ADVANCES IN COINTEGRATION AND SUBSET CORRELATION HEDGING METHODS

Marcos M. López de Prado[1]
RCC at Harvard University

First draft: May 2008
Last update: February 2011
Working paper

ABSTRACT

After reviewing some well-known hedging algorithms, we introduce two new procedures, called DFO and MMSC. The former is a cointegration method that estimates the hedging weights that are most likely to deliver a hedging error absent of unit root. The latter studies the geometry of the hedging errors and estimates a hedging vector such that its subsets are as orthogonal as possible to the error. Results indicate that DFO produces estimates similar to the ECM method, but more stable. Likewise, MMSC estimates are similar to PCA but more stable. BTCD estimates cannot be related to any of the aforementioned methodologies.

Keywords: Hedging portfolios, robustness, portfolio theory, stationarity, ECM, ADF, KPSS, PCA, BTCD, MMSC.

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

---

[1] Postdoctoral Research Fellow. Contact details at http://www.realcolegiocomplutense.harvard.edu

1.- INTRODUCTION

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, as the goal consists on defining a mimicking portfolio with minimum tracking error. The methods discussed here 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, like OLSD, ECM and MVP, 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 BTCD procedure is presented, which accepts any number of lags, regressors and forecast 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 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 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. The consequences are 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.

This paper is organized as follows. Section 2 reviews several hedging methods, pointing out their virtues and pitfalls. Section 3 introduces the DFO model. Section 4 introduces the MMSC model. Section 5 applies the seven hedging methodologies to pairs of index futures, pointing out which should be preferred in terms of robustness among those comparable. Section 6 outlines our conclusions. The Appendices complement the mathematical apparatus involved in these methods.

2.- 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 others.

We will assume that the set of instruments to form the hedge are predefined. If that's not the situation faced by the modeler, the selection could be done applying a standard factor selection algorithm[2] on the hedging procedures presented.

The hedging problem is posed in the following terms. Let $P_{1,t}$ represent the market value at time t of a portfolio we wish to hedge, with $t=1,\dots,T$. Unless otherwise stated, ΔP is assumed to be an invariant transformation, however the methods presented here can be applied to other invariant transformations. Provided a set of $n=2,\dots,N$ variables (instruments or portfolios) available for building a hedge, the hedging problem consists in computing the vectors of weights ω that is optimal according to a particular method. The combined position of portfolio plus hedge, $S_t = P_{1,t} + \sum_{n=2}^{N} \omega_n P_{n,t}$, is denoted spread, and $\varepsilon(h) = S_{T+h} - S_T$ is the hedging error after h observations.

2.1.- REGRESSION WEIGHTS

There are three basic OLS approaches to estimating a hedge: Differences, Levels, and Error Correction.

2.1.1.- 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, $n=1,\dots,N$ 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 (α) is statistically insignificant. The goodness of the fit can be evaluated through the adjusted R^2. $\omega_n = -\beta_n$, $n=2,\dots,N$.

In summary, this approach may be applied under the conditions that $\alpha \approx 0$ and ε is iid, with $\varepsilon \rightarrow N(0,\sigma_\varepsilon^2)$, $E\varepsilon_t\varepsilon_s = 0, \forall t \neq s$. This is extremely restrictive, as we are assuming 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[3]. In other words, this model fails to impose any condition on the behavior of the cumulative hedging errors, $\varepsilon(h) = S_{T+h} - S_T$, 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 ω in different directions.

---

[2] 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 ADF and KPSS stats in those analyses.

[3] S_t could be I(1).

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

2.1.2.- LEVELS (OLSL)

The target is to solve $P_{1,t} = \sum_{n=2}^{N} \beta_n P_{n,t} + S_t$, with $\omega_n = -\beta_n$, $n = 2, \dots, N$. 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), discussed in coming sections.

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 ω, 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 is not separated from the observed levels in the levels equation, the weights may not accurately capture the equilibrium relationship. That inconvenience is addressed by the ECM.

2.1.3.- 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

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

where p_1, p_2 are the natural logarithms of market values P_1, P_2, and γ must be tested to be positive ($H_0: \gamma \leq 0$). The spread is characterized by the weightings $(\omega_1, \omega_2) = (1, -K)$, where $K = e^{\frac{\beta_0}{\gamma}}$. 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 substantially dissimilar from the DFO and BTCD methods, discussed later.

2.2.- MINIMUM VARIANCE PORTFOLIOS (MVP)

First introduced by Markowitz (1952), they consist in solving the basic quadratic optimization problem, with a single linear constraint in equality[4]. Their popularity has grown ever since, with studies as recent as Clarke, de Silva and Thorley (2011) or Scherer (2010).

Let V be the covariance of matrix ΔP, 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).

---

[4] A constraint is needed to exclude the zero-weightings solution.

$$\begin{aligned} &\underset{\beta}{\text{Min}} \quad \beta' V \beta \\ &\text{s.t.} \quad \beta' a = 1 \end{aligned} \tag{2}$$

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

$$\begin{aligned} F1: \frac{\partial L(\beta,\lambda)}{\partial \beta} &= V\beta - \lambda a = 0 \\ F2: \frac{\partial L(\beta,\lambda)}{\partial \lambda} &= \beta' a - 1 = 0 \end{aligned} \tag{3}$$

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

Thus, $\beta = \frac{V^{-1} a}{a' V^{-1} a}$, and $\omega_j = \frac{\beta_j}{\beta_1}, j = 1, \ldots, N$, 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.

$$\begin{vmatrix} \frac{\partial^2 L(\beta,\lambda)}{\partial \beta^2} & \frac{\partial^2 L(\beta,\lambda)}{\partial \beta \partial \lambda} \\ \frac{\partial^2 L(\beta,\lambda)}{\partial \lambda \partial \beta} & \frac{\partial^2 L(\beta,\lambda)}{\partial \lambda^2} \end{vmatrix} = \begin{vmatrix} V' & -a' \\ a & 0 \end{vmatrix} = a'a \geq 0 \tag{4}$$

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).[5] 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). The approach presents similar caveats as the regression of differences (OLSD).

2.3.- 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.

[5] This is the 'fully invested portfolio constraint'. It is somewhat arbitrary to choose a vector of 1s, but any other number will simply re-scale the solution.

Let V be the NxN covariance of matrix ΔP, where the first column represents the covariances against the portfolio to be hedged. The target is to compute the vector of weightings β 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 β such that $W^{*'}\beta = 0_m$, where $W^{*'}$ is the transposed eigenvector matrix after having 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.

$W^{*'}\beta = 0_m$ is an homogeneous system with infinite non-trivial solutions, because $\mathrm{Rank}\left[W^{*'}\right] = m < N$. In order to find a single solution, we impose $\beta_i = 1$ on the N-m last columns. Let be $W^{*'*}$ the mxm matrix which results from moving the last N-m columns (numeraires) from $W^{*'}$ to the right side of the equation, which we denote $W^{*'-*}$. β^* is the submatrix of β 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, $\omega_j = \frac{\beta_j}{\beta_1}, j = 1, \ldots, N$.

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[6]. It could be argued however that, having the smallest variance (in differences), the stationarity of the eigenvectors with smallest eigenvalues is a minor concern[7]. This makes of PCA a valid, consistent method of hedging.

2.4.- BOX-TIAO CANONICAL DECOMPOSITION (BTCD)

The following method is derived from Box and Tiao (1977). Let's consider an AR(1) model on each time series of a set of variables, $j = 1, \ldots, J$, where j=1 corresponds to the portfolio to be hedged.

$$P_{t,j} = \sum_{i=1}^{J} \beta_{i,j} P_{t-1,i} + \varepsilon_{t,j} \tag{5}$$

We can fit a VAR(1) model[8,9] for the entire set: $\hat{\beta} = \left(P_{t-1}' P_{t-1}\right)^{-1} P_{t-1}' P_t$. For J=1, this reduces to $E\left[P_t^2\right] = E\left[\left(\beta P_{t-1}\right)^2\right] + E\left[\varepsilon_t^2\right]$. Box-Tiao defined a measure of predictability,

---

[6] The components we remained 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).

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

[8] As the explanatory variables are the same in each equation, the Multivariate Least Square is equivalent to the Ordinary least squares (OLS) estimator applied to each equation separately, as shown by Zellner (1962).

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

$\lambda = \frac{E_{t-1}[P_t^2]}{E[P_t^2]} = 1 - \frac{E[\varepsilon_t^2]}{E[P_t^2]}$, as a proxy for the mean reversion parameter of the Orstein-Uhlenbeck stochastic process. When λ is small, $E[\varepsilon_t^2]$ dominates $E_{t-1}[P_t^2]$ and P_t is almost pure noise. When λ is large, $E_{t-1}[P_t^2]$ dominates $E[\varepsilon_t^2]$ and P_t is almost perfectly predictable.

Likewise, we can derive a similar measure of predictability for 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 Jx1 vector Ω such that λ_Ω is minimized, i.e. $\underset{\Omega}{\text{Min}} \quad \frac{\Omega'\beta'\Gamma\beta\Omega}{\Omega'\Gamma\Omega}$. This is equivalent to solving the generalized eigenvalue problem in $\lambda_\Omega \in \Re$ characterized by $\det\left(\lambda_\Omega\Gamma - \beta'\Gamma\beta\right) = 0$.[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 $D = \left(C^{-1}\right)'AC^{-1}$, 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.[13] The same argument can be used to find the maximum value of $R(D,z)$.

---

[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, Γ can be defined as a covariance matrix of P_t.

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

[13] For a succinct proof, consider $\begin{array}{ll} \underset{x}{\text{Max}} & x'Ax \\ \text{s.t.} & x'x = 1 \end{array}$, where A is symmetric. Take derivatives on its Lagrangian $L(x) = x'Ax + \tilde{\lambda}(x'x - 1)$. The first order necessary condition, $\frac{\partial L(x)}{\partial x} = x'(A + A') + 2\tilde{\lambda}x' = 0 \Rightarrow Ax = \lambda x$, where $\lambda = -2\tilde{\lambda}$. $\frac{\partial L(x)}{\partial \tilde{\lambda}} = 0 \Rightarrow x'x = 1$. Furthermore, $x'Ax = x'\lambda x = \lambda$, thus all critical points (and extreme values in particular) are derived from computing the eigenvectors of $x'Ax$, and the stationary values from the respective eigenvalues.

Assuming that Γ is positive definite, the solution is $\Omega^* = \Gamma^{-\frac{1}{2}} z$, where z is the eigenvector corresponding to the smallest eigenvalue of the matrix $\Gamma^{-\frac{1}{2}} \beta' \Gamma \beta \Gamma^{-\frac{1}{2}}$.[14] Once Ω^* 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.

This procedure is not limited to using the 1-lagged forecasted variables as the only explanatory variables. Next, we generalize this method to any forecasting setup[15]:

1. Fit $\hat{\beta}$ on the forecasting equation.
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 Cholesky's decomposition on $\left(P_t' P_t\right)$ by applying $\left(P_t' P_t\right)^{-\frac{1}{2}} = W \Lambda^{-\frac{1}{2}} W'$.
6. Compute a PCA on $\left(P_t' P_t\right)^{-\frac{1}{2}} \left(\hat{P}_t' \hat{P}_t\right) \left(P_t' P_t\right)^{-\frac{1}{2}}$, which is the matrix D of the standard Rayleigh quotient.
7. Determine $\Omega^* = \left(P_t' P_t\right)^{-\frac{1}{2}} z$, where z is the eigenvector associated to the smallest eigenvalue (λ_Ω).
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 λ_Ω.
9. Set a unit position on the portfolio to be hedged (j=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 Γ and β 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 Γ (d'Aspremont, 2008), though not satisfactory solution seems available at the moment.

3.- DICKEY-FULLER OPTIMAL (DFO)

The target is to find a vector of weights ω 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

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

[15] This allows adding an intercept and additional lags to our specification.

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).

3.1.- 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 β, $DF = \frac{\hat{\beta}-1}{\sigma_{\hat{\beta}}}$, 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,

$$\widehat{DF} = \frac{\frac{\sigma_{\Delta S}\rho_{\Delta S, L(S)}}{\sigma_{L(S)}}}{\frac{\sigma_{\Delta S}}{\sigma_{L(S)}}\sqrt{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}}} = \frac{\rho_{\Delta S, L(S)}}{\sqrt{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}}} \tag{6}$$

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.

3.2.- DF STAT MINIMIZATION

The previous epigraph has shown how to estimate the DF stat directly. Considering that **S** is a linear function of ω, and DF a function of **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 ω that delivers a hedge with minimum DF:

$$\underset{\substack{\omega \\ \omega_0 = 1}}{Min} \frac{\rho_{\Delta s(\omega), L(s(\omega))}}{\sqrt{\frac{1-\rho^2_{\Delta s(\omega), L(s(\omega))}}{T-2}}} \tag{7}$$

The Appendix 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 2.1.1. For N sufficiently small, ω can be reliably computed through a brute force, grid search algorithm.

A similar procedure could be devised on the KPSS test for stationarity. We leave to the Appendix the discussion of the details.

4.- MINI-MAX SUBSET CORRELATION (MMSC)

Of the approaches discussed in the previous Section, three seemed particularly interesting. DFO searched for the linear combination of positions that exhibit the strongest mean reversion. PCA and BTCD looked deep into the geometry of the hedging set, and identified uncorrelated sources of variability responsible for most of the risk (or Principal 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 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 weightings such that the maximum correlation of any leg or subset of legs to the spread is minimal (thus the name Mini-Max Subset Correlation).

4.1.- MOTIVATION

Suppose a n-legged spread, characterized by its weights, ω, and the covariance matrix of value changes, V. The spread's risk can be decomposed in terms of its legs' contributions as

$$\sigma_{\Delta S}^2 = \omega' V \omega = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_i\ \sigma_{i,j}\,\omega_j \ = \sum_{i=1}^{n} \omega_i\ \sigma_{i,\Delta S} \\ = \sigma_{\Delta S} \sum_{i=1}^{n} \omega_i\ \sigma_i\ \rho_{i,\Delta S} \tag{8}$$

Therefore, the spread's risk is a weighted average of the instruments' standard deviations, where the weightings are a function of the correlations of each instrument to the spread.

$$\sigma_{\Delta S} = \sum_{i=1}^{n} \omega_i\ \sigma_i\ \rho_{i,\Delta S} \tag{9}$$

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

$$\omega_i\ \sigma_i\ \rho_{i,\Delta S} = \frac{\sigma_{\Delta S}}{n}, \forall i \tag{10}$$

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

4.2.- 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 weightings 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 $\mathbf{D}^{*}$ as the result of removing from matrix $\mathbf{D}$ any column $i \mid k < D_i' D_i < n$, where $1 \leq k < n$[16]. Likewise, $N^{*} = \sum_{i=1}^{k<n} \binom{n}{i} + 1 \leq N$.

4.3.- SUBSET COVARIANCE MATRIX (B)

Let $\mathbf{B}$ be a (N^{*}xN^{*}) matrix, $B_{p,q} = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i D^{*}_{i,p} \sigma_{i,j} D^{*}_{j,q} \omega_j = \omega' D^{*'} V D^{*} \omega$, V is the covariance matrix of ΔP, and $\sigma_{i,j}$ represents the covariance of changes between instruments i and j, $p=1, \ldots, N^{*}$, $q=1, \ldots, N^{*}$.

4.4.- SUBSET CORRELATION MATRIX (C)

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

4.5.- MAXIMUM SUBSET CORRELATION (MSC)

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

4.6.- MAELOC SPREAD

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

$$\begin{aligned} &\min_{\omega} \quad \mathrm{Max}\left\{ \left| C_{p,N^{*}} \right| \right\},\ p = 1, \ldots, N^{*}-1 \\ &\text{s.t.} \quad \omega_1 = 1 \end{aligned} \tag{11}$$

The spread's first leg corresponds to the portfolio we wish to hedge, and its weight 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 the Appendix.

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.

---

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

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

3. **Customizable**: The algorithm converges in presence of any number of weights constrained.
4. **Control over lead-lag effects**: The interval used to compute changes, on which the covariance matrix is computed, 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 weightings of the Maeloc-spread in order to provide a hedge.

4.7.- 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,

$$MLC = \operatorname{Min}\left\{ \left| C_{p,q} \right| \right\},\ p = 1,\ldots,N^{*}\text{-}1,\ q = 1,\ldots,N^{*}\text{-}1 \tag{12}$$

We must stress that MLC is computed after the Maeloc-spread has been determined. Figure 1 illustrates how the **C** matrix is divided into two regions where MSC and MLC are to be found. Consider a **C** matrix of a spread with 3 legs. A low MSC (maximum of the green 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 yellow area).

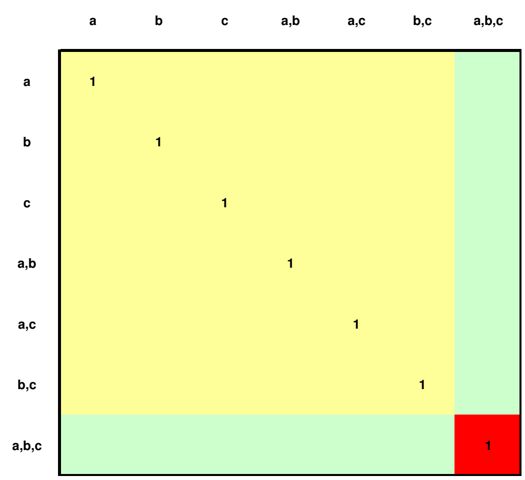

Figure 1 – The MSC and MLC regions of the C matrix

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 weights 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.

5.- EMPIRICAL RESULTS

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 January 1st 2008 to January 31st 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 sample by equidistant observation weights. We are now ready to apply the procedures.

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. Results are estimated over the entire sample (LR) and sequentially re-estimated every session over the prior 5 sessions (SR). LR w2 is the hedge estimated over the entire sample. Avg (SR w2) is the average value of the hedge as estimated every day over the previous 5 sessions. StDev (SR w2) is the standard deviation of the same. t-Stat is the fraction of the prior two. StDev (d1 SR w2) is standard deviation of the weights change from one day to the next. 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 (there will be no overlap between both samples).

Portf. 1	Portf. 2	LR w1	LR w2	Avg (Spr w2)	StDev (Spr w2)	t-Stat (Spr w2)	StDev (d1 Spr w2)	StDev (d2 Spr w2)	StDev (d3 Spr w2)	StDev (d4 Spr w2)	StDev (d5 Spr 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	2.48	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	0.83	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

Portf. 1	Portf. 2	LR w1	LR w2	Avg (Spr w2)	StDev (Spr w2)	t-Stat (Spr w2)	StDev (d1 Spr w2)	StDev (d2 Spr w2)	StDev (d3 Spr w2)	StDev (d4 Spr w2)	StDev (d5 Spr 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

Portf. 1	Portf. 2	LR w1	LR w2	Avg (Spr w2)	StDev (Spr w2)	t-Stat (Spr w2)	StDev (d1 Spr w2)	StDev (d2 Spr w2)	StDev (d3 Spr w2)	StDev (d4 Spr w2)	StDev (d5 Spr 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	48.17	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	49.96
NX1 Index	ES1 Index	1	11.07	8.00	178.92	0.04	271.81	255.08	248.39	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

Portf. 1	Portf. 2	LR w1	LR w2	Avg (Spr w2)	StDev (Spr w2)	t-Stat (Spr w2)	StDev (d1 Spr w2)	StDev (d2 Spr w2)	StDev (d3 Spr w2)	StDev (d4 Spr w2)	StDev (d5 Spr 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

Portf. 1	Portf. 2	LR w1	LR w2	Avg (Spr w2)	StDev (Spr w2)	t-Stat (Spr w2)	StDev (d1 Spr w2)	StDev (d2 Spr w2)	StDev (d3 Spr w2)	StDev (d4 Spr w2)	StDev (d5 Spr w2)
DM1 Index	ES1 Index	1	-0.94	-0.94	0.03	-35.67	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	-22.08	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

Port 1	Port 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	-0.53	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	-0.66	0.20	-3.33	0.26	0.25	0.25	0.25	0.25
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	-0.51	0.16	-3.60	0.15	0.22	0.22	0.22	0.22
Z 1 Index	ES1 Index	1	-1.31	-1.31	0.38	-4.16	0.32	0.32	0.32	0.32	0.32
EO1 Index	ES1 Index	1	-1.09	-1.38	0.22	-7.32	0.26	0.30	0.30	0.30	0.30
RTA1 Index	ES1 Index	1	-1.45	-1.13	0.42	-2.68	0.36	0.60	0.60	0.60	0.60
NX1 Index	ES1 Index	1	-0.03	-0.66	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

Port 1	Port 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	-0.65	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	-5.88	0.05	0.11	0.15	0.17	0.15
GX1 Index	ES1 Index	1	-4.12	-4.27	0.98	-4.34	0.46	0.73	0.91	1.07	1.15
CF1 Index	ES1 Index	1	-1.16	-1.15	0.23	-5.08	0.15	0.17	0.24	0.28	0.26
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	0.16	0.21	0.24	0.26
RTA1 Index	ES1 Index	1	-1.36	-1.32	0.22	-6.05	0.06	0.10	0.11	0.13	0.14
NX1 Index	ES1 Index	1	-1.16	-1.16	0.23	-5.15	0.15	0.20	0.24	0.27	0.26
FA1 Index	ES1 Index	1	-1.40	-1.46	0.16	-8.13	0.07	0.06	0.05	0.10	0.11

Table 7 – MMSC hedge

Short run Avg DFO and Avg ECM vectors are very close, although DFO seems to provide more robust results. Likewise, short run Avg MMSC and Avg PCA vectors are very similar, with MMSC delivering more robust estimates. These similarities are not surprising. 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 weights that are most orthogonal to the hedging error (in terms of positions and subsets of positions).

The previous remarks are confirmed by computing the correlation between procedures on the daily re-estimated hedging vectors (out-of-sample).

DM1_ES	OLS	ECM	MVP	PCA	[illegible]	[illegible]	MMSC
OLS	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

NQ1_ES	OLS	ECM	MVP	PCA	[illegible]	[illegible]	MMSC
OLS	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

VG1_ES	OLS	ECM	MVP	PCA	[illegible]	[illegible]	MMSC
OLS	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

GX1_ES	OLS	ECM	MVP	PCA	[illegible]	[illegible]	MMSC
OLS	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

[illegible]	OLS	ECM	MVP	PCA	[illegible]	[illegible]	MMSC
OLS	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

[illegible]	OLS	ECM	MVP	PCA	[illegible]	[illegible]	MMSC
OLS	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

[illegible]	OLSD	ECM	MVP	PCA	BTCD	DFO	MMSC
OLSD	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
BTCD	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
DFO	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

[illegible]	OLSD	ECM	MVP	PCA	BTCD	DFO	MMSC
OLSD	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
BTCD	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
DFO	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

[illegible]	OLSD	ECM	MVP	PCA	BTCD	DFO	MMSC
OLSD	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
BTCD	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
DFO	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

[illegible]	OLSD	ECM	MVP	PCA	BTCD	DFO	MMSC
OLSD	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ECM	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MVP	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]
PCA	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]	[illegible]
BTCD	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
DFO	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]
MMSC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

Table 8 – Correlation matrices of the time series of hedging vectors

In summary, Table 8 indicates that we should prefer DFO over ECM and MMSC over PCA, as they yield very similar results with the first of each couple delivering more stable estimates. MVP is the most unstable of all procedures, and OLSD's theoretical inconsistencies make it an unreliable choice. BTCD is another robust procedure that is not highly correlated with DFO or MMSC. Hence, we advocate for BTCD, DFO and MMSC as efficient, mutually different hedging procedures.

6.- CONCLUSIONS

We have introduced two novel hedging procedures, DFO and MMSC. Historical backtests show that DFO delivers estimates close to those derived by ECM, although the estimates from the first are more stable over time. For the same reason, we should prefer MMSC estimates over PCA's. 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).

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.

$$P_{1,t} = KP_{2,t} \tag{13}$$

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

$$p_{1,t} = \beta_0 + \beta_1 p_{2,t} + \beta_2 p_{2,t-1} + \alpha_1 p_{1,t-1} + \varepsilon_t \tag{14}$$

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

$$p_1^* = \beta_0 + \beta_1 p_2^* + \beta_2 p_2^* + \alpha_1 p_1^* \tag{15}$$

which leads to

$$p_1^* = \frac{\beta_0}{1-\alpha_1} + \frac{\beta_1 + \beta_2}{1-\alpha_1} p_2^* \tag{16}$$

and sets the general equilibrium conditions as

$$k = \frac{\beta_0}{1-\alpha_1} \qquad \beta_1 + \beta_2 = 1 - \alpha_1 \tag{17}$$

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:

$$\Delta p_{1,t} = \beta_0 + \beta_1 \Delta p_{2,t} + \gamma\left(p_{2,t-1} - p_{1,t-1}\right) + \varepsilon_t \tag{18}$$

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., γ ought to be positive). This has the important consequence that a test of significance on γ should be one-tailed, with $H_0: \gamma \leq 0$.

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 $\Delta p_{1,t} = \beta_0 + \beta_1 \Delta p_{2,t} + \gamma\left(p_{2,t-1} - k - p_{2,t-1}\right)$, 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 weightings $(\omega_1, \omega_2) = (1, -K)$, where $K = e^{\frac{\beta_0}{\gamma}}$.

A.2.- DERIVATIVES OF THE DF STATISTIC

A.2.1.- FIRST DERIVATIVE

$$\sigma_{\Delta S}^2 = \sum_{j=1}^{I}\sum_{k=1}^{I} \omega_j \; \omega_k \, \sigma_{\Delta P_j, \Delta P_k} = \omega_i^2 \sigma_i^2 + 2\omega_i \sum_{j \neq i}^{I} \omega_j \sigma_{\Delta P_i, \Delta P_j} + \sum_{j \neq i}^{I}\sum_{k \neq i}^{I} \omega_j \; \omega_k \, \sigma_{\Delta P_j, \Delta P_k} \tag{19}$$

$$\sigma_{L(S)}^2 = \omega_i^2 \sigma_i^2 + 2\omega_i \sum_{j \neq i}^{I} \omega_j \sigma_{L(P_i), L(P_j)} + \sum_{j \neq i}^{I}\sum_{k \neq i}^{I} \omega_j \; \omega_k \, \sigma_{L(P_j), L(P_k)} \tag{20}$$

$$\sigma_{\Delta S, L[S]} = \sum_{j=1}^{I}\sum_{k=1}^{I} \omega_j \; \omega_k \, \sigma_{\Delta P_j, L(P_k)} = \omega_i \sum_{j \neq i}^{I} \omega_j \left(\sigma_{\Delta P_i, L(P_j)} + \sigma_{\Delta P_j, L(P_i)}\right) + \sum_{j \neq i}^{I}\sum_{k \neq i}^{I} \omega_j \; \omega_k \, \sigma_{\Delta P_j, L(P_k)} \tag{21}$$

$$\frac{\partial \sigma_{\Delta S, L(S)}}{\partial \omega_i} = \sum_{j \neq i}^{I} \omega_j \left(\sigma_{\Delta P_i, L(P_j)} + \sigma_{\Delta P_j, L(P_i)}\right) \tag{22}$$

$$\frac{\partial \sigma_{\Delta S}}{\partial \omega_i} = \frac{1}{2}\left[\sigma_{\Delta S}^2\right]^{-\frac{1}{2}} \frac{\partial \sigma_{\Delta S}^2}{\partial \omega_i} \tag{23}$$

$$\frac{\partial \sigma_{\Delta S}^2}{\partial \omega_i} = 2\omega_i\, \sigma_{\Delta P_i}^2 + 2\sum_{j \neq i}^{I} \omega_j \sigma_{\Delta P_i, \Delta P_j} \tag{24}$$

$$\frac{\partial \sigma_{L(S)}}{\partial \omega_i} = \frac{1}{2}\left[\sigma_{L(S)}^2\right]^{-\frac{1}{2}} \frac{\partial \sigma_{L(S)}^2}{\partial \omega_i} \tag{25}$$

$$\frac{\partial \sigma_{L(S)}^2}{\partial \omega_i} = 2\omega_i\, \sigma_{L(P_i)}^2 + 2\sum_{j \neq i}^{I} \omega_j \sigma_{L(P_i), L(P_j)} \tag{26}$$

$$\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i} = \frac{\partial \sigma_{\Delta S}}{\partial \omega_i}\sigma_{L(S)} + \frac{\partial \sigma_{L(S)}}{\partial \omega_i}\sigma_{\Delta S} \tag{27}$$

$$\frac{\partial \rho_{\Delta S, L(S)}}{\partial \omega_i} = \frac{\frac{\partial \sigma_{\Delta S, L(S)}}{\partial \omega_i}\sigma_{\Delta S}\sigma_{L(S)} - \frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i}\rho_{\Delta S, L(S)}}{\sigma_{\Delta S}^2 \sigma_{L(S)}^2} \tag{28}$$

$$\begin{aligned}
\frac{\partial \widehat{DF}}{\partial \omega_i} &= \frac{\partial \rho_{\Delta S, L(S)}}{\partial \omega_i} \frac{\sqrt{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}} + \frac{1}{(T-2)}\left(\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}\right)^{-\frac{1}{2}} \rho_{\Delta S, L(S)}^2}{\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}} \\
&= \frac{\partial \rho_{\Delta S, L(S)}}{\partial \omega_i}\left(\left(\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}\right)^{-\frac{1}{2}}\right. \\
&\quad \left. + \frac{1}{(T-2)}\left(\frac{1-\rho_{\Delta S, L(S)}^2}{T-2}\right)^{-\frac{3}{2}} \rho_{\Delta S, L(S)}^2\right)
\end{aligned} \tag{29}$$

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

$$\frac{\partial^2 \sigma_{\Delta S, L(S)}}{\partial {\omega_i}^2} = 0 \tag{30}$$

$$\frac{\partial\left[\sigma_{\Delta S}^2 \sigma_{L(S)}^2\right]}{\partial \omega_i} = 2\left[\sigma_{\Delta S}\sigma_{L(S)}\right]\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial \omega_i} \tag{31}$$

$$\frac{\partial^2\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i{}^2} = \frac{\partial^2\sigma_{\Delta S}}{\partial\omega_i{}^2}\sigma_{L(S)} + \frac{\partial^2\sigma_{L(S)}}{\partial\omega_i{}^2}\sigma_{\Delta S} + 2\,\frac{\partial\sigma_{L(S)}}{\partial\omega_i}\frac{\partial\sigma_{\Delta S}}{\partial\omega_i} \tag{32}$$

$$\frac{\partial^2\sigma_{\Delta S}}{\partial\omega_i{}^2} = -\frac{1}{8}\left[\sigma_{\Delta S}^2\right]^{-2}\left(\frac{\partial\sigma_{\Delta S}^2}{\partial\omega_i}\right)^2\frac{\partial^2\sigma_{\Delta S}^2}{\partial\omega_i{}^2} \tag{33}$$

$$\frac{\partial^2\sigma_{\Delta S}^2}{\partial\omega_i{}^2} = 2\sigma_{\Delta P_i}^2 \tag{34}$$

$$\frac{\partial^2\sigma_{L(S)}}{\partial\omega_i{}^2} = -\frac{1}{8}\left[\sigma_{L(S)}^2\right]^{-2}\left(\frac{\partial\sigma_{L(S)}^2}{\partial\omega_i}\right)^2\frac{\partial^2\sigma_{L(S)}^2}{\partial\omega_i{}^2} \tag{35}$$

$$\frac{\partial^2\sigma_{L(S)}^2}{\partial\omega_i{}^2} = 2\sigma_{L(P_i)}^2 \tag{36}$$

$$\begin{aligned}\frac{\partial^2\rho_{\Delta S,L(S)}}{\partial\omega_i{}^2} = \left(\sigma_{\Delta S}\sigma_{L(S)}\right)^{-4}\Bigg[\Bigg(&\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i}\frac{\partial\sigma_{\Delta S,L(S)}}{\partial\omega_i}\\ &-\frac{\partial^2\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i{}^2}\rho_{\Delta S,L(S)}\\ &-\frac{\partial\rho_{\Delta S,L(S)}}{\partial\omega_i}\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i}\Bigg)\sigma_{\Delta S}^2\sigma_{L(S)}^2\\ &-\frac{\partial\left[\sigma_{\Delta S}^2\sigma_{L(S)}^2\right]}{\partial\omega_i}\Bigg(\frac{\partial\sigma_{\Delta S,L(S)}}{\partial\omega_i}\sigma_{\Delta S}\sigma_{L(S)}\\ &-\frac{\partial\left[\sigma_{\Delta S}\sigma_{L(S)}\right]}{\partial\omega_i}\rho_{\Delta S,L(S)}\Bigg)\Bigg]\end{aligned} \tag{37}$$

$$\frac{\partial^2 \widehat{DF}}{\partial \omega_i^2} = \frac{\partial^2 \rho_{\Delta S, L(s)}}{\partial \omega_i^2} \left(\left(\frac{1-\rho^2_{\Delta S,L(s)}}{T-2} \right)^{-\frac{1}{2}} + \frac{1}{(T-2)} \left(\frac{1-\rho^2_{\Delta S,L(s)}}{T-2} \right)^{-\frac{3}{2}} \rho^2_{\Delta S,L(s)} \right)$$
$$+ \frac{\partial \rho_{\Delta S,L(s)}}{\partial \omega_i} \left(\left(\frac{1-\rho^2_{\Delta S,L(s)}}{T-2} \right)^{-\frac{3}{2}} \rho_{\Delta S,L(s)} \frac{\partial \rho_{\Delta S,L(s)}}{\partial \omega_i} + \left(\frac{3}{(T-2)} \left(\frac{1-\rho^2_{\Delta S,L(s)}}{T-2} \right)^{-\frac{5}{2}} 2\rho_{\Delta S,L(s)} \frac{\partial \rho_{\Delta S,L(s)}}{\partial \omega_i} \right) \rho^2_{\Delta S,L(s)} + \frac{2}{T-2} \rho_{\Delta S,L(s)} \frac{\partial \rho_{\Delta S,L(s)}}{\partial \omega_i} \left(\frac{1-\rho^2_{\Delta S,L(s)}}{T-2} \right)^{-\frac{3}{2}} \right) \quad (38)$$

A.3.- GRADIENT OPTIMIZATION OF MAELOC SPREADS

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

$$\rho_{\Delta S, w_j \Delta P_j}(w_i + \Delta w_i) = \rho_{\Delta S, w_j \Delta P_j}(w_i) + \frac{\partial \rho_{\Delta S, w_j \Delta P_j}}{\partial w_i} \Delta w_i + \frac{1}{2} \frac{\partial^2 \rho_{\Delta S, w_j \Delta P_j}}{\partial w_i^2} \Delta w_i^2 \quad (39)$$

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. $\sigma^2_{\Delta S} = \sum_{k=1}^{N} \sum_{j=1}^{N} w_k w_j \sigma_{kj} = w_i^2 \sigma_i^2 + 2 w_i \left[\sum_{j \neq i}^{N} w_j \sigma_{ij} \right] + \sum_{k \neq i}^{N} \sum_{j \neq i}^{N} w_k w_j \sigma_{kj}$
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. $\frac{\partial \sigma_{\Delta S, w_i \Delta P_i}}{\partial w_i} = 2 w_i \sigma_i^2 + \sum_{j \neq i}^{N} w_j \sigma_{ij} = \frac{\sigma_{\Delta S, w_i \Delta P_i}}{w_i} + w_i \sigma_i^2$

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

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

And since $\frac{\partial\rho_{\Delta S,w_i\Delta P_i}}{\partial w_i}=\frac{\sigma_i}{\sigma_{\Delta S}}\left[1-\rho_{\Delta S,w_i\Delta P_i}^2\right]$, we can use w_i to regulate $\partial\rho_{\Delta S,w_i\Delta P_i}$ by applying $\partial w_i=\partial\rho_{\Delta S,w_i\Delta P_i}\frac{\sigma_{\Delta S}}{\sigma_i\left[1-\rho_{\Delta S,w_i\Delta P_i}^2\right]}$. $w_i^*=w_i+\partial w_i$, 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 $\rho_{\Delta S,w_i\Delta P_i}=\left|\rho_{\Delta S,\Delta P_i}\right|$.[18]

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

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

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

[18] This can be seen from the linear relation $\Delta S=\Delta Pw$.

Therefore, $\partial w_i = \partial \rho_{\Delta S,w_j\Delta P_j} \dfrac{\sigma_{\Delta S}}{\sigma_i\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]}$. 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. $$= \frac{\sigma_i\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]}{\sigma_{\Delta S}}\rho_{\Delta S,w_i\Delta P_i} + \frac{\sigma_i}{\sigma_S}\left[1-\rho^2_{S,w_iP_i}\right]\rho_{S,w_jP_j} =$$

$$= \frac{\sigma_i}{\sigma_S}\left(\rho_{S,w_jP_j} + \rho_{S,w_iP_i}\left[\rho_{ij} - 2\rho_{S,w_jP_j}\rho_{S,w_iP_i}\right]\right)$$

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

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

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

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

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

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

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,

$$\Delta\rho_{\Delta S,w_j\Delta P_j} - \frac{\sigma_i}{\sigma_{\Delta S}}\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right]\Delta w_i + \frac{1}{2}\frac{\sigma_i^2}{\sigma^2_{\Delta S}}\left[\rho_{\Delta S,w_j\Delta P_j} + \rho_{\Delta S,w_i\Delta P_i}\left(2\rho_{ij} - 3\rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right)\right]\Delta w_i^{\,2} = 0 \tag{40}$$

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

$$
\begin{aligned}
a &= \frac{1}{2}\frac{\sigma_i^2}{\sigma_{\Delta S}^2}\left[\rho_{\Delta S,w_j\Delta P_j} + \rho_{\Delta S,w_i\Delta P_i}\left(2\rho_{ij} - 3\rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right)\right] \\
b &= -\frac{\sigma_i}{\sigma_S}\left[\rho_{ij} - \rho_{\Delta S,w_j\Delta P_j}\rho_{\Delta S,w_i\Delta P_i}\right] \\
c &= \Delta\rho_{\Delta S,w_j\Delta P_j}
\end{aligned}
\tag{41}
$$

The solution is two roots, $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, of which we use the one which produces the smallest $\left|\Delta w_i\right|$, i.e. $\Delta w_i \begin{cases} \frac{-b + \sqrt{b^2 - 4ac}}{2a} & \text{if } b \geq 0 \\ \frac{-b - \sqrt{b^2 - 4ac}}{2a} & \text{if } b < 0 \end{cases}$.[19]

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' weightings Ω and the ith-column of matrix subset definition D^*, D_i^*. **X** is the matrix of instruments' series. Let be $\mathbf{P}^*$ 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 $q \cdot S = P^* w$.[20]

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

---

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

[20] 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 $q>0$.

Δw_i is the change to subset i's weighting that returns the desired $\rho_{S,w_j P_j}$. $w_i^* = w_i + \Delta w_i = w_i\left(1 + \frac{\Delta w_i}{w_i}\right)$. In order to backpropagate Δw_i into instruments' weightings Ω, $\Omega \bullet D_i^* w_i^* = \left(1 + \frac{\Delta w_i}{w_i}\right)\Omega \bullet D_i^* w_i = (1 + \Delta w_i)\Omega \bullet D_i^* w_i$. In other words, we simply need to scale by $(1 + \Delta w_i)$ the weight on any instrument involved in subset i.[21]

A.3.5.- STEP SIZE

At every iteration, we want to reduce the exposure to the subset that produces the $MSC = \text{Max}\left\{\left|\rho_{S,w_j P_j}\right|\right\}$, $j = 1,\ldots,N^*\text{-}1$. Let's say that $j \,|\, MSC = \left|\rho_{S,w_j P_j}\right|$. 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 $k>1$[22]. We'll define $\overline{C} = \frac{1}{N^* - 1}\sum_{i=1}^{N^*-1} \rho_{S,w_i P_i}$ as the target for the next iteration, and $\Delta\rho_{S,w_j P_j} = \overline{C} - \rho_{S,w_j P_j}$.[23]

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.

A.4.- KWIATKOWSKI, PHILLIPS, SCHMIDT AND SHIN (KPSS)

KPSS (Kwiatkowski, Phillips, Schmidt and Shin, 1992) is used for testing a null hypothesis that an observable time series is stationary around a deterministic trend. The series is expressed as the sum of deterministic trend, random walk, and stationary error, and the test is the LM test of the hypothesis that the random walk has zero variance. It complements the DF and ADF tests, as KPSS' null hypothesis is for stationarity, while the former tests have the unit root as null hypothesis.

$$LM = \frac{\sum_{t=1}^{T}\left(\sum_{i=1}^{t} \hat{\varepsilon}_i\right)^2}{T^2 \hat{\sigma}^2} \qquad (42)$$

Where $\hat{\varepsilon}$ are the estimated OLS residuals obtained after detrending S_t with an intercept (or an intercept and time trend), and $\hat{\sigma}^2$ is the long run variance estimator, $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 + \frac{2}{T}\sum_{\tau}^{l} \varphi_{\tau,l} \sum_{t=\tau+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-\tau}$. KPSS use the Bartlett kernel, as suggested

---

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

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

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

by Newey and West (1987), whereby $\varphi_{\tau,l} = 1 - \frac{\tau}{l+1}$. For consistency, it is necessary that $l \rightarrow \infty$ as $T \rightarrow \infty$. The bandwidth l can be selected following Newey and West (1994), although $l = O\left(T^{1/2}\right)$ is usually enough (Maddala and Kim, 2004, p.121).

For low dimensional hedging portfolios, a vector ω that maximizes LM can be estimated by brute force grid search.

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.
- 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.
- 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): "*U*nit 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. (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.