Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification

Prayut Jain [1] and Shashi Jain [2,*

[1]   Department of Mathematics, Indian Institute of Science, Bengaluru 560012, India; prayutjain@gmail.com
[2]   Department of Management Studies, Indian Institute of Science, Bengaluru 560012, India
*   Correspondence: shashijain@iisc.ac.in

Abstract: The Hierarchical risk parity (HRP) approach of portfolio allocation, introduced by Lopez de Prado (2016), applies graph theory and machine learning to build a diversified portfolio. Like the traditional risk-based allocation methods, HRP is also a function of the estimate of the covariance matrix, however, it does not require its invertibility. In this paper, we first study the impact of covariance misspecification on the performance of the different allocation methods. Next, we study under an appropriate covariance forecast model whether the machine learning based HRP outperforms the traditional risk-based portfolios. For our analysis, we use the test for superior predictive ability on out-of-sample portfolio performance, to determine whether the observed excess performance is significant or if it occurred by chance. We find that when the covariance estimates are crude, inverse volatility weighted portfolios are more robust, followed by the machine learning-based portfolios. Minimum variance and maximum diversification are most sensitive to covariance misspecification. HRP follows the middle ground; it is less sensitive to covariance misspecification when compared with minimum variance or maximum diversification portfolio, while it is not as robust as the inverse volatility weighed portfolio. We also study the impact of the different rebalancing horizon and how the portfolios compare against a market-capitalization weighted portfolio.

Keywords: machine learning for portfolio; covariance misspecification; superior predictive ability; NIFTY

1. Introduction

Many of the present day portfolio optimization techniques are based on the mean-variance optimization framework that was developed by Markowitz (1952). Due to the practical challenges associated with forecasting the mean returns, the prevalent popular portfolio risk optimization techniques require only the forecast of covariance of returns. Some of the notable risk-based portfolio allocation methods that rely only on covariance forecasts are the minimum variance Clarke et al. (2006), maximum diversification Choueifaty and Coignard (2008), equal risk budget Leote et al. (2012), and equal risk contribution Maillard et al. (2010).

The most well known and common estimator for the forecast of covariance of returns is the sample-based covariance. It is calculated from the time series of historical returns. For a covariance matrix of size N there needs to be at least $N(N+1)/2$ independent and identically distributed (iid) returns observations to estimate the sample-based forecast. Therefore, in order to construct a covariance matrix of returns for 50 assets, one would ideally need at the least 5 years of daily returns time series, with the hope that they are iid data. There is ample evidence that asset returns exhibit heteroskedasticity with volatility clustering, and also that the correlation structures do not remain invariant for such

long periods (Zakamulin 2015; Lopez de Prado 2016). There are broadly two major directions of work to address the above concern. The first approach is related to the development of better covariance forecast models. Some of the notable works in this direction are the shrinkage estimation of covariance matrix proposed by Ledoit and Wolf (2003) and the exponentially weighted covariance matrix that was popularized by Riskmetrics (1996). The sophisticated dynamic conditional correlation (DCC) model by Engle (2002), where the persistence in the variance and correlation dynamics is achieved by using a GARCH(1,1) type model is one of the most popular multivariate GARCH models for covariance forecasts. Another popular multivariate GARCH model is the constant conditional correlation (CCC) propsed by Bollerslev et al. (1990), where unlike the DCC-GARCH one uses a constant conditional correlation. The advantage of the CCC model is its easy estimation, although with an assumption that conditional correlations are time-invariant.

Hierarchical Risk Parity (HRP), as proposed in Lopez de Prado (2016), uses graph theory and machine learning algorithms to infer the hierarchical relationships between the assets which are then directly utilized for portfolio diversification. This approach, therefore, constitutes the second, more recent, direction of work to circumvent issues related to covariance matrix forecasts. Most of the traditional risk-based optimal allocations require the inversion of the covariance matrix, a step that is avoided in HRP. This provides an additional advantage to HRP, as the inversion of ill-conditioned matrices that is required in most risk-based portfolios can add significant estimation errors. The technique is extended in Raffinot (2017) where different methods for hierarchical clustering are employed and the robustness and performance of these algorithms with respect to traditional risk-based portfolios are studied.

Zakamulin (2015) investigated the impact of the various covariance matrix forecasting methodologies on the performance of minimum variance and target volatility strategies. The study however does not pay attention to the performance of other popular risk-based allocation methodologies with these forecasting techniques. The impact of covariance matrix misspecification on the optimal weights that result from different risk-based optimization methods is reported in Ardia et al. (2017). The paper, however, does not study the impact of covariance matrix misspecification on portfolio performance, but rather just the portfolio weights. In Trucíos et al. (2019), the performance of one step ahead covariance estimates from various covariance forecasting methods was empirically studied using several performance metrics. In Cesarone and Colucci (2018), a CVaR-based ERC portfolio was introduced and its performance against other risk and capital diversification was empirically studied. While Raffinot (2017) showed better performance for HRP and its variants when compared to traditional risk-based allocation techniques, the study does not account for the impact of covariance misspecification on the outcomes due to the use of possible inferior covariance matrix forecasting methods.

We would also like to refer readers to the growing literature on the use of regularization techniques for improving out-of-sample performance of risk-based portfolios, with some notable works being Brodie et al. (2009); Fastrich et al. (2015) and Carrasco and Noumon (2011). As portfolios can be evaluated using multiple performance criteria, Sawik (2012) introduces a multi-objective portfolio model. A review of other robust optimization methods and their applications is provided in Gabrel et al. (2014).

The objectives of this paper are two-fold. The first objective is to empirically study whether there are covariance matrix forecasting methodologies that provide superior performance for both traditional risk-based, and machine learning-based portfolios. This is achieved by looking at the out-of-sample performance of the portfolios, constructed using covariance matrix obtained from different forecasting methodologies, at the daily, weekly, and monthly forecasting horizon. The second objective is to study if the more sophisticated machine learning algorithms provide a better portfolio performance when compared with the traditional risk-based portfolios that are constructed using appropriate covariance forecasting methodology. For both the objectives, we use the stationary bootstrap-based superior

predictive ability (SPA) test proposed in Hansen (2005). The SPA test has been designed to evaluate whether an observed excess performance is significant or could have occurred by chance.

This paper is organized as follows. Section 2 describes the various risk-based portfolio allocation methods considered in the paper, while Section 3 describes the covariance forecast models. Section 4 explains HRP, the machine learning based portfolio allocation approach, we consider in this work. In Section 5 we describe the Data used and Methodology followed for the out-of-sample performance evaluations. We present our empirical results in Section 6 and Section 7 contains some concluding remarks.

2. Risk-Based Portfolios

We define a generic portfolio in a market with N risky assets by the $N \times 1$ vector of portfolio weights $\vec{w} \equiv (w_1, \ldots, w_N)'$. The $N \times N$ covariance matrix of the $N \times 1$ arithmetic returns $\vec{r} \equiv (r_1, \ldots, r_N)'$ forecasted for the desired holding horizon is denoted by Σ. We consider long only portfolios in our entire analysis. The constraints can be defined as

$$\mathcal{C} \equiv \{\vec{w} \in \mathbb{R}_+^N | w' \vec{\mathbb{1}}_N = 1\},$$

where $\vec{\mathbb{1}}_N$ is $N \times 1$ vectors of ones.

The study considers the following traditional risk-based portfolios

2.1. Minimum Variance Portfolio (MVP)

The weights for a minimum variance portfolio is obtained by solving the following quadratic optimization problem

$$\vec{w_{\min}} \equiv \underset{\vec{w} \in \mathcal{C}}{\text{argmin}}\ \vec{w}' \Sigma \vec{w} \tag{1}$$

2.2. Inverse Volatility Weighted Portfolio (IVWP)

Let $\sigma \equiv \sqrt{\text{diag}(\Sigma)}$ be the $N \times 1$ vector of standard deviation of arithmetic returns, then the inverse volatility weighted portfolio assigns the following weights to the N assets

$$\vec{w_{iv}} = \left(\frac{\frac{1}{\sigma_1}}{\sum_{j=1}^N \frac{1}{\sigma_j}}, \cdots, \frac{\frac{1}{\sigma_N}}{\sum_{j=1}^N \frac{1}{\sigma_j}} \right) \tag{2}$$

2.3. Equal Risk Contribution Portfolio (ERC)

This strategy assigns weights such that each asset contributes equally to the overall portfolio volatility. If we denote the $\%RC_i$ as the percentage risk contribution of the ith asset, then

$$\%\text{RC}_i = \frac{w_i [\Sigma w]_i}{\vec{w}' \Sigma \vec{w}}.$$

In the ERC portfolio, $\%RC$ for all the assets is equivalent to $\frac{1}{N}$. The weights can be calculated by solving

$$\vec{w_{erc}} \equiv \underset{\vec{w} \in \mathcal{C}}{\text{argmin}} \left\{ \sum_{i=1}^N (\%RC_i - \frac{1}{N})^2 \right\} \tag{3}$$

2.4. Maximum Diversification Portfolio (MDP)

This is a strategy where we maximize the diversification ratio of the portfolio, which is the ratio of weighted average of stock volatility and portfolio volatility,

$$\text{DR}(\vec{w}) = \frac{\vec{w}' \sigma}{\sqrt{\vec{w}' \Sigma \vec{w}}} \tag{4}$$

The weights for maximum diversification portfolio that was first proposed by Choueifaty and Coignard (2008) is then obtained by solving

$$\vec{w_{md}} \equiv \underset{\vec{w} \in \mathcal{C}}{\text{argmin}} \{-\text{DR}(\vec{w})\} \tag{5}$$

2.5. Market-Capitalization-Weighted Portfolio (MCWP)

A market-capitalization-weighted portfolio is constructed by assigning the following portfolio weights,

$$\vec{w_{mc}} \equiv \left(\frac{M_1}{\sum_{j=1}^{N} M_j}, \cdots, \frac{M_N}{\sum_{j=1}^{N} M_j} \right), \tag{6}$$

where M_i is the market capitalization of the ith asset at the time of asset allocation. Market-cap-weighted portfolios serve as an important benchmark, as under a standard interpretation of the Capital Asset Pricing Model; a *market* portfolio is automatically Sharpe ratio maximized. Note that Hsu (2004) empirically show that market cap weighted portfolios can be sub-optimal.

3. Covariance Matrix Forecasting Methods

Given the time series of T past returns, $\vec{r_{t-T}}, \ldots, \vec{r_{t-1}}$, we want to forecast the covariance of returns, Σ_t, of $\vec{r_t}$. We discuss below the methods used in our study for the forecast of the covariance matrix.

3.1. Sample-Based Covariance (SMPL)

Following the notations of Zakamulin (2015) we assume that the vector of daily asset return is given by

$$\vec{r_t} = \mu_t + \varepsilon_t$$

where ε_t is the vector of white noise on day t such that $\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = \vec{0_N}$, where $\vec{0_N}$, is $N \times 1$ vector of zeros. To estimate the sample based covariance matrix, we use the rolling window of T historical log returns. The covariance matrix on day t is given by

$$\hat{\Sigma}_t = \mathbb{E}[\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}] \approx \frac{1}{T} \sum_{i=t-T}^{t-1} \varepsilon_i \varepsilon_i' \tag{7}$$

Because of the time aggregation property of log-returns the covariance matrix for weekly and monthly returns projection can be obtained as the sum of the iterated 1 day ahead covariance predictions.

3.2. Exponentially Weighted Moving Average (EWMA)

This estimator is designed to focus more on the recent past returns, a method that was popularized by the Riskmetrics (1996). The exponentially weighted covariance matrix is estimated by using the following recursion:

$$\hat{\Sigma}_t = (1 - \lambda) \varepsilon_{t-1} \varepsilon_{t-1}' + \lambda \hat{\Sigma}_{t-1},$$

where based on the recommendation of the RiskMetrics group a decay constant $\lambda = 0.94$ for daily returns is used. We calculate the forecast for weekly and monthly EWMA covariance matrix by multiplying the daily covariance matrix by the number of days in the subsequent week and month, respectively.

3.3. Dynamic Conditional Correlation GARCH (DCC-GARCH)

The DCC GARCH proposed by Engle (2002), models two latent processes, the conditional variance D_t^2, and the conditional correlation R_t, where both $D_t, R_t \in \mathbb{R}^{N \times N}$, and D_t is diagonal. The conditional covariance is then given by:

$$\hat{\boldsymbol{\Sigma}}_t = D_t R_t D_t \tag{8}$$

The elements of conditional variance D_t are modelled using the univariate GARCH(1,1) model so as to incorporate the conditional heteroskedasticity. This can be compactly written as

$$D_t^2 = \boldsymbol{\omega} + \boldsymbol{\kappa} \odot \boldsymbol{\varepsilon_{t-1}} \boldsymbol{\varepsilon'_{t-1}} + \boldsymbol{\lambda} \odot D_{t-1}^2, \tag{9}$$

where $\boldsymbol{\omega} \equiv \mathrm{diag}(\omega_1, \ldots, \omega_N)$, similarly $\boldsymbol{\kappa} \equiv \mathrm{diag}(\kappa_1, \ldots, \kappa_N)$, $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_N)$, and $\odot$ is the Hadamard product.

The conditional correlation is modelled as

$$\begin{aligned} R_t &= \mathrm{diag}(Q_t)^{-\frac{1}{2}} Q_t \mathrm{diag}(Q_t)^{-\frac{1}{2}} \\ Q_t &= (1 - \alpha - \beta)\bar{Q} + \alpha \varepsilon_{t-1} \varepsilon'_{t-1} + \beta Q_{t-1}, \end{aligned} \tag{10}$$

where $\bar{Q}$ is the unconditional correlation matrix of ε.

The parameter estimation is done via a two stage optimization, where first the parameters, $\boldsymbol{\omega}, \boldsymbol{\kappa}, \boldsymbol{\lambda}$ that maximize the log-likelihood of the conditional variance are determined. In the second stage, the values of α, and β that maximize the log likelihood of the conditional correlation are determined while taking into account the results from stage one. We use the log-returns for calibrating the model, as then using the time aggregation property of log-returns the covariance matrix for weekly and monthly returns projection can be obtained as the sum of the iterated 1 day ahead covariance predictions.

4. Hierarchical Risk Parity (HRP)

The traditional risk-based portfolios are sensitive to the accuracy of the forecasted covariance matrix (see Ardia et al. 2017; Zakamulin 2015). When the assets are highly correlated, there is a greater need for diversification. However, for highly correlated returns, the condition number of the covariance matrix, i.e., the ratio between its maximal and minimal eigenvalues is large. The weights calculated when the covariance matrix has a high condition number can have large estimation errors, as their calculations involve inversion of the covariance matrix. Therefore, the benefit of diversification in such a case cannot be materialized, due to the large estimation errors for the portfolio. This as De Prado (2018) refers, is known as the Markowitz's curse.

The Hierarchical Risk Parity (Lopez de Prado 2016) approach addresses the problems of traditional risk-based portfolio optimisation by using the covariance matrix without inverting it. In essence, HRP calculates the inverse volatility weights for groups of similar assets, that are iteratively scaled down as one moves to even smaller sub-groups until each asset forms a subgroup. The algorithm operates in three stages. The first step involves determining the hierarchical relationships between the assets using a recursive cluster formation scheme. The clusters are formed using correlations to identify similar groups of assets that are successively merged until one large cluster. The next stage involves the quasi-diagonalisation of the covariance matrix by rearranging rows and columns based on the information from the first stage. The aim of the second stage is to achieve a more

diagonal representation of the covariance matrix with high correlations placed close to each other and therefore, the diagonal. Quasi-diagonalisation ensures that similar investments are grouped together and dissimilar ones are kept fairly apart. After this quasi-diagonalisation of the covariance matrix, weights are distributed using inverse variance allocation between sub-groups that are obtained by recursively bisecting the rearranged covariance matrix from the second stage. We here detail the three stages of the HRP algorithm.

4.1. Clustering

Clustering is a partitioning technique to group data points based on their characteristics. In the case of HRP, the correlation coefficient is used as the characteristic to measure the similarity between time series, and therefore to cluster assets that have similar time series. HRP uses an agglomerative nesting for clustering, where initially all the individual assets behave as a separate cluster. Then, on the basis of their correlation, they start forming bigger clusters until all the similar assets are clustered together. First, a suitable distance metric is defined as:

$$d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$$

where $d_{i,j}$ is the correlation-distance index between i^{th} and j^{th} asset and $\rho_{i,j}$ is the corresponding Pearson's correlation coefficient. Matrix $D \equiv d_{ij} i, j = 1, \ldots, N$ defined in such a way will be an appropriate metric space (see De Prado 2018 for proof). Next, a matrix that defines Euclidean distance between any two columns of D is defined as $\tilde{D}$, whose elements are

$$\tilde{d}_{ij} = \sqrt{\sum_{n=1}^{N}(d_{n,i} - d_{n,j})^2} \; i, j = 1, \ldots, N.$$

Agglomerative clustering starts with every asset representing a single cluster. At each step, the closest two clusters are merged into one. The measure of dissimilarity between the clusters is known as the linkage criterion.

There are three different agglomerative clustering linkage criteria that are used in this study.

- Single Linkage: The single linkage (SL) clustering method keeps the distance between two clusters as the minimum of the distance between any two points in the clusters such that:

$$d_{C_i, C_j} = \min_{x,y}\{\tilde{D}(x,y) | x \in C_i, y \in C_j\}$$

  This method is simpler to implement but sensitive to outliers and might result in long chained clusters Raffinot (2017).

- Average Linkage: In the average linkage (AL) technique the distance is defined by the average of the distance between any two points in the clusters. For clusters C_i, C_j:

$$d_{C_i, C_j} = \text{mean}_{x,y}\{\tilde{D}(x,y) | x \in C_i, y \in C_j\}$$

- Ward's Method: The most popularly used method is Ward's method (Ward 1963). It says that the distance between two clusters is how much the sum of squared errors will increase when they are merged:

$$d_{C_i, C_j} = \frac{m_i m_j}{m_i + m_j} \left\| c_i - c_j \right\|.$$

  where m_i, m_j are the cluster sizes and c_i, c_j are the center of the clusters C_i, C_j. It starts at zero, and then grows as clusters merge.

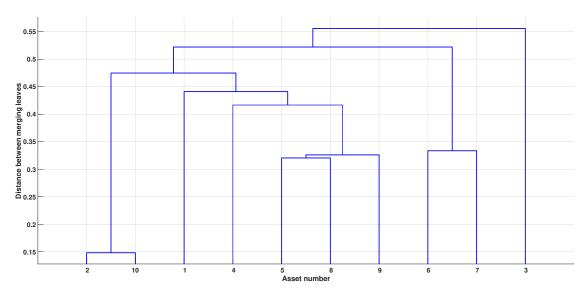Figure 1 gives a schematic of the outcome of agglomerative clustering of the assets.

Figure 1. A Sequence of clusters formed in hierarchical clustering represented as a dendrogram with y-axis representing the distance between the two merging leaves.

4.2. Quasi Diagonalisation

This step of the HRP algorithm, rearranges the covariance matrix using the information from the clustering algorithm. It places the assets with high correlations adjacently and close to the matrix diagonal, making sure that similar assets are placed together. It allows us to allocate weights optimally following an inverse-volatility allocation described below.

4.3. Recursive Bisection

The weights are allocated by inverse-volatility technique between two clusters and are scaled down as each cluster is recursively bisected until a single asset is left in each cluster. The algorithm for recursive bisection has the following steps (see Lopez de Prado 2016 for details):

1. Initialize a list of assets in the portfolio with $L = \{L_0\}$, with $L_0 = \{n\}_{n=1,...,N}$.
2. Initialize a vector of weights as $w_i = 1, i = 1, \ldots, N$.
3. Stop if $|L_i| = 1$, $\forall L_i \in L$
4. For each $L_i \in L$ such that $|L_i| > 1$

    *   Bisect L_i into two subsets, $L_i^1 \cup L_i^2 = L_i$, where $|L_i^1| = int[\frac{1}{2}|L_i|]$
    *   Calculate the variance of L_i^j, $j = 1, 2$ as $\tilde{V}_i^j = \tilde{w}_i^j V_i^j \tilde{w}_i^{j\prime}$, where V_i^j is the covariance matrix of the elements within cluster j, and

    $$\tilde{w}_i^j = \frac{tr[V_i^j]^{-1}}{\sum_i tr[V_i^j]^{-1}},$$

    which is the inverse volatility weight for the elements of the cluster.
    *   Compute the split factor $\alpha_i = 1 - \frac{\tilde{V}_i^1}{\tilde{V}_i^1 + \tilde{V}_i^2}$
    *   Rescale allocations w_n by a factor of $\alpha_i \forall n \in L_i^1$
    *   Rescale allocations w_n by a factor $(1 - \alpha_i) \forall n \in L_i^2$

5. Loop to Step 2.

5. Data and Methodology

The optimal weights in a portfolio depend on the general level of correlation between the assets of the investment universe Schumann (2013), and the specific composition of the investment universe Bertrand and Lapointe (2018). We try and capture different correlation and composition structures by creating five different universes as summarized in Table 1. We use the individual stocks that comprise the NIFTY 50 index of the National Stock Exchange (NSE) in India, to create these sub-universes. The first universe includes the top 10 stocks in the financial sector, the second includes the top 10 stocks by market capitalization (as of December 2016), the third and fifth universe contain randomly selected individual stocks from NIFTY 50, and the fourth universe contains individual stocks from the energy sector. The composition of the five universes is listed in Appendix A.

Table 1. List of universes constructed: N denotes the number of assets considered in each universe. The Min, Med, Max reports the minimum, median and maximum values of the unconditional volatilities, respectively, and the unconditional pairwise correlations in percentage for each universe.

#	Universe	N	Volatility			Correlation		
			Min	Med	Max	Min	Med	Max
1	Financial Services Sector	10	14.43	23.97	44.10	−13.01	04.90	87.32
2	Top 10 Market−Cap	10	14.43	21.88	29.52	−16.46	08.76	44.84
3	Random 10	10	−16.46	08.76	44.84	−06.38	14.47	38.99
4	Energy sector	10	−06.38	14.47	38.99	02.30	17.04	32.73
5	Random 10	10	17.81	23.24	51.28	−07.30	06.72	28.70

For each dataset, we divide the observations into an estimation period and an evaluation period:

$$t = \underbrace{-T+1, \ldots, 0}_{\text{estimation period}}, \quad \underbrace{1, \ldots, n}_{\text{evaluation period}}$$

We use the daily adjusted closing prices of the individual stocks from November 2010 to December 2016 (a total of $T = 1525$) observations for the estimation period. The parameters of the three covariance forecast models, as described in Section 3, are estimated using the first T inter-day observations. We consider following three cases for portfolio rebalancing, (a) daily $h = 1$, (b) weekly $h = 5$, and (c) monthly $h = 20$. For daily rebalancing, we need to forecast, given the returns up until time t, and the model parameters calibrated using the rolling window of T past observations[1], the covariance of returns on the $t + 1$th day. To obtain the weekly $t + 5$ and monthly $t + 20$ covariance forecasts, we sum the iterated 1-step ahead covariance predictions using the parameters calibrated for t. Iterated sum of daily covariance forecasts to obtain weekly and monthly forecasts is possible if we work with log returns, because of its time aggregation property. The h-period forecast of covariance matrix of the log returns is then converted to h-period covariance matrix of linear returns (See Appendix B for detailed explanation) for calculating the optimal weights. This conversion is essential as only weighted sum of linear (as opposed to log) individual assets' returns is equal to the portfolio return.

5.1. Intra-Day Realized Covariance Estimator

During the evaluation period, we calculate the out-of-sample realized portfolio performance based on different risk measures as described in Section 5.2. In order to compute the realized performance, we use the minute by minute intra-day prices (400 data points per day) collected from NSE for the period starting from 2 January 2017 until 31 December 2017[2]. For each of the n evaluation dates,

---

[1] In practice, typically, a rolling window smaller than the number of observations in the estimation period is used. We here, however, use the longest possible rolling window.
[2] The dataset are available from the author upon request.

the intra-day returns data is constructed artificially by fitting a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) on the available data in order to obtain returns on an equally spaced time grid of $m = 200$ intra-day points between 9:15 and 15:30 IST for all the assets. The intra-day returns is then defined as $\vec{r}_{t,i} \equiv \log(\frac{\vec{P}_{t,i}}{\vec{P}_{t,i-1}})$, $i = 1, \ldots, m$ where $\vec{P}_{t,i} \in \mathbb{R}^{N \times 1}$, is a vector of asset prices. It is reasonable to assume that $\mathbb{E}[\vec{r}_{t,i} | \mathcal{F}_{t-1}] \approx \vec{0}_N$ and that intra-day returns have no autocorrelation for moderately large values of m (see Hansen and Lunde 2005). The relationship between the log intra-day returns and the daily returns is given by,

$$\vec{r}_t = \sum_{i=1}^{m} \vec{r}_{t,i},$$

The realized daily covariance Σ_t can then be estimated as

$$
\begin{aligned}
\Sigma_t &\equiv \text{cov}(\vec{r}_t) \\
&= \text{cov}(\sum_{i=1}^{m} \vec{r}_{t,i}) \\
&= \sum_{i=1}^{m} \text{cov}(\vec{r}_{t,i}) \quad (11) \\
&\approx \sum_{i=1}^{m} \mathbb{E}\left[\tilde{r}_{t,i}\vec{r}'_{t,i} | \mathcal{F}_{t-1}\right] \quad (12) \\
&= \mathbb{E}\left[\sum_{i=1}^{m} \tilde{r}_{t,i}\vec{r}'_{t,i} | \mathcal{F}_{t-1}\right], \quad (13)
\end{aligned}
$$

where the equality in Equation (11) is due to the assumption of absence of autocorrelations in the returns time series, Equation (12) is the outcome of the assumption that the expected value of intra-day returns is nearly equal to zero. We therefore use $\Sigma_t^{OC} \equiv \sum_{i=1}^{m} \tilde{r}_{t,i}\vec{r}'_{t,i}$ as the estimator for realized intra-day covariance. As the NSE stock market is not open 24 h, the intra-day covariance misses out the covariance contribution from the time market closes until it opens on the next working day. We follow the approach of Martens (2002) and Koopman and Hol Uspensky (2002), where a scaling factor is used to convert intra-day volatility to obtain a measure of volatility for the whole day. The scaling factor for returns of the i-th stock is computed as

$$c_i = 1 + \frac{\hat{\sigma}^2_{oc,i}}{\hat{\sigma}^2_{co,i}},$$

where $\hat{\sigma}^2_{co,i}$, is the variance in the close to open log returns for the $i-$th stock, and $\hat{\sigma}^2_{oc,i}$ is the corresponding open to close variance of log returns measured in the evaluation period. Let $\vec{c} \equiv \{c_1, \ldots, c_N\}'$, then the measure for daily covariance from the intra-day return is obtained as

$$\Sigma_t \equiv \sum_{i=1}^{m} \hat{r}_{t,i}\vec{r}'_{t,i} \quad (14)$$

where $\vec{r}_{t,i} \equiv \sqrt{\vec{c}} \odot \vec{r}_{t,i}$.

The realized variance is next used to evaluate the performance of the portfolios based on the measures described below.

5.2. Portfolio Risk Measures

To assess the out-of-sample performance of the different portfolio strategies, and different covariance forecast methods, we use the following risk measures. We use these risk measures to define the loss functions for the superior predictive ability test that is described in Section 5.3.

(1) Portfolio variance: We use the total daily variance of the portfolio as the first measure of performance. The realized variance of the portfolio is given by,

$$\sigma_t^2(\vec{w}_t) = \vec{w}_t \boldsymbol{\Sigma}_t \vec{w}_t',$$

(15)

where $\vec{w}_t$ is the vector of weights obtained for a particular choice of portfolio allocation technique and covariance forecast $\hat{\boldsymbol{\Sigma}}_t$. $\boldsymbol{\Sigma}_t$ is the realised covariance matrix obtained from the intra-day returns data for $t = 1, ..., n$, as described in Section 5.1. A higher realized value of portfolio variance is an indicator of bad performance, and therefore we can directly use portfolio variance as a loss-function for the SPA test.

(2) Conditional Value-at-Risk (CVaR), also known as the expected shortfall, is a measure of risk, which is defined as (see Acerbi and Tasche 2002). Let X be the profit-loss of a portfolio on a specified time horizon and let $\alpha \in (0, 1)$ be some specified probability level. The expected α shortfall of the portfolio is defined as

$$\text{ES}^{(\alpha)}(X) = -\frac{1}{\alpha} \left(\mathbb{E}\left[X \vec{1}_{X \leq x^{(\alpha)}} \right] - x^{(\alpha)} (\mathbb{P}\left[X \leq x^{(\alpha)} \right] - \alpha) \right),$$

where $\mathbb{P}$ is the appropriate probability measure and

$$x^{(\alpha)}(X) = \sup\{x | \mathbb{P}[X \leq x] \leq \alpha\}.$$

Note that the corresponding value-at-risk (VaR) is given by

$$\text{VaR}^{(\alpha)}(X) = -x^{(\alpha)}(X).$$

We first compute the out-of-sample realized intra-day returns for the constructed portfolio $X_i = \vec{w}_t' \vec{r}_{t,i}$, $i = 1, \ldots, m$ and then sort it according to increasing profits $X_{1:m} \leq \ldots \leq X_{m:m}$ and approximate the number of α elements in the sample by $s = [m\alpha] = \max\{v | v \leq m\alpha, m \in \mathbb{N}\}$. Then the set of worst case losses corresponding to parameter α would be represented by the least s outcomes$\{X_{1:m}, \ldots, X_{s:m}\}$. VaR of the portfolio would be $-X_{s:m}$, and the expected shortfall can be estimated as

$$\text{ES}_m^{(\alpha)}(X) = -\frac{\sum_{i=1}^s X_{i:m}}{s}.$$

Again as higher CVaR values are indicators of bad performance, we use the CVaR values directly as loss function for the SPA test.

(3) Herfindahl Index (H^*) of percentage risk contribution: The normalized Herfindahl index is an indicator of concentration risk. It takes the value between 0 and 1, where 0 signifies a perfectly diversified portfolio. It is calculated as:

$$H^*(\%RC(\vec{w}_t)) = \frac{\sum_{i=1}^N (\%RC_i)^2 - \frac{1}{N}}{1 - \frac{1}{N}}$$

where $(\%RC_i) = \frac{(\hat{w}_t)_i [\boldsymbol{\Sigma}_t \vec{w}]_i}{\vec{w} \boldsymbol{\Sigma}_t \vec{w}'}$. As greater value of the index reflects greater risk concentration and therefore we use the index directly as one of our loss functions for the SPA test.

(4) Diversification Ratio (DR): It is computed as defined in Equation (4). In order to compute the realized DR. we use the portfolio weights computed using the forecasted covariance matrix, and the covariance matrix in the equation is substituted with the realized covariance matrix. It gives the measure of diversification in the portfolio and takes values ≥ 1. As we know, a higher diversification ratio is a better performance indicator; we use $-DR$ as our loss function.

(5) Sharpe Ratio: The Sharpe ratio, also called reward-to-variability ratio, is a measure of excess return per unit of deviation. It is defined as

$$SR = \frac{\mathbb{E}(\tilde{r}^p_t - r_f)}{\sigma^p_t}$$

where $\tilde{r}^p_t$ are the portfolio returns and r_f is the risk free rate, and σ^p_t is the standard deviation of excess returns of the portfolio. The portfolio variance is calculated using the intra-day returns as given by Equation (15). For our SPA test we calculate the weekly realised Sharpe ratio. As increasing Sharpe ratio implies reduced losses, we use $-SR$ as loss function for the SPA test.

5.3. Test for Superior Predictive Ability

In our study we want to evaluate whether a particular benchmark model is significantly outperformed by other models, while taking into account the large number of models that are being compared. Let $k = 0, \ldots, l$ be the models being considered, with $k = 0$ being the chosen benchmark model and $k = 1, \ldots, l$ are the models the benchmark is being compare against. Each model leads to a sequence of daily losses, $L_{k,t}$, $t = 1, \ldots, n$, where the losses are chosen as the realized portfolio variance, CVaR, $H^*(\%RC)$, and the negative of Diversification Ratio, as described in Section 5.2. The relative performance variables are defined as

$$X_{k,t} \equiv L_{0,t} - L_{k,t} \qquad k = 1, ..., l \; t = 1, \ldots, n$$

Let $\vec{X}_t = (X_{1,t}, \ldots, X_{l,t})'$ be a vector of relative performances and if $\mu = \mathbb{E}(\vec{X}_t)$, $\mu \in R^{l \times 1}$ is defined, our null hypothesis is

$$H_0 : \mu \leq \vec{0},$$

that is, the benchmark model is not inferior to any of the alternative models when the objective is to minimize the expectation of the loss function considered.

The SPA test is based on the test statistic,

$$T^{\text{SPA}}_n = max_{k=1,...,l} \frac{\bar{X}_k}{\hat{\omega}_{kk}},$$

where $\bar{X}_k = \frac{1}{n} \sum_{t=1}^n X_{k,t}$ and $\hat{\omega}^2_{kk}$ is the consistent estimator of $\omega^2_{kk} \equiv \lim_{n \to \infty} \text{var}(\sqrt{n} \bar{X}_k)$ and thus T^{SPA}_n represents the largest test statistic of relative performance. We want to find if T^{SPA}_n is too large for it to be plausible that $\mu \leq \vec{0}$. This is achieved through the SPA test where the distribution of T^{SPA}_n is estimated under the null hypothesis and the critical value of T^{SPA}_n is obtained.

Under the assumptions that $\vec{X}_t$ is stationary and has well defined moments (see Gonçalves and de Jong 2003 for the necessary assumptions and Hansen and Lunde 2005 for the justification of the assumptions), it is known that the distribution of $\sqrt{n}(\bar{\vec{X}} - \mu)$ converges to a multivariate normal distribution with mean $\vec{0}$ and covariance $\Omega \equiv \lim_{n \to \infty} \mathbb{E}[n(\bar{\vec{X}} - \mu)(\bar{\vec{X}} - \mu)']$. This result can be used to determine the distribution of T^{SPA}_n, however, as n is practically not large enough relative to l it is not possible to obtain the $l \times l$ covariance matrix Ω. One has to then rely on the stationary bootstrap of Politis and Romano (1994) to estimate the distribution of T^{SPA}_n.

5.4. Stationary Bootstrap Based Implementation

We obtain B bootstrap re-samples, $(\vec{X}^*_{b,1}, \ldots, \vec{X}^*_{b,n})$, $b = 1, ..., B$, using the stationary bootstrap approach of Politis and Romano (1994). The bootstrapped re-samples are then used to estimate ω^2_{kk} and the distribution for T^{SPA}_n. First we calculate the sample averages, $\bar{X}^*_b = \frac{1}{n} \sum_{t=1}^n X^*_{b,t}$ and next estimate $\hat{\omega}^2_{kk} \equiv \frac{n}{B} \sum_{b=1}^B (\bar{X}^*_{b,k} - \bar{X}_k)^2$ from the bootstrapped re-samples, as the empirical distribution of $n^{1/2} \vec{X}^*_b$ converges to the true asymptotic distribution of $n^{1/2} \vec{X}$, (see Gonçalves and de Jong 2003). As we seek

the distribution of T_n^{SPA} under the null hypothesis, we must recentre the bootstrap variables about the true value of μ. As we do not have a true value for μ, we can use the three estimates proposed in Hansen (2005), i.e.,

$$\hat{\mu}_k^l = \min(\bar{X}_k, 0), \quad \hat{\mu}_k^c = \bar{X}_k \vec{1}_{\{\bar{X}_{k,n} \leq -A_{k,n}\}} \quad and \quad \hat{\mu}_k^u = 0,$$

where $A_{k,n} \equiv \frac{1}{4} n^{-1/4} \hat{\omega}_{kk}$ is the correction factor[3]. Now we redefine our performance variables for each bootstrapped re-sample as $\bar{Z}_{b,k}^{*,i} = \bar{X}_{b,k}^* - g_i(\bar{X}_k)$, for $i = l, c, u$, where $g_l(x) \equiv \max(x, o)$, $g_c(x) \equiv x \cdot \vec{1}_{\{x > -A_{k,n}\}}$, and $g_u(x) \equiv x$. Hence we can approximate the distribution of T_n^{SPA} by the empirical distribution,

$$T_{b,n}^{\mathrm{SPA}*,i} = \max_{k=1,\ldots,l} \frac{n^{1/2} \bar{Z}_{b,k}^{*,i}}{\hat{\omega}_{kk}}, \quad b = 1, \ldots B, i = l, c, u$$

and calculate the p-values as $\hat{p}^i \equiv B^{-1} \sum_{b=1}^{B} 1_{\{T_{b,n}^{\mathrm{SPA}*,i} > T_n^{\mathrm{SPA}}\}}$ for $i = l, c, u$. We reject the null hypothesis for small p-values. Here, the three p-values obtained for $i = l, c, u$ corresponds to the consistent value of the true p-value(p^c), and lower and upper bound for the true p-values (p^l, p^c respectively).

6. Results

The aim of the paper is to study two major objectives, the first is for a given portfolio allocation method (for different rebalancing horizons), is there a benchmark covariance forecast method that is not inferior to the other methods considered. The second objective is to determine based on risk objectives, allocation methods that are not inferior to other allocation methods for different rebalancing horizons. For both objectives, we study whether the outcomes are consistent for different rebalancing frequencies of the portfolios.

6.1. Superior Method for Forecasting Covariance Matrix

The covariance forecast models considered for this study are the SMPL, EWMA and DCC-GARCH, details of which are described in Section 3. The portfolio allocation methods considered and the corresponding loss functions that were used for the SPA test are reported in Table 2.

Table 2. The portfolio allocation method and corresponding loss function to determine benchmark covariance forecast models.

Category	Method Name	Loss Function Used for SPA
Traditional Risk-Based	MVP	$\sigma_t(\vec{w}_t)$
	IVWP	$H^*(\%RC(\vec{w}_t))$
	ERC	$H^*(\%RC(\vec{w}_t))$
	MDP	$-DR(\vec{w}_t)$
Machine Learning	HRP (SL)	$\sigma_t(\vec{w}_t)$
	HRP (AL)	$\sigma_t(\vec{w}_t)$
	HRP (Ward)	$\sigma_t(\vec{w}_t)$

In order to compute the out-of-sample loss we perform the following steps:

1. Forecast the covariance matrix $\hat{\mathbf{\Sigma}}_t$ using the three approaches for the appropriate forecast horizon.

---

[3] It ensures that $\lim_{n \to \infty} P(\hat{\mu}_k^c = 0 \mid \mu_k = 0) = 1$ and $\lim_{n \to \infty} P(\bar{Z}_{b,k,n}^* \leq 0 \mid \mu_k < 0) = 1$ which is important for the consistency, as the models with $\mu_k < 0$ do not influence the asymptotic distribution of T_n^{SPA}, see Hansen (2005).

2. Compute the portfolio weights, $\vec{w}_t$, using the above covariance matrix for the portfolio allocation method being considered.
3. Compute using the intra day data, *realized* returns and realized covariance matrix Σ_t.
4. Use (2) and (3) to compute the time series of *realized* losses L_{kt}, using the appropriate loss function for the allocation method (as provided in Table 2), for each covariance forecast methodology.
5. For different choices of benchmark covariance forecast models, compute the *p*-values for the null hypothesis, which is that a chosen model is as good as any other model.

The results from the SPA test for the case of daily rebalancing in the form of *p*-values (we only report p^c-values, as p^l, and p^u are not significantly different) is reported in Table 3. The *p*-values correspond to the null hypothesis that a chosen model is as good as any other model. A low *p*-value (we take a value ≤ 0.05) rejects the null hypothesis, which implies that the chosen model cannot be considered as a benchmark model and is inferior to the other models being considered.

Table 3. *p*-values from the SPA test when DCC, SMPL, and EWMA are chosen as the benchmark model respectively for the different portfolio allocation techniques. The corresponding loss functions considered for each portfolio allocation technique are provided in Table 2.

Benchmark Method	Univ 1	Univ 2	Univ 3	Univ 4	Univ 5
MVP					
DCC	1.000	0.108	1.000	1.000	1.000
SMPL	0.185	1.000	0.095	0.090	0.178
EWMA	0.136	0.034	0.123	0.087	0.202
IVWP					
DCC	1.000	0.039	1.000	0.096	1.000
SMPL	0.251	1.000	0.010	1.000	0.005
EWMA	0.106	0.513	0.048	0.034	0.009
ERC					
DCC	1.000	1.000	1.000	0.221	1.000
SMPL	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EWMA	<0.0001	<0.0001	0.084	1.000	<0.0001
MDP					
DCC	1.000	1.000	1.000	1.000	1.000
SMPL	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EWMA	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)					
DCC	1.000	1.000	1.000	1.000	1.000
SMPL	0.004	<0.0001	0.0001	0.002	<0.0001
EWMA	0.090	<0.0001	0.001	0.321	<0.0001
HRP (AL)					
DCC	1.000	1.000	1.000	0.093	1.000
SMPL	0.083	0.001	<0.0001	1.000	<0.0001
EWMA	0.122	<0.0001	<0.0001	<0.0001	<0.0001
HRP (Ward)					
DCC	1.000	1.000	1.000	0.467	1.000
SMPL	0.061	0.055	<0.0001	0.141	<0.0001
EWMA	0.085	0.121	0.003	1.000	<0.0001

The results show that for both, machine learning based HRP variants and the traditional risk-based portfolios, DCC GARCH can be considered as the benchmark model in majority of the universes. In few universes other covariance forecast methodologies can result in not inferior performance,

especially when HRP (Ward), IVWP and MVP are used for portfolio allocation. However, for MDP, amongst the forecast methods considered, only DCC-GARCH results in superior performance.

We reach almost similar conclusions for weekly and monthly rebalancing, as the results (not reported here) are not significantly different from that for daily rebalancing.

6.2. Benchmark Allocation Methods for Different Portfolio Performance Objectives

From Section 6.1 it is evident that DCC-GARCH can be used as benchmark model for the traditional risk-based allocation methods as well as machine learning based allocation methods, as it provides performance that is not inferior to any other covariance forecast model in most of the universes considered. We now try to determine if there are benchmark allocation methods whose performance are not inferior to the other models when different risk objectives are considered as loss functions. Unless specified otherwise, we use DCC-GARCH to forecast the covariance matrix for all the allocation methods. The market-cap weighted portfolio is also included in the study as it serves as proxy for passive investment strategies.

6.2.1. Out-of-Sample Portfolio Variance

We first study the out-of-sample daily realized portfolio variance for the different allocation models considered. Table 4 reports the corresponding *p*-values for the null hypothesis, that the performance of the portfolio constructed using the benchmark allocation method is not inferior to the performance from other allocation methods. The results reported in the table are for the case of daily rebalancing of the portfolio. Clearly, MVP, designed with the objective to minimize portfolio variance, does not perform well as a benchmark model. We find that at least one of the variants of HRP has a large *p*-value in every universe. The performance of the HRP (SL) can not be considered inferior to other models in any universe, when DCC-GARCH is used for covariance estimation. For this case, when SMPL, an inferior covariance estimation method, is used we find that there is no single allocation method that is not discarded in one of the universes as a benchmark model. IVWP seems most robust amongst the traditional risk-based portfolios and HRP (AL) amongst the machine learning based portfolios, as both of them cannot be considered inferior to other models in three out of five universes.

Table 4. p^c-values for different portfolio allocation benchmark models considered when DCC-GARCH is used for estimating the covariance matrix and the portfolio is rebalanced daily. The highlighted cells are outcomes for variants of HRP.

Realized Portfolio Variance with DCC-GARCH					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	<0.0001	1.0000	<0.0001	<0.0001
ERC	<0.0001	0.0030	<0.0001	0.0450	<0.0001
MDP	<0.0001	<0.0001	<0.0001	0.0100	<0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	1.0000	0.0570	0.2020	0.2970	0.1460
HRP (AL)	0.0130	1.0000	0.2320	1.0000	0.0470
HRP (Ward)	0.0440	<0.0001	0.0670	0.4760	1.0000
Realized Portfolio Variance with SMPL					
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.587	1.000	<0.0001	0.2400
ERC	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MDP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	1.0000	0.0180	<0.0001	<0.0001	0.1680
HRP (AL)	0.1890	1.000	<0.0001	1.0000	0.0020
HRP (Ward)	0.0560	0.1200	<0.0001	0.006	1.0000

We next study, whether different portfolio rebalancing frequencies can affect the choice of benchmark model for minimizing the portfolios out-of-sample daily realized variance. Table A2 in Appendix C reports the *p*-values for the null hypothesis which is that a chosen benchmark allocation method is not inferior in minimizing the portfolio variance, when the portfolio is rebalanced weekly and monthly respectively. Again, DCC-GARCH is used to forecast the covariance matrix for the weekly and monthly time horizons, and the realized intra-day asset returns and covariances are used to measure the realized portfolio variance. With a weekly rebalancing frequency, amongst the risk-based portfolios, ERC and IVWP are not inferior to other allocation methods in three and two out of five universes respectively. With monthly rebalancing they are not inferior in one and two out of five universes respectively. With the machine learning based allocations, we see that with longer forecasting and rebalancing horizons, the fraction of universes in which a variant of HRP did not have an inferior performance goes down.

A summary of the realized annual volatilities of the portfolios constructed using the above allocation methods with daily rebalancing is reported in Table 5. A few observations that can be made are, firstly the market-cap weighted portfolios have the highest volatilities. Secondly, the volatilities of risk-based and machine learning based portfolios are in similar range, although when DCC GARCH is used, HRP variants have the minimum portfolio variance in all the universes. Finally an inferior covariance estimation model, in this case SMPL, results in higher volatilities. Even for this case, the HRP variants have minimum out-of-sample volatility in each of the universe considered, except Universe 2 where ERC has the lowest volatility.

Table 5. The out-of-sample realized annual portfolio volatilities for different covariance forecast models when they were rebalanced daily.

Annual Portfolio Volatility with DCC GARCH								
Universe	MVP	IVWP	ERC	MDP	MWP	HRP (SL)	HRP (AL)	HRP (Ward)
1	11.08%	11.31%	10.77%	10.84%	17.52%	10.36%	10.58%	10.49%
2	9.70%	9.39%	9.26%	9.78%	10.60%	9.26%	9.21%	9.45%
3	11.05%	11.45%	11.40%	11.47%	16.21%	11.13%	11.03%	11.22%
4	11.78%	11.86%	11.82%	11.86%	13.00%	11.76%	11.65%	11.54%
5	9.54%	9.83%	9.93%	10.62%	11.56%	9.36%	9.47%	9.46%
Annual Portfolio Volatility with SMPL								
1	11.34%	11.64%	11.09%	11.58%	17.52%	10.31%	10.36%	10.38%
2	10.26%	9.42%	9.41%	10.44%	10.60%	9.48%	9.52%	9.59%
3	11.26%	11.47%	11.46%	11.60%	16.21%	11.25%	11.19%	11.40%
4	11.56%	11.80%	11.78%	11.89%	13.00%	11.59%	11.54%	11.44%
5	9.87%	9.95%	10.17%	11.43%	11.56%	9.77%	9.88%	9.74%

Finally, we look at the realized annual volatilities of the portfolios when different rebalancing horizons were considered. The results are reported in Table A3 in Appendix C. An observation that can be made here is that the portfolio volatility increases with increasing rebalancing horizon. The increment is not linear with a greater relative change in the volatility, moving from daily to weekly, than from weekly to monthly. The exception is the market-cap weighted portfolio whose volatility marginally reduced (in average) while moving from daily to weekly rebalancing. The greatest increment in volatility in expectation over the universes with increasing rebalancing horizon is that for MVP, followed by the variants of HRP.

6.2.2. Out-of-Sample Weekly CVaR

Expected shortfall is a widely used coherent risk measure, especially for computing capital reserves for unforeseen losses. We next look at the out-of-sample realized weekly CVaR, using the intra day returns of the portfolios constructed using different allocation methods. Table 6 reports the

p-values for the null hypothesis, that a chosen benchmark model, in expectation, has lower realized expected shortfall value than others. From the reported values, it is clear that when MVP and MDP are considered as benchmark models, the null hypothesis is rejected due to low *p*-values. IVWP and ERC are the only two risk-based portfolios that have significant *p*-values in at least a couple of universes. However, the results show that at least one of the variants of HRP has a large *p*-value in each of the universe. The performance of the HRP (SL) is not inferior to other models considered in four out of five universes. HRP (AL) is not inferior in three out of five universes, while HRP (Ward) only came out as not inferior in two out of five universes. With an inferior forecast model for covariance, the only risk-based portfolio that does not have inferior performance is IVWP, which cannot be considered inferior in three out of five universes. With larger covariance misspecification, HRP (SL) does not perform well, however, HRP (Ward) comes out as a benchmark model in four out of five universes.

Table 6. p^c-values for different portfolio allocation benchmark models considered when weekly expected shortfall is taken as the loss function and the portfolio is rebalanced daily. The highlighted cells are outcomes for variants of HRP.

Realized 5-Day CVaR with DCC-GARCH					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	0.0160	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.0870	1.0000	<0.0001	<0.0001
ERC	<0.0001	0.2170	0.8150	<0.0001	<0.0001
MDP	<0.0001	<0.0001	<0.0001	0.0310	<0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	1.0000	0.0960	0.1510	0.1760	<0.0001
HRP (AL)	<0.0001	1.0000	0.1260	1.0000	<0.0001
HRP (Ward)	<0.0001	0.0290	<0.0001	0.3270	1.0000
Realized 5-day CVaR with SMPL					
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.0860	1.000	<0.0001	1.000
ERC	<0.0001	<0.0001	0.1470	<0.0001	<0.0001
MDP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	0.357	<0.0001	<0.0001	0.005	0.040
HRP (AL)	1.000	<0.0001	<0.0001	1.0000	<0.0001
HRP (Ward)	0.325	1.000	<0.0001	0.234	0.493

How does portfolio rebalancing frequency affect the choice of benchmark model for minimizing the portfolios out-of-sample weekly CVaR values? Table A4 in Apendix C reports the *p*-values for weekly and monthly rebalancing frequencies with DCC-GARCH used to forecast the covariance matrix. Amongst the risk-based portfolios, ERC's performance is not inferior to other models in four out of five universes, while IVWP's performance is not inferior in three universes, for weekly rebalancing with DCC GARCH used to forecast the covariance matrix. HRP variants are not inferior in four out of five universes, with HRP (SL) still performing (in terms of fraction of universes it is not inferior) better when compared to HRP (AL) and HRP (Ward). With monthly rebalancing both ERC and IVWP come out as not inferior choice in three out of five universes, while the null hypothesis with a variant of HRP considered as a benchmark model is not rejected in just two out of five universes.

6.2.3. Out-of-Sample Herfindahl Index and Diversification Ratio

While minimizing portfolio variance and expected shortfall are seen as outcomes of better diversification, we next study directly the extent of out-of-sample diversification using the Herfindahl index of the realized percentage risk contribution, and the realized diversification ratio of the portfolio. Table 7 reports the *p*-values corresponding to different choices of the portfolio allocation methods considered as a benchmark models with daily rebalancing. Clearly MDP, designed with the objective

to maximize portfolio diversification, does not perform well as a benchmark model, with the null hypothesis being rejected in all the universes. Only ERC and IVWP have large *p*-values in three and two out of five universes, respectively. With weekly rebalancing (see Table A5 in Appendix C) this becomes four and three out of five universes, respectively while with monthly, it is again three and two out of five universes, respectively. IVWP can be considered as the benchmark model when the objective is to minimize the Herfindahl index computed based on the realized percentage risk contribution of the underlying assets. In two out of five universes, ERC is not inferior to others when the loss function is taken as $H^*(\%RC)$. The conclusions for the choice of benchmark model with objective to minimize the Herfindhal index remains the same when different rebalancing horizons are considered, as reported in Table A5.

Table 7. p^c-values for different choices of portfolio allocation benchmark models when DCC-GARCH is used for estimating the covariance matrix and the portfolio is rebalanced daily.

Realized Diversification Ratio with DCC-GARCH					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	<0.0001	1.0000	<0.0001	1.0000
ERC	1.0000	1.0000	<0.0001	1.0000	0.0210
MDP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Realized $H^*(\%RC)$ with DCC-GARCH					
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	1.0000	0.1120	1.0000	0.2220	1.0000
ERC	0.0220	1.0000	<0.0001	1.0000	<0.0001
MDP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

6.2.4. Out-of-Sample Sharpe Ratios

We have so far tried to identify if there are benchmark models that perform in a way that is *not inferior* to other models with respect to purely risk driven objectives. We now bring in realized portfolio returns into our analysis by comparing the performance of different models when the objective is to maximize the Sharpe ratios. We take as loss function for our SPA test the negative value of the realized weekly Sharpe ratios. The weekly Sharpe ratios are computed using the intra-day returns. Table 8 reports the *p*-values for the null hypothesis, which is that a chosen benchmark model is not inferior to others when it comes to maximizing the Sharpe ratios. When DCC GARCH is used for forecasting the covariance matrix, there are candidates from both risk-based portfolios and machine learning-based portfolios whose performance can not be considered inferior to other models in most of the universes. The minimum variance portfolio can be considered as a benchmark model in four out of five universes. Market-cap weighted portfolios also are not inferior to other methods in four out of five universes. Amongst the ML based portfolios, HRP (SL) can be considered as a benchmark model in almost all the universes for this case.

When SMPL is used for forecasting the covariance matrix, the performance of both traditional risk-based portfolios and machine learning-based portfolios is significantly impacted, which is clear from the corresponding *p*-values. However, the market-capitalization weighted portfolio, which does not require a covariance forecast, can be considered not inferior to other models in all the universes.

Table 8. p^c-value for different choices of benchmark models when the loss function considered is negative of weekly Sharpe ratio and the portfolio is rebalanced daily.

Realized Weekly Sharpe Ratio with DCC GARCH					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	1.000	1.000	0.692	0.074	0.011
IVWP	0.043	0.065	0.145	1.000	0.080
ERC	0.013	0.004	0.166	0.854	0.105
MDP	0.001	<0.0001	1.000	0.771	0.095
MWP	0.011	0.635	0.103	0.189	1.000
HRP (SL)	0.122	0.226	0.192	0.467	0.060
HRP (AL)	0.044	0.755	0.045	0.137	0.080
HRP (Ward)	0.073	0.090	0.117	0.297	0.079
Realized weekly Sharpe Ratio with SMPL					
MVP	1.000	0.003	0.002	0.062	0.002
IVWP	0.125	0.001	0.122	0.617	0.033
ERC	0.005	<0.0001	0.150	0.804	0.043
MDP	0.003	<0.0001	1.000	1.000	0.045
MWP	0.156	1.000	0.158	0.192	1.000
HRP (SL)	0.403	<0.0001	0.031	0.076	0.009
HRP (AL)	0.497	<0.0001	0.018	0.003	0.008
HRP (Ward)	0.356	<0.0001	0.015	0.001	<0.0001

The impact of rebalancing frequencies on the choice of the benchmark model, whose weekly out-of-sample Sharpe ratios are not lower than the other models, is reported in Table 9. An observation that can be made is that with longer rebalancing horizons, the relative performance of machine learning based portfolios becomes inferior. While the HRP (SL) could be considered to be not inferior in almost all the universes when the portfolio was rebalanced daily; with weekly and monthly rebalancing this reduces to just two. The relative performance of the minimum variance portfolio, on the other hand, improves, with a *p*-values close to one in most of the universes. Another interesting observation is that in this case, the relative performance of the market-cap weighted portfolio also deteriorates with longer rebalancing horizons. It should be noted that we comment only on the *relative* performance of the models for different rebalancing horizons considered. It should *not* be inferred, for instance, that HRP gives the highest Sharpe ratio when the portfolio is rebalanced daily, but rather if a choice of weekly rebalancing has been made, MVP would in expectation provide not inferior Sharpe ratio than HRP.

For the sake of completeness, we provide the summary of the realized annual Sharpe ratios for the different portfolio strategies for the year 2017. Table 10 provides the realized Sharpe ratios when DCC-GARCH and SMPL are respectively used for covariance forecast, while the portfolio is rebalanced daily. We see that the realized Sharpe ratios of MVP and MDP amongst the traditional risk-based portfolios are significantly affected by the covariance misspecification. The results for IVWP and ERC appear more robust in presence of covariance misspecification, a result consistent with the findings of Ardia et al. (2017). The machine learning-based portfolios, as expected from previous experiments, perform better with DCC-GARCH. However, an inferior covariance estimator, does not as significantly affect the outcomes as it does for MVP. Market-cap weighted portfolio is outperformed in majority of the universes only by IVWP, when the portfolio is daily rebalanced.

Table 9. p^c-value for different choices of benchmark models when the loss function considered is negative of weekly Sharpe ratio and different rebalancing horizons are considered. The covariance matrix forecast is made using DCC-GARCH.

Realized Weekly Sharpe Ratio with Weekly Rebalancing					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	1.000	1.000	1.000	1.000	0.606
IVWP	0.027	0.102	0.004	0.592	0.174
ERC	0.007	0.056	0.003	0.743	0.206
MDP	<0.0001	0.004	0.022	0.551	0.093
MWP	0.001	0.242	0.015	0.081	1.000
HRP (SL)	0.046	0.054	0.002	0.697	0.017
HRP (AL)	0.028	0.083	<0.0001	0.798	0.023
HRP (Ward)	0.026	0.063	0.001	0.507	0.096
Realized Weekly Sharpe Ratio with Monthly Rebalancing					
MVP	1.000	1.000	1.000	1.000	0.508
IVWP	0.002	0.014	0.006	0.331	0.122
ERC	0.001	0.004	0.001	0.356	0.065
MDP	0.000	0.000	0.005	0.227	0.015
MWP	0.002	0.071	0.010	0.047	1.000
HRP (SL)	0.004	0.001	0.003	0.189	0.018
HRP (AL)	0.002	0.002	0.002	0.174	0.024
HRP (Ward)	0.002	0.000	0.018	0.119	0.251

Table 10. The out-of-sample realized annual Sharpe ratio for the portfolios considered when they were rebalanced daily.

	Annual Sharpe Ratio with DCC GARCH							
Universe	**MVP**	**IVWP**	**ERC**	**MDP**	**MWP**	**HRP (SL)**	**HRP (AL)**	**HRP (Ward)**
1	2.382	2.476	2.485	2.007	1.768	2.521	2.241	2.431
2	2.384	2.935	2.839	2.365	2.860	2.747	2.813	2.611
3	1.267	2.080	2.041	1.972	0.696	1.714	1.653	1.684
4	0.151	1.728	1.672	1.401	1.553	1.376	1.233	1.243
5	1.315	2.384	2.301	2.020	2.659	2.011	2.038	2.034
	Annual Sharpe Ratio with SMPL							
1	1.012	2.171	2.174	1.738	1.768	2.007	2.043	1.996
2	0.781	2.680	2.503	1.879	2.860	2.018	2.125	2.114
3	0.493	1.907	1.861	1.796	0.696	1.317	1.278	1.352
4	0.176	1.716	1.664	1.375	1.553	1.265	1.332	1.281
5	0.558	2.246	2.115	1.679	2.659	1.659	1.644	1.638

The realized annual Sharpe ratio for different rebalancing horizons, when DCC-GARCH is used for covariance forecast, is reported in Table 11. The Sharpe ratios improve for most of the allocation methods while moving from daily to monthly rebalancing, except for the market-capitalization weighted portfolio. The most significant improvement in the Sharpe ratios is for MVP, followed by the three variants of HRP. Overall, with longer horizons for rebalancing, MVP performs the best, while the performance of the variants of HRP, IVWP and ERC are similar for the dataset we consider. For our dataset, MDP and MWP perform comparatively poor (in that order), when the objective is to maximize the Sharpe ratio.

Table 11. The out-of-sample realized annual Sharpe ratio for the portfolios considered when they were rebalanced daily.

Universe	MVP	IVWP	ERC	MDP	MWP	HRP (SL)	HRP (AL)	HRP (Ward)
Annual Sharpe Ratio with Weekly Rebalancing								
1	3.156	2.720	2.752	2.401	1.662	2.658	2.587	2.591
2	3.591	3.188	3.076	2.557	2.659	2.925	3.047	2.906
3	2.333	2.343	2.309	2.247	0.503	2.068	2.039	2.079
4	1.789	2.106	2.046	1.739	1.444	2.144	2.153	2.015
5	2.741	2.737	2.633	2.258	2.511	2.663	2.730	2.892
Annual Sharpe Ratio with Monthly Rebalancing								
1	3.054	2.754	2.767	2.266	1.711	2.782	2.878	2.614
2	3.494	3.194	3.069	2.489	2.570	3.045	2.914	3.099
3	2.384	2.319	2.277	2.197	0.274	2.087	2.083	2.177
4	2.055	2.089	2.034	1.753	1.266	2.198	1.995	1.979
5	2.783	2.684	2.545	2.053	2.389	2.752	2.752	2.750

If an inferior covariance forecast method is used, the above inference can change significantly for longer rebalancing horizons. This is illustrated in Table 12 which reports the annual Sharpe ratios of a portfolio rebalanced monthly using the covariance forecasts obtained from SMPL. For the dataset that we consider, IVWP which is more robust to covariance misspecification, performs the best in most of the universes. The other allocation methods that are not significantly affected are the variants of HRP and ERC. For the dataset considered, MVP and MDP appear to be most significantly affected by inferior covariance forecasts.

Table 12. The annual Sharpe ratio for the different allocation methods when the portfolio is rebalanced monthly and SMPL is used to forecast the covariance matrix.

Universe	MVP	IVWP	ERC	MDP	MWP	HRP (SL)	HRP (AL)	HRP (Ward)
Annual Sharpe Ratio with Monthly Rebalancing Using SMPL								
1	2.431	2.526	2.532	2.057	1.711	2.826	2.726	2.291
2	2.489	3.041	2.861	2.175	2.570	2.808	2.910	2.853
3	2.057	2.248	2.207	2.155	0.274	2.054	2.051	2.090
4	1.978	2.065	2.019	1.747	1.266	2.086	2.078	2.064
5	2.368	2.637	2.489	1.978	2.389	2.493	2.487	2.621

7. Conclusions

We have compared the out-of-sample performance of portfolios constructed using traditional risk-based allocation methods with those constructed using machine learning methods. We summarize the outcomes of the different experiments performed.

7.1. Choice of Covariance Estimator

As the forecasted covariance matrix plays an important role in risk-based allocation methods, we first determined whether there were covariance forecasting methods that led to a superior out-of-sample performance of the different portfolio allocation strategies. For each portfolio allocation method, we used an appropriate risk objective and measured using an SPA test which of the covariance forecast methodologies resulted in a superior performance on that objective. The risk objective chosen for a portfolio allocation method was the one closest to the objective it was trying to optimize on. For instance, for MVP the chosen performance measure was portfolio variance (lower the better), while for maximum diversification portfolio it was the diversification ratio (higher the better). As portfolio weights in HRP are not calculated by optimizing a particular risk objective, we used the variance of

the portfolio as the objective for measuring superior performance for HRP. The following were the key observations from the analysis of the results.

- The performance of all the portfolios, with respect to their corresponding objectives, is superior when DCC GARCH is used for forecasting the covariance in majority of the universes.
- MDP appears most sensitive to the choice of better covariance forecast methodology.

7.2. Portfolio Variance

When the objective is to minimize portfolio variance, it turns out MVP whose weights are estimated by minimizing the in-sample portfolio variance does not have a superior out-of-sample performance. Other notable outcomes can be summarized as

- HRP variants are superior in the majority of universes when the objective is to minimize out-of-sample variance.
- With a poor covariance estimator, IVWP and HRP (AL) are superior methods in majority of the universes.
- With longer rebalancing horizons, HRP variants have superior performance in majority of the universes with a few exceptions where IVWP and ERC have in expectation lower portfolio variance.

7.3. Expected Shortfall

From our experiments on out-of-sample 5 day CVaR we can make the following observations:

- HRP variants are superior when it comes to minimizing out-of-sample CVaR. Amongst them, HRP (SL) performs relatively better in majority of the universes.
- With inferior covariance estimates, IVWP and HRP (Ward) result in superior performance in majority of the universes.
- With longer rebalancing horizons ERC consistently results in superior performance when it comes to minimizing out-of-sample expected shortfall.

7.4. Herfindhal Index and Diversification Ratio

While minimizing the Herfindhal index or maximizing Diversification ratio are not the end goal of a portfolio manager, they do serve as indicators for a diversified portfolio and lower portfolio concentration risk. While MDP is designed to maximize the in-sample diversification ratio, we find that its out-of-sample performance is inferior to other allocation methods in all the universes. For these two objectives, HRP also comes out as an inferior choice. The main observations from our analysis can be summarized as

- ERC followed by IVWP are superior in majority of universes when the objective is to maximize out-of-sample diversification ratio.
- IVWP followed by ERC are superior in majority of universes when the objective is to minimize the Herfindhal index of %*RC*.
- These results are consistent for different rebalancing horizons.

7.5. Sharpe Ratio

The outcomes of the experiments for superior performance when it comes to maximizing out-of-sample weekly Sharpe ratios can be summarized as follows:

- With daily rebalancing, and covariance estimated using DCC GARCH, many allocation methods including MVP, IVWP, MWP, and HRP results in *not inferior* performance when it comes to maximizing out-of-sample weekly Sharpe ratios.

- With an inferior covariance estimate Market Weighted portfolio has superior performance in majority of the universes, followed by IVWP.
- With increasing rebalancing horizons MVP clearly has superior performance in majority of the universes.

7.6. Strengths and Weaknesses

We summarize the strength and weakness of the different allocation methods.

- MVP: We see that out-of-sample performance of MVP is poor when it comes to minimizing the portfolio variance or expected shortfall. Its performance is good when it comes to maximizing Sharpe ratio, especially when the portfolio is supposed to be rebalanced less frequently. However, its performance with respect to Sharpe ratio is highly sensitive to covariance misspecification.
- IVWP: IVWP has superior performance when it comes to maximizing the out-of-sample Herfindhal index. It also has lower out-of-sample portfolio variance and expected short fall amongst the risk-based portfolios, especially when an inferior covariance estimator is used. However, with a superior covariance estimator, it often is not the best choice for most risk objectives.
- ERC: ERC has superior performance when it comes to maximizing the out-of-sample diversification ratio. It also appears to be the best choice for minimizing expected shortfall when the portfolio is not rebalanced often. It seems to have inferior performance when it comes to maximizing Sharpe ratio with longer rebalancing horizons.
- MDP: For our dataset, MDP had inferior performance for most objectives in the majority of the universes. It seems most sensitive to misspecification in a covariance matrix.
- MWP: Market weighted portfolios did not perform well with objectives of portfolio variance and expected shortfall minimization. They, however, showed up as superior models for maximizing weekly Sharpe ratios, especially when an inferior covariance matrix estimator is available. With longer rebalancing horizons and good covariance forecasts models, they result in inferior Sharpe ratios when compared to the other methods considered.
- HRP: HRP variants have superior performance when it comes to realized portfolio variance and expected shortfall when DCC GARCH is used to forecast the covariance matrix. They also are not inferior when the objective is to maximize weekly Sharpe ratio, when the portfolio is rebalanced daily and DCC GARCH is used to forecast the covariance matrix. They do seem to be sensitive to the choice of the covariance forecast model. The performance of different variants of HRP seems similar, although it might appear that HRP (SL) has a slight edge for our dataset. It is important to note that the key strength of HRP lies when the portfolio is constructed with many underlying assets, where inversion of the covariance matrix and good estimation of correlations becomes challenging. We have considered portfolios with only ten constituents; however, it would be interesting to see the outcomes of further studies carried out for portfolios with larger numbers of underlying assets.

Appendix A. Portfolio Composition

The composition of the universes constructed out of the constituents of NIFTY 50 index of National Stock Exchange, India are provided in Table A1.

Table A1. Ticker name of the constituents of each of the universes constructed for our study.

Universe	Ticker Name									
1	AXISBANK	BANKBARODA	HDFCBANK	ICICIBANK	INDUSINDBK	KOTAKBANK	SBIN	TCS	YESBANK	HDFC
2	AXISBANK	HDFCBANK	HDFC	ICICIBANK	INFY	ITC	KOTAKBANK	LT	RELIANCE	TCS
3	BHARTIARTL	CIPLA	GAIL	ITC	MARUTI	NTPC	POWERGRID	TATAMOTORS	TATASTEEL	SUNPHARMA
4	BHEL	BPCL	GAIL	NTPC	ONGC	POWERGRID	RELIANCE	TATAPOWER	COALINDIA	HINDALCO
5	YESBANK	RELIANCE	ICICIBANK	IDEA	INFY	NTPC	CIPLA	HDFCBANK	WIPRO	ZEEL

Appendix B. Converting Covariance Matrix of Log Returns to Linear Returns

The covariance matrix of log returns can be converted approximately to the covariance matrix of linear returns following the approach of Meucci (2001). We denote the logarithmic and linear returns for asset i by

$$r^i_{t,\tau} = ln\left(\frac{P^i_{t+\tau}}{P^i_t}\right), \qquad R^i_{t,\tau} = \frac{P^i_{t+\tau}}{P^i_t} - 1,$$

respectively, where τ is the time horizon and P^i_t is the price at time t. Now, let $\vec{M}_\tau$ be the vector of expected values of linear returns of N assets, and $\vec{S}_\tau$ be the corresponding covariance matrix that we wish to determine, i.e.,

$$\vec{M}_\tau = \mathbb{E}(\vec{R}_{t,\tau}), \qquad \vec{S}_\tau = cov(\vec{R}_{t,\tau}).$$

Using the relation $\vec{R}_{t,\tau} = e^{\vec{r}_{t,\tau}} - 1$, and under the assumption that returns are log-normally distributed Meucci (2001) shows that for assets with index i, and j

$$
\begin{aligned}
\mathbb{E}(R^i_{t,\tau}) &= \int R^i_{t,\tau}(r)\phi(r)dr = \int \left(e^{r^i_{t,\tau}} - 1\right)\phi(r)dr, \\
\mathbb{E}(R^i_{t,\tau}R^j_{t,\tau}) &= \int R^i_{t,\tau}(r)R^j_{t,\tau}(r)\phi(r)dr = \int \left(e^{r^i_{t,\tau}} - 1\right)\left(e^{r^j_{t,\tau}} - 1\right)\phi(r)dr \\
&= \int e^{r^i_{t,\tau}+r^j_{t,\tau}}\phi(r)dr - \int e^{r^i_{t,\tau}}\phi(r)dr - \int e^{r^j_{t,\tau}}\phi(r)dr + 1,
\end{aligned}
\tag{A1}
$$

where $\phi(x)$ is the probability density function of the standard normal random variable. Let the covariance matrix of logarithmic returns be Σ_τ and the expected return of logarithmic return be $\mu_\tau = E(\vec{r}_{t,\tau})$. Then under the above assumptions the expected linear return of asset i is

$$M^i_\tau = e^{\mu^i_\tau + \frac{1}{2}\Sigma^{ii}_\tau} - 1,$$

and the covariance between asset i and j is

$$S^{ij}_\tau = \mathbb{E}(R^i_{t,\tau}R^j_{t,\tau}) - M^i_\tau M^j_\tau = e^{\mu^i_\tau + \mu^j_\tau + \frac{1}{2}(\Sigma^{ii}_\tau \Sigma^{jj}_\tau)}\left(e^{\Sigma^{ij}_\tau} - 1\right).$$

Appendix C. Tables Containing *p*-Values for Weekly and Monthly Rebalancing

This appendix consists of tables that are not included in the results section to make the section easier to follow. These tables contains the *p*-values for the mentioned risk measures for weekly and monthly rebalanced portfolios.

Table A2. p^c-value of different benchmark portfolios based on the out-of-sample portfolio variance for different rebalancing horizons. The covariance forecast model is taken as DCC-GARCH.

Realized Portfolio Variance with Weekly Rebalancing					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	0.1080	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.0130	1.0000	<0.0001	1.0000
ERC	0.5710	0.4700	<0.0001	1.0000	<0.0001
MDP	<0.0001	<0.0001	<0.0001	0.0990	< 0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	1.0000	1.0000	0.0010	0.1230	0.2730
HRP (AL)	0.0240	0.4650	<0.0001	0.5420	0.0070
HRP (Ward)	0.0110	<0.0001	<0.0001	0.3430	0.5160
Realized Portfolio Variance with Monthly Rebalancing					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.0020	1.0000	<0.0001	0.3380
ERC	<0.0001	0.0130	<0.0001	1.0000	<0.0001
MDP	<0.0001	<0.0001	<0.0001	0.1720	< 0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	<0.0001	1.0000	<0.0001	0.0160	0.8380
HRP (AL)	1.0000	<0.0001	<0.0001	0.0320	1.0000
HRP (Ward)	<0.0001	0.0560	<0.0001	0.0730	0.3610

Table A3. The out-of-sample realized annual volatilities for different rebalancing horizons when DCC GARCH is used for covariance estimation.

Realized Annual Portfolio Volatility with Weekly Rebalancing								
Universe	**MVP**	**IVWP**	**ERC**	**MDP**	**MWP**	**HRP (SL)**	**HRP (AL)**	**HRP (Ward)**
1	11.16%	11.76%	11.15%	11.03%	16.65%	10.81%	11.47%	11.54%
2	10.38%	9.48%	9.36%	9.88%	10.56%	9.48%	9.45%	9.77%
3	11.35%	11.51%	11.48%	11.60%	16.31%	11.45%	11.41%	11.56%
4	11.93%	11.89%	11.86%	11.89%	13.01%	11.73%	11.66%	11.53%
5	10.00%	9.92%	10.03%	10.80%	11.52%	9.72%	9.82%	9.77%
Realized Annual Portfolio Volatility with Monthly Rebalancing								
1	11.65%	11.71%	11.10%	10.98%	16.74%	10.69%	10.53%	11.59%
2	10.57%	9.52%	9.40%	9.95%	10.65%	9.57%	9.56%	9.48%
3	11.40%	11.63%	11.60%	11.73%	16.44%	11.64%	11.59%	11.66%
4	11.88%	12.00%	11.96%	12.01%	13.02%	12.19%	11.97%	11.91%
5	10.35%	10.11%	10.25%	11.11%	11.63%	9.87%	9.87%	9.89%

Table A4. p^c-value of different benchmark portfolios based on the out-of-sample CVaR for different rebalancing horizons. The covariance forecast model is taken as DCC-GARCH.

Realized 5-day CVaR with Weekly Rebalancing					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	0.1040	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.3420	1.0000	<0.0001	0.4200
ERC	0.3040	1.0000	0.4330	0.4850	<0.0001
MDP	<0.0001	<0.0001	<0.0001	0.3170	< 0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	1.0000	0.6970	0.0320	0.0650	0.0600
HRP (AL)	0.0330	0.5570	0.0450	1.0000	<0.0001
HRP (Ward)	0.0190	0.0260	<0.0001	0.4130	1.0000
Realized 5-Day CVaR with Monthly Rebalancing					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
MVP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IVWP	<0.0001	0.5400	1.0000	<0.0001	0.0800
ERC	<0.0001	0.4900	0.3000	1.0000	<0.0001
MDP	<0.0001	<0.0001	<0.0001	0.5200	< 0.0001
MWP	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HRP (SL)	<0.0001	1.0000	0.0200	0.0500	0.0200
HRP (AL)	1.0000	<0.0001	0.0200	0.4700	0.0500
HRP (Ward)	<0.0001	<0.0001	0.0100	0.5200	1.0000

Table A5. p^c-value for different choices of benchmark model based on the out-of-sample $H^*(\%RC)$ and diversification ratios, when different rebalancing horizons are considered. The covariance forecast model is taken as DCC-GARCH.

Realized Diversification Ratio with Weekly Rebalancing					
Benchmark Method	**Univ 1**	**Univ 2**	**Univ 3**	**Univ 4**	**Univ 5**
IVWP	<0.0001	<0.0001	1.0000	0.0940	1.0000
ERC	1.0000	1.0000	<0.0001	1.0000	0.1960
Realized Diversification Ratio with Monthly Rebalancing					
Benchmark Method	Univ 1	Univ 2	Univ 3	Univ 4	Univ 5
IVWP	<0.0001	<0.0001	1.0000	0.0100	1.0000
ERC	0.1810	1.0000	<0.0001	1.0000	0.0270
Realized $H^*(\%RC)$ with Weekly Rebalancing					
Benchmark Method	Univ 1	Univ 2	Univ 3	Univ 4	Univ 5
IVWP	1.0000	0.1060	1.0000	0.2240	1.0000
ERC	0.0260	1.0000	<0.0001	1.0000	<0.0001
Realized $H^*(\%RC)$ with Monthly Rebalancing					
Benchmark Method	Univ 1	Univ 2	Univ 3	Univ 4	Univ 5
IVWP	1.0000	1.0000	1.0000	0.1740	1.0000
ERC	0.0260	0.4400	<0.0001	1.0000	<0.0001

References

Acerbi, Carlo, and Dirk Tasche. 2002. Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes* 31: 379–88. [CrossRef]

Ardia, David, Guido Bolliger, Kris Boudt, and Jean-Philippe Gagnon-Fleury. 2017. The impact of covariance misspecification in risk-based portfolios. *Annals of Operations Research* 254: 1–16. [CrossRef]

Bertrand, Philippe, and Vincent Lapointe. 2018. Risk-based strategies: The social responsibility of investment universes does matter. *Annals of Operations Research* 262: 413–29. [CrossRef]

Bollerslev, Tim, 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *Review of Economics and statistics* 72: 498–505. [CrossRef]

Brodie, Joshua, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. 2009. Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences 106*: 12267–72. [CrossRef] [PubMed]

Carrasco, Marine, and Nérée Noumon. 2011. *Optimal Portfolio Selection Using Regularization*. Technical Report. State College: Citeseer.

Cesarone, Francesco, and Stefano Colucci. 2018. Minimum risk versus capital and risk diversification strategies for portfolio construction. *Journal of the Operational Research Society* 69: 183–200. [CrossRef]

Choueifaty, Yves, and Yves Coignard. 2008. Toward maximum diversification. *Journal of Portfolio Management* 35: 40. [CrossRef]

Clarke, Roger G., Harindra de Silva, and Steven Thorley. 2006. Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management* 33: 10–24. [CrossRef]

De Prado, Marcos Lopez. 2018. *Advances in Financial Machine Learning*. Hoboken: John Wiley & Sons.

Engle, Robert. 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20: 339–50.

Fastrich, Björn, Sandra Paterlini, and Peter Winker. 2015. Constructing optimal sparse portfolios using regularization methods. *Computational Management Science* 12: 417–34. [CrossRef]

Gabrel, Virginie, Cécile Murat, and Aurélie Thiele. 2014. Recent advances in robust optimization: An overview. *European Journal of Operational Research* 235: 471–83. [CrossRef]

Gonçalves, Sılvia, and Robert de Jong. 2003. Consistency of the stationary bootstrap under weak moment conditions. *Economics Letters* 81: 273–78. [CrossRef]

Hansen, Peter Reinhard. 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* 23: 365–80.

Hansen, Peter R., and Asger Lunde. 2005. A forecast comparison of volatility models: Does anything beat a garch (1, 1)? *Journal of Applied Econometrics* 20: 873–89. [CrossRef]

Hsu, Jason C. 2004. Cap-weighted portfolios are sub-optimal portfolios. *Journal of Investment Management* 4: 1–10. [CrossRef]

Koopman, Siem Jan, and Eugenie Hol Uspensky. 2002. The stochastic volatility in mean model: Empirical evidence from international stock markets. *Journal of applied Econometrics* 17: 667–89. [CrossRef]

Ledoit, Olivier and Michael Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603–21. [CrossRef]

Leote, Raul, Xiao Lu, and Pierre Moulin. 2012. Demystifying equity risk-based strategies: A simple alpha plus beta description. *Journal of Portfolio Management* 38: 56–70. [CrossRef]

Lopez de Prado, Marcos. 2016. Building diversified portfolios that outperform out-of-sample. *Journal of Portfolio Management* 1–31. [CrossRef]

Maillard, Sébastien, Thierry Roncalli, and Jérôme Teïletche. 2010. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management* 36: 60–70. [CrossRef]

Markowitz, Harry. 1952. Portfolio selection. *The Journal of Finance* 7: 77–91.

Martens, Martin. 2002. Measuring and forecasting s&p 500 index-futures volatility using high-frequency data. *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 22: 497–518.

Meucci, Attilio. 2001. *Common Pitfalls in Mean-Variance Asset Allocation*. London: Wilmott Magazine.

Politis, Dimitris N., and Joseph P. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association 89*: 1303–13. [CrossRef]

Raffinot, Thomas. 2017. Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management* 44: 89–99. [CrossRef]

Riskmetrics, 1996. J.P. Morgan Technical Document. Available online: http://www.jpmorgan.com/RiskManagement/RiskMetrics/RiskMetrics.html (accessed on 29 June 2019).

Sawik, Bartosz. 2012. Downside risk approach for multi-objective portfolio optimization. In *Operations Research Proceedings 2011*. Berlin/Heidelberg: Springer, pp. 191–96.

Schumann, Enrico. 2013. Take-the-Best in Portfolio Selection. Available online: SSRN2214376 (accessed on 10 March 2019).

Trucíos, Carlos, Mauricio Zevallos, Luiz K. Hotta, and André A. P. Santos. 2019. Covariance prediction in large portfolio allocation. *Econometrics* 7: 19. [CrossRef]

Ward, Joe H., Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58: 236–44. [CrossRef]

Zakamulin, Valeriy. 2015. A test of covariance-matrix forecasting methods. *Journal of Portfolio Management* 41: 97. [CrossRef]