toward Maximum Diversification"*, The Journal of Portfolio Management, Fall.
- d'Aspremont, Alexandre (2008): *"Identifying small mean reverting portfolios"*, Quantitative Finance (forthcoming).
- Dickey, D.A. and W.A. Fuller (1979): *"Distribution of the Estimators for Autoregressive Time Series with a Unit Root"*, Journal of the American Statistical Association, 74, pp. 427–431.
- Easley, D., M. López de Prado and M. O'Hara (2012): *"Flow Toxicity and Liquidity in a High Frequency World"*, Review of Financial Studies, forthcoming. http://ssrn.com/abstract=1695596
- Elliott, G., Rothenberg, T. J. & J. H. Stock (1996): *"Efficient Tests for an Autoregressive Unit Root"*, Econometrica, Vol. 64, No. 4., pp. 813–836.
- Engle, R. and C. Granger (1987): *"Cointegration and error correction: Representation, estimate and testing"*, Econometrica, 55, 251-276.
- Gosh, Asim (1993): *"Hedging with stock index futures: Estimation and forecasting with error correction model"*, The Journal of Futures Markets, October, p.743
- Grinold, R., R. Kahn (1999): *"Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk"*, McGraw-Hill, 2nd Edition.
- Johansen, S. (1991): *"Cointegration and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models"*, Econometrica, Vol.59, No.6, November, pp.1551-1580.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992): *"Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root"*, Journal of Econometrics 54, 159–178.
- Littermann, R. and J. Sheinkman (1991): *"Commons Factors Affecting Bond Returns"*, Journal of Fixed Income, June, pp 54-61.
- López de Prado, M. and C. Rodrigo (2004): *"Invertir en Hedge Funds"*, Ed. Díaz de Santos, Madrid.
- Lord, R. and A. Pelsser (2007): *"Level-slope-curvature: Fact or artifact?"*, Applied Mathematical Finance, Vol. 14, No. 2, May, pp. 105-130.
- Luo, Y., R. Cahan, J. Jussa, Z. Chen and M. Álvarez (2011): *"Minimum Variance: Exposing the 'magic'"*, Global Markets Quantitative Research, Deutsche Bank.
- Maddala, G.S. and I.M. Kim (2004): *"Unit roots, cointegration and structural change"*, Cambridge.
- Maillard, S., T. Roncalli and J. Teiletche (2010): *"On the properties of equally-weighted risk contribution portfolios"*, The Journal of Portfolio Management, Vol. 36, No. 4, pp. 60-70.
- Markowitz, H. (1952): *"Portfolio selection"*, Journal of Finance, 7, pp. 77-91.
- Meucci, A. (2005): *Risk and Asset Allocation*, Springer, pp. 445-450.

- Meucci, A. (2010): *“Managing Diversification”*, Risk, 22 (5), pp. 74-79.
- Moulton, P. and A. Seydoux (1998): *“Using Principal Components Analysis to Structure Butterfly Trades”*, Global Relative Value Research, Deutsche Bank.
- Newey, W. K. and K. West (1987): *"A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix"*, Econometrica, Econometric Society, vol. 55(3), May, pp. 703-08.
- Newey, W. K. and K. West (1994): *“Automatic lag selection in covariance matrix estimation”*, Review of Economic Studies 61, 631-653.
- Phillips, P.C.B and P. Perron (1988): *"Testing for a Unit Root in Time Series Regression"*, Biometrika, 75, 335–346
- Said, S.E. and D. A. Dickey (1984): *“Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order”*. Biometrika 71, 599–607
- Scherer, B. (2010): *“A New Look At Minimum Variance Investing”*, Working paper, SSRN: http://ssrn.com/abstract=1681306.
- Steeley, J. M. (1990): *“Modelling the dynamics of the term structure of interest rates”*, Economic and Social Review, 21, pp. 337–361.
- Stevens, G. (1998): *“On the Inverse of the Covariance Matrix in Portfolio Analysis”*, The Journal of Finance, Vol LIII, No. 5.
- Vidyamurthy, V. (2004): *“Pairs trading: Quantitative methods and analysis”*, Wiley Finance.
- Zellner (1962): *“An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”*, Journal of the American Statistical Association, Vol. 57, No. 298 (June), pp. 348-368.