ABSTRACT

In this work we present a new framework for modelling portfolio dynamics and how to incorporate this information in the portfolio selection process. We define drivers for asset and portfolio dynamics, and their optimal selection. We introduce the new Commonality Principle, which gives a solution for the optimal selection of portfolio drivers as being the common drivers. Asset dynamics are modelled by PDEs and approximated with Neural Networks, and sensitivities of portfolio constituents with respect to portfolio common drivers are obtained via Automatic Adjoint Differentiation (AAD). Information of asset dynamics is incorporated via sensitivities into the portfolio selection process. Portfolio constituents are projected into a hypersurface, from a vector space formed by the returns of common drivers of the portfolio. The commonality principle allows for the necessary geometric link between the hyperplane formed by portfolio constituents in a traditional setup with no exogenous information, and the hypersurface formed by the vector space of common portfolio drivers, so that when portfolio constituents are projected into this hypersurface, the representations of idiosyncratic risks from the hyperplane are kept at most in this new subspace, while systematic risks representations are added via exogenous information as part of this common drivers vector space. We build a sensitivity matrix, which is a similarity matrix of the projections in this hypersurface, and can be used to optimize for diversification on both, idiosyncratic and systematic risks, which is not contemplated on the literature. Finally, we solve the convex optimization problem for optimal diversification by applying a hierarchical clustering to the sensitivity matrix, avoiding quadratic optimizers for the matrix properties, and we reach over-performance in all experiments with respect to all other out-of-sample methods.

Keywords Portfolio Optimization · Portfolio Management · Partial Differential Equations · Neural Networks · Sensitivity Analysis · Diversification · Idiosyncratic risk · Systematic risk · Commonality Principle · Common Drivers · Causality · Hierarchical Clustering · Automatic Adjoint Differentiation · Hierarchical Risk Parity · Geometry · Vector Space · Risk Management · Manifold

1 Introduction and Background Literature

1.1 Portfolio Theory and Diversification

In the field of portfolio management, there have been an extensive list of portfolio selection methods since H. Markowitz in 1952 introduced Modern Portfolio Theory [1]. The concept of diversification starts from Markowitz, and his question: What explains diversification? Together with Tobin (1958) created the two-factor model (Expected value and Variance) and the mean-variance framework, that relies on investors’ rationality and risk aversion [2]. Then, the Capital Asset Pricing Model (CAPM) is derived from the two-factor approach, by Sharpe (1964) [3] and Lintner (1965) [4], as an equilibrium for the mean-variance framework starting the asset pricing literary movement. Lintner, developed formal proofs for the justification of the use of variance as risk measure, by connecting the assumptions of uncertainty and rational behavior, using expected utility theory, probability, market risk conditions and assumptions on the market competition and expectations [5]. The next movement comes from the empirical analysis of the two-factor approach, the Efficient Capital Markets Hypothesis (EMH) and Rational Expectations. Fama (1970) develops a framework to test
Hierarchical Sensitivity Parity

(EMH) \(^6\). Empirical evidence suggests CAPM delivers poor results for diversification. Others, like Roll (1977), argue that CAPM is not testable for diversification as market portfolio is unobserved \(^7\).

Further developments come from Merton (1973), with a dynamic extension of the traditional CAPM from Sharpe and Lintler \(^8\). Investment opportunity evolves in time by means of stochasticity of the risk-free rate and Markov theorem allows for the first two moments explain portfolio selection, both with restrictive assumptions in continuity, free-arbitrage, and others. This movement further developed in all theory behind derivatives pricing, however there are strong empirical evidence that suggest all these models have unrealistic assumptions \(^9\).

1.2 Machine Learning, Time Series analysis, Predictability

ML models have been extensively used to detect latent factors that drive market returns. Dimensionality reduction, clustering and regression techniques are applied to come up with distance matrices that can outperform in a convex portfolio optimization setting. References include M. Gautier et al. with a survey of clustering financial time series \(^10\). M. Lopez de Prado Hierarchical Risk Parity method, in which hierarchical structure of the correlation matrix is used to improve portfolio selection \(^11\). M. Avellaneda applies hierarchical PCA and spectral properties of the correlation matrix to portfolio construction \(^12\). On the other hand, many developments have been carried out in regime switch detection and change point detection and the benefits on Portfolio Management. Kim et al. introduced regime switching models \(^13\). Nystrup et al. construct a regime dependent portfolio and show that forecasting regimes and incorporating this into the process can improve risk-adjusted returns \(^14\).

Y. Li et al introduce network topology to portfolio optimization \(^15\). V. Satone et al. creates Fund2Vec, a new way of using embeddings for building portfolios of mutual funds \(^16\). S. Sharifi et al \(^17\), Laloux et al. \(^18\), Plerou et al.\(^19\), Pafka et al. \(^20\), focused on Random Matrix Theory (RMT) for Portfolio Optimization. RMT is used to improve results by denoising and detuning the correlation matrix.

Bayesian inference for portfolio management was introduced by F. Black and R. Litterman in 1991 \(^21\). Since then, work have been carried by Bevan and Winkelmann \(^22\), Satchell and Scowcroft \(^23\), Krishman and Mains (2005) , with a BL extension \(^24\). Bayesian Networks have added structure to the Bayesian inference paradigm like in R. Rebonato and A. Denev \(^25\). These are techniques that incorporate causality to the portfolio selection.

2 Framework Description

2.1 Sensitivities and asset dynamics

Our method focuses on asset dynamics that can be modelled by unknown PDEs with independent variables as drivers. These PDEs are approximated with Neural Networks with time series data and sensitivities of assets with respect to the drivers are obtained from these Neural Networks.

The method we use for obtaining the sensitivities is focused on Derivative-based local methods, and particularly on Adjoin modelling and Automated Differentiation (AAD). AAD is a common method for extracting derivatives and system sensitivities. In financial applications we can mention, B. Huge and A. Savine , that introduced Differential Machine Learning for derivative pricing and hedging \(^26\). Sensitivity analysis with Neural Networks in other sciences include, Ozesmi and Ozesmi (1999) \(^27\), Garson (1991) \(^28\), Olden et al. (2004) \(^29\), Scardi and Harding (1999) \(^30\). A full survey can be found in Jaime Pizarroso, Jose Portela, Antonio Munoz, R package for time series \(^31\).

If we model financial assets as first-order PDEs with \(y\) being the asset and \(x\) a vector of drivers causing its dynamics. This PDE is unknown and analytically unsolvable and its solution could be represented as:

\[
y (t) = F\left( \frac{\partial y(t)}{\partial x_1(t)}, \frac{\partial y(t)}{\partial x_2(t)}, \ldots, \frac{\partial y(t)}{\partial x_N(t)}, \frac{\partial x_1}{\partial t}, \frac{\partial x_N}{\partial t}, \frac{\partial y}{\partial t}, x_1, \ldots, x_N \right) \tag{1}
\]

\(F\) can be non-linear, in the case of being linear it can be expressed as:

\[
y (t) = x_1(t) \frac{\partial y(t)}{\partial x_1(t)} + x_2(t) \frac{\partial y(t)}{\partial x_2(t)} + \cdots + x_N(t) \frac{\partial y(t)}{\partial x_N(t)} + \frac{\partial y}{\partial t} + \frac{\partial x_1}{\partial t} + \cdots + \frac{\partial x_N}{\partial t} \tag{2}
\]
Hierarchical Sensitivity Parity

If we represent it in discrete time, it can incorporate lagged values of the asset and drivers. In the linear case we have:

\[
Asset_t = \gamma_1 Asset_{t-1} + \gamma_2 Asset_{t-2} + \cdots + \gamma_m Asset_{t-m} + \beta_1 Driver_{1,t-1} + \beta_2 Driver_{1,t-2} + \cdots + \beta_n Driver_{n,t-n} + \cdots + \alpha_1 Driver_{2,t-1} + \alpha_2 Driver_{2,t-2} + \cdots + \alpha_k Driver_{2,t-k} + \cdots + \mu_1 Driver_{N,t-1} + \mu_2 Driver_{N,t-2} + \cdots + \mu_p Driver_{N,t-p} + \varepsilon
\]  

(3)

Or, with a non-linear function:

\[
Asset_t = F(Asset_{t-1},\ldots, Asset_{t-m}, Driver_{1,t-1},\ldots, Driver_{n,t-n},\ldots, Driver_{N,t-1},\ldots, Driver_{N,t-p}) + \varepsilon
\]

(4)

with \( F \) non-linear

We do not know these functions, but we can approximate them with time series data and a Neural Network, because they are universal approximators, as demonstrated in G. Cybenko [32]. Below we show an example of the functional approximator and its components (Figure 1 for illustration), following A. Muñoz San Roque [33]:

\[
d[k] = g\left(d^{(k-1)}, u^{(k)}, \varepsilon^{(k-1)}\right) + \varepsilon[k]
\]

(5)

Where:

- \( g \) is a nonlinear function
- \( d[k] \) is the output of the process in time \( k \)
- \( d^{(k-1)} = [d[k-1], d[k-2], \ldots]^T \)
  is a vector containing the output of the process in times \( k-1 \) and backwards.
- \( u^{(k)} = [u[k-1], u[k-2], \ldots]^T \)
  is a vector containing the exogenous entries at times \( k \) and backwards.
- \( \varepsilon^{(k)} = [\varepsilon[k-1], \varepsilon[k-2], \ldots]^T \)
  is a vector with the WN realizations at the output at time \( k-1 \) and backwards.
- \( \varepsilon[k] \)
  is the WN realization at the output at time \( k \).

Different configurations of Networks can be used to approximate different modelling PDEs. For the document we only focus on feed-forward networks, and sensitivities of the assets with respect to their drivers are obtained via AAD, following the methodology by B.Huge and A. Savine for a Vanilla Net [26].
Hierarchical Sensitivity Parity

2.2 Commonality Principle for Portfolio Drivers (Common Drivers)

This principle is one of our main contributions and is key for the methodology. We define the concept of optimal driver for an asset as:

- Optimal in persistence, this being the amount of time it stays as a driver.
- Optimal in probability of causality, as causality cannot be guarantee we speak in terms of probabilities. An optimal driver must maximize the probability of causing the assets dynamics.

We define as Specific Drivers, the optimal drivers for individual assets. In portfolio selection, specific drivers tend to not be shared between portfolio constituents because of the rational search for diversification, but it always happens. We demonstrate that optimal drivers for a portfolio, optimal choice both in terms of persistence and probability of causality for a portfolio, are the specific drivers that are shared the most among all portfolio constituents, that are repeatedly selected as specific drivers on the most number of constituents. We call this the Commonality Principle.

The commonality principle for portfolio common drivers, not only gives the optimal choice for portfolio drivers among any set of drivers, it also makes possible the geometric connection between two vector sub-spaces that allows for the method to reach idiosyncratic plus systematic diversification. We now proof the commonality principle by proving the two optimals and show these geometric connection.

2.2.1 Proof of Persistence in the Commonality Principle. Modern Portfolio Theory

We can make use of modern portfolio theory to proof that optimal choice for portfolio drivers in terms of persistence are the most common drivers from all specific drivers’ constituents.

Drivers contribute to the risks of an asset by causing its dynamics, and we can separate risks in idiosyncratic and systematic.

Specific drivers will contribute to both types of risks. The majority of specific drivers from a particular portfolio constituent will contribute to its own idiosyncratic risk, which does not affect other constituents. But also, for a particular constituent, some specific drivers may contribute to the systematic risk of this individual asset, and these drivers will contribute to the systematic risk of other constituents too. But, for sure the amount of systematic risk explained by all specific drivers from all constituents is not maximal almost surely, because there is an important amount of specific drivers focused on idiosyncratic risks. We are not constraining the problem to look for drivers of systematic risks only.

When looking for the portfolio drivers that follow the commonality principle, we are restricting ourselves to look only for the specific drivers that are most common as being specific among all portfolio constituents. In this case, we do maximize the proportion of systematic risk from all portfolio constituents, that is explained by any possible subset of drivers, by selecting the common drivers as the portfolio drivers, the proportion of idiosyncratic risks explained is minimized and systematic is maximized. Focusing on systematic risks have the following advantages:

- Systematic risks of the portfolio are the most persistent. We therefore guarantee optimality in drivers persistence for the common drivers selection, which are causing the maximum possible amount of this type of risks.
- This focus on systematic risks is one of the causes of the competitive advantage of our method respect the competitors that focus only on idiosyncratic diversification.

2.2.2 Proof of Probability of Causality in the Commonality Principle

There is no guarantee of causality, so we deal with probability of causality. We can prove that the maximum probability of causality for a portfolio, given any possible selection of drivers, is reached in the optimal choice of drivers following the commonality principle, with the most repeated drivers as specific drivers among all portfolio constituents.

- Probability of causality for an asset or a portfolio, given a set of drivers, are defined as the probability that the dynamics of that asset or portfolio are caused by this set of drivers.
- For each constituent, 1, . . . , N of a portfolio and its set of specific drivers (M specific drivers for Asset 1, W for Asset N, etc.) there is a probability of causality, probability of causing the dynamics at p time steps in the future given by the following vector x:

\[
P (\text{Asset}_{1+p} | \text{Driver}_{A_1, \ldots, A_M}) \leq X_1, \ldots, P (\text{Asset}_{N+p} | \text{Driver}_{Z_1, \ldots, Z_W}) \leq X_N \tag{10}\]

- At a portfolio level we have:

\[
P (\text{Portfolio}_{t+p} | \text{Driver}_{C, \ldots, C_M}) \leq Y \tag{11}\]
Hierarchical Sensitivity Parity

with M common drivers C.

- To prove the principle for causality we need to verify the following proposition is true almost surely. We will show, it is, only in the special case if we focus at a portfolio level, which is exactly our only interest for portfolio selection:

\[
P(\text{Portfolio}_{t+p}|\text{Driver}C, \ldots C_M) > P(\text{Portfolio}_{t+p}|\text{Driver}A_1, \ldots A_M) + \cdots + P(\text{Portfolio}_{t+p}|\text{Driver}Z_1, \ldots Z_W)
\]  

(12)

- For any given set of drivers to choose from, we need to verify that the probability of causality of the portfolio conditional on the common drivers is greater than the sum of the probabilities of causalities of the portfolio conditional on each specific drivers subset (optimal drivers for individual constituents). We omit the proof for being self-evident that if this previous formula is correct for any set of drivers to choose from, the general case were the probability of causality of the portfolio conditional on common drivers is greater too than the probability of causality of the portfolio conditional on all combinations (optimal and not-optimal) of this entire set of drivers for selection, will also be correct, for any given set of drivers.

- For a particular portfolio, with Asset 1 having specific drivers \(A_1, \ldots, A_M\), Asset 2 having \(B_1, \ldots, B_K, \ldots, \) up to Asset N having \(Z_1, \ldots, Z_W\). For the proof we make use of Reichenbach (1956) concept of Common Cause Principle (CCP) \(^{[34]}\). From this reference, we explain the principle. Suppose that events A and B are positively probabilistically correlated:

\[
p(A \cap B) > p(A)p(B) \tag{13}
\]

Reichenbach’s Common Cause Principle says that when such a probabilistic correlation between A and B exists, this is because one of the following causal relations exists: A is a cause of B; B is a cause of A; or A and B are both caused by a third factor, C. In the last case, the common cause C occurs prior to A and B, and must satisfy the following four independent conditions:

\[
p(A \cap B|C) = p(A|C)p(B|C) \tag{14}
\]

\[
p(A \cap B|\overline{C}) = p(A|\overline{C})p(B|\overline{C}) \tag{15}
\]

\[
p(A|C) > p(A|\overline{C}) \tag{16}
\]

\[
p(B|C) > p(B|\overline{C}) \tag{17}
\]

\(\overline{C}\) denotes the absence of event C (the negation of the proposition that C happens) and it is assumed that neither C nor \(\overline{C}\) has probability zero. Line (14) says that A and B are conditionally independent, given C. In Reichenbach’s terminology, C screens A off from B. Line (15) says that \(\overline{C}\) also screens A off from B. Lines (16) and (17) say that A and B are more probable, conditional on C, than conditional on the absence of C. These inequalities are natural consequences of C being a cause of A and of B. Together, conditions (14) through (17) mathematically entail (13). The common cause can thus be understood to explain the correlation in (13)

- For the general (CCP) case that the correlated effects are random variables like ours. Suppose X and Y are RV that are correlated, i.e., there exist \(x_i\) and \(y_j\) such that

\[
p(X = x_i \cap Y = y_j) \neq p(X = x_i)p(Y = y_j) \tag{18}
\]

Then there exists a set of variables \(Z_1, \ldots, Z_M\) such that each variable is cause of X and Y, and

\[
p(X = x_i \cap Y = y_j|Z_1 = z_{k_1}, \ldots, Z_M = z_{k_m}) = p(X = x_i|Z_1 = z_{k_1}, \ldots, Z_M = z_{k_m})p(Y = y_j|Z_1 = z_{k_1}, \ldots, Z_M = z_{k_m}) \tag{19}
\]

In our case, \(P(\text{Asset}_1_{t+p}|\text{Driver}A_1, \ldots A_M), \ldots, P(\text{Asset}_N_{t+p}|\text{Driver}Z_1, \ldots Z_W),\) correlation between individual assets and specific drivers is maximal. For \(P(\text{Portfolio}_{t+p}|\text{Driver}A_1, \ldots A_M), \ldots, P(\text{Portfolio}_{t+p}|\text{Driver}Z_1, \ldots Z_W),\) correlation between portfolio and specific drivers is not maximal. Correlation between Portfolio and Common Drivers is maximal, because we can show that:
Hierarchical Sensitivity Parity

\[
\text{Average}(\text{Corr}(\text{Portfolio}_{t+p}, \text{CommonDriver}_{1}, \ldots, \text{CommonDriver}_{M}))
\]
\[
= \frac{\sum_{i=1}^{M} \text{corr}(\text{Portfolio}_{t+p}, \text{CommonDriver}_{Ci})}{M}
+ \sum_{i=1}^{W} \text{corr}(\text{Portfolio}_{t+p}, \text{SpecificDriver}_{Zi})
\]
\[
> \frac{\sum_{i=1}^{M} \text{corr}(\text{Portfolio}_{t+p}, \text{All Specific Drivers}_{i})}{M + \cdots + W}
\]
\[
= \text{Average}(\text{Corr}(\text{Portfolio}_{t+p}, \text{All drivers from entire Set}_{i}))
\]

The set of correlations between portfolio and common drivers will always be maximal on average, respect other cases such as the set of correlations between portfolio and specific drivers, and these two respect the set of all drivers. It can be proved too that, given a portfolio of assets, a portfolio of common drivers and a portfolio of portfolios of specific drivers from portfolio constituents:

\[
\text{Portfolio of Assets}_{t+p} = \sum_{i=1}^{N} w_i \text{Asset}_{it+p}
\]
\[
\text{Portfolio of Common Drivers}_{t} = \sum_{i=1}^{M} q_i \text{CommonDrivers}_{it}
\]
\[
\text{Portfolio of Specific Drivers}_{t} = \sum_{i=1}^{N} w_i \left( \sum_{k=1}^{P} z_k \text{SpecificDriver}_{P_i} \right)
\]
\[
\text{Portfolio of All Drivers}_{t} = \sum_{i=1}^{NN} t_i \text{Driver}_{i}
\]

The following is true for all weights and elements from any set of drivers:

\[
\text{Corr}(\text{Portfolio of Assets}_{t+p}, \text{Portfolio of Common Drivers}_{t})
\]
\[
> \text{Corr}(\text{Portfolio of Assets}_{t+p}, \text{Portfolio of Specific Drivers}_{t})
\]
\[
> \text{Corr}(\text{Portfolio of Assets}_{t+p}, \text{Portfolio of All Drivers}_{t})
\]

• We have proved that we are in the maximum for correlation in the case of common drivers selection, and by applying the Reichenbach Common Cause Principle, we are in the maximum case for probability of causality. Because, either the common drivers are the candidates that are causing the portfolio dynamics with highest probability, or they share common factors that makes this probability of causality the maximum. This cannot be said with any other selection of drivers because we lack for the maximum in correlation.

There are other ways to proof the probability of causality part of the commonality principle (Conditional Probabilities for example). For the document we think one is sufficient.

2.3 Diversification and Geometry

In this section we show how this method provides some competitive advantage respect to all other methods, by keeping the most of idiosyncratic diversification and adding systematic diversification.

Correlation and Co-variance matrices are introduced in the process of portfolio optimization because their elements represent geometric projections between asset returns in a hyperplane of some vector space. In Figure 2, we show
a representation of 3 assets and two correlations as cosines. In this hyperplane, or vector subspace, formed by this type of matrices, portfolio optimization techniques search for optimal diversification only reaching idiosyncratic levels following Modern Portfolio Theory [1], [3]. On the other hand, we can incorporate exogenous information to the portfolio optimization process, in our case in the form of drivers. Sensitivities of assets with respect to a set of drivers, are the coordinates that allows us to project these assets into a new hypersurface from a vector space formed by these drivers. In Figure 3, we show a simplification of this hypersurface by a yellow arrow. We can see assets projections, computed from the sensitivities in colored dash. Instead of searching for diversification on the traditional vector subspace, hyperplane, with correlations and co-variances, one can search for diversification in the hypersurface, in the vector space formed by these drivers. Figure 3 is the case were these drivers are the common drivers in line with the commonality principle.

The issue here is that, depending on the exogenous variables you choose to form this hypersurface, different forms and properties will arise for this hypersurface, where assets are going to be projected so that we can optimize for diversification in this new manifold, different sources of diversification will be represented on this manifold depending on your choice of exogenous variables, or drivers based on our framework. The Commonality Principle will allows us to choose the set of exogenous variables, the common drivers, that makes possible the geometric link between the hyperplane and the hypersurface, that allows to keep the maximum amount of idiosyncratic risk representation from the hyperplane. And at the same time, it adds the systematic risk representation implicit on the hypersurface formed by the common drivers and its exogenous nature, which also are optimal in persistence and probability of causality for the portfolio.

This is the key input of the principle, common drivers selection to the portfolio optimization process, that gives competitive advantage with respect to any other method that do not use the commonality principle. Allowing to keep the maximum amount of idiosyncratic diversification, plus a different amount of systematic diversification on the portfolio optimization process in this N-dimensional hypersurface of N common drivers for the portfolio. Which explains the substantial improvement in performance with respect to the rest of methods.

The Commonality Principle is a more general concept that applies irrespective of the type of geometry used, Euclidean, Riemannian, Network Graphs, etc.

3 Methodology

We now describe the different steps in the methodology and add a section on each called model flexibility, to described adaptable choices and parameters to improve performance. There is a timeline subsection at the end that describes our model configuration for the experiments in this document.

3.1 Selection of Common Drivers

From a set of M drivers with M > N, N being the number of constituents of the portfolio. For each constituent we rank correlations respect to all drivers for different lags and time horizons, with a threshold that depends on the lag. We select drivers that have passed the thresholds the greatest amount of times among all portfolio constituents. We now show the algorithm:

\[ \forall \ Asset_i, SpecificDriver \_i \ i = 1, \ldots N, \ \forall \ Driver \_j \ j = 1, \ldots M \] (26)
Equation (27), states that for a driver to be a specific driver for a particular constituent it has to have correlations above thresholds $T_1$ and $T_0$ for respective lags. Equation (28) and (29) are the formulation for the problem of finding the set of $(i+1)$ elements that have a greater value than a threshold from other sets of elements. Equation (30) is adapting the set $A$ from the previous formulation to our problem, because we want from all drivers of the full drivers set, those that are simultaneously specific for the greatest number of portfolio constituents. $K$ is a parameter of choice that indicates the maximum number of common drivers to select for the model implementation. $B_k$ will be the $k$ common drivers optimally chosen. Optimal in terms of passing the thresholds and being repeated the maximum number of times among portfolio constituents. Common Drivers are selected from past data up to the selection time, and left fixed for present and future portfolio re-balancing. For this document, we focused on 6 and 12 months past daily windows for common driver selection. We also decided to update common drivers selection every 6 months (with those past windows), and left the selection fixed for the subsequent 6 months.

Model flexibility: The Modeler or Portfolio Manager has flexibility in choosing $K$. She also has flexibility in the common drivers optimal selection in that, if the number of common drivers that pass all thresholds and that are repeated among all constituents is greater than $K$, which happens often. In this case she can choose $K$ of them by looking at the maximum in correlation, or she can make her own rational selection of $K$ among this subset of winners based on prior knowledge. She also has all the freedom to choose the window length for common drivers selection. And the most important point, is that she can adapt at any time the frequency of updates for common drivers selection, to tackle regimes, non-stationarity, market conditions, changes in the portfolio at any level including risk aversion, etc.
Hierarchical Sensitivity Parity

3.2 Portfolio Re-balancing

This section is performed in full on every re-balancing time, in this particular order, and once per re-balance, with the latest common drivers selection up to that time. For this document, we have assumed that portfolio re-balance is carried out once per month, with past data, and weights kept fixed for the subsequent 30 calendar days. Also, for our case we leave fixed the common drivers selection for 6 months onward, so 6 re-balancing times are performed with the same past selection.

3.2.1 Network Architecture

On each re-balancing time, and for each portfolio constituent, we fit a Neural Network with the constituent as output, and common drivers as inputs, with time series returns data from the period we want to approximate the dynamics and extract sensitivities. We fit multitude architectures at this stage based on different configurations in terms of number of layers, neurons, fitting window length, lags between output and inputs. For this document we focus always on past data prior the re-balancing point in time. We train the architectures on the fitting period and evaluate the fit with an error measure like Mean Square Error (MSE). We select the optimal architecture for this re-balance based on this metric. For this document we used a Multi-layer Perceptron. Finally, each portfolio constituent will have an optimal architecture for that re-balancing time.

Model flexibility: The user can choose different error measures, also different types of Neural Networks such as Recurrent Neural Networks. She can even use a Neural Graph if she has some prior modelling knowledge about the constituent dynamics. Fitting window is another flexible parameter. She can opt for choosing the best fit from all configurations, or truncate the space of architectures selection by choosing only the ones with LAG 0, or 1, or both.

3.2.2 Sensitivity Analysis

For each portfolio constituent, an optimal architecture is selected, and AAD with Tensorflow is used to approximate the partial derivatives of the constituent with respect to the common drivers (sensitivities) as in [26] for the vanilla net case. These sensitives are discrete functions with a value for each time-step of the fitting period. To obtain a metric of each sensitivity for that period we use the mean of these values for this document. However, the choice of different functions to resume this sensitivity information for the method can improve the performance.

Model flexibility: A different function for a metric that can resume the sensitivity values for the fitting period can be used, instead of the mean, like another statistic, trending measures (linear fit, non-linear fit), second derivative approximations. This can be beneficial for extracting sensitivity dynamics for the future with an increase on our method performance.

3.2.3 Sensitivity Distance Matrix

Once all sensitivities are obtained for all portfolio constituent with respect to the common drivers, we need to incorporate them into a problem of Portfolio Management. Following what we have described on section 2.3 Diversification and Geometry. We represent all portfolio constituents into an N-dimensional space, with N being the number of common drivers (same for all), and coordinates being the sensitivities of the portfolio constituents respect to each of the common drivers.

In this space of common drivers sensitivities coordinates, in which all portfolio constituents are represented and for which we compute the euclidean distance matrix. We call it sensitivity matrix, and represents similarity of the projections of portfolio constituents with respect to the hypersurface formed by the vector space of common drivers. The Commonality Principle allows us to make the link from a geometric point of view, between a hyperplane of a vector space of portfolio constituents, were only idiosyncratic risks are accessible, to a hypersurface from a vector space of common drivers, were systematic risks are accessible, and, because of the drivers being common from the perspective of the commonality principle, idiosyncratic risks are still accessible in this hypersurface, and the amount of idiosyncratic risk accessible is optimal, because the drivers are optimally common with respect to the principle. Any other case delivers less access to idiosyncratic risk, for any other amount of systematic risk accessible from the hypersurface, from the exogenous information, in which we project portfolio constituents. This meaning, we could change drivers, or exogenous variables, and change the amount of systematic risk accessible, but the idiosyncratic one will remind optimal if and only if drivers are selected based on the commonality principle.

Now that we have described the geometry, the sensitivity matrix, by measuring similarity of projections in that hypersurface, will allows us to search for optimal diversification. Just like correlation matrix allows us for searching optimal idiosyncratic diversification in the hyperplane. In our case, by looking for a linear combination of constituents, a portfolio, that can be represented in that hypersurface of common drivers on a location that makes all risks, idiosyncratic
and systematic, represented in that hypersurface of common drivers, as orthogonal as possible, will make the optimal portfolio. Optimal portfolio with optimal weights, and optimal diversification for idiosyncratic risks, and sub-optimal diversification for the systematic risks. In comparison with the traditional case of the hyperplane, which is the case for the correlation matrix, where the optimal portfolio is a linear combination of the constituents that makes its location such as idiosyncratic risks only are optimally diversified.

3.2.4 Convex Optimization via Hierarchical Clustering

We need to solve a convex optimization problem to search for optimal diversification in that hypersurface, and for that we have a sensitivity matrix in which similarity means, similar projection with respect to common drivers, projection that, once again, incorporates both sources of risks. To solve the convex optimization problem with the sensitivity matrix, we perform a hierarchical clustering on the sensitivity matrix. For our solution to the convex optimization problem, we leverage on the work by M. Lopez de Prado, Hierarchical Risk Parity method [11]. He applies a hierarchical clustering on a correlation matrix, extracts a new distance matrix based on those hierarchies, and solves the convex optimization problem for portfolio selection and this new matrix, with two numerical methods, Quasi-diagonalization and Recursive Bisection. In our case, our framework, setup, the distance matrix and the information that incorporates are all different. But for the solution to the convex optimization problem with our sensitivity matrix, we find it quite useful. Just like hierarchies on the correlation matrix are proved to improve the optimization solution when applied like in [11], and these are hierarchies on projections between assets, the same logic must apply to the projections in our setup.

We obtain a new matrix from a hierarchical clustering applied to our sensitivity matrix, and we apply the subsequent Quasi-diagonalization and Recursive Bisection numerical methods just like in HRP. For our leverage on his solution we name the method: Hierarchical Sensitivity Parity.

Model flexibility: Other distances measures can be used apart from the euclidean. Other solutions to the convex optimization problem can be proposed for the sensitivity matrix. We use standard numerical methods to convert the sensitivity matrix into an approximated one that is positive semi-define before we apply the hierarchical clustering. Other transformations that can make the sensitivity matrix to keep its information while having the properties of a correlation matrix (including the Hermitian property, positive semi-definiteness, and positive real-valued leading diagonal) may improve the part of the solution that is leveraged from HRP. But also, other transformations on the sensitivity matrix may lead to the successful implementations of other optimization methods.

3.3 Timeline for the document experiments

We provide here a guideline of how we have configured the method in our experiments for this document.
Hierarchical Sensitivity Parity

| SXXP (Portfolio EU)                          | SPX (Portfolio USA)                          |
|----------------------------------------------|----------------------------------------------|
| ASML HOLDING NV                              | GENERAL ELECTRIC CO                           |
| LVMH MOET HENNESSY LOUIS                     | GOLDMAN SACHS GROUP INC                       |
| SAP SE                                       | APPLE INC                                     |
| SIEMENS AG-REG                               | NVIDIA CORP                                   |
| L’OREAL                                      | DOVER CORP                                    |
| SANOFI                                       | FORD MOTOR CO                                 |
| ALLIANZ SE-REG                               | ORACLE CORP                                   |
| SCHNEIDER ELECTRIC SE                        | PACKAGING CORP OF AMERICA                     |
| TELEFONICA SA                                | MCDONALD’S CORP                               |
| BANCO SANTANDER SA                           | PFIZER INC                                    |
| INTL CONSOLIDATED AIRLINE-DI                 | SCHLUMBERGER LTD                              |
| REPSONL SA                                   | BLACKROCK INC                                 |
| INDRA SISTEMAS SA                            | PHILIP MORRIS INTERNATIONAL                    |
| GRIFOLS SA                                   | EQUINIX INC                                   |

Table 1: Portfolios for experiments. One for Europe, other for USA

The process is implemented so that Common Drivers are selected from past data with the selection method from a variable window (6 or 12 month in our case). These candidates are kept fixed for the next 6 month of portfolio re-balances. From this point of common drivers selection onward, on each re-balance, sensitivities are computed with respect to those common drivers for all portfolio constituents. Sensitivities are obtained with the optimal architecture as stated above, with a fitting window of past data up to the re-balancing date (optimal lag, window fit, and NN parameters, we select the best case with optimal MSE). Sensitivities are computed for all constituents, sensitivities averages for the fitting period are recorded (as in section 3.2.2), euclidean distance matrix for all constituents is computed, with coordinates for each constituent being the sensitivity averages of the constituent respect the common drivers. All of this for that re-balancing date. Hierarchical Clustering, Quasi-diagonalization and Recursive Bisection are used to solve for the weights. Finally, these weights solution are kept fixed for the next 30 days (1 month). One month after, a new re-balance is carried out with this configuration, but same common drivers selection. 6 months after the first common drivers selection was performed, new common drivers are selected, and left fixed for the next 6 months. Re-balances from that point onward are carried out as before but with respect these new common drivers until next selection 6 months after again.

This is the method configuration we used for the experiments.

Model flexibility: Portfolio Manager or Modeler can adjust this timeline at any time. She can adapt frequencies of rebalances, common drivers updates, windows for selection, etc. This is of course beneficial in that our method can be adapted to market conditions, regimes, news, fat-tail events, prior knowledge, views, etc.

4 Implementation

For the experiment we use the previous methodology, timeline, and the following datasets. We use two portfolios of 14 equities, from SP500 and Stoxx600, well diversified in terms of sectors. We can see the portfolios in Table 1.

We run a back testing, 100% out-of-sample, all models and computations are done with past data and solutions are left fixed for present and future portfolio decisions, up to a new re-balancing point or common driver selection, when again, past data is used for computations and solutions are left fixed again. The same applies to all other competitors methods, so that experiments are kept realistic.

We focus on the two portfolios, and we show different performances for many different techniques including our solution. We show how different portfolio managers would have performed with this portfolios of equities with many different methods. Equity names are kept fixed, for all methods they must be invested in all of them. Performance is measured as if they would have started all on the 01/06/2020 and end on the 01/12/2021. For our method, there are 3 points in time where common drivers are selected, 01/06/2020, 01/01/2021, and 01/07/2021. All re-balances are performed on the 1st of each month for all methods. Weights are kept fixed for 30 days afterwards.

We have included many different mean variance optimization techniques (Maximum Sharpe, Minimum Volatility, Quadratic Utility, Target Return, etc), and HRP method from Lopez de Prado. All of them with the same past window for the optimal choice of weights. We have used the same past window on each re-balancing for the fittings of our
Hierarchical Sensitivity Parity

![Markowitz Methods](image)

Figure 5: Top Mean Variance Methods, 1/N. NA V starting from 01/06/2020, showing subperiod (02/2021-12/2021).

|            | Max Sharpe Mark | Min Vol Mark | QU Mark | 1/N |
|------------|-----------------|--------------|---------|-----|
| Return     | 49%             | 44%          | 49%     | 50% |
| Vol (Ann)  | 16%             | 15%          | 17%     | 17% |
| Sharpe     | 3,061           | 3,002        | 2,920   | 3,006 |

Table 2: USA Portfolio: Top Mean Variance Methods, 1/N. Return, Risk and Sharpe for all period: 01/06/2020 – 01/12/2021

neural networks and sensitivities extraction, which is all what we need for the next steps in our method to get optimal weights.

For the sake of clarity, we only show the top performers always, which always include our method, and HRP of Lopez de Prado. These means that, we run multitude of experiments with different methods but the ones that we do not show are under-performing this ones, most by far.

4.1 Portfolio USA

We run experiments for all methods stated before together with 1/N (equal-weighted), with different constraints (ie. for target return) and a maximum of 10% weight for any individual (1/N=7.1%) to avoid concentration. NAVs are computed with an initial investment of 100, from 01/06/2020 up to 01/12/2021.

On table 2, we show return, volatility (Annualized) and Sharpe for the Mean Variance winners, and 1/N only, for the entire period 01/06/2020 to 01/12/2021. On Figure 5, we show NAVs from 02/2021 to 12/2021 for these models only.

Now we include our method. On table 3, we include the top previous mean variance performer, Max Sharpe, the 1/N, and HRP from Lopez de Prado. We also include the best performing versions of our method, HSP 6m LAG1 SELECT, HSP 6m LAG0 OPT, HSP 6m LAG1 OPT, which are all over-performing the rest of methods. There are three versions of the model, because we have flexibility in some of the model parameters. We have two choices: between 6 or 12 months for the past window in the common drivers selection. We have other 3 choices: we can choose between LAG 1, or LAG 0, or both, for the entire period, for the re-balancing times, when architectures are fitted on past data in order to compute sensitivities. We have the option to choose the best based on MSE from only architectures with LAG 1 between output and inputs, or LAG 0, or the best in MSE terms from all (LAG0 and LAG1). But this is fixed for all sample period, so LAG0, means optimal architectures are chosen with LAG 0 for all months. This is to be consistent, but we wanted to see how sensible the performance of the model is to the LAG selection in the fitting step. Last 2 choices are that, from the pool of common drivers that pass all thresholds and are repeated among all constituents, because this are still a great number, we select the winners common drivers either OPT, which means based on correlation value, or SELECT, which means based on rational from a junior portfolio manager.
Hierarchical Sensitivity Parity

Figure 6: USA Portfolio: Top Mean Variance Methods, 1/N, HRP Lopez de Prado, HSP for different model flexibility choices. NAV starting from 01/06/2020.

|               | HSP 6m LAG 1 SELECT | HSP 6m LAG 0 OPT | HSP 6m LAG 1 OPT | 1/N | HRP Lopez de Prado |
|---------------|---------------------|------------------|------------------|-----|------------------|
| Return        | 54%                 | 55%              | 54%              | 50% | 52%              |
| Vol (Ann)     | 17%                 | 17%              | 17%              | 17% | 17%              |
| Sharpe        | 3,157               | 3,340            | 3,116            | 3.0 | 2,954            |

Table 3: USA Portfolio: 1/N, HRP Lopez de Prado, HSP for different model flexibility choices. Return, Risk and Sharpe for all period: 01/06/2020 – 01/12/2021

So we have 3+2+2 = 7 choices of models and we show the two top performers, which are better than the rest. We are only left with 7 choices because we wanted to do so for the experiments. But the model has much more flexibility, which means that there is much more room for improvement. The key point from this results is that, we have exceed the expectations. The model is better to all competitors, with a limited configuration for experiments. This means there is much more room to get even better results. To put into perspective, we could have tried more windows for common driver selection and not just 6, 12 months. More importantly, we decided to leave the common drivers selection fixed for the subsequent 6 months. But we could improve much more performance if we increase the frequency of the updates and makes it more dynamic, for example update common drivers on each monthly re-balance. We can change the fitting window on each re-balance to adapt to market conditions instead of a fixed window length. We can select better drivers from the pool of common drivers with better prior knowledge, in the SELECT case. These are simple changes to the model configuration. I have not include the extensive list of model flexibility from previous section, just simple configuration changes that are not contemplated in the experiments and can be optimized for a even better performance.

For example, the yellow case, HSP 6m LAG1 SELECT, will be a 6-month past period for selecting common drivers at the 3 dates mentioned before, LAG=1 for output vs input in the fitting of the neural network and sensitivity extraction for all re-balances. SELECT is the method for selecting the winners among the common drivers in those 3 dates, based on a junior portfolio manager experience, no other numerical analysis carried out for this rational.

On Figure 6, we show NAVs for the entire period, and Figure 7, a sub-period from 01/02/2021 to 01/12/2021. It is relevant to mention that the last common driver selection is on 01/07/2021, 5 months before the last under-performing for all models that we can see on November 2021 onward (Figure 7). It is quite likely that 5 months after the last common drivers update these optimal selection could have changed, and if we had updated them more often than every 6 months, it is likely we would not see that loss in our model. Even though, all models fall except HSP 6m LAG 0 OPT.

4.2 Portfolio EU

Everything stays the same in these EU experiments as of the US case, the only thing it changes is the portfolio constituents and time series accordingly. Our method is still the best performer in EU case respect the rest of competitors.
In Table 4, we show the top Mean Variance case, 1/N, HRP from Lopez de Prado, and two choices from our method, HSP 6m LAG 1 OPT, which means 6 month window for common driver selection, LAG 1 for the NN fitting on re-balances, OPT, means Optimal choice of winners from common drivers subset based on maximum correlation. HSP 6m LAG 0 & 1 SELECT, in this case SELECT means winners were selected based on junior portfolio manager rational, and optimal NN architecture for fitting was selected based on MSE from the pool of all LAGs, 0 and 1, on all re-balances (See Appendix A, for a sample table with all architectures in one re-balance time and one constituent). Performance on Table 4 is for the entire period, 01/06/2020 – 01/12/2021.

Figure 8 shows NAVs of our competitor’s only, best mean variance cases, 1/N and HRP. Figure 9 shows the NAVs of our method together with the best of the competitors.
Hierarchical Sensitivity Parity

Figure 9: EU Portfolio: Top Mean Variance Methods, 1/N, HRP Lopez de Prado, HSP for different model flexibility choices. NAV starting from 01/06/2020, showing sub-period

|                | Min vol | Target Ret | 1/N | HRP | HSP 6m LAG 1 OPT | HSP 6m LAG0 & 1 SELECT |
|----------------|---------|------------|-----|-----|-----------------|------------------------|
| Return         | 22%     | 16%        | 25% | 30% | 34%             | 30%                    |
| Vol (Ann)      | 17%     | 16%        | 18% | 19% | 21%             | 21%                    |
| Sharpe         | 1.3014  | 0.9688     | 1.3740 | 1.5242 | 1.6494 | 1.433            |

Table 4: EU Portfolio: Top Mean Variance Methods, 1/N, HRP Lopez de Prado, HSP for different model flexibility choices. Return, Risk and Sharpe for all period: 01/06/2020 – 01/12/2021.

We can see how our method is better than all other methods. Same comments on the US experiment section are applicable to this section, there is plenty of room for easy improvement in performance, because we use the same limited configuration as in US case, and by optimizing more this we have a lot of room for better results (windows for common drivers selection, selection rational, frequency of drivers update, fitting windows, etc.).

5 Conclusion

We have presented a method for portfolio selection/optimization that incorporates information about assets and portfolio dynamics into the process improving performance in all experiments respect to out-of-sample state-of-the-art methods. This is achieved by keeping the most of idiosyncratic diversification, which is the only focus for most competitors, as well as adding systematic diversification. This is possible thanks to the geometric connection that the Commonality Principle provides, one of the main contributions from this document, by giving the common drivers as optimal choice for portfolio drivers. This allows to keep in a hypersurface of common drivers vector space, the optimal amount of idiosyncratic risk while adding systematic risk via exogenous common drivers with optimal persistence and probability of causality of the portfolio dynamics.

Another contribution is the way asset and portfolio dynamics are introduced into a problem of portfolio optimization. It is the first time, AAD with a Neural Network is used to extract sensitivities from financial assets and incorporate them into the portfolio optimization process. For this we use a sensitivity matrix, in which we are able to keep both sources of diversification, and incorporate the sensitivity information into a problem of searching for optimal diversification.

Another contribution is the method we use for solving the convex optimization problem for optimal diversification, by applying a Hierarchical Clustering on the sensitivity matrix, to leverage on the existing solution from Lopez de Prado, HRP [11]. This is the first time that, a new way to represent the problem of optimal diversification arises in the form of
Hierarchical Sensitivity Parity

a new distance matrix, and hierarchical clustering, plus numerical methods from [11] are used to solve the optimization problem. This opens the door to new ways of redefining diversification, to come up with a new distance matrix that, like in our case, if hierarchies maintain the properties of the problem can be solved with this solution.

We have showed for all the experiments we ran that our method is the top performer. Also, how can we easily improve this performance by changing the frequency of common drivers selection, updating for example on every monthly re-balance instead of every 6 months, or in a regime change. Also, adapting the fitting window length, selecting common drivers winners with a better rational if required. These are simple changes that can only improve performance even more in this experiments’ datasets and others. Finally, we have added a model flexibility section with all the main points, the user has flexibility to adapt the method, for any changing conditions and get more upside in performance.

The model is robust, consistent, has proved to outperform and we have stated the reasons of the outperformance. Even if we could have run many more experiments, we outperform all cases and we show how to improve them more in these and other datasets. It is also flexible, with parameters and modelling choices that can tackle problems that arise in portfolio management such as back-test overfitting, non-stationary, fat tails, etc. some of them described in M. Lopez de Prado [35].

For future work, we can include the suggestions in the model flexibility points from the methodology section. It worth mentioning the application of other Neural Networks such as Recurrent, Neural Graphs. Also, apply other distances for the sensitivity matrix. Trying to explore the Commonality Principle in other geometries such as Network Graphs. Try to find other ways to select better winners from the set of common drivers based on other analysis that could improve the systematic diversification of the model.

Acknowledgments

This was supported by Rentamarkets, and by Miguel de Felipe Roman, Junior Quantitative Analyst at Rentamarkets, for helping on running some of the experiments.

6 APENDIX A: Examples of Common Drivers Selections and Re-balancing Fittings

| SPX 1 6m OPT    | SPX 2 6m OPT     | SPX 3 6m OPT    |
|-----------------|-----------------|-----------------|
| MSCI INDIA      | S&P 500 HEALTH CARE IDX | DOW JONES INDUS. AVG |
| USD-NOK RR 25D 3M | S&P 500 CONS STAPLES IDX | S&P 500 INDEX |
| USD-SEK RR 25D 3M | ISHARES MSCI USA QUALITY FAC | MSCI WORLD |
| IBEX 35 INDEX   | ISHARES MSCI USA SIZE FACTOR | MSCI Daily TR Net World |
| S&P 500 HEALTH CARE IDX | ISHARES MSCI USA MIN VOL FAC | S&P 500 HEALTH CARE IDX |
| S&P 500 CONS STAPLES IDX | MSCI World Quality Pr $ | ISHARES MSCI USA QUALITY FAC |
| STXE 600 Utilities EUR | World Size Tilt | ISHARES MSCI USA SIZE FACTOR |
| STXE 600 Telecomm EUR | MSCI WORLD Min Vol PR | ISHARES MSCI USA MIN VOL FAC |
| MSCI EM LATIN AMERICA | MSCI World ESG | MSCI WORLD Min Vol PR |
| MSCI World High Dividend | MSCI WORLD REAL EST | MSCI World ESG |
| MSCI WORLD/HLTH CARE | MSCI WORLD/HC & STPL | MSCI Daily Net TR World |

Table 5: Start and end dates of past windows for the 3 points in time (1,2,3 in first column) of the common drivers selection. 6 and 12 month widows on 01/06/2020, 01/01/2021 and 01/07/2021.

Table 6: For the 3 dates of common drivers selection, for SP500 and 6 month past window, we have OPT selection for common drivers winners, which means maximal correlation for final selection among the set of common drivers.
Hierarchical Sensitivity Parity

**SPX 1 6m SELECT**
- USD SWAP SEMI 30/360 10Y
- EUR-CZK X-RATE
- CHF-JPY X-RATE
- BUONI POLIENNALI DEL TES
- UK Gilts 30 Year
- US Generic Govt 10 Yr
- MSCI INDIA
- U.S. Treasury
- USD-NOK RR 25D 3M
- USD-SEK RR 25D 3M
- NASDAQ COMPOSITE
- MSCI US REIT INDEX
- Japanese Yen Spot
- Indian Rupee Spot
- U.S. TIPS
- 1-3 Year EU

**SPX 2 6m SELECT**
- Generic 1st 'FV' Future
- BONOS Y OBLIG DEL ESTADO
- NASDAQ COMPOSITE
- EUR SWAP ANN (VS 6M) 10Y
- U.S. TIPS
- MSCI World Momentum Pri$ USD
- MSCI WORLD/REAL EST

**SPX 3 6m SELECT**
- DOW JONES INDUS. AVG
- Generic 1st 'S ' Future
- S&P 500 INDUSTRIALS IDX
- SOYBEAN FUTURE Nov21
- ISHARES MSCI USA VALUE FACTO
- ISHARES MSCI USA SIZE FACTOR
- MSCI WORLD VALUE INDEX
- World Size Tilt
- MSCI WORLD/INDUSTR

Table 7: For the 3 common drivers dates selection, 6 month past window and SELECT case, which means junior PM rational for final selection of winners among the common drivers set.

**SPX 1 1y OPT**
- MSCI INDIA
- USD-NOK RR 25D 3M
- USD-SEK RR 25D 3M
- S&P 500 HEALTH CARE IDX
- S&P 500 UTILITIES IDX
- STXE 600 Utilities EUR
- STXE 600 Telcomm EUR
- ISHARES MSCI USA MIN VOL FAC
- MSCI World High Dividend
- MSCI WORLD/HLTH CARE
- MSCI WORLD/CON STPL
- MSCI WORLD/UTILITY

**SPX 2 1y OPT**
- NASDAQ 100 STOCK INDX
- MSCI INDIA
- USD-NOK RR 25D 3M
- S&P 500 HEALTH CARE IDX
- S&P 500 UTILITIES IDX
- STXE 600 Telcomm EUR
- STXE 600 Food&Bevrg EUR
- ISHARES MSCI USA MIN VOL FAC
- MSCI WORLD/HLTH CARE
- MSCI WORLD/CON STPL
- MSCI WORLD/UTILITY

**SPX 3 1y OPT**
- DOW JONES INDUS. AVG
- S&P 500 FINANCIALS INDEX
- S&P 500 INDUSTRIALS IDX
- S&P 500 MATERIALS INDEX
- ISHARES MSCI USA VALUE FACTO
- ISHARES MSCI USA QUALITY FACTO
- ISHARES MSCI USA SIZE FACTOR
- ISHARES MSCI USA MIN VOL FAC
- MSCI WORLD VALUE INDEX
- World Size Tilt
- MSCI WORLD Min Vol PR
- MSCI World High Dividend
- MSCI World ESG

Table 8: For 3 dates of common drivers selection, 12 months past window and OPT method for final selection of winners among common drivers set.
Hierarchical Sensitivity Parity

**Table 9:** For 3 dates of common drivers selection, 12 months past window and SELECT method for final selection of winners among the common drivers set.

![Table image](image-url)

**References**

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

[2] James Tobin. Liquidity preference as behavior towards risk. *Review of Economic Studies*, 25(2):65–86, 1958.

[3] William Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[4] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfoliols and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.

[5] John Lintner. Security prices, risk, and maximal gains from diversification. *Journal of Finance*, 20(4):587–615, 1965.
[6] Eugene Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, 1970.

[7] Richard Roll. A critique of the asset pricing theory’s tests part i: On past and potential testability of the theory. *Journal of Financial Economics*, 4(2):129–176, 1977.

[8] Robert Merton. Theory of rational option pricing. *Bell Journal of Economics*, 4(1):141–183, 1973.

[9] Espen Gaarder Haug and Nassim Nicholas Taleb. Option traders use (very) sophisticated heuristics, never the black–scholes–merton formula. *Journal of Economic Behavior & Organization*, 77(2):97–106, 2011.

[10] Gautier Marti, Sébastien Andler, Frank Nielsen, and Philippe Donnat. Clustering financial time series: How long is enough?, 2016.

[11] Marcos López de Prado. Building diversified portfolios that outperform out-of-sample. *The Journal of Portfolio Management*, 42:59–69, 2016.

[12] Marco Avellaneda and Juan Andrés Serur. Hierarchical pca and modeling asset correlations, 2020.

[13] Chang-Jin Kim, Jeremy Piger, and Richard Startz. Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, 143(2):263–273, April 2008.

[14] Peter Nystrup, Henrik Madsen, and Erik Lindström. Dynamic portfolio optimization across hidden market regimes. *Quantitative Finance*, 18(1):83–95, 2018.

[15] Yan Li, Xiong-Fei Jiang, Yue Tian, Sai-Ping Li, and Bo Zheng. Portfolio optimization based on network topology. *Physica A: Statistical Mechanics and its Applications*, 515(C):671–681, 2019.

[16] Vipul Satone, Dhruv Desai, and Dhagash Mehta. Fund2vec: Mutual funds similarity using graph learning, 2021.

[17] S. Sharifi, M. Crane, A. Shamaie, and H. Ruskin. Random matrix theory for portfolio optimization: a stability approach. *Physica A: Statistical Mechanics and its Applications*, 335(3):629–643, 2004.

[18] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Random matrix theory and financial correlations. Science & Finance (CFM) working paper archive 500053, Science & Finance, Capital Fund Management, January 1999.

[19] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6), Jun 2002.

[20] Szilard Pafka, Marc Potters, and Imre Kondor. Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization, 2004.

[21] Fischer Black and Robert Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.

[22] Andrew Bevan, Andrew Bevan, and Kurt Winkelmann. Using the black-litterman global asset allocation model: Three. In *Years of Practical Experience.” Fixed Income Research, Goldman, Sachs & Company*, 1998.

[23] S Satchell and A Scowcroft. A demystification of the Black–Litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management*, 1(2):138–150, September 2000.

[24] Hari P. Krishnan and Norman E. Mains. Black-litterman + 1: Generalizing to the two factor case. Risk, July 2005.

[25] Riccardo Rebonato and Alexander Denev. Coherent Asset Allocation and Diversification in the Presence of Stress Events. MPRA Paper 30534, University Library of Munich, Germany, April 2011.

[26] Brian Huge and Antoine Savine. Differential machine learning, 2020.

[27] Stacy L. Özesmi and Uygar Özesmi. An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, 116:15–31, 1999.

[28] G. David Garson. A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, 9(3):399–434, 1991.

[29] Julian D Olden, Michael K Joy, and Russell G Death. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178(3):389–397, 2004.

[30] Michele Scardi and Lawrence W. Harding. Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecological Modelling*, 120:213–223, 1999.

[31] J. Pizarroso, J. Portela, and A. Muñoz. Neuralsens: Sensitivity analysis of neural networks, 2021.

[32] George V. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2:303–314, 1989.
Hierarchical Sensitivity Parity

[33] Antonio Muñoz San Roque. *Aplicación de técnicas de redes neuronales artificiales al diagnostico de procesos industriales*. PhD thesis, Universidad Pontificia Comillas, 1996.

[34] Hans Reichenbach. *The Direction of Time*. Dover Publications, 1956.

[35] Marcos López de Prado and David H. Bailey. The false strategy theorem: A financial application of experimental mathematics. *The American Mathematical Monthly*, 128(9):825–831, 2021.