Volatility Timing: why risk-based rules outperform naïve diversification*

Redouane Elkamhi† Chanik Jo‡ Jacky Lee§ Marco Salerno¶

February, 2022

Abstract

DeMiguel et al. (2009b) show that estimation error dwarfs diversification benefits resulting in naïve diversification (1/N) dominating mean-variance portfolios. We illustrate the necessary and sufficient conditions for risk-based allocation rules to be optimal in a mean-variance framework. We show empirically that many common datasets satisfy such conditions, making these rules preferred to mean-variance in the presence of estimation error. Our out-of-sample tests show that these rules outperform both mean-variance and 1/N. Further, we show that clustering the data using machine learning enhances the diversification benefits of these rules by making the data closer to the required conditions for optimality.

Keywords: Portfolio Choice, Asset Allocation, Machine Learning, Clustering

JEL Classification: G11

*We would like to thank Raymond Kan, Gang Li, Edna Lopez Avila and Sheikh Sadik for helpful comments and suggestions. This article is for informational purposes only and should not be construed as legal, tax, investment, financial, or other advice. The views and opinions expressed here are those of the authors alone and do not necessary reflect the views of their employers and their affiliates.

†Rotman School of Management, Toronto, ON M5S 3E6, Canada; redouane.elkamhi@rotman.utoronto.ca

‡CUHK Business School, Shatin, N.T., Hong Kong; chanikjo@cuhk.edu.hk

§Healthcare of Ontario Pension Plan Trust Fund, Toronto, ON M3N 3A1, Canada; jlee5@hoopp.com

¶Rotman School of Management, Toronto, ON M5S 3E6, Canada; marco.salerno@rotman.utoronto.ca, Corresponding author.

Conflict-of-interest disclosure statement

Redouane Elkamhi

I have no potential conflicts to disclose.

Chanik Jo

I have no potential conflicts to disclose.

Jacky Lee

I have no potential conflicts to disclose.

Marco Salerno

I have no potential conflicts to disclose.

Estimating expected returns from time series of realized stock return data is very difficult...the estimates of variances or covariances from the available time series will be much more accurate than the corresponding expected return estimates

Robert Merton (1980)

1 Introduction

Diversification is one of the most celebrated concepts in financial economics. As Harry Markowitz elegantly said: "Diversification is the only free lunch". Academics and practitioners have developed a myriad of theoretical models in the last 70 years to exploit the benefits of diversification, and to achieve better portfolios' performances. However, theoretical models have to be estimated, and there is a trade-off between estimation error and theoretical optimality. Many theoretical models require the estimation of expected returns which have been notoriously difficult to estimate (e.g. Merton, 1980; Black, 1993).[1] Therefore, both academics and practitioners have developed portfolio allocation rules that seek diversification based solely on the covariance matrix (risk-based rules), which can be estimated more accurately than expected returns (e.g. Engle, 2002; Engle and Colacito, 2006). Examples of such rules include risk-parity (Qian, 2011) and volatility timing (e.g. Moreira and Muir, 2017; Kirby and Ostdiek, 2012; Fleming, Kirby, and Ostdiek, 2001, 2003).[2]

This paper makes the following contributions: (1) it provides a statistical test for the necessary and sufficient conditions under which inverse volatility rules (Kirby and Ostdiek, 2012) – whose implied weights are proportional to the inverse of the assets' volatilities to the power of a constant γ – are equivalent to the optimal mean-variance portfolio; (2) it demonstrates that such necessary and sufficient conditions are satisfied for many datasets empirically; (3) it presents empirical evidence that these rules perform exceptionally well out-

[1] Black (1993) writes: "Estimating expected returns is hard. Daily data hardly help at all. Only longer time periods help. We need decades of data for accurate estimates of *average* expected return. We need such a long period to estimate the average that we have little hope of seeing changes in expected returns".

[2] These rules have recently been used by practitioners as shown by the popularity of risk parity funds (for example, Bridgewater, AQR Risk Parity Fund, Invesco Risk Parity, BlackRock Market Advantage, etc.) and also risk parity ETFs (for example, RPAR Risk Parity ETF and Horizons global risk parity ETF).

of-sample, confirming the predictions from our test of the necessary and sufficient conditions; and (4) it develops a clustering methodology based on machine learning to enhance the benefits of diversification for such rules by exploiting some of the properties of the inverse volatility rules, which will be made clearer below.[3]

It is well known that there is a trade-off between estimation error and theoretical optimality. An allocation rule that is optimal from a theoretical point of view but subject to large estimation error might underperform a rule that is theoretically sub-optimal but has little or no estimation error. In our analysis of the inverse volatility rules, we focus on two questions: (a) how much do the inverse volatility rules reduce the impact of estimation error compared to the optimal mean-variance portfolio (tangency portfolio)? And (b) how far are the inverse volatility rules from the theoretically optimal portfolio? To address the former question, we first show the analytical distribution of the tangency portfolio weights under estimation error.[4] We then use Monte Carlo simulations to show that estimation error is greatly reduced when an investor needs to estimate only the volatilities and omits expected returns and correlations, which is consistent with the evidence in DeMiguel, Garlappi, Nogales, and Uppal (2009a) and others.[5] As expected, our results illustrate that the weights of the inverse volatility (IVol) portfolio are more stable and robust because the IVol rule is less affected by estimation error compared to the tangency (TAN) and the global minimum variance (GMV) portfolios.[6] We address the latter question of how far the inverse volatility rules are from the theoretically optimal portfolio by developing a statistical test of the necessary and sufficient conditions for the weights of the tangency portfolio to be statistically equivalent to the weights of the inverse volatility rules. Using a confidence interval of 5%,

[3]A detailed definition of the inverse volatility rules is provided in Section 4.

[4]We rely on results that appeared in the mathematical finance literature (see for example, Bodnar and Okhrin, 2011).

[5]Given the large literature addressing estimation error in portfolio choices, we provide a separate review in Section 1.1.

[6]Even if our goal is to evaluate the reduction in estimation error that the IVol has compared to the TAN portfolio, we report the results for the GMV portfolio as well because it allows us to visualize the estimation error coming from expected returns, correlations, and volatilities separately. In fact, the TAN portfolio requires all 3 components to be estimated, the GMV portfolio requires only correlations and volatilities, and the inverse volatility portfolio only the volatilities.

we find that for the majority of datasets considered in this study, there exists an inverse volatility portfolio for a given γ whose weights are statistically equivalent to those of the mean-variance optimal portfolio.

Next, we conduct an out-of-sample analysis on the performance of inverse volatility portfolio rules. Our goal is to evaluate how they compare to naïve diversification (1/N rule), mean-variance optimized portfolios as well as the norm-constrained portfolio of DeMiguel et al. (2009a) (henceforth DGNU1), which uses only the information from the variance-covariance matrix. We use 16 different datasets in this study including US equity portfolios (e.g. Fama-French 25 portfolios based on size and book-to-market), international equity portfolios (e.g. the Fama-French portfolios for Europe and Japan), a mix of equities, bonds, commodities portfolios built using data from Bloomberg, a dataset of individual stocks, and a large set of 76 anomalies in the cross-section of equity returns.[7] As in DeMiguel et al. (2009b), we rely on different metrics to evaluate the performance of the various allocation rules: Sharpe ratio, Certainty Equivalent Returns (CEQ), and turnover ratios.

We show that inverse volatility rules strongly outperform the 1/N rule in terms of Sharpe ratios. For example, the inverse volatility rule always delivers a higher Sharpe ratio than the 1/N rule and is statistically superior in 11 out of the 16 datasets considered in this paper. In the remaining 5 cases, the difference between the Sharpe ratios of the inverse volatility rule and 1/N rules is not statistically different from zero. We demonstrate that these differences in Sharpe ratios lead to large differences in wealth accumulation over time even after controlling for the same level of risk (volatility). For example, we show that investing \$1 in 1984 in the portfolios of equities, fixed income, and commodities using the 1/N rule would result in a wealth of $\approx$ \$50 in 2019 while investors' wealth would be almost four times greater ($\approx$ \$190) using the inverse volatility rule. The performance of the inverse

[7] We start from a sample of more than 200 anomalies, we keep only the ones whose returns are statistically different from zero according to the methodology in Hou et al. (2018) and obtain a final sample of 76 anomalies. The portfolio of anomalies is of particular interest to both academics and practitioners. By analyzing the portfolio of 76 anomalies, this paper contributes also to the construction of these "value-added" strategies in addition to the design of strategic asset allocation.

volatility rules with respect to the DGNU1 portfolio varies across datasets. In 9 out of 16 datasets, the DGNU1 portfolio performs similarly to the inverse volatility portfolios (i.e., exhibiting statistically the same Sharpe ratios) while in the remaining cases, the DGNU1 yields statistically superior Sharpe ratios.

Our contribution in this study is to show that we can develop a statistical test (i.e., the J-Test discussed in Section 3.1) that can give us an insight into whether the inverse volatility portfolios are theoretically as optimal as the mean-variance portfolio once we account for estimation error. We show that for many datasets this is indeed the case and, as expected, our empirical results back up this finding by revealing that inverse volatility rules perform well out-of-sample across different datasets. Overall, the out-of-sample performance that we document provides empirical support to the results of our statistical test. Our contribution is not to show that the inverse volatility portfolios supersede any portfolio construction techniques available. We acknowledge that there might be other portfolio construction techniques based on the sole use of the covariance matrix that can outperform inverse volatility portfolios by leveraging the reduction in estimation error due to the omission of expected returns.

The inverse volatility portfolio (IVol) also has the property that it is equivalent to the tangency portfolio when assets have the same Sharpe ratios and equal pairwise correlations. If these two assumptions (same Sharpe ratios and equal pairwise correlations) were observed empirically, then the inverse volatility portfolio would be equivalent to the tangency portfolio but would be considerably less affected by estimation error. While the assumption of equal Sharpe ratios across assets is not necessarily true, there is evidence that various asset classes have similar Sharpe ratios over a long period.[8] The assumption of equal pairwise correlations is more restrictive and it does not find empirical support. We ask the question: can we 'group'

[8] For example, Van Binsbergen (2020) shows the same Sharpe ratio of 0.22 for both S&P 500 and 10-year government bond from 1970-2021. Wright et al. (2014) find that the hypothesis of equality for the Sharpe ratios of 18 iShares ETFs cannot be rejected at the 1% level. Ardia and Boudt (2015) find that the hypothesis of equality for the Sharpe ratios using hedge fund returns from 2008-2012 cannot be rejected at the 1% level for 84.22% of all pairs of two Sharpe Ratios.

the data such they are closer to such an assumption? In this paper, we show that this is indeed possible. Specifically, we propose a clustering methodology that clusters together assets that are highly correlated between each other and exhibit a low correlation with assets in other clusters. The assets in the same cluster all have high correlations between each other therefore using the inverse volatility portfolio would be close to the tangency portfolio (i.e., theoretically optimal). We derive a closed-form solution for the weights of our clustering methodology and we show analytically that it is identical to the tangency portfolio when assets follow the two assumptions of equal Sharpe ratios and same pairwise correlations.

The next step is to decide how to cluster the data and then evaluate whether our clustering methodology works empirically. To cluster the data, we use machine learning algorithms to create buckets of assets that are homogeneous within-groups (high correlation) and heterogeneous between-groups (low correlation). Therefore, our methodology brings the data closer to the assumption of equal pairwise correlations across assets because we build inverse volatility portfolios using the assets within each cluster, which are highly correlated between each other (i.e., correlations all closer to one for an asset within a cluster). Once we have the 'within clusters' portfolios, we build an inverse volatility portfolio across clusters, which are close to being uncorrelated (i.e., correlations are close to zero between clusters).[9] We note that clustering the data should improve the performance of the portfolio when there is a wide variety of pairwise correlations across assets. For example, if all N assets are uncorrelated, then there are as many clusters as the number of assets and we should not expect improvements from clustering the data. Similarly, if all the N assets have a high correlation then they would all belong to the same (unique) cluster and we, again, should not see benefits from clustering.

Our empirical results show that using machine learning techniques to cluster the data improves the out-of-sample performance of the portfolio in terms of Sharpe ratios. Specif-

[9]At the time of rebalancing a portfolio, our methodology clusters the data using only historical information and, therefore, it is fully out-of-sample and does not suffer from a look-ahead bias.

ically, we compare the performance of inverse volatility portfolio rules with and without clustering the data. We use different clustering techniques and confirm that our results are robust across various algorithms, although the result is stronger for hierarchical clustering. Our findings show that clustering enhances the performance especially when there is a large number of assets to cluster and a large dispersion in the correlations between assets.[10] In most cases, the differences in Sharpe ratios between a portfolio built on clustered vs. unclustered data are large and confirm that clustering the data can provide benefits to investors when using the inverse volatility rule. For example, the dataset with the 76 anomalies – which has both a large number of assets and a high average dispersion of correlations – benefits from hierarchical clustering with increases in the Sharpe ratio in excess of 10%.

1.1 Literature Review

There is a large literature on portfolio allocation that focuses on designing optimal portfolios while minimizing the impact of estimation error. We can broadly divide the literature into Bayesian and non-Bayesian methods. We first discuss the literature that uses non-Bayesian methods, and we explain how this article contributes to it. We then review the strand of the literature that relies on Bayesian approaches, and we discuss how we apply some of these methods into our study.

First, some non-Bayesian studies achieve a trade-off between optimality and reducing the impact of estimation error by introducing structure on the data in order to obtain more stable estimates of expected returns and covariance matrix (DeMiguel et al., 2014).[11] Another approach is to combine portfolios to balance the trade-off between bias and estimation error. Examples of papers that follow this approach include Tu and Zhou (2011), Kan and Zhou (2007), Kan et al. (2021), and Anderson and Cheng (2016). Brandt et al. (2009) model port-

[10] We build a measure of the average correlation dispersion to evaluate how similar assets are.

[11] For example, DeMiguel et al. (2014) use a VAR model to capture the serial dependence of stock returns and use this information to tilt the portfolios' weights to improve the portfolios' performance. A similar approach is followed by the literature on forecasting the covariance matrix (Engle, 2002; Engle and Colacito, 2006, and many others).

folio weights as a function of firms' characteristics, and show that their methodology exhibits an out-of-sample performance which is improved over the standard mean-variance approach that models the distribution of returns without taking into account firms' characteristics. Other authors focused on the timing of strategies. For example, Kirby and Ostdiek (2012) and Moreira and Muir (2017) propose two simple timing strategies, which update portfolios' weights based on estimated changes in conditional volatilities.

While this literature enhances our understanding of the estimation of expected returns and covariances, our study provides both empirical evidence and a theoretical reason that relying solely on the variance-covariance matrix can be an effective method to greatly reduce estimation error and still achieve a good out of sample performance. Furthermore, we provide a statistical test to check whether we should expect inverse volatility portfolios to perform well on a given dataset.

Second, we move to the strand of the literature using a Bayesian approach to build optimal portfolios and reduce the effect of estimation error. Diffuse priors and Shrinkage are common methods used to make estimates of expected returns more stable. Shrinkage (Neyman, 1961) "shrinks" sample means towards a common "objective" mean. This reduces the variance of estimates and therefore the estimation error (Jobson and Korkie, 1981). Popular linear shrinkage methodologies include Jorion (1986), Ledoit and Wolf (2003), Ledoit and Wolf (2004a) Ledoit and Wolf (2004b). More recently, there are also non-linear shrinkage methodologies proposed. For example, Ledoit and Wolf (2017) shrink each of the sample covariance matrix's eigenvalues individually and determine the optimal shrinkage intensity for each based on its magnitude. Barroso and Saxena (2019) use the out-of-sample errors to compute the shrinkage parameters which require minimal assumptions about the data generating process. Pedersen et al. (2021) show that shrinking the correlations towards zero leads to better estimates not only of the variance-covariance matrix but also of the expected returns. The alternative to shrinkage is to form an informative prior using economic beliefs to reduce the arbitrariness of the statistical methods. In Tu and Zhou (2010), the priors are based on economic objectives such as the ratio of expected returns over variance. Pástor and Stambaugh (2000) and Pástor (2000) use investors' beliefs in an asset-pricing model to calculate the shrinkage factor and the common "objective".

MacKinlay and Pástor (2000) impose a restriction on the expected returns and variance-covariance estimates. They show that investors can achieve better expected returns estimates by jointly estimating expected returns using an asset pricing model and accounting for model misspecification (i.e. missing factors). Jagannathan and Ma (2003) show that imposing a short-sale constraint (i.e. requiring weights to be non-negative) improves the performance of mean-variance portfolios and also is equivalent to shrinking the extreme elements of the covariance matrix. DeMiguel et al. (2009a) propose portfolios that can be interpreted as resulting from shrinking the portfolio-weight vector instead of shrinking the moments of asset returns. They provide a Bayesian interpretation for their proposed portfolios and also for the portfolio by Jagannathan and Ma (2003). Different from the traditional Bayesian portfolio choice literature, they consider investors having a prior belief on the portfolio weights rather than asset-return distribution. We rely on this literature and estimate the covariance matrix using the Shrinkage estimator by Ledoit and Wolf (2004a,b). We present our empirical results when the covariance matrix is estimated using this Shrinkage estimator as well as the sample-based estimator.

Last, clustering techniques have been used recently in asset management. For example, de Prado (2016) uses hierarchical clustering to group assets iteratively and then allocates weights based on the inverse variance rule. We show the conditions under which clustering data – not limited to hierarchical clustering – can improve the performance of the inverse volatility rule. We develop a new clustering technique to build potentially more robust portfolios leveraging the properties of the inverse volatility rule and we provide empirical evidence that our proposed clustering methodology can improve the out-of-sample performance of inverse volatility portfolios.

The rest of the paper is organized as follows. Section 2 provides a motivation for the use of risk-based allocation rules. Section 3 develops a statistical test and evaluates whether the inverse volatility portfolios are statistically equivalent to the mean-variance optimal portfolio on various datasets. Section 4 provides the empirical results for the comparison of risk-based allocation rules with respect to naïve diversification as well as other portfolios that rely solely on the use of covariances. Last, Section 5 motivates the use of clustering and shows the related empirical results. Section 6 concludes.

2 Estimation Error and Performance

This section studies the effect of estimation error on the performance of various allocation rules: (1) the optimal tangency portfolio, which requires the estimation of both the expected excess returns and the variance-covariance matrix, (2) the global minimum variance portfolio, which requires solely the estimation of the variance-covariance matrix, and (3) the inverse volatility portfolio, which requires the estimation only of the volatilities. We elect to study the inverse volatility allocation rule because it is one of the risk-based allocation rules that we use for our empirical analysis in Section 4. The goal of this section is to highlight that the error caused by the estimation of expected returns is far larger than the error caused by the estimation of correlations and volatilities.[12]

2.1 Modeling estimation error and 3 benchmark portfolios

We follow Kan and Zhou (2007) and we consider the portfolio choice of an investor who chooses a portfolio amongst N risky assets and one risk-free asset. The investor does not know the true values of excess returns (μ) and variance-covariance matrix (Σ) but estimates

[12]Several studies have shown that estimates of expected returns are subject to large estimation errors and lead to inferior performance ex-post (e.g., Michaud, 1989; Merton, 1980; Chopra and Ziemba, 2013; Best and Grauer, 1991; Broadie, 1993).

them using their sample-based counterparts

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t \qquad\qquad \widehat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \widehat{\mu})(r_t - \widehat{\mu})' \tag{1}$$

where r_t is the vector of excess returns for the N risky assets at time t, and T is the length of the estimation window. Under the assumption that the returns of the N risky assets (r_t) are jointly normal, it is well-known that $\widehat{\mu}$ and $\widehat{\Sigma}$ are distributed as follows

$$\widehat{\mu} \sim \mathcal{N}(\mu, \Sigma/T) \qquad\qquad \widehat{\Sigma} \sim \frac{1}{T}\mathcal{W}_N(T-1, \Sigma) \tag{2}$$

where $\mathcal{N}(\cdot)$ denotes the normal distribution with mean μ and variance-covariance matrix Σ/T, and $\mathcal{W}(\cdot)$ denotes the Wishart distribution with $T-1$ degrees of freedom and covariance matrix Σ.

In the presence of estimation error, the optimal weights for the tangency portfolio $\hat{w}_{TAN}$ are known to be[13]

$$\hat{w}_{TAN} = \frac{\widehat{\Sigma}^{-1}\widehat{\mu}}{\mathbf{1}_N'\widehat{\Sigma}^{-1}\widehat{\mu}} \tag{3}$$

where $\mathbf{1}_N$ is a vector of ones.

Bodnar and Okhrin (2011) derive the characteristic functions of $(T\widehat{\Sigma})^{-1}\widehat{\mu}$ as follows.

$$\phi(t) = \int_0^{\infty} exp(i\frac{t'\Sigma^{-1}\mu}{z} - \frac{t'\Sigma^{-1}t}{2Tz^2})f_{\chi^2_{T-N}}(z)\phi_{F_{N-1,T-N+1,Ts}}(\frac{i(N-1)t'\Sigma^{-1}t}{2T(T-N+1)z^2})dz \tag{4}$$

where $f_{\chi^2_{T-N}}(\cdot)$ denotes the densities of χ^2 with $T-N$ degrees of freedom. $\phi_{F_{N-1,T-N+1,Ts}}(\cdot)$ denotes the characteristic function of the non-central F-distribution with $N-1$ and $T-N+1$ degrees of freedom and non-centrality parameter Ts. From Johnson et al. (1995), the

[13]This follows from the standard optimal portfolio choice problem of Markowitz (1952) and it has been solved by many authors. For example, see Kan and Zhou (2007) for a mathematical derivation.

characteristic function can be written as

$$\phi_{F_{N-1,T-N+1,Ts}}\left(\frac{i(N-1)t'\Sigma^{-1}t}{2T(T-N+1)z^2}\right) = exp\left(-\frac{Ts}{2}\right)\sum_{j=0}^{\infty}\frac{(Ts/2)^j}{j!}\times {}_1F_1\left(\frac{N-1}{2}+j, -\frac{T-N+1}{2}; \frac{t'\Sigma^{-1}t}{2Tz^2}\right) \tag{5}$$

where $s = \mu' R_1 \mu$ and $R_1 = \Sigma^{-1} - \Sigma^{-1}\mathbf{1}_N\mathbf{1}'_N\Sigma^{-1}/\mathbf{1}'_N\Sigma^{-1}\mathbf{1}_N$. ${}_1F_1(a, b; x)$ is the confluent hypergeometric function of the first kind defined as $\frac{\Gamma(b)}{\Gamma(a)}\sum_{i=0}^{\infty}\frac{\Gamma(a+i)}{\Gamma(b+i)}\frac{x^i}{i!}$. However, Okhrin and Schmid (2006) prove that the estimator for the weights for the tangency portfolio does not possess finite moments.[14] Therefore, we rely on simulations to study the estimation error for the weights of the tangency portfolio.

The two portfolios that we compare with the tangency portfolio are the inverse volatility portfolio and the global minimum variance portfolio. The weights of the inverse volatility portfolio $\hat{w}_{IVol}$ are defined as

$$\hat{w}_{IVol} = \frac{1/\bar{\sigma}}{\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i}} \tag{6}$$

where $\widehat{\sigma}_i$ is the variance estimate of asset i, and $1/\bar{\sigma} = \left[\, 1/\widehat{\sigma}_1 \;\; 1/\widehat{\sigma}_2 \;\; \cdots \;\; 1/\widehat{\sigma}_N \,\right]$ is a vector with the reciprocal of the N asset volatilities.

The optimal weights of the global minimum variance portfolio $\hat{w}_{GMV}$ are

$$\hat{w}_{GMV} = \frac{\widehat{\Sigma}^{-1}\mathbf{1}_N}{\mathbf{1}'_N\widehat{\Sigma}^{-1}\mathbf{1}_N} \tag{7}$$

The question that we address in the next section is how errors in expected returns and volatilities affect the estimation error of the 3 aforementioned portfolios.

2.2 Visualizing the Estimation Error

We investigate how severe the estimation error is by comparing the Sharpe ratios that the 3 aforementioned allocation rules would generate if investors did not know the true values of μ and Σ. Our procedure can be described as follows. We simulate the expected excess returns

[14] Okhrin and Schmid (2006) show that the vector of the estimated weights for the tangency portfolio is infinity while the moments larger than one are not defined.

and variance-covariance matrix ($\widehat{\mu}$ and $\widehat{\Sigma}$) using Equation (2). For every simulation i, we have independent draws $\widehat{\mu}_i$ and $\widehat{\Sigma}_i$. Using the allocation rules defined by Equations (3), (6) and (7), we obtain the optimal weights $\widehat{w}_{i,TAN}$, $\widehat{w}_{i,GMV}$, $\widehat{w}_{i,IVol}$. When investors use allocation rule j and there is estimation error, the expected excess returns ($\mu_{j,P}$) and standard deviation ($\sigma_{j,P}$) of the portfolio are:

$$\widehat{\mu}_{j,P} = \widehat{w}_j' \mu \qquad\qquad \widehat{\sigma}_{j,P} = \widehat{w}_j' \, \Sigma \, \widehat{w}_j \tag{8}$$

where $\widehat{w}_j$ are the weights chosen by the investors following allocation rule j (e.g. global minimum variance). It follows that the Sharpe ratio of allocation rule j subject to estimation error is:

$$\widehat{SR}_j = \frac{\widehat{\mu}_{j,P}}{\widehat{\sigma}_{j,P}} = \frac{\widehat{w}_j' \mu}{\widehat{w}_j' \, \Sigma \, \widehat{w}_j} \tag{9}$$

If investors knew the true values of μ and σ, the true Sharpe ratio of allocation rule j would be:

$$SR_j = \frac{w_j' \mu}{w_j' \, \Sigma \, w_j} \tag{10}$$

where w_j are the weights for rule j when the investors know the true values of μ and Σ.

In Figure 1, we plot the portfolios performance in terms of expected excess returns $\mu_{j,P}$ and covariances $\sigma_{j,P}$ for 5,000 simulations when investors are subject to estimation error. On the y-axis, the figure shows the average portfolio excess return while the x-axis shows its standard deviation. In all panels, the green solid line shows the efficient frontier using the true mean excess return μ and variance-covariance matrix Σ. All rates, variances, and volatilities reported are at a monthly frequency and the estimation window used is $T = 60$ months (5 years).[15] The true volatility of each asset i (σ_i) is the same across all simulations.[16] In all panels, assets are assumed to have the same Sharpe ratios. That is, for any two assets

[15]The effect of the length of the estimation window is studied further below in Internet Appendix C.

[16]We set the true volatility by randomly drawing from a uniform distribution between 10% and 40% for a fixed seed so that the true volatility is the same across simulations. We also ran untabulated robustness tests for different ranges (e.g. random draw from a uniform distribution between 15% to 30%, normal distribution with mean 20% and 5% standard deviation). The results are qualitatively the same.

i, j, we assume that $\mu_i/\sigma_i = \mu_j/\sigma_j = 1/12$. Since our simulations are done with a monthly frequency, assuming a monthly Sharpe ratio of 1/12 corresponds to annualized Sharpe ratio of approximately 28.87%. This assumption might seem ad-hoc but it is justified by an empirical observation. See for example, Van Binsbergen (2020), Wright et al. (2014), and Ardia and Boudt (2015).

[Insert Figure 1 here]

The takeaways from Figure 1 can be summarized as follows: (i) the tangency portfolio presents a much wider dispersion than the other two portfolios, thus confirming the results from the literature (Michaud, 1989; Merton, 1980; DeMiguel et al., 2009b) that errors in expected returns greatly affect the performance and stability of the assets' weight; (ii) the severity of the dispersion in asset weights increases when we increase the number of assets as shown by comparing Panel A (5 assets) and Panel B (15 assets).[17]

3 When is inverse volatility equivalent to the tangency portfolio?

This section is organized as follows. Section 3.1 discusses a test for the necessary and sufficient conditions under which inverse volatility rules are equivalent to the mean-variance portfolio. Section 3.2 describes the empirical datasets used in this study. Section 3.3 provides a formal definition of the various allocation rules considered in this study. Last, in Section 3.4, we apply such a test to the empirical datasets and discuss the results.

[17] The fact that the performance of the TAN portfolio is negatively affected by the increase in the number of assets has been shown empirically by several studies (e.g. Brandt et al., 2009; Simaan, 1997).

3.1 Necessary and sufficient conditions for optimality of IVolγ portfolios

In this section, we discuss the necessary and sufficient conditions for the γ-inverse volatility portfolio to be equivalent to the tangency portfolio (MVO).[18] The test is based on the Generalized Method of Moments (GMM). Let μ and Σ be the mean and covariance matrix of N risky assets. We can write the weights of the tangency portfolio and generalized inverse volatility portfolio of the N assets as follows

$$w_{MVO} = \frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu}, \tag{11}$$

$$w_{IVol\gamma} = \frac{D^{-\gamma}1_N}{1_N'D^{-\gamma}1_N}, \tag{12}$$

where $D^\gamma \equiv \text{Diag}(\Sigma)^{\frac{\gamma}{2}} = \text{Diag}(\sigma_1^\gamma, \ldots, \sigma_N^\gamma)$ is a diagonal matrix of the volatility to the power of γ of the N assets.

Proposition 1. *A necessary and sufficient condition for $w_{MVO} = w_{IVol\gamma}$ is that the cross product of $\Sigma^{-1}\mu$ and $D^{-\gamma}1_N$ is a vector of zeros (i.e., $\Sigma^{-1}\mu \times D^{-\gamma}1_N = 0_N$).*

We provide the proof of Proposition 1 in Internet Appendix A.

The next step is to develop a statistical test to evaluate the equivalence between tangency and inverse volatility portfolios.

Proposition 2. *Let w_{MVO} and $w_{IVol\gamma}$ be defined as in Equation (11) and Equation (12). The null hypothesis that the inverse volatility portfolio for a given γ is equivalent to the tangency portfolio can be tested using the following statistics*

$$J = T(P'\hat{c})'(P'\hat{V}(\hat{c})P)^{-1}(P'\hat{c}) \xrightarrow{d} \chi^2_{N-1}, \tag{13}$$

where P is an $N(N-1)$ orthonormal matrix with its columns orthogonal to 1_N, $\hat{c} = \hat{D}^\gamma\hat{\Sigma}^{-1}\hat{\mu}$, $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T R_t$, $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^T (R_t - \hat{\mu})(R_t - \hat{\mu})'$, $\hat{D}^\gamma = Diag(\hat{\Sigma})^{\frac{\gamma}{2}}$, and R_t is a vector of returns,

[18] We are immensely grateful to Raymond Kan for his suggestion and help on the derivation of this test.

and $\hat{V}(\hat{c})$ is a consistent estimator of $V(\hat{c})$.

We provide the proof of Proposition 2 in Internet Appendix A.

The null hypothesis H_0 of the test in Equation (13) is that the inverse volatility portfolio for a given γ is equivalent to the tangency portfolio. If the test cannot reject H_0, it means that the weights of IVolγ and MVO portfolios are statistically indistinguishable. What is the intuition behind our test? We know that the tangency portfolio is the portfolio that maximizes the Sharpe ratio, and any portfolio different from the tangency portfolio achieves a lower in-sample Sharpe ratio. However, we ought to know whether the difference with the tangency portfolio is statistically significant and our test provides an answer to this question.

3.2 Description of the datasets

We consider 16 different datasets which we group into Multi-Assets datasets, which contain both equities, fixed and commodities, Equities-Only datasets, and a set of anomalies in the cross-section of equity returns.

For the Equities-Only datasets, we use various portfolios of domestic and international equities. We consider various industry portfolios of domestic equities based on the classification of Kenneth French: the 10 Industry portfolios (10 Industries), the 17 Industry portfolios (17 Industries), the 30 Industry portfolios (30 Industries), and the 49 Industry portfolios (49 Industries) from Kenneth French's website.[19] These portfolios allow us to evaluate the robustness of our results across various definitions of industries as well as a various number of assets. We also consider portfolios sorted on several firms' characteristics. We download data from Kenneth French's website for the 25 portfolios sorted on size and book-to-market (25 Size and B/M), and the 25 portfolios sorted on size and operating profitability (25 Size and Operating Prof). We also consider the 12 portfolios based on the Fama and French (2015) 5-factor model (12 FF5 Portfolios): 3 portfolios sorted on book-to-market (low, medium, high), 3 portfolios sorted on investment (low, medium, high), 3 portfolios sorted on operat-

[19] The data can be downloaded from Kenneth French's website at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

ing profitability (low, medium, high), and 3 portfolios sorted on size (low, medium, high). Last, in the domestic equity group, we consider 11 portfolios based on the q-factor model of Hou, Xue, and Zhang (2015) and Hou, Mo, Xue, and Zhang (2019) (11 q-Portfolios): 3 portfolios sorted on ROE, 3 portfolios sorted on investment over asset ratio, 3 portfolios sorted on expected growth, and 2 portfolios sorted on size. Last, we also consider a dataset of individual stocks. Specifically, we select 75 random individual U.S. stocks, which traded continuously on the market between 1967 and 2019 to have a balanced panel of data.

For the international portfolios of equities, we consider the 25 portfolios sorted on size and book-to-market using Japanese stocks (25 Japan Size and B/M), the 25 portfolios sorted on size and book-to-market using European stocks (25 Europe Size and B/M), the 25 portfolios sorted on size and book-to-market using Asian stocks excluding Japan (25 Asia (exc. Japan) Size and B/M). All these portfolios are available from Kenneth French's website.

For the Multi-Asset Portfolios, we consider a set of assets that include equities, fixed income, and commodities. We build this set by using two large indices of equities, S&P 500 and Russell 2000. These two indices cover well both the large cap (S&P 500) and the small cap (Russell 2000). For the fixed income indices, we use the total returns for 10- and 30-year US government bonds from Global Financial data – tickers TRUSG10M and TRUSG30M. For commodities, we use four indices: the S&P GSCI Total Return Index (ticker _GTCD) and the Goldman Sachs Commodity Price Index (ticker _SPGSCID) which are both composite indices of various commodities returns, the S&P GSCI Livestock Index (ticker _SPGSLVD) and S&P GSCI Industrial Metals Index (ticker _SPGSIND) which capture the returns of Livestock and Metal commodities. We build a set called "Equity, Fixed Income, Commodities, and 10 S&P Industries" by adding 10 S&P industry portfolios from Bloomberg (tickers SPTRSC10 Index, SPTRSC20 Index, and so on) to the set of equities, fixed, and commodities. We add this set to increase the number of assets and, as we show in Section 5, show that when the number of assets is sufficiently large and dispersed in terms of correlations, clustering the data can improve the out-of-sample performance of the inverse volatility rule.

[Insert Table 1 here]

Last, we consider 76 anomalies in the cross-section of equity returns, which are also known as "absolute return strategies", as datasets. The list of the anomalies can be found in Internet Appendix D. We compiled the set of 76 anomalies as follows. Following Hou, Xue, and Zhang (2018), we only consider anomalies that have been constructed using value-weighted returns, NYSE-breakpoints and that generate excess returns that are statistically different from zero. If the authors of the original study publish the data, we download the anomalies from their websites. Otherwise, we build them ourselves. All the anomalies studied in this paper produce excess returns that are statistically significant from zero for the time period from 1981 to 2019. In Internet Appendix D, we list all the anomalies as well as the reference paper that we followed to replicate them or the paper of the authors from whom we obtained the data.

For all the aforementioned datasets, we calculate excess returns with respect to the risk-free rate, which is downloaded from Kenneth French's website. A list of the portfolios described above and the details of when the data start and end are provided in Table 1.

3.3 Description of the allocation rules

In our empirical tests, we benchmark against the equally weighted portfolio (1/N) as well as the norm-constrained portfolio of DeMiguel, Garlappi, Nogales, and Uppal (2009a). The 1/N portfolio – which gives the same weight to all assets and does not require the estimation of expected returns or variance-covariance matrix – has been shown to be robust across many datasets and outperforms mean-variance optimization (DeMiguel, Garlappi, and Uppal, 2009b). In our empirical tests, we report the results for the tangency portfolio (MVO) as well as various inverse volatility portfolios for various levels of gamma. For ease of notation, we label the inverse volatility rule for a given γ as the IVolγ rule. We describe the MVO and IVolγ rules below.

Regarding the MVO Rule, at each rebalancing date, t investors choose the optimal weights $w_{t,MVO}$ to maximize the quadratic utility function

$$w_{t,MVO} = \arg\max_{w_t} w_t'\widehat{\mu}_t - \frac{\lambda}{2} w_t'\widehat{\Sigma}_t w_t$$

where $\widehat{\mu}_t$ is the estimated vector of expected returns, $\widehat{\Sigma}_t$ is the estimated variance-covariance matrix, and λ is the risk aversion coefficient. The solution to the above problem is[20]

$$w_{t,MVO} = \frac{\widehat{\Sigma}_t^{-1}\widehat{\mu}_t}{\mathbf{1}_N'\widehat{\Sigma}_t^{-1}\widehat{\mu}_t}$$

In our empirical results, we calculate the norm-constrained portfolio following DeMiguel et al. (2009a). We label this portfolio "DGNU1" as we use the 1-norm constraint with the threshold parameter calibrated by maximizing portfolio return in the previous period. We choose the 1-norm constraint because it contains the minimum variance portfolio with short-sales constraints as a special case when the threshold parameter is equal to 1.[21]

A key point is how to estimate the expected returns (for the MVO portfolio only) and covariance matrix. In our main results, we use sample-based estimates of expected returns. For the covariance matrix, we use a "robust" estimate using the Shrinkage estimator of Ledoit and Wolf (2004b).[22] We elect to use the Shrinkage estimator of Ledoit and Wolf (2004b) because it has been shown to provide better out-of-sample results compared to the sample-based estimator. For the datasets that contain less than 60 portfolios, we use an estimation window of 60 months. For those datasets with more than 60 assets (76 Anomalies, and Individual Stocks), we use a longer estimation window of 120 months.

[20]This is a well-known solution and we omit the derivation in the interest of brevity. We refer to Best (2010) for a textbook treatment of the derivation of this result.

[21]We prefer the 1-norm to the 2-norm constraint because the 2-norm constraint is very close to the 1/N portfolio. As DeMiguel et al. (2009a) write "we would expect that the 2-norm-constrained portfolios will, in general, remain relatively close to the 1/N portfolio and thus will assign a positive weight to all assets."

[22]Ledoit and Wolf (2004b) show that the largest (smallest) sample eigenvalues are systematically biased upwards (downwards). They develop a shrinkage estimator that corrects this bias by pulling down (up) the largest (smallest) eigenvalues toward the mean of all sample eigenvalues.

IVolγ Rule. According to the γ-inverse volatility portfolio (Kirby and Ostdiek, 2012), the weights of each asset are proportional to the inverse of their volatility to the power of γ. Formally, the optimal weights of the inverse variance portfolio $w_{t,IVol}$ are

$$w_{t,IVol} = \frac{\left(1/diag(\sqrt{\widehat{\Sigma}_t})\right)^{\gamma}}{\sum_{i=1}^{N} \frac{1}{(\widehat{\sigma}_{t,i})^{\gamma}}}$$

where $1/diag(\sqrt{\widehat{\Sigma}_t})$ is a vector with the reciprocal of the assets' volatilities. Last, all allocation rules are rebalanced monthly.

3.4 Using the J-test empirically

We apply the test in Equation (13) to our datasets and report the results in Table 2. In Table 2, we report the p-value of the test (row "J p-val" for each dataset) as well as the in-sample Sharpe ratio of the IVolγ rule for various levels of γ (row "SR"). First, we note that for $\gamma = 0$ – which is the same as the 1/N rule – , the IVol0 shows a p-value that is always lower than 5% with the exception of three cases (30 Industries p-value = 0.087, 25 Japan Size and B/M p-value = 0.08, and Individual Stocks p-value of 0.475). This implies that our test rejects the null hypothesis that 1/N is equivalent to the tangency portfolio at a 5% significance level for the majority of datasets considered in this study. Our result demonstrates via a statistical test that naïve diversification is sub-optimal compared to the tangency portfolio. Interestingly, for many datasets increasing the value of γ also increases the p-value of the test, and often the test cannot reject the null hypothesis. This means that we cannot reject the null hypothesis that IVolγ is equivalent to the tangency portfolio greater than or equal to 1. This is surprising because it means that, for various levels of γ, investing in the IVolγ rule is not statistically different from the optimal tangency portfolio and it requires fewer parameters to be estimated (i.e. volatilities only). Therefore, our results provide a justification for why the inverse volatility rules are known to perform well out-of-sample (e.g., Kirby and Ostdiek, 2012) and outperform mean-variance portfolios.

A special observation is required for the dataset of Individual Stocks. The p-value of the test is close to 1, strongly suggesting that the test cannot reject the null hypothesis of equivalence between mean-variance and inverse volatility portfolios. The p-values for this dataset of Individual Stocks are much higher than the p-value for the datasets formed on portfolios suggesting that the higher idiosyncratic volatility – which affects individual stocks much more than portfolios – makes the mean-variance portfolio statistically equivalent not only to the inverse volatility portfolios but also to 1/N. In other words, when building a portfolio of individual stocks, a simple equally weighted portfolio is not statistically different from a mean-variance optimal portfolio.

Overall, Table 2 shows that IVolγ rules are closer to theoretical optimality compared to the 1/N rule. Of course, our test is conducted in-sample and it does not take into account the effect of estimation error. The 1/N rule does not suffer from estimation error while IVolγ rules have some estimation error due to the estimation of the assets' volatilities. Does the estimation error dwarf the benefits of diversification from the IVolγ rules? We demonstrate in the next section that this is not the case.

[Insert Table 2 and Figure 2 here]

Our test can also be used to visualize what the optimal γ should be from an in-sample analysis. Figure 2 plots the p-values from the J-test described in Section 3.1 for the three datasets that have a combination of equities, fixed, and commodities. Panel A plots the p-value for the set of Equities, Fixed, and Commodities which exhibits a bell curve with a maximum of around 2.8. Although for no value of γ the p-value is higher than 5% (i.e. the tests always reject the hypothesis of equivalence with the tangency portfolio), the figure suggests that for $\gamma = 2.8$, the IVolγ strategy would have historically performed better than IVol0 or IVol1 which exhibit a p-value of almost zero. Panel B and Panel C show qualitatively similar results when we plot the p-value for the "Equity and Fixed Income" and "Equity, Fixed Income, Commodities and 10 S&P Industries" datasets. We highlight that this analysis is conducted in-sample and it provides a visualization of the statistical test using the available history of data. We address the out-of-sample performance of the various strategies in Section 4.

4 Out-of-Sample Tests

4.1 Empirical Out-of-Sample Performance

Our aim is to study the performance of the inverse volatility rules across the various datasets described in Section 3.2. We start by analyzing the performance of the Sharpe ratios. Table 3 presents the monthly Sharpe ratios.[23] The column "Dataset" contains the name of the dataset. The various allocation rules are: (1) the "1/N" rule (naïve diversification); (2) the tangency portfolio from Mean-Variance Optimization (MVO); (3) the 1-norm-constrained portfolio developed in DeMiguel et al. (2009a); and (4) the IVolγ rule for $\gamma \in \{1, 2, 3, 4, 5\}$. In parenthesis, we report the p-value of the test for the difference between Sharpe ratios of each allocation rule with respect to 1/N, which is the naïve diversification benchmark. We develop our own GMM test for the equality of Sharpe ratios, which is described in Internet Appendix B. While Memmel (2003)'s test – which is commonly used (e.g., DeMiguel et al., 2009b) – requires the data to be normally distributed, the test presented here does not require such an assumption and is therefore more general.

Table 3 shows that inverse volatility rules strongly outperform the 1/N rule in terms of the Sharpe ratio. Specifically, the IVol1 always delivers a higher Sharpe ratio than the 1/N rule, and it is statistically superior – at a 5% significance level – in 11 out of 16 datasets considered in this study. In Internet Appendix F, we show that these differences in Sharpe ratios lead to large differences in wealth accumulation over time even after controlling for the same level of risk (volatility). An interesting point is to connect the results from Table 2 with those

[23]We describe the performance measures in Internet Appendix E.

of Table 3. For example, in Table 2, we showed that for the dataset "Equities, Fixed and Commodities", our J-test showed that the in-sample γ that would yield the highest Sharpe ratio is 2.8. Interestingly, Table 3 confirms this: the SR for IVol3 (i.e., using a $\gamma = 3 \approx 2.8$) is the largest and it is statistically different from the SR of the 1/N portfolio. For higher values of γ, the p-value of the test for the equality between 1/N and the IVolγ starts to increase, thus showing that we cannot reject the null hypothesis of equality in Sharpe ratios. Indeed, for $\gamma = 5$, the SR is 0.337 – which is greater than the 0.235, the SR of 1/N – but the p-value of the test for the null hypothesis that the two Sharpe ratios are statistically equal cannot be rejected at the 5% confidence level (i.e, p-value = 0.066).

Overall, our findings provide empirical support to the statistical test results discussed in Section 3. We showed in Table 2 that for many datasets there is no statistical difference between the weights of the optimal mean-variance portfolio (TAN) and the inverse volatility rules (IVolγ). In the out-of-sample tests presented in Table 3, we show that the inverse volatility rules are not only equivalent to the tangency portfolio but they outperform it because of lower estimation error as they do not require the estimation of expected returns and correlations.

Our results also provide another insight into why inverse volatility rules perform better than 1/N. In Table 2, we showed that the p-value for the test of the null hypothesis that the 1/N rule is equivalent to the tangency portfolio is lower than the p-value of the test using inverse volatility rules. This means that our test rejects with higher confidence the null hypothesis that 1/N is optimal in a mean-variance framework compared to the null hypothesis that inverse volatility rules are optimal. Our out-of-sample tests show that the inverse volatility rules outperform 1/N even in the presence of estimation error (which is absent in the 1/N rule). This implies that in the majority of cases the diversification benefits from the inverse volatility rules outweigh the drawbacks caused by estimation error.

[Insert Table 3 and Table 4 here]

How do inverse volatility portfolios fare with respect to recently developed portfolios that use only covariances as input? We address this question in Table 4 where we report the Sharpe ratios of the various allocation rules again but this time we report the p-value (in parenthesis) for the test of the null hypothesis that the Sharpe ratio of a given allocation rule is equivalent to the DGNU1 portfolio. The DGNU1 is the 1-norm-constrained portfolio developed in DeMiguel et al. (2009a), which is closely related to the global minimum variance portfolio but with a norm-constraint on the asset weights. This portfolio has the advantage of not using expected returns, thus reducing estimation error. In general, the DGNU1 portfolio performs very well and it vastly outperforms both 1/N as well as the mean-variance portfolio (MVO). The next logical question to address is whether the DGNU1 portfolio supersedes inverse volatility portfolios. The answer to such a question largely depends on which dataset we consider. For example, for the "Equities and Fixed Income" dataset, the Sharpe ratios of the DGNU1 portfolio and the inverse volatility portfolios are statistically identical as shown by the high p-values (in parenthesis). That is, our analysis shows that the data cannot reject the null hypothesis test that the Sharpe ratios are equal between a given inverse volatility portfolio and the DGNU1 portfolio. For other datasets, such as the "76 Anomalies" dataset, the DGNU1 portfolio performs better than the inverse volatility portfolios as shown by the low p-values.

Overall, in 9 out of 16 datasets the DGNU1 portfolio performs similarly to the inverse volatility portfolios (i.e., exhibiting statistically the same Sharpe ratios) while in the remaining cases, the DGNU1 yields statistically superior Sharpe ratios. As we discussed above, our contribution is not to show that the inverse volatility portfolios supersede any portfolio construction techniques available. We acknowledge that there might be other portfolio construction techniques based on the sole use of the covariance matrix that can outperform inverse volatility portfolios by leveraging the reduction in estimation error due to the omission of expected returns. However, our contribution in this study is to show that we can develop a statistical test (i.e., the J-Test discussed in Section 3.1) that can give us an insight into whether the inverse volatility portfolios are theoretically as optimal as the mean-variance portfolio once we account for estimation error.

After the analysis of Sharpe ratios, we turn our attention to certainty equivalent returns (CEQ), which we study in Table 5. For ease of readability, we multiply the CEQs by 100. Despite not being the result of an optimization problem that maximizes the quadratic utility function, Table 5 shows that inverse volatility rules achieve an overall good performance in terms of CEQs. The CEQs of inverse volatility portfolios are either statistically superior or identical to those of 1/N in all datasets with the exception of the dataset of "76 Anomalies" for which IVolγ rules exhibit lower CEQs than 1/N.

We study the turnovers in Panel A of Table 6. An allocation rule that has lower turnover leads to lower transaction costs and, therefore, higher net of fees returns. The turnovers of inverse volatility rules are on average lower than that of the 1/N rule. For example, the IVol1 rule applied on the set of "76 Anomalies" leads to a turnover that is 7% lower than that of the 1/N rule as shown by the fact that the ratio of the turnover of IVol1 over the turnover of 1/N is 0.93. This result combined with the fact that inverse volatility rules also deliver higher Sharpe ratios implies that the return net of transaction costs of the IVolγ rules are considerably higher than that of the 1/N rule. Notably, the turnover of the DGNU1 portfolio – which performs very well in terms of Sharpe ratios and CEQs – exhibits large turnover ratios compared to the inverse volatility portfolios in the majority of datasets. For example, using the set of "Individual Stocks", the ratio of the inverse volatility portfolios' turnover with respect to the 1/N rule ranges between 0.78 and 0.95 depending on the value of γ while it is equal to 2.46 for the DGNU1 portfolio. This means that the ratio of the turnover between the DGNU1 and inverse volatility portfolios ranges between 2.59 (=2.46/0.95) and 3.15 (=2.46/0.78). Last, consistent with the findings in DeMiguel et al. (2009b), Panel A of Table 6 shows that the turnovers of the tangency portfolio (TAN) are larger than the turnovers of the 1/N rule for almost all datasets.

[Insert Table 5 and Table 6 here]

In Panel B of Table 6, we show the annualized return-gain of each allocation rule with respect to the 1/N rule in terms of Sharpe ratios. The return-gain is defined in Equation (E.4) and we assume that there are proportional transaction costs as described in Internet Appendix E. The inverse volatility rules have a positive return-gain in the vast majority of cases. This means that the 1/N rule would have to deliver additional net of transaction costs returns in order to perform as well as these rules. Specifically, the return-gains are positive for all datasets with the exception of the "76 Anomalies" and the set of "Equity, Fixed Income". Overall, the results in Panel B of Table 6 combined with the results in Table 3 show that IVolγ portfolios not only deliver a higher Sharpe ratio before transaction costs but also an even greater Sharpe ratio when investors take into account transaction costs. For most datasets, the 1/N rule would have to deliver an extra $\approx 1\%$ per year, after transaction costs, in order to perform as well as the inverse volatility rules in terms of the Sharpe ratio.

5 Does clustering improve the performance?

5.1 Clustering with Inverse Volatility Rule

We have shown so far that the IVolγ portfolios outperform naïve diversification and mean-variance optimized portfolios and, more importantly, we built a statistical test (i.e., the J test described in Section 3) to check whether the inverse volatility rules are equivalent to the mean-variance portfolio from a statistical point of view. For the development of the test, we leveraged the properties of the inverse volatility rules portfolios. In this section, we conduct a similar analysis: we ask whether we can use a property of the inverse volatility portfolio – which we discuss below – in order to increase its out-of-sample performance via clustering. We describe our intuition in two steps: (i) we discuss the property of the inverse volatility portfolio that makes it optimal – even in the absence of estimation error – from a mean-variance portfolio when assets have the same Sharpe ratio and cross-correlations are all equal; (ii) while there is empirical evidence that various assets have similar Sharpe ratios over a long period (e.g., Van Binsbergen, 2020), we do not find evidence that pairwise correlations are the same across assets which makes the inverse volatility rule different from the mean-variance portfolio in the absence of estimation error; (iii) we describe how clustering via machine learning can bring the data closer to the required assumption of equal pairwise correlations, thus making the inverse volatility portfolio close to the theoretically optimal mean-variance portfolio.

First, we show that under the assumption that Sharpe ratios across assets are equal and if pairwise correlations are the same then the inverse volatility portfolio converges to the optimal tangent portfolio. Formally, assume that we have N assets with a vector of excess returns and covariance matrix equal to μ and Σ, respectively. Let us define a matrix with the volatilities on the diagonal

$$D \equiv \text{Diag}(\Sigma)^{1/2} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_N \end{bmatrix}$$

Proposition 3. *Let (i) $\bar{s}$ be the Sharpe ratio common to all assets such that $\mu_i/\sigma_i = \bar{s}$ for each asset $i \in \{1, 2, \cdots N\}$, and (ii) and let the correlations across any two assets i, j be equal to the constant ρ (i.e., $\rho_{i,j} = \rho$ for each i, j, where $\rho_{i,j}$ is the correlation between assets i and j). Then the inverse volatility portfolio is equivalent to the optimal mean-variance portfolio (i.e., $w_{IVol1} = w_{MVO}$, where $w_{IVol1} = \frac{D^{-1}1_N}{1_N' D^{-1} 1_N}$ and $w_{MVO} = \frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu}$*

Proof. Using the assumption (ii) from Proposition 3 that pairwise correlations are all equal, let us rewrite the correlation matrix Q as

$$Q = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & \cdots & 1 \end{bmatrix} = \rho \mathbf{1}_N \mathbf{1}'_N + (1-\rho)\mathbf{I}_N$$

where $\mathbf{I}_N$ is the identity matrix, and $\mathbf{1}_N$ is a column vector of ones. We can write the variance-covariance matrix as

$$\Sigma = D \cdot Q \cdot D \tag{14}$$

.

By definition of the inverse volatility portfolio (see Equation (12)), we have that the weights w_{IVol1} are proportional to $D^{-1} \cdot \mathbf{1}_N$. We also know that the weights of the tangency portfolio w_{TAN} are proportional to $\Sigma^{-1} \cdot \mu$ as it can be seen from Equation (11). We can therefore write $\Sigma^{-1} \cdot \mu = D^{-1} \cdot Q^{-1} \cdot D^{-1} \cdot \mu$ using Equation (14). Formally, we can write the weights of the $IVol1$ and MVO portfolios as follows

$$w_{IVol} \propto D^{-1} \cdot \mathbf{1}_N$$
$$w_{TAN} \propto D^{-1} \cdot Q^{-1} \cdot D^{-1} \cdot \mu$$

To prove that $w_{TAN} = w_{IVol}$ it is sufficient to show that (a) μ is proportional to $D \cdot \mathbf{1}_N$ and (b) that $Q^{-1}\mathbf{1}_N$ is proportional to $\mathbf{1}_N$.

To prove (a) (i.e., that μ is proportional to $D\mathbf{1}_N$), we use the assumption (i) that Sharpe ratios across assets are equal to $\bar{s}$ and we can write $\mu = \bar{s}D\mathbf{1}_N$. This proves that $\mu \propto D\mathbf{1}_N$ when Sharpe ratios are equal across assets.

To prove (b) (i.e., that $Q^{-1}\mathbf{1}_N$ is proportional to $\mathbf{1}_N$), we proceed in two steps.

First, we show that one of the eigenvectors of Q is $\mathbf{1}_N$. Consider the matrix $B \equiv$

$Q + (\rho - 1)\mathbf{I}_N$ whose elements are all equal to ρ. Note that

$$B\mathbf{1}_N = \begin{bmatrix} \rho & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & \cdots & \rho \end{bmatrix} \mathbf{1}_N = \begin{bmatrix} N\rho \\ N\rho \\ N\rho \\ N\rho \end{bmatrix} = N\rho \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = N\rho\mathbf{1}_N$$

which proves that $\mathbf{1}_N$ is an eigenvector of B. Let $\mathbf{x}$ be an eigenvector of Q and λ_x be its associated eigenvalue. It follows that

$$\begin{aligned} &Q\mathbf{x} = \lambda_x \mathbf{x} \\ &(Q + (\rho - 1)\mathbf{I}_N)\mathbf{x} = \lambda_x \mathbf{x} + (\rho - 1)\mathbf{I}_N \mathbf{x} \implies \\ &(Q + (\rho - 1)\mathbf{I}_N)\mathbf{x} = (\lambda_x + \rho - 1)\mathbf{x} \implies \\ &B\mathbf{x} = (\lambda_x + \rho - 1)\mathbf{x} \end{aligned} \tag{15}$$

The last step in Equation (15) follows from the definition of the matrix B. Equation (15) shows that the matrices Q and B have the same eigenvectors. Therefore, since $\mathbf{1}_N$ is an eigenvector of B, it is also an eigenvector of Q.

Second, using the fact the pairwise correlations are constant, we can write $Q = \rho\mathbf{1}_N\mathbf{1}_N' + (1 - \rho)\mathbf{I}_N$. Since one of the eigenvectors of Q is $\mathbf{1}_N$ then we can write $Q\mathbf{1}_N = \lambda\mathbf{1}_N$, where λ is the associated eigenvalue. It follows that $Q^{-1}\mathbf{1}_N = \frac{1}{\lambda}\mathbf{1}_N$. This proves that, when pairwise correlations are constant across assets, $Q^{-1}\mathbf{1}_N \propto \mathbf{1}_N$. □

5.2 How can clustering be useful in conjunction with the Inverse Volatility rule?

Having proven that theoretically when Sharpe ratios and pairwise correlations are equal across assets, the inverse volatility portfolio is equivalent to the optimal tangency portfolio in a mean-variance framework, we check whether the two assumptions hold empirically. These two assumptions are clearly violated when we look at the data but there are two considerations: (1) many assets have similar Sharpe ratios. For example, Wright et al. (2014) find that the hypothesis of equality for the Sharpe ratios of 18 iShares ETF cannot be rejected at the 1% level. Ardia and Boudt (2015) find that the hypothesis of equality for the Sharpe ratios using hedge fund returns from 2008 to 2012 cannot be rejected at the 1% level for 84.22% of all pairs of two Sharpe Ratios. Van Binsbergen (2020) shows the same Sharpe ratio of 0.22 for both S&P 500 and 10-year government bond from 1970-2021. (2) it is possible, as we show below, to group together assets such that they have similar correlations with each other. On the latter point, in this section, we propose a methodology to build a portfolio by leveraging clustering techniques and the properties of the inverse volatility portfolio.

We can summarize our proposal as follows: (i) first we cluster the data such that their pairwise correlations are close to being equal and build inverse volatility portfolios within each cluster; (ii) in a second stage, we build a portfolio across clusters in which they all exhibit low pairwise correlations across them. Our goal is to cluster together assets that have a high correlation with each other. Having high correlations between assets within a cluster brings the data closer to the feature of equal pairwise correlations, which would make inverse volatility portfolios identical to the tangency portfolio from an optimality standpoint – but with the added benefit of lower estimation error. Furthermore, assets belonging to different clusters will have low correlations which again makes the data closer to the assumption of equal pairwise correlations. Our results highlight that machine learning does not have to be a black-box but we can uncover the reasons that lead it to improve the performance of diversification.[24]

[24]Our proposed clustering methodology differs from de Prado (2016) in that it can be used with various clustering methodologies while de Prado (2016) is bounded to hierarchical clustering.

5.2.1 How can investors build portfolios using clustering?

While we provide the intuition behind our clustering methodology above, we provide a formal description in this section. The machine learning literature provides a great number of clustering techniques. The common goal of the various algorithms is to generate a set of clusters, where each cluster is distinct from each other cluster, while the assets within each cluster are similar to each other. We consider four different methodologies: hierarchical clustering with Single link, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and K-means hierarchical clustering with Ward link.

The clustering methodology can be summarized as follows and it is the same for all algorithms: (1) compute the distance matrix based on the correlation matrix of the assets; (2) choose the optimal number of clusters based on the Silhouette score for the chosen clustering algorithm; (3) cluster the data using the chosen algorithm and the optimal number of clusters from (2).

First, the distance matrix is a matrix that provides information on the similarity of the various assets. We use the correlation matrix to build the distance matrix based on the Euclidean distance. Formally, we denote Q the $N \times N$ correlation matrix and DM the $N \times N$ distance matrix, where N is the number of assets. The element i, j of DM is calculated as

$$DM_{i,j} = \sqrt{\sum_{n=1}^{N} (Q_{n,i} - Q_{n,j})^2} \tag{16}$$

Equation (16) is the Euclidean distance between the correlations of asset i and asset j. We use this measure because elements that have a high correlation between each other and similar correlations with the other assets will have a low distance. One could argue that we could simply use the correlation matrix as a measure of distance. However, this would imply that the distance between two assets i and j uses only the information on the correlation between i and j while the expression in Equation (16) uses also the information about the correlations of assets i and j with the other assets. For ease of notation, we omitted the time

t subscript in Equation (16) but it is worth noting that we re-calculate the distance matrix monthly (i.e., at each portfolio rebalancing date).[25]

Second, it is important to have a methodology that can endogenously determine the number of clusters without requiring user input. This is to ensure that the methodology is truly out-of-sample and it is not the result of data-mining or fine-tuning of the parameters. Each algorithm possesses tuning parameters that when changed lead to a different number of clusters being detected. For example, the DBSCAN algorithm requires us to specify the maximum distance between two assets for one to be considered as in the same cluster as the other, hierarchical clustering requires the definition of a threshold that affects the number of clusters, etc. We choose the tuning parameter of each algorithm by maximizing the Silhouette coefficient. An example will help to clarify the methodology.

Let us assume that we have a set of 5 assets and that, after grouping them using hierarchical clustering with Single linkage, we obtain the dendrogram shown in Figure 3. The y-axis contains the distance measure between clusters while the x-axis contains the assets (5 assets, numbered from 0 to 4). If we choose a value of 0.7 as the threshold to determine the number of clusters, we will have 2 clusters. If we choose a threshold of 0.2, we will have as many clusters as the number of assets. We develop a methodology, which is described below, that does not require any threshold and it uses the full information in the dendrogram in order to cluster the data.

We choose the threshold that maximizes the Silhouette coefficient. The Silhouette coefficient is calculated as $(MNC - MIC)/max\{MIC, MNC\}$ where MIC is the mean intra-cluster distance and MNC is the mean nearest-cluster distance. The Silhouette coefficient varies from -1 to 1 and it has an intuitive interpretation: when it is close to 1, it means that the observations in the current cluster are "far" from its closest cluster; a value of -1 indicates that the observations in the current cluster are closer to a different cluster (i.e.

[25]For the datasets that contain less than 60 portfolios, we use an estimation window of 60 months. For those datasets with more than 60 assets (76 Anomalies, 100 Size and B/M, and 100 Size and Operating Prof), we use a longer estimation window of 120 months.

they are assigned to the wrong cluster). In other words, the Silhouette coefficient measures how well the data are clustered therefore we choose this methodology instead of an arbitrarily chosen threshold. Specifically, at each time t, we compute the dendrogram using the sample-based correlation matrix, and we search for the threshold value that maximizes the Silhouette coefficient. While this example is applied to hierarchical clustering, the same intuition can be used for the other algorithms.

[Insert Figure 3 here]

Once the clusters have been formed, we need a methodology to build a portfolio from the clustered data. We again use the dendrogram in Figure 3 to help us illustrate our methodology. Let us assume that, following the clustering methodology described above, we have two clusters: cluster 1 contains the assets 0, 1, and 4 while cluster 2 contains the assets 2 and 3. Using the chosen allocation rule (i.e. IVol1), we form a portfolio *within each cluster* using the sample-based estimates of the assets' volatilities within each cluster. We call these two portfolios $CP1$ (Cluster Portfolio 1) and $CP2$ (Cluster Portfolio 2). The weights of each asset i belonging to cluster portfolio j are labelled as w_i^{CPj}. We then move to calculating a portfolio *between clusters*. We first calculate the covariance matrix between the two cluster portfolios in our example, $CP1$ and $CP2$, by building a time-series of returns for cluster portfolio CPj using the optimal weights (w_i^{CPj} for each asset i in CPj). This step gives us a time series of returns for each cluster portfolio which we use to calculate the covariance matrix between clusters. We then use our chosen allocation rule and the *between clusters* covariance matrix to form the *between clusters* portfolios. Each cluster portfolio is assigned a weight $\bar{w}^{CPj}$. Last, we calculate the weight of each single asset taking into account both the *within cluster* and *between clusters* portfolio weights. For any asset i that belongs to cluster portfolio j, the optimal weight accounting for both *within cluster* and *between clusters* allocations is

$$w_i = \bar{w}^{CPj} \cdot w_i^{CPj}$$

As we explained above, our methodology clusters the data using only historical information (i.e. the sample-based covariance matrix), therefore it is fully out-of-sample.

5.2.2 An example of how our proposed clustering works

We provide an example of why our clustering methodology works when assets have equal pairwise correlations and Sharpe ratios. Our goal is to compare the weights of the tangency portfolio with the weights implied by our clustering methodology. Let us describe the setup of our example. Let us assume that we have N assets that can be partitioned into two clusters A and B. We can express their expected returns and variance-covariance matrix as follows

$$\mu_a = [\mu_A, \mu_B]' \tag{17}$$

$$\Sigma_a = \begin{bmatrix} \Sigma_A & 0 \\ 0 & \Sigma_B \end{bmatrix} \tag{18}$$

where μ_i is the vector of expected returns of the assets in cluster i, and Σ_i is the variance-covariance matrix of the assets in cluster i, for $i \in \{A, B\}$. The weights of the tangency portfolio are as easily calculated as

$$w_{TAN} \propto \Sigma_a^{-1}\mu_a = \begin{bmatrix} \Sigma_A^{-1} & 0 \\ 0 & \Sigma_B^{-1} \end{bmatrix} \cdot \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix} = \begin{bmatrix} \Sigma_A^{-1}\mu_A \\ \Sigma_B^{-1}\mu_B \end{bmatrix} \tag{19}$$

where the first equality follows from the property of the partitioned matrices. Under the assumption that Sharpe ratios are the same and equal to k, and assets have equal pairwise correlations, we can write for each $i \in \{A, B\}$

$$\mu_i = kV_i\mathbf{1}_i \tag{20}$$

$$\Sigma_i = V_iC_iV_i \tag{21}$$

where V_i is a diagonal matrix with the volatilities of assets in cluster i on its diagonal, $\mathbf{1}_i$ is a $N \times 1$ vector of ones, C_i is the correlation matrix of the assets in cluster i. Substituting Equation (20) and Equation (21) in Equation (19), we can rewrite the weights of the tangency portfolio as

$$w_{TAN} \propto k \begin{bmatrix} V_A^{-1} C_A^{-1} \mathbf{1}_A \\ V_B^{-1} C_B^{-1} \mathbf{1}_B \end{bmatrix} \tag{22}$$

The next step is to derive an expression for the weights of the assets following our clustering methodology and then compare them to the weights described in Equation (22). According to our clustering methodology, the first step is to build "within clusters" portfolios, which we label w_{IVolA} and w_{IVolB} in our example with two clusters

$$w_{IVolA} \propto V_A^{-1} \mathbf{1}_A \implies w_{IVolA} = f_A V_A^{-1} \mathbf{1}_A \tag{23}$$

$$w_{IVolB} \propto V_B^{-1} \mathbf{1}_B \implies w_{IVolB} = f_B V_B^{-1} \mathbf{1}_B \tag{24}$$

where f_A and f_B are two constants that simply regularize the weights of the portfolios such that they sum up to one.

We now need to build the "between clusters" portfolio and, to do that, we need to calculate the variance-covariance matrix between the two portfolios $IVolA$ and $IVolB$, which can be expressed as

$$\begin{aligned} \Sigma_{AB} &= \begin{bmatrix} w'_{IVolA} & 0 \\ 0 & w'_{IVolB} \end{bmatrix} \cdot \Sigma_a \cdot \begin{bmatrix} w_{IVolA} & 0 \\ 0 & w_{IVolB} \end{bmatrix} \\ &= \begin{bmatrix} w'_{IVolA} & 0 \\ 0 & w'_{IVolB} \end{bmatrix} \cdot \begin{bmatrix} V_A C_A V_A & 0 \\ 0 & V_B C_B V_B \end{bmatrix} \cdot \begin{bmatrix} w_{IVolA} & 0 \\ 0 & w_{IVolB} \end{bmatrix} \\ &= \begin{bmatrix} f_A^2 \mathbf{1}'_A C_A \mathbf{1}_A & 0 \\ 0 & f_B^2 \mathbf{1}'_B C_B \mathbf{1}_B \end{bmatrix} \end{aligned} \tag{25}$$

The weights of the "between clusters" portfolios are

$$w_{AB} = \begin{bmatrix} \frac{1}{f_A^2 \mathbf{1}'_A C_A \mathbf{1}_A} \\ \frac{1}{f_B^2 \mathbf{1}'_B C_B \mathbf{1}_B} \end{bmatrix} \tag{26}$$

In our example, w_{AB} is a 2×1 vector but in general it will be a vector equal to the number of clusters. Combining the weights of the "between" and "within" clusters portfolios yields the weights of each asset which, in our example, can be written as

$$w_{cluster} = \begin{bmatrix} \frac{1}{f_A^2 \mathbf{1}'_A C_A \mathbf{1}_A} \otimes w_{IVolA} \\ \frac{1}{f_B^2 \mathbf{1}'_B C_B \mathbf{1}_B} \otimes w_{IVolB} \end{bmatrix} \tag{27}$$

where $w_{cluster}$ is a vector containing the weights of the N assets, which is formed by stacking together the weights of the assets in clusters A and B. The intuition for the expression describing $w_{cluster}$ is simple. We multiply the weights of each "within" cluster portfolio (i.e., w_{IVolA} and w_{IVolB}) by the weights assigned to them by the "between" cluster portfolio.

Since we have closed-form solutions for both the tangency portfolio and our clustering methodology, we can now compare their weights and see when they deviate from each other. To visualize the error, we use the following example. We use 5 assets, 3 of which are in the first cluster (e.g., cluster A), and two are in the second cluster (e.g., cluster B). We assume their Sharpe ratios are equal to 0.5 and their volatilities are defined by the following vector $[0.1, 0.12, 0.15, 0.05, 0.07]$, where the first 3 volatilities belong to the assets in cluster A and the remaining 2 belong to the assets in cluster B. In the base case scenario, we assume that the correlation between clusters (ρ_{btw}) is zero (i.e., assets that belong to different clusters are uncorrelated) and the correlation within clusters is 0.9 (ρ_{win}). Using these assumptions and the closed-form solutions from Equation (22) and Equation (27), we can calculate the weights of each asset using the tangency portfolio and our clustering methodology. To evaluate how much the two methodologies deviate from each other, we calculate the sum of squared deviations between the two weights: $SSD = |w_{TAN} - w_{cluster}|_2$, where $|\cdot|_2$ indicates the norm-2 operator.

In Figure 4, we analyze how changing the correlation between clusters ρ_{btw} and within clusters ρ_{win} affects the sum of squared deviations SSD. In Panel A, we fix all parameters to the base case scenario and we vary the correlation between clusters ρ_{btw}. As expected, when ρ_{btw} is close to zero, the deviations between our clustering methodology and the tangency portfolio are close to zero as well. In other words, our clustering methodology is equivalent to the tangency portfolio. However, as ρ_{btw} increases, then the sum of squared deviations SSD increases, showing that our clustering methodology deviates from the tangency portfolio. In Panel B, we repeat the exercise but this time we vary the correlation within clusters (ρ_{win}). Again, our results confirm our intuition. When ρ_{win} is close to one, the deviations between our clustering methodology and the tangency portfolio are close to zero, thus confirming that when the *within clusters* correlation is high then our methodology converges to the tangency portfolio. As ρ_{win} decreases, the sum of squared deviations SSD increases, showing that our clustering methodology is not effective because it deviates from the tangency portfolio. Overall, our results show that if we are able to cluster together assets that are highly correlated and separate those that have low correlations, then we can use the inverse volatility rule – which is subject to much less error than mean-variance – and still achieve an optimal portfolio from a theoretical point of view.

[Insert Figure 4 here]

5.2.3 When is clustering expected to help?

Before presenting the empirical results on the performance of the various allocation rules using machine learning, we discuss when investors should expect clustering to work best. Intuitively, clustering the data should work well when there is a large number of clusters, and each cluster contains assets that are similar to each other and can be grouped together. We should not expect clustering to work well when the total number of assets is small or when the assets are either all different between each other or all the same. For example, if all M assets are uncorrelated, then there are as many clusters as the number of assets, and diversifying based on assets or clusters would be the same. Also, if all the M assets have a high correlation (e.g. pairwise correlation of 0.99), then they all belong to the same (unique) cluster and we, again, should not see benefits from clustering. It follows that clustering the data should bring the most benefit when assets have dispersed correlations.

In order to evaluate this dispersion in the correlations of the assets, we build a proxy as follows. For each rebalancing date t, the correlation matrix C_t is calculated using the excess returns from $t-m$ to $t-1$, where m is the estimation window. We define SD_t as the standard deviation of the off-diagonal elements of C_t. We then average across all rebalancing dates to obtain the average correlation dispersion:

$$\text{Avg Correlation Dispersion} = \frac{1}{T}\sum_{t=1}^{T} SD_t \tag{28}$$

Our datasets include portfolios of equities, equities and bonds, and anomalies in the cross-section of equity returns. Intuitively, we should expect that correlations between assets are more dispersed for the portfolios of equities and bonds rather than the portfolios of equities only. This is because, as shown in Fama and French (1993) and many other authors, the risk-factors driving equity returns are different from those driving bonds returns.[26] The average correlation dispersion measure defined in Equation (28) is consistent with this intuition. Table 7 presents the average correlation dispersion for the datasets studied in this paper. Consistent with the intuition that the average correlation dispersion should be higher for datasets containing assets from various asset classes, the datasets with the highest average correlation dispersion are the Multi-Assets datasets which contain a mix of equities, fixed income, and commodities. The set of anomalies (76 Anomalies) also exhibits a high dispersion (0.2831). In the set of 76 anomalies, there is a considerable variation: some anomalies are

[26] For example, Fama and French (1993) argue that equity returns are driven by a market factor, a size factor, and a book-to-market factor while for bond returns there are two factors, the term premium and the default premium.

highly correlated between each other as they capture a similar feature (e.g. different ways of designing a momentum anomaly) while some others are not correlated between each other. This structure contributes to the 76 anomalies being the 4th highest dataset when ranked by the average correlation dispersion. The remaining assets are all equity portfolios and they exhibit low average correlation dispersion. Indeed, the 49 Industry portfolio (49 Industries) is the next dataset with the highest dispersion in correlations and it achieves a dispersion of 0.1752 which is less than half of the dispersion for the dataset of Equities, Fixed Income, and Commodities. The 11 portfolios based on the q-factors of Hou et al. (2015, 2019) are the dataset with the lowest average correlation dispersion (0.0508).

[Insert Table 7 here]

5.3 The benefits of clustering: empirical evidence

In this section, we present our empirical findings on the performance of the inverse volatility rule (IVol1) when data are clustered using the methodology described above. Table 8 reports the ratio between the Sharpe ratio of the IVol1 rule using clustered data and the Sharpe ratio of the same rule using unclustered data. We sort the datasets in Table 8 based on their average correlation dispersion in descending order.

Table 8 shows that clustering – especially hierarchical clustering – works well for the set of 76 Anomalies. Using the hierarchical clustering with Single linkage, the Sharpe ratio of IVol1 using clustered data is 11% higher than when IVol1 is applied to the unclustered data. The set of 76 Anomalies is particularly suited to clustering because it possesses a large number of assets and, at the same time, a high average dispersion in correlations. To understand the importance of having a reasonably large number of assets, we look at the two datasets of “Eq., Fixed Income, Comm.” and “Eq., F.I., Comm. and 10 S&P”. Both datasets are diversified across equities, fixed income and commodities. The only difference is that in the “Eq., F.I., Comm. and 10 S&P”, we add 10 S&P industries from Bloomberg to increase the total number of assets. The results on the benefits of clustering between these two datasets are in stark contrast between each other. For "Eq., F.I., Comm. and 10 S&P", the performance is considerably improved while for "Eq., Fixed Income, Comm.", clustering leads to a decrease in performance. Both datasets exhibit large average dispersion in correlation and the main difference between the two lies in the number of assets: "Eq., Fixed Income, Comm." has 8 assets in total (2 equities, 2 fixed income, and 4 commodities indices) while "Eq., F.I., Comm. and 10 S&P" contains 18 assets (12 equities, 2 fixed income, and 4 commodities indices).

[Insert Table 8 here]

Consistent with our description of Avg Correlation Dispersion defined in Equation (28), Table 8 shows that, as the Avg Correlation Dispersion decreases, the benefits of clustering are mostly marginal or disappear. Overall, our findings show that clustering improves the performance of risk-based allocation rules in terms of the Sharpe ratio when there is enough variation in the correlation between the assets (i.e. when the Avg Correlation Dispersion is high).

6 Conclusion

In this paper, we first develop a statistical test to evaluate whether applying inverse volatility rules is equivalent to using a tangency portfolio (i.e., optimal mean-variance portfolio) in the presence of estimation error for both expected returns and variance-covariance matrix. Specifically, we develop a statistical test for the null hypothesis that the inverse volatility portfolios are equivalent to the tangency portfolio and then apply it to many datasets. Our results show that in the majority of the cases, there is no statistical difference between the optimal mean-variance portfolio and the inverse volatility portfolios, thus providing an explanation for why inverse volatility rules perform well empirically.

Furthermore, the statistical test developed here can be used to guide our decision of whether or not to use inverse volatility rules on a given dataset: if historical returns show that inverse volatility rules are equivalent to the tangency portfolio, they are likely to perform well out-of-sample, assuming that the distribution of returns remains the same. We confirm this intuition empirically and, consistent with previous literature, our results show that inverse volatility rules outperform both the naïve diversification (1/N) rule and mean-variance optimized portfolios.

Finally, we propose a clustering methodology that leverages the properties of the inverse volatility portfolio to improve the benefits of diversification of such portfolio and ultimately achieve a higher portfolio's performance. Our clustering methodology is guided by the fact that the weights of the inverse volatility portfolio are exactly equal to the weights of the tangency portfolio when assets have the same Sharpe ratios and equal pairwise correlations. Our clustering methodology groups the assets such that the assets within each cluster are as close as possible to this condition. We show that in many datasets considered in this study, our clustering methodology improves performance relative to the "unclustered" data.

References

Anderson, Evan W, and Ai-Ru Cheng, 2016, Robust bayesian portfolio choices, *The Review of Financial Studies* 29, 1330–1375.

Ardia, David, and Kris Boudt, 2015, Testing equality of modified sharpe ratios, *Finance Research Letters* 13, 97–104.

Asness, Clifford, and Andrea Frazzini, 2013, The devil in hml's details, *The Journal of Portfolio Management* 39, 49–68.

Asness, Clifford S, Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Barbee Jr, William C, Sandip Mukherji, and Gary A Raines, 1996, Do sales–price and debt–equity explain stock returns better than book–market and firm size?, *Financial Analysts Journal* 52, 56–60.

Barroso, Pedro, and Konark Saxena, 2019, Lest we forget: learn from out-of-sample forecast errors when optimizing portfolios, *Review of Financial Studies, forthcoming* .

Barth, Mary E, John A Elliott, and Mark W Finn, 1999, Market rewards associated with patterns of increasing earnings, *Journal of Accounting Research* 37, 387–413.

Best, Michael J, 2010, *Portfolio optimization* (CRC Press).

Best, Michael J, and Robert R Grauer, 1991, On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results, *The Review of Financial Studies* 4, 315–342.

Black, Fischer, 1993, Estimating expected return, *Financial Analysts Journal* 49, 36–38.

Blitz, David, Joop Huij, and Martin Martens, 2011, Residual momentum, *Journal of Empirical Finance* 18, 506–521.

Bodnar, Taras, and Yarema Okhrin, 2011, On the product of inverse wishart and normal distributions with applications to discriminant analysis and portfolio theory, *Scandinavian Journal of Statistics* 38, 311–331.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *The Journal of Finance* 62, 877–915.

Brandt, Michael W, Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns, *The Review of Financial Studies* 22, 3411–3447.

Broadie, Mark, 1993, Computing efficient frontiers using estimated parameters, *Annals of Operations Research* 45, 21–58.

Chan, Konan, Louis KC Chan, Narasimhan Jegadeesh, and Josef Lakonishok, 2001a, Earnings quality and stock returns, Technical report, National Bureau of Economic Research.

Chan, Louis KC, Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *The Journal of Finance* 51, 1681–1713.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis, 2001b, The stock market valuation of research and development expenditures, *The Journal of Finance* 56, 2431–2456.

Chopra, Vijay K, and William T Ziemba, 2013, The effect of errors in means, variances, and covariances on optimal portfolio choice, in *Handbook of the Fundamentals of Financial Decision Making: Part I*, 365–373 (World Scientific).

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

DeMiguel, Victor, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal, 2009a, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science* 55, 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009b, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The Review of Financial Studies* 22, 1915–1953.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal, 2020, A transaction-cost perspective on the multitude of firm characteristics, *The Review of Financial Studies* 33, 2180–2222.

DeMiguel, Victor, Francisco J Nogales, and Raman Uppal, 2014, Stock return serial dependence and out-of-sample portfolio performance, *The Review of Financial Studies* 27, 1031–1073.

Desai, Hemang, Shivaram Rajgopal, and Mohan Venkatachalam, 2004, Value-glamour and accruals mispricing: One anomaly or two?, *The Accounting Review* 79, 355–385.

Eisfeldt, Andrea L, and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *The Journal of Finance* 68, 1365–1406.

Engle, Robert, 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business & Economic Statistics* 20, 339–350.

Engle, Robert, and Riccardo Colacito, 2006, Testing and valuing dynamic correlations for asset allocation, *Journal of Business & Economic Statistics* 24, 238–253.

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

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 2001, The economic value of volatility timing, *The Journal of Finance* 56, 329–352.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 2003, The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* 67, 473–509.

Foster, George, Chris Olsen, and Terry Shevlin, 1984, Earnings releases, anomalies, and the behavior of security returns, *The Accounting Review* 59, 574–603.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Hand, John RM, and Jeremiah Green, 2011, The importance of accounting information in portfolio optimization, *Journal of Accounting, Auditing & Finance* 26, 1–34.

Heston, Steven L, and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* 38, 297–331.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2019, Which factors?, *Review of Finance* 23, 1–35.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2018, Replicating anomalies, *The Review of Financial Studies* 33, 2019–2133.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *The Journal of Finance* 58, 1651–1683.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48, 65–91.

Jobson, J Dave, and Bob M Korkie, 1981, Performance hypothesis testing with the sharpe and treynor measures, *The Journal of Finance* 36, 889–908.

Johnson, Norman L, Samuel Kotz, and Narayanaswamy Balakrishnan, 1995, *Continuous Univariate Distributions, volume 2*, volume 289 (John wiley & sons).

Jorion, Philippe, 1986, Bayes-stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.

Kan, Raymond, Xiaolu Wang, and Guofu Zhou, 2021, Optimal portfolio choice with estimation risk: No risk-free asset case, *Management Science, forthcoming* .

Kan, Raymond, and Guofu Zhou, 2007, Optimal portfolio choice with parameter uncertainty, *Journal of Financial and Quantitative Analysis* 42, 621–656.

Kirby, Chris, and Barbara Ostdiek, 2012, It's all in the timing: Simple active portfolio strategies that outperform naïve diversification, *Journal of Financial and Quantitative Analysis* 47, 437–467.

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, 603–621.

Ledoit, Olivier, and Michael Wolf, 2004a, Honey, I shrunk the sample covariance matrix, *The Journal of Portfolio Management* 30, 110–119.

Ledoit, Olivier, and Michael Wolf, 2004b, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* 88, 365 – 411.

Ledoit, Olivier, and Michael Wolf, 2017, Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks, *The Review of Financial Studies* 30, 4349–4388.

Loughran, Tim, and Jay W Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Analysis* 46, 1629–1650.

MacKinlay, Craig A, and L'uboš Pástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *The Review of Financial Studies* 13, 883–916.

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

Memmel, Christoph, 2003, Performance hypothesis testing with the sharpe ratio, *Finance Letters* 1, 21–23.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323 – 361.

Michaud, Richard O, 1989, The markowitz optimization enigma: Is "optimized" optimal?, *Financial Analysts Journal* 45, 31–42.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *The Journal of Finance* 72, 1611–1644.

Neyman, Jerzy, 1961, Proceedings of the fourth berkeley symposium on mathematical statistics and probability.

Okhrin, Yarema, and Wolfgang Schmid, 2006, Distributional properties of portfolio weights, *Journal of Econometrics* 134, 235–256.

Pástor, L'uboš, 2000, Portfolio selection and asset pricing models, *The Journal of Finance* 55, 179–223.

Pástor, L'uboš, and Robert F Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

Pedersen, Lasse Heje, Abhilash Babu, and Ari Levine, 2021, Enhanced portfolio optimization, *Financial Analysts Journal* 77, 124–151.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *The Journal of Finance* 63, 921–945.

Qian, Edward, 2011, Risk parity and diversification, *The Journal of Investing* 20, 119–127.

Simaan, Yusif, 1997, Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model, *Management Science* 43, 1437–1446.

Sloan, Richard G, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *The Accounting Review* 71, 289–315.

Tu, Jun, and Guofu Zhou, 2010, Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty, *Journal of Financial and Quantitative Analysis* 45, 959–986.

Tu, Jun, and Guofu Zhou, 2011, Markowitz meets talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics* 99, 204–215.

Tuzel, Selale, 2010, Corporate real estate holdings and the cross-section of stock returns, *The Review of Financial Studies* 23, 2268–2302.

Van Binsbergen, Jules H, 2020, Duration-based stock valuation: Reassessing stock market performance and volatility, Technical report, National Bureau of Economic Research.

Wright, John, SC Yam, and Siu Pang Yung, 2014, A test for the equality of multiple sharpe ratios, *Journal of Risk* 16, 3–21.

Figures

Figure 1. The effect of estimation error. This figure depicts the performance of three different allocation rules under the effect of estimation error. The three allocation rules plotted are: the optimal mean-variance tangency portfolio (TAN), the global minimum variance portfolio (GMV), and the inverse volatility portfolio (IVol). The y-axis displays the average portfolio excess return μ_P while the x-axis shows its standard deviation σ_P. In each panel, the green solid line shows the efficient frontier using the true mean excess return μ and variance-covariance matrix Σ. The blue triangles show the performance in terms of expected excess returns (μ_P) and volatility (σ_P) of 5,000 simulated tangency (TAN) portfolios when investors estimate the mean excess return and variance covariance matrix, $\widehat{\mu}$ and $\widehat{\Sigma}$, using an estimation window of 60 months. The yellow diamonds show the performance for the global minimum variance (GMV) portfolios, and the red squares show the performance for the inverse volatility (IVol) portfolios. In each panel, the legend reports the name of the series and the median Sharpe ratio from the 5,000 simulations. Panel A and Panel B show the simulations when there are 5 and 15 uncorrelated assets, respectively. In all panels, assets are assumed to have the same Sharpe ratios.

Panel A | Panel B

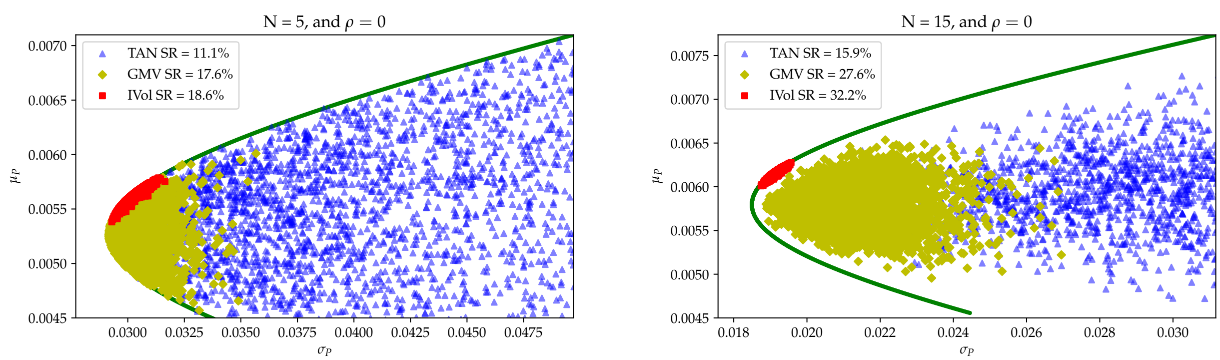
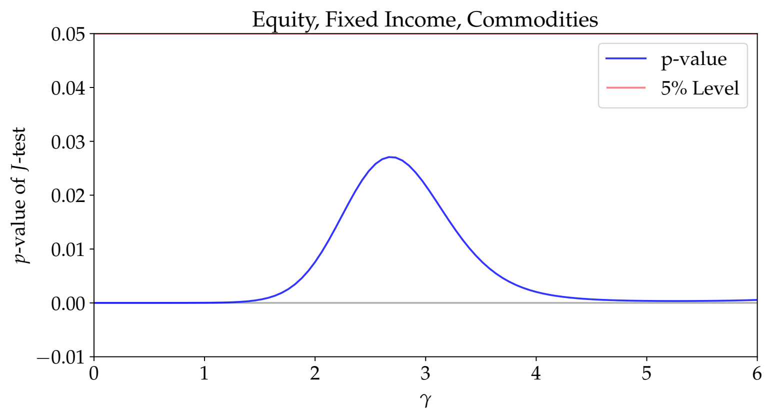
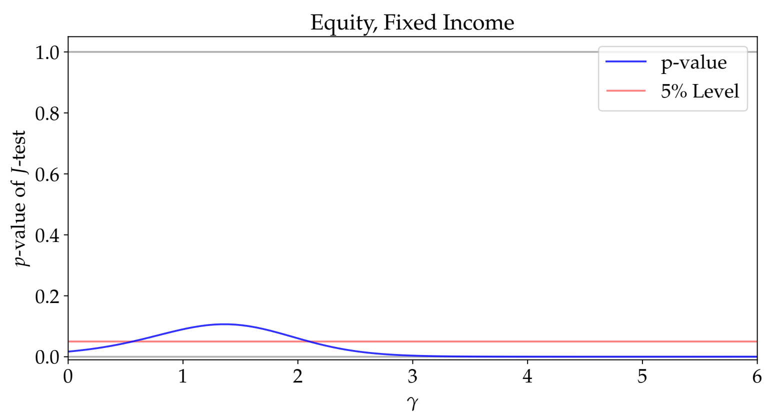


Figure 2. Visualize the J test. This figure shows the p-value from the J-test described in Section 3.1, which tests the null hypothesis that the portfolio IVolγ is equivalent to the mean-variance portfolio for a given γ. The IVolγ rule is described in Section 3.3.

Panel A



Panel B



Panel C

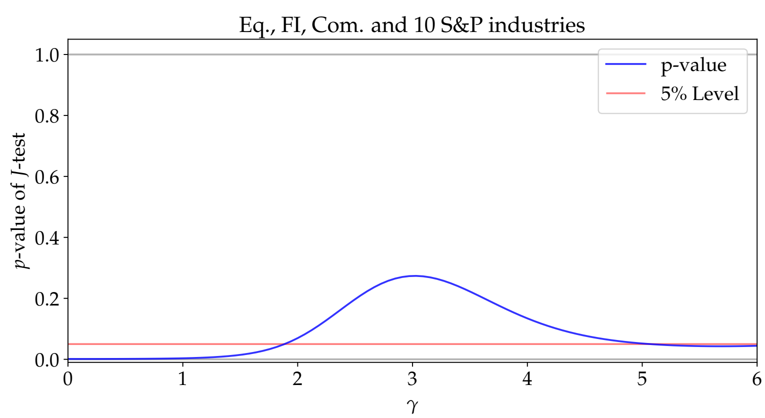


Figure 3. A dendrogram. This figure shows how hierarchical clustering groups together various assets to form clusters.

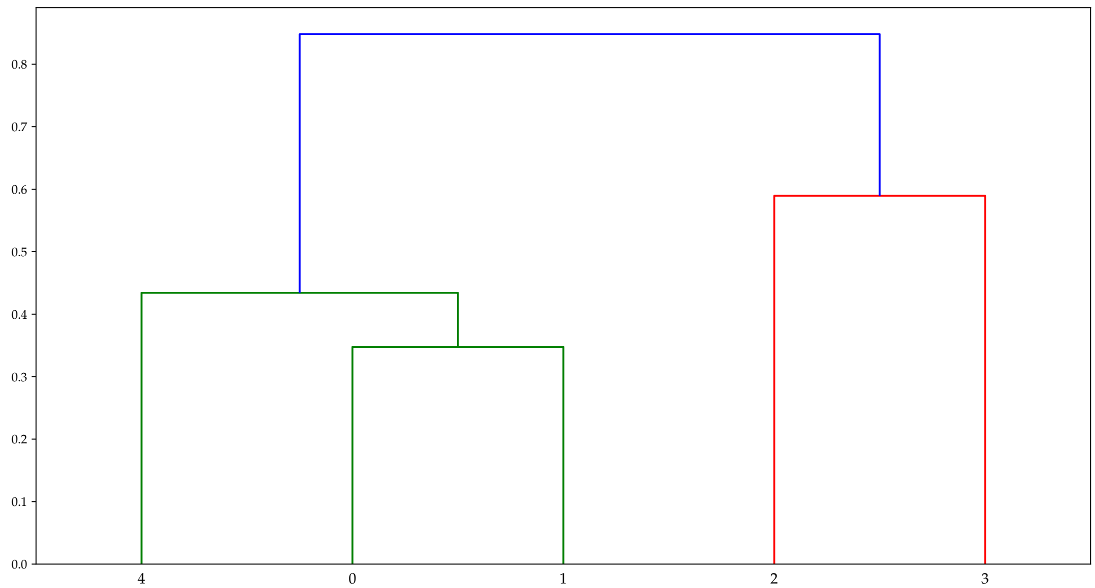
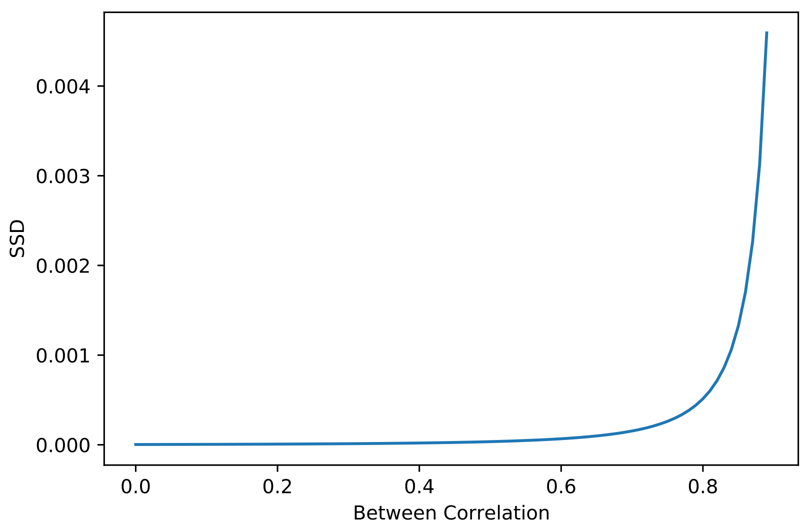
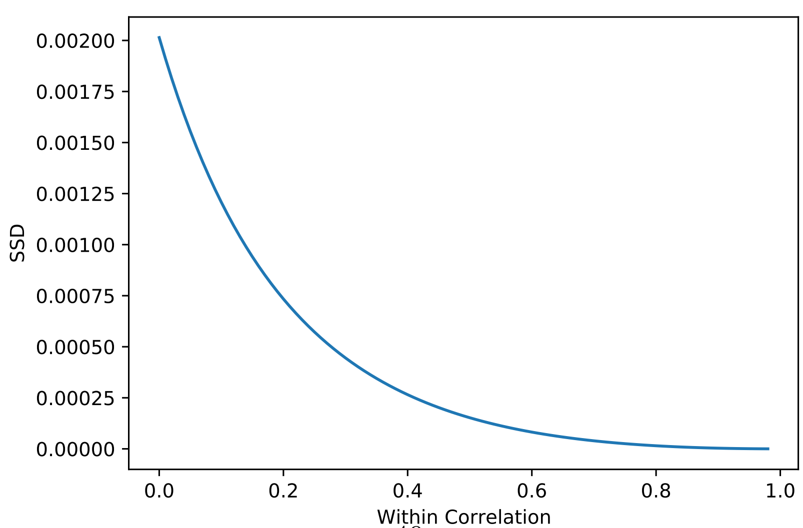


Figure 4. Clustering. The two figures below show how much our clustering methodology differs from the optimal tangency portfolio for various levels of correlations. We assume that there are 5 assets, 3 of which are in the first cluster (e.g., cluster A) and two are in the second cluster (e.g., cluster B). We assume their Sharpe ratios are equal to 0.5 and their volatilities are defined by the following vector $[0.1, 0.12, 0.15, 0.05, 0.07]$, where the first 3 volatilities belong to the assets in cluster A and the remaining 2 belong to the assets in cluster B. In the base case scenario, we assume that the correlation between clusters (ρ_{btw}) is zero (i.e., assets that belong to different clusters are uncorrelated) and the correlation within clusters is 0.9 (ρ_{win}). Using these assumptions and the closed-form solutions from Equation (22) and Equation (27), we calculate the weights of each asset using the tangency portfolio and our clustering methodology and the sum of squared deviations as $SSD = |w_{TAN} - w_{cluster}|_2$, where $|\cdot|_2$ indicates the norm-2 operator. In Panel A, we fix all parameters to the base case scenario and we vary the correlation between clusters ρ_{btw}. In Panel B, we repeat the exercise but this time we vary the correlation within clusters (ρ_{win}).

Panel A



Panel C



Tables

Table 1. The datasets are listed below. For a detailed description and discussion of the datasets, please see Section 3.2.

Datasets	**Dates**
Equity, Fixed Income, Commodities	1979 - 2019
Equity, Fixed Income, Commodities and 10 S&P Industries	1994 - 2019
Equity and Fixed Income	1979 - 2019
76 Anomalies	1981 - 2019
Individual Stocks	1967 - 2019
10 Industries	1960 - 2019
17 Industries	1960 - 2019
30 Industries	1960 - 2019
49 Industries	1960 - 2019
25 Size and B/M	1960 - 2019
25 Size and Operating Profitability	1963 - 2019
11 q-Portfolios	1971 - 2019
12 FF5-Portfolios	1971 - 2019
25 Europe Size and Book-to-market	1990 - 2019
25 Japan Size and Book-to-market	1990 - 2019
25 Asia Size and Book-to-market	1990 - 2019

Table 2. J-test and in-sample Sharpe ratios. For a given level of γ, this table shows the in-sample Sharpe ratios of the IVolγ portfolio and the p-value from the J-test of the null hypothesis that the portfolio IVolγ is equivalent to the mean-variance portfolio. The portfolio rule IVolγ and the J-test are described in Section 3.

			Values of γ						
			1/N ($\gamma = 0$)	**1**	**2**	**3**	**4**	**5**	**6**
Multi-Asset	Eq., Fixed Income, Comm.	J p-val	0.000	0.000	0.008	0.022	0.002	0.000	0.001
		SR	0.232	0.277	0.318	0.329	0.314	0.295	0.281
	Eq., F.I., Comm. and 10 S&P	J p-val	0.001	0.003	0.070	0.274	0.134	0.052	0.044
		SR	0.248	0.281	0.329	0.376	0.374	0.329	0.289
	Equity, Fixed Income	J p-val	0.017	0.090	0.060	0.004	0.000	0.000	0.000
		SR	0.319	0.334	0.332	0.317	0.299	0.285	0.276
	76 Anomalies	J p-val	0.000	0.000	0.000	0.019	0.325	0.834	0.986
		SR	0.350	0.365	0.367	0.361	0.352	0.343	0.335
Equity Only	Individual Stocks	J p-val	0.475	1.000	0.989	0.986	1.000	0.742	0.992
		SR	0.187	0.192	0.196	0.200	0.202	0.202	0.200
	10 Industries	J p-val	0.045	0.075	0.136	0.237	0.371	0.512	0.632
		SR	0.232	0.237	0.242	0.246	0.250	0.253	0.256
	17 Industries	J p-val	0.018	0.021	0.039	0.090	0.205	0.384	0.579
		SR	0.211	0.218	0.225	0.232	0.239	0.245	0.251
	30 Industries	J p-val	0.087	0.098	0.162	0.313	0.546	0.770	0.905
		SR	0.211	0.217	0.223	0.229	0.235	0.241	0.247
	49 Industries	J p-val	0.005	0.004	0.013	0.081	0.348	0.734	0.940
		SR	0.205	0.212	0.218	0.225	0.232	0.239	0.246
	100 Size and B/M	J p-val	0.000	0.000	0.001	0.289	0.962	1.000	1.000
		SR	0.215	0.221	0.225	0.228	0.231	0.233	0.235
	100 Size and Operating Prof	J p-val	0.000	0.000	0.007	0.569	0.994	1.000	1.000
		SR	0.216	0.220	0.223	0.225	0.227	0.229	0.230
	25 Size and B/M	J p-val	0.000	0.000	0.000	0.000	0.010	0.136	0.463
		SR	0.216	0.221	0.225	0.229	0.231	0.233	0.234
	25 Size and Operating Prof	J p-val	0.000	0.000	0.000	0.004	0.067	0.315	0.652
		SR	0.213	0.217	0.220	0.222	0.224	0.226	0.227
	11 q-Portfolios	J p-val	0.000	0.000	0.000	0.000	0.000	0.005	1.000
		SR	0.167	0.172	0.178	0.183	0.188	0.192	0.195
	12 FF5-Portfolios	J p-val	0.000	0.000	0.000	0.000	0.001	0.015	0.076
		SR	0.207	0.209	0.211	0.213	0.214	0.216	0.217
	25 Europe Size and B/M	J p-val	0.000	0.000	0.008	0.146	0.578	0.894	0.982
		SR	0.147	0.148	0.149	0.150	0.151	0.152	0.153
	25 Japan Size and B/M	J p-val	0.080	0.099	0.240	0.544	0.832	0.960	0.993
		SR	0.063	0.063	0.063	0.064	0.064	0.064	0.065
	25 Asia Size and B/M	J p-val	0.000	0.000	0.005	0.161	0.630	0.915	0.986
		SR	0.132	0.133	0.134	0.135	0.136	0.137	0.138

Table 3. Comparison of Sharpe ratios and test against 1/N. This table presents the monthly Sharpe ratios of various allocation rules across different datasets. The column "Dataset" contains the datasets, which are described in Table 1. The various allocation rules are: (1) the "1/N" rule (naïve diversification); (2) the tangency portfolio from Mean-Variance Optimization (MVO); (3) the 1-norm-constrained portfolio (DGNU1) developed in DeMiguel et al. (2009a); and (4) the IVolγ rule for $\gamma \in \{1, 2, 3, 4, 5\}$. In parentheses, we report the p-value of the difference between the Sharpe ratio of each allocation rule from the 1/N rule, which is computed using the test described in Internet Appendix B.

Dataset	**1/N**	**MVO**	**DGNU1**	**IVol1**	**IVol2**	**IVol3**	**IVol4**	**IVol5**
Eq., Fixed Income	0.235	0.342	0.343	0.285	0.330	0.352	0.350	0.337
Comm.		(0.030)	(0.003)	(0.000)	(0.000)	(0.002)	(0.017)	(0.066)
Eq., Fixed Income,	0.189	0.205	0.296	0.222	0.258	0.294	0.317	0.320
Comm. and 10 S&P		(0.826)	(0.076)	(0.001)	(0.002)	(0.005)	(0.017)	(0.051)
Eq., Fixed Income	0.321	0.257	0.350	0.339	0.341	0.331	0.318	0.307
		(0.053)	(0.243)	(0.138)	(0.414)	(0.768)	(0.946)	(0.761)
76 Anomalies	0.296	0.538	0.566	0.310	0.313	0.307	0.298	0.286
		(0.000)	(0.000)	(0.090)	(0.258)	(0.567)	(0.958)	(0.692)
Individual Stocks	0.203	0.037	0.180	0.206	0.209	0.211	0.214	0.214
		(0.003)	(0.552)	(0.553)	(0.502)	(0.491)	(0.520)	(0.582)
10 Industries	0.229	0.088	0.271	0.237	0.244	0.250	0.254	0.256
		(0.003)	(0.108)	(0.000)	(0.001)	(0.002)	(0.004)	(0.010)
17 Industries	0.209	0.042	0.256	0.218	0.227	0.234	0.240	0.245
		(0.005)	(0.111)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
30 Industries	0.211	0.073	0.266	0.220	0.229	0.236	0.242	0.248
		(0.001)	(0.106)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
49 Industries	0.203	0.108	0.243	0.212	0.220	0.228	0.234	0.240
		(0.089)	(0.252)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
25 Size and B/M	0.217	-0.002	0.300	0.224	0.230	0.234	0.237	0.239
		(0.000)	(0.004)	(0.000)	(0.000)	(0.001)	(0.001)	(0.003)
25 Size and Operating Prof	0.199	0.057	0.308	0.205	0.211	0.215	0.219	0.223
		(0.025)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
11 q-Portfolios	0.163	0.293	0.306	0.168	0.174	0.178	0.182	0.185
		(0.182)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
12 FF5-Portfolios	0.213	0.157	0.291	0.216	0.220	0.223	0.225	0.228
		(0.431)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
25 Europe Size and B/M	0.157	-0.046	0.281	0.165	0.172	0.179	0.185	0.190
		(0.005)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
25 Japan Size and B/M	0.085	0.055	0.119	0.087	0.089	0.091	0.093	0.094
		(0.721)	(0.409)	(0.111)	(0.118)	(0.128)	(0.141)	(0.157)
25 Asia Size and B/M	0.118	0.019	0.196	0.120	0.121	0.123	0.125	0.127
		(0.238)	(0.019)	(0.175)	(0.141)	(0.113)	(0.090)	(0.073)

Table 4. Comparison of Sharpe ratios and test against DGNU1. This table presents the monthly Sharpe ratios of various allocation rules across different datasets. The column "Dataset" contains the datasets, which are described in Table 1. The various allocation rules are: (1) the "1/N" rule (naïve diversification); (2) the tangency portfolio from Mean-Variance Optimization (MVO); (3) the 1-norm-constrained portfolio (DGNU1) developed in DeMiguel et al. (2009a); and (4) the IVolγ rule for $\gamma \in \{1, 2, 3, 4, 5\}$. In parentheses, we report the p-value of the difference between the Sharpe ratio of each allocation rule from the DGNU1 portfolio, which is computed using the test described in Internet Appendix B.

Dataset	**DGNU1**	**1/N**	**MVO**	**IVol1**	**IVol2**	**IVol3**	**IVol4**	**IVol5**
Eq., Fixed Income	0.343	0.235	0.342	0.285	0.330	0.352	0.350	0.337
Comm.		(0.003)	(0.977)	(0.040)	(0.488)	(0.609)	(0.774)	(0.846)
Eq., Fixed Income,	0.296	0.189	0.205	0.222	0.258	0.294	0.317	0.320
Comm. and 10 S&P		(0.076)	(0.140)	(0.171)	(0.423)	(0.951)	(0.589)	(0.573)
Eq., Fixed Income	0.350	0.321	0.257	0.339	0.341	0.331	0.318	0.307
		(0.243)	(0.004)	(0.465)	(0.362)	(0.200)	(0.126)	(0.092)
76 Anomalies	0.566	0.296	0.538	0.310	0.313	0.307	0.298	0.286
		(0.000)	(0.542)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Individual Stocks	0.180	0.203	0.037	0.206	0.209	0.211	0.214	0.214
		(0.552)	(0.007)	(0.494)	(0.423)	(0.351)	(0.295)	(0.265)
10 Industries	0.271	0.229	0.088	0.237	0.244	0.250	0.254	0.256
		(0.108)	(0.000)	(0.174)	(0.253)	(0.334)	(0.408)	(0.461)
17 Industries	0.256	0.209	0.042	0.218	0.227	0.234	0.240	0.245
		(0.111)	(0.000)	(0.185)	(0.283)	(0.396)	(0.515)	(0.626)
30 Industries	0.266	0.211	0.073	0.220	0.229	0.236	0.242	0.248
		(0.106)	(0.000)	(0.165)	(0.239)	(0.327)	(0.425)	(0.526)
49 Industries	0.243	0.203	0.108	0.212	0.220	0.228	0.234	0.240
		(0.252)	(0.007)	(0.363)	(0.486)	(0.622)	(0.767)	(0.914)
25 Size and B/M	0.300	0.217	-0.002	0.224	0.230	0.234	0.237	0.239
		(0.004)	(0.000)	(0.007)	(0.010)	(0.012)	(0.013)	(0.013)
25 Size and Operating Prof	0.308	0.199	0.057	0.205	0.211	0.215	0.219	0.223
		(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
11 q-Portfolios	0.306	0.163	0.293	0.168	0.174	0.178	0.182	0.185
		(0.000)	(0.889)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
12 FF5-Portfolios	0.291	0.213	0.157	0.216	0.220	0.223	0.225	0.228
		(0.002)	(0.052)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
25 Europe Size and B/M	0.281	0.157	-0.046	0.165	0.172	0.179	0.185	0.190
		(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
25 Japan Size and B/M	0.119	0.085	0.055	0.087	0.089	0.091	0.093	0.094
		(0.409)	(0.395)	(0.431)	(0.451)	(0.470)	(0.486)	(0.499)
25 Asia Size and B/M	0.196	0.118	0.019	0.120	0.121	0.123	0.125	0.127
		(0.019)	(0.041)	(0.018)	(0.018)	(0.018)	(0.019)	(0.020)

Table 5. Comparison of Certainty Equivalent Returns. This table presents the monthly Certainty Equivalent Returns (CEQ) of various allocation rules across different datasets. The column "Dataset" contains the datasets, which are described in Table 1. The various allocation rules are: (1) the "1/N" rule (naïve diversification); (2) the tangency portfolio from Mean-Variance Optimization (MVO); (3) the 1-norm-constrained portfolio (DGNU1) developed in DeMiguel et al. (2009a); and (4) the IVolγ rule for $\gamma \in \{1, 2, 3, 4, 5\}$. For ease of reading, we multiply the CEQs by 100. In parentheses, we report the p-value of the difference between the CEQ of each allocation rule from the 1/N rule, which is computed using the test described in DeMiguel, Garlappi, and Uppal (2009b).

Dataset	1/N	MVO	DGNU1	IVol1	IVol2	IVol3	IVol4
Eq., Fixed Income	0.433	-1.303	0.536	0.516	0.555	0.555	0.536
Comm.		(0.002)	(0.203)	(0.030)	(0.101)	(0.239)	(0.400)
Eq., Fixed Income	0.344	-3.027	0.442	0.439	0.509	0.534	0.515
Comm. and 10 S&P		(0.000)	(0.542)	(0.031)	(0.077)	(0.181)	(0.350)
Eq., Fixed Income,	0.667	0.472	0.602	0.664	0.632	0.593	0.562
		(0.053)	(0.273)	(0.935)	(0.596)	(0.391)	(0.287)
76 Anomalies	0.358	0.420	0.344	0.338	0.316	0.294	0.275
		(0.545)	(0.835)	(0.155)	(0.062)	(0.027)	(0.016)
Individual Stocks	0.406	-663.637	0.324	0.426	0.440	0.447	0.445
		(0.000)	(0.624)	(0.460)	(0.510)	(0.587)	(0.687)
10 Industries	0.519	-5.758	0.637	0.553	0.578	0.595	0.605
		(0.000)	(0.248)	(0.006)	(0.011)	(0.024)	(0.047)
17 Industries	0.432	-1466.417	0.591	0.482	0.522	0.552	0.573
		(0.000)	(0.202)	(0.000)	(0.001)	(0.002)	(0.005)
30 Industries	0.441	-841.488	0.623	0.490	0.530	0.561	0.585
		(0.000)	(0.200)	(0.001)	(0.002)	(0.003)	(0.005)
49 Industries	0.396	-3.421	0.555	0.449	0.491	0.525	0.552
		(0.000)	(0.314)	(0.000)	(0.000)	(0.001)	(0.002)
25 Size and B/M	0.457	-12.619	0.758	0.499	0.529	0.549	0.562
		(0.000)	(0.028)	(0.000)	(0.001)	(0.003)	(0.007)
25 Size and Operating Prof	0.360	-34.509	0.771	0.401	0.435	0.462	0.483
		(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
11 q-Portfolios	0.048	-204.184	0.876	0.101	0.148	0.188	0.220
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
12 FF5-Portfolios	0.448	-81.325	0.743	0.467	0.484	0.498	0.511
		(0.000)	(0.007)	(0.003)	(0.003)	(0.003)	(0.003)
25 Europe Size and B/M	0.166	-260.064	0.728	0.212	0.253	0.289	0.321
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
25 Japan Size and B/M	-0.259	-422.250	0.023	-0.229	-0.204	-0.183	-0.165
		(0.000)	(0.192)	(0.006)	(0.007)	(0.009)	(0.012)
25 Asia Size and B/M	-0.241	-1135.74	0.330	-0.218	-0.195	-0.173	-0.152
		(0.000)	(0.006)	(0.010)	(0.008)	(0.006)	(0.004)

Table 6. Comparison of Turnovers. This table presents the ratio of the monthly turnover of each allocation rule with respect to the 1/N rule (Panel A) as well as the return-gain with respect to the 1/N strategy in terms of the Sharpe ratio (Panel B), which is the extra return the 1/N rule needs to provide in order that the Sharpe ratio of the 1/N strategy equals that of each allocation rule in the presence of transactions costs as described in Internet Appendix E. The column "Datasets" contains the datasets, which are described in Table 1. The various allocation rules are: (1) the "1/N" rule (naïve diversification); (2) the tangency portfolio from Mean-Variance Optimization (MVO); (3) the 1-norm-constrained portfolio (DGNU1) developed in DeMiguel et al. (2009a); and (4) the IVolγ rule for $\gamma \in \{1, 2, 3, 4, 5\}$.

Dataset	**1/N**	**MVO**	**DGNU1**	**IVol1**	**IVol2**	**IVol3**	**IVol4**	**IVol5**
			Panel A: Turnover ratios					
Eq., Fixed Income Comm.	1.00	2.50	0.82	0.96	0.88	0.77	0.65	0.54
Eq., F.I. , Comm. and 10 S&P	1.00	3.53	1.22	0.96	0.92	0.85	0.77	0.67
Eq., Fixed Income	1.00	1.00	0.81	0.92	0.78	0.63	0.49	0.38
76 Anomalies	1.00	2.42	1.50	0.93	0.86	0.81	0.76	0.71
Individual Stocks	1.00	246.92	2.46	0.95	0.90	0.86	0.82	0.78
10 Industries	1.00	29.11	2.10	0.98	0.96	0.94	0.92	0.90
17 Industries	1.00	220.87	2.98	0.97	0.94	0.91	0.89	0.86
30 Industries	1.00	1304.24	3.87	0.95	0.92	0.89	0.87	0.84
49 Industries	1.00	34.38	4.30	0.95	0.91	0.88	0.86	0.83
25 Size and B/M	1.00	1811.94	6.87	0.98	0.96	0.94	0.92	0.90
25 Size and Operating Prof	1.00	380.05	8.22	0.99	0.97	0.96	0.94	0.93
11 q-Portfolios	1.00	355.49	6.40	1.00	1.00	0.99	0.99	0.98
12 FF5-Portfolios	1.00	1740.97	5.92	0.96	0.91	0.87	0.83	0.79
25 Europe Size and B/M	1.00	8135.87	7.96	0.98	0.96	0.94	0.92	0.91
25 Japan Size and B/M	1.00	2068.50	8.10	0.97	0.95	0.93	0.91	0.89
25 Asia Size and B/M	1.00	32862.23	6.24	0.99	0.98	0.97	0.96	0.95

			Panel B: Annualized Return-Gain					
Eq., Fixed Income Comm.	0.00%	3.28%	3.31%	1.50%	2.89%	3.59%	3.53%	3.12%
Eq., F.I. , Comm. and 10 S&P	0.00%	0.58%	4.01%	1.21%	2.57%	3.91%	4.80%	4.91%
Eq., Fixed Income	0.00%	-1.96%	0.94%	0.58%	0.63%	0.31%	-0.09%	-0.43%
76 Anomalies	0.00%	4.02%	4.51%	0.23%	0.27%	0.18%	0.02%	-0.18%
Individual Stocks	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
10 Industries	0.00%	-4.87%	2.19%	0.43%	0.78%	1.06%	1.28%	1.43%
17 Industries	0.00%	-6.02%	2.66%	0.53%	1.00%	1.40%	1.73%	1.99%
30 Industries	0.00%	-13.18%	3.18%	0.52%	0.99%	1.41%	1.79%	2.11%
49 Industries	0.00%	-3.71%	2.43%	0.54%	1.02%	1.46%	1.86%	2.23%
25 Size and B/M	0.00%	-9.27%	5.32%	0.44%	0.78%	1.04%	1.23%	1.36%
25 Size and Operating Prof	0.00%	-9.22%	7.08%	0.40%	0.75%	1.05%	1.31%	1.52%
11 q-Portfolios	0.00%	20.63%	11.23%	0.42%	0.81%	1.16%	1.45%	1.67%
12 FF5-Portfolios	0.00%	-8.93%	4.47%	0.20%	0.39%	0.56%	0.71%	0.85%
25 Europe Size and B/M	0.00%	-7.58%	7.58%	0.46%	0.89%	1.28%	1.63%	1.94%
25 Japan Size and B/M	0.00%	6.35%	2.25%	0.16%	0.31%	0.43%	0.54%	0.63%

(Continued on next page)

Dataset	1/N	MVO	DGNU1	IVol1	IVol2	IVol3	IVol4	IVol5
25 Asia Size and B/M	0.00%	-3.87%	6.02%	0.11%	0.24%	0.37%	0.50%	0.65%

Table 7. Dispersion in assets' correlations. For each dataset, we report the average dispersion in correlations, which is described in details in Section 5. For each rebalancing date t, the correlation matrix C_t is calculated using the excess returns from $t-m$ to $t-1$, where m is the estimation window (60 months or 120 months when a dataset has more than 60 assets). We define SD_t as the standard deviation of the off-diagonal elements of C_t. We then average across all rebalancing dates to obtain the average correlation dispersion: Avg Correlation Dispersion $= \frac{1}{T}\sum_{t=1}^{T} SD_t$.

	Datasets	**Avg Correlation Dispersion**
Multi-Assets	Equity, Fixed Income	0.4370
	Eq., F.I., Comm. and 10 S&P	0.3658
	Eq., Fixed Income, Comm.	0.3612
	76 Anomalies	0.2831
Equity Only	49 Industries	0.1752
	10 Industries	0.1638
	17 Industries	0.1612
	30 Industries	0.1611
	Individual Stocks	0.1442
	25 Size and B/M	0.0966
	25 Japan Size and B/M	0.0887
	25 Size and Operating Prof	0.0802
	25 Asia Size and B/M	0.0756
	25 Europe Size and B/M	0.0729
	12 FF5-Portfolios	0.0577
	11 q-Portfolios	0.0508

Table 8. Clustering and Sharpe Ratios. This table shows the improvement in the Sharpe ratios of the inverse volatility rule when assets are grouped into heterogeneous clusters using various clustering algorithms. We report the proportion of the Sharpe ratio of IVol1 when assets are clustered with respect to the Sharpe ratio of the same rule with unclustered data. The clustering methodology is described in details in Section 5. The column "HC Single" clusters the data using hierarchical clustering with the Single linkage, the column DBSCAN uses the clustering algorithm "Density-Based Spatial Clustering of Applications with Noise (DBSCAN)", the column "Kmeans" uses the K-means algorithm, and the column "HC Ward" uses the hierarchical clustering with the Ward linkage. In parentheses, we report the p-value – computed using the test described in Internet Appendix B – of the difference between the Sharpe ratio when using clustering and the Sharpe ratio without clustering.

	HC Single	**DBSCAN**	**Kmeans**	**HC Ward**
Eq., Fixed Income, Comm.	0.92	0.90	0.93	0.87
	(0.33)	(0.00)	(0.06)	(0.00)
Eq., F.I., Comm. and 10 S&P	1.15	1.24	1.14	1.03
	(0.35)	(0.00)	(0.17)	(0.00)
Equity, Fixed Income	0.99	0.99	0.99	0.99
	(0.13)	(0.00)	(0.13)	(0.00)
76 Anomalies	1.11	1.06	0.98	1.24
	(0.57)	(0.00)	(0.60)	(0.00)
10 Industries	1.05	1.03	1.04	1.03
	(0.22)	(0.00)	(0.60)	(0.00)
17 Industries	1.10	1.09	1.06	1.06
	(0.07)	(0.00)	(0.09)	(0.00)
30 Industries	1.18	1.08	1.08	1.05
	(0.00)	(0.00)	(0.25)	(0.00)
49 Industries	1.14	0.81	1.04	1.05
	(0.08)	(0.00)	(0.11)	(0.00)
Individual Stocks	0.98	1.02	1.00	1.04
	(0.87)	(0.00)	(0.52)	(0.00)
25 Size and B/M	0.97	1.00	1.02	0.98
	(0.50)	(0.00)	(0.98)	(0.00)
25 Size and Operating Prof	0.98	1.02	1.02	0.98
	(0.72)	(0.00)	(0.58)	(0.00)
11 q-Portfolios	1.00	0.88	0.88	1.00
	(0.94)	(0.00)	(0.00)	(0.00)
12 FF5-Portfolios	1.00	0.97	0.95	1.01
	(0.99)	(0.00)	(0.39)	(0.00)
25 Europe Size and B/M	0.85	0.96	1.00	0.98
	(0.10)	(0.00)	(0.50)	(0.00)
25 Japan Size and B/M	0.89	0.84	0.92	0.92
	(0.50)	(0.00)	(0.27)	(0.00)
25 Asia Size and B/M	0.94	1.07	1.02	0.97
	(0.60)	(0.00)	(0.39)	(0.00)

Internet Appendix A Proof for statistical test for equality of between IVolγ and mean-variance portfolio

In this section, we provide the proofs of Proposition 1 and Proposition 2.

We begin by proving Proposition 1.

Proof. We begin with the proof of the necessary condition (i.e., $w_{TAN} = w_{IVol\gamma}$ implies that $\Sigma^{-1}\mu \times D^{-\gamma}1_N = 0_N$). From Equation (11) and (12), for $w_{TAN} = w_{IVol\gamma}$ it needs to be that

$$\frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu} = \frac{D^{-\gamma}1_N}{1_N'D^{-\gamma}1_N} \tag{A.1}$$

which we can re-write as

$$\Sigma^{-1}\mu = \frac{1_N'\Sigma^{-1}\mu}{1_N'D^{-\gamma}1_N}D^{-\gamma}1_N \tag{A.2}$$

The above equation implies that the vectors $\Sigma^{-1}\mu$ and $D^{-\gamma}1_N$ are proportional to each other (i.e., they are parallel). Therefore, their cross product satisfies the condition $\Sigma^{-1}\mu \times D^{-\gamma}1_N = 0_N$. This completes the proof of the necessary condition.

As for the sufficient condition, $\Sigma^{-1}\mu \times D^{-\gamma}1_N = 0_N$ implies that there exists a constant $k \in \mathbb{R}$ such that

$$\Sigma^{-1}\mu = kD^{-\gamma}1_N \tag{A.3}$$

Given the fact that $w_{TAN} = \frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu}$, it follows that

$$w_{TAN} = \frac{\Sigma^{-1}\mu}{1_N'\Sigma^{-1}\mu} = \frac{kD^{-\gamma}1_N}{1_N'kD^{-\gamma}1_N} = \frac{D^{-\gamma}1_N}{1_N'D^{-\gamma}1_N} = w_{IVol\gamma} \tag{A.4}$$

where the second equality follows from using Equation (A.3). This completes the proof of sufficient condition. □

Next, we prove Proposition 2

Proof. Proposition 1 implies that for w_{MVO} to be equivalent to $w_{IVol\gamma}$, the following condition must hold

$$D^{\gamma}\Sigma^{-1}\mu \times 1_N = 0_N. \tag{A.5}$$

It is known that the above condition is equivalent to

$$P'D^{\gamma}\Sigma^{-1}\mu = 0_{N-1} \tag{A.6}$$

where P is an $N(N-1)$ orthonormal matrix with its columns orthogonal to 1_N.

Let $c = D^{\gamma}\Sigma^{-1}\mu$. In order to test this necessary and sufficient condition, we can look at the sample counterpart of c, defined as

$$\hat{c} = \hat{D}^{\gamma}\hat{\Sigma}^{-1}\hat{\mu}, \tag{A.7}$$

where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$, $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'$, $\hat{D}^{\gamma} = \text{Diag}(\hat{\Sigma})^{\frac{\gamma}{2}}$, and R_t is a vector of returns.

To derive a statistical test of the necessary and sufficient condition, we first find the limiting distribution of $\hat{c}$. Under the assumption that R_t is stationary and ergodic with finite fourth moments, it is known that the limiting distribution is

$$\sqrt{T}(\hat{c} - c) \xrightarrow{d} N(0_N, V(\hat{c})), \tag{A.8}$$

and we can construct a test of the null hypothesis $H_0 : P'c = 0_{N-1}$ using

$$J = T(P'\hat{c})'(P'\hat{V}(\hat{c})P)^{-1}(P'\hat{c}) \xrightarrow{d} \chi^2_{N-1}, \tag{A.9}$$

where $\hat{V}(\hat{c})$ is a consistent estimator of $V(\hat{c})$.

The last step is to derive an explicit expression of $V(\hat{c})$. Since $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$ and $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'$ are simply GMM estimators of μ and Σ in an exactly identified system, the asymptotic distribution of $\hat{\theta} = [\hat{\mu}',\ \text{vec}(\hat{\Sigma})']'$ is given by

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0_{N+N^2}, S_0), \tag{A.10}$$

where

$$S_0 = \sum_{j=-\infty}^{\infty} E[g_t g'_{t+j}], \tag{A.11}$$

and

$$g_t = \begin{bmatrix} R_t - \mu \\ \text{vec}((R_t - \mu)(R_t - \mu)' - \Sigma) \end{bmatrix}. \tag{A.12}$$

Let us denote $c_1 \equiv \Sigma^{-1}\mu$ and $\hat{c}_1 = \hat{\Sigma}^{-1}\hat{\mu}$. Given that

$$\frac{\partial c_1}{\partial \mu} = \Sigma^{-1}, \tag{A.13}$$

$$\frac{\partial c_1}{\partial \text{vec}(\Sigma)'} = -\mu'\Sigma^{-1} \otimes \Sigma^{-1}, \tag{A.14}$$

and using the delta method, we can show that

$$\sqrt{T}(\hat{c}_1 - c_1) \xrightarrow{d} N\left(0_N, \sum_{j=-\infty}^{\infty} E[h_{1t}h'_{1,t+j}]\right), \tag{A.15}$$

where

$$h_{1t} = \Sigma^{-1}(R_t - \mu) - \mu'\Sigma^{-1}(R_t - \mu)\Sigma^{-1}(R_t - \mu) + \Sigma^{-1}\mu. \tag{A.16}$$

Denote $c_2 \equiv [\sigma_1^2, \ldots, \sigma_N^2]'$ and $\hat{c}_2 = [\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2]'$. The limiting distribution of $\hat{c}_2$ is

$$\sqrt{T}(\hat{c}_2 - c_2) \xrightarrow{d} N\left(0_N, \sum_{j=-\infty}^{\infty} E[h_{2t}h'_{2,t+j}]\right), \tag{A.17}$$

where

$$h_{2t} = \text{Diag}((R_t - \mu)(R_t - \mu)') - c_2. \tag{A.18}$$

Since $c_i = c_{1i}\sqrt{c_{2i}^{\gamma}}$, we can use the delta method to obtain

$$\sqrt{T}(\hat{c} - c) \xrightarrow{d} N\left(0_N, \sum_{j=-\infty}^{\infty} E[q_t q'_{t+j}]\right), \tag{A.19}$$

where the i-th element of q_t is given by

$$q_{it} = \sigma_i^{\gamma} e'_i h_{1t} + \gamma \frac{c_{1i} e'_i D^{2(\gamma-1)} h_{2t}}{2\sigma_i^{\gamma}}, \tag{A.20}$$

and e_i is an N-vector with its i-th element equals to one and zero otherwise. More compactly, we can write

$$q_t = D^{\gamma} h_{1t} + \frac{\gamma}{2}\text{Diag}(c_1) D^{-\gamma} D^{2(\gamma-1)} h_{2t}. \tag{A.21}$$

Under the i.i.d. assumption on R_t, we have

$$V(\hat{c}) = E[q_t q'_t]. \tag{A.22}$$

In addition, when R_t is i.i.d. normal, we have

$$\begin{aligned}
E[h_{1t}h'_{1t}] &= \Sigma^{-1} + (\mu'\Sigma^{-1}\mu)\Sigma^{-1} + 2\Sigma^{-1}\mu\mu'\Sigma^{-1} - 2\Sigma^{-1}\mu\mu'\Sigma^{-1} + \Sigma^{-1}\mu\mu'\Sigma^{-1} \\
&= (1 + \mu'\Sigma^{-1}\mu)\Sigma^{-1} + \Sigma^{-1}\mu\mu'\Sigma^{-1}, && \text{(A.23)} \\
E[h_{2t}h'_{2t}] &= C, && \text{(A.24)} \\
E[h_{1t}h'_{2t}] &= -E[\mu'\Sigma^{-1}(R_t - \mu)\Sigma^{-1}(R_t - \mu)h'_{2t}] = -G + \Sigma^{-1}\mu c'_2, && \text{(A.25)} \\
&= -2\text{Diag}(\mu), && \text{(A.26)}
\end{aligned}$$

where $C_{ij} = 2\sigma_{ij}^2$ and $G_{ij} = e'_i\Sigma^{-1}\mu\sigma_j^2 + 2e'_i e_j \mu_j$. It follows that an explicit expression for $V(\hat{c})$ is

$$\begin{aligned}
\hat{V}(\hat{c}) = E[\hat{q}_t\hat{q}'_t] &= \hat{D}^{\gamma} E[\hat{h}_{1t}\hat{h}'_{1t}]\hat{D}^{\gamma} - 2\gamma\text{Diag}(\hat{\mu})\hat{D}^{2(\gamma-1)}\text{Diag}(\hat{c}_1) + \\
&+ \frac{1}{4}\gamma^2 \text{Diag}(\hat{c}_1)\hat{D}^{\gamma-2}\,\hat{C}\,\hat{D}^{\gamma-2}\text{Diag}(\hat{c}_1) && \text{(A.27)}
\end{aligned}$$

A consistent estimator of $V(\hat{c})$ can simply be calculated by using the sample counterparts

$$\begin{aligned}
V(\hat{c}) = E[q_t q'_t] &= D^{\gamma} E[h_{1t}h'_{1t}] D^{\gamma} - 2\gamma\text{Diag}(\mu) D^{2(\gamma-1)}\text{Diag}(c_1) + \\
&+ \frac{1}{4}\gamma^2 \text{Diag}(c_1) D^{\gamma-2}\, C\, D^{\gamma-2}\text{Diag}(c_1) && \text{(A.28)}
\end{aligned}$$

□

Internet Appendix B Testing the equality of Sharpe ratios

In this section, we describe the test that we use to evaluate whether two time-series of returns have statistically different Sharpe ratios. This is an extension of the test developed by Jobson and Korkie (1981) as corrected by Memmel (2003). While Memmel (2003)'s test requires returns to be normal, the test described here is more general as it does not require the data to be normal.[27]

Let $r_t = [r_{1t}, r_{2t}]' \sim N(\mu, \Sigma)$ be the excess return of two assets at time t, where $\mu = [\mu_1, \mu_2]'$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

Let $SR_i = \frac{\mu_i}{\sigma_i}$ be the Sharpe ratio of an asset i. We denote the sample counterparts of the true variable x with $\widehat{x}$ (e.g. $\widehat{\mu}_1$ is the sample mean for asset 1).

First, we derive the distribution of $\widehat{\delta} = \widehat{SR}_1 - \widehat{SR}_2$. We assume that r_t is stationary with finite fourth moments. Let $\widehat{\varphi} = [\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1, \widehat{\sigma}_2]'$. Since $\widehat{\varphi}$ is the GMM estimator of φ with moment conditions

$$E\left[g(\varphi)\right] = E\begin{bmatrix} r_{1t} - \mu_1 \\ r_{2t} - \mu_2 \\ (r_{1t} - \mu_1)^2 - \sigma_1^2 \\ (r_{2t} - \mu_2)^2 - \sigma_2^2 \end{bmatrix} = 0_4$$

then

$$\sqrt{T}(\widehat{\varphi} - \varphi) \xrightarrow{d} N\left(0_4, \sum_{j=-\infty}^{+\infty} E\left[g_t(\varphi) g_{t+j}(\varphi)'\right]\right)$$

Using the delta method, the asymptotic distribution of $\widehat{\delta} = \widehat{SR}_1 - \widehat{SR}_2$ is

$$\sqrt{T}(\widehat{\delta} - \delta) \xrightarrow{d} N\left(0, \sigma_{\Delta SR}\right)$$

where $\sigma_{\Delta SR} = \sum_{j=-\infty}^{+\infty} E\left[(h_{1,t} - h_{2,t})(h_{1,t+j} - h_{2,t+j})\right]$ and $h_{i,t} = \frac{r_{it} - \mu_i}{\sigma_i} - \frac{\mu_i}{\sigma_i}\frac{(r_{it} - \mu_i)^2 - \sigma_i^2}{2\sigma_i^2}$. It follows that the statistic Z to test whether the difference in the Sharpe ratios $\widehat{SR}_1 - \widehat{SR}_2$ is given by

$$Z = \frac{\widehat{SR}_1 - \widehat{SR}_2}{\sqrt{1/T \cdot \widehat{\sigma}_{\Delta SR}}} \xrightarrow{d} N\left(0, 1\right) \tag{B.1}$$

Note that the Z statistic described in Equation (B.1) converges to the one described in Memmel (2003) under the assumption that r_t is i.i.d. and bivariate normally distributed.

[27] We are extremely grateful to Raymond Kan for sharing his notes on the derivation of this test.

Internet Appendix C Estimation Window and Distribution of out-of-sample Sharpe Ratios

The estimation window can have a large effect on the performance of different allocation rules. Longer estimation windows might allow for more precise estimates of the parameters and improve the performance of any allocation rule. In this section, we study how different estimation windows affect the 3 aforementioned allocation rules (TAN, GMV, and IVol). Furthermore, as we argued above, there is a trade-off between minimizing estimation error and optimality of a portfolio. If a portfolio is theoretically optimal but subject to large estimation error, it might perform worse than a portfolio that is sub-optimal theoretically but subject to low estimation error. In this section, we also show what is the length of the estimation window that is required for the TAN portfolio (which is optimal in a mean-variance framework) to perform as well as a suboptimal strategy (GMV or IVol).[28]

Figure C.1 shows the Sharpe ratio of a given allocation rule as a function of the estimation window T. The y-axis displays the monthly Sharpe ratio $\widehat{SR}_j$, which is defined in Equation (9), and the x-axis shows the estimation window T, expressed in months. For a fixed estimation window T, we simulate 10,000 portfolios with optimal weights calculated using the estimated mean excess return and variance-covariance matrix, $\widehat{\mu}$ and $\widehat{\Sigma}$. As in Section 2.2, assets are assumed to have the same Sharpe ratios, and the true volatility of each asset is randomly drawn from a uniform distribution between 10% and 40%. The red solid line shows the median Sharpe ratio of the tangency portfolio ($\widehat{SR}_{TAN}$) for the 10,000 simulations, the blue dashed line displays the median Sharpe ratio for the GMV portfolios ($\widehat{SR}_{GMV}$) and the green dotted line depicts the median Sharpe ratio for the IVol portfolios ($\widehat{SR}_{IVol}$). The shaded areas contain the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles from the 10,000 simulations.

In Panel A of Figure C.1, we consider the case of 5 uncorrelated assets. As shown by the shaded area, the distribution of out-of-sample Sharpe ratios for the TAN portfolio is much wider than that of the GMV or IVol portfolios. This is consistent with the intuition from Section 2.2 showing that the TAN portfolio is subject to considerably larger estimation error compared to GMV and IVol. Also, the median Sharpe ratio for the TAN portfolio becomes superior to that of GMV portfolio only when the estimation window is greater than 1,500 months ($\approx$ 125 years). The Sharpe ratios of the TAN and IVol portfolios if one knew the true parameters would be approximately 18.63% as shown by the convergence for $T \to \infty$.[29] For the GMV portfolio, the true Sharpe ratios would be 18.07%. Notably, the difference in the Sharpe ratios would be small between the 3 different allocation rules if investors knew the true parameters. The results for the case of 15 uncorrelated assets are shown in Panel B of Figure C.1 and they are qualitatively similar to those in Panel A for the case of 5 assets.

[Insert Figure C.1 here]

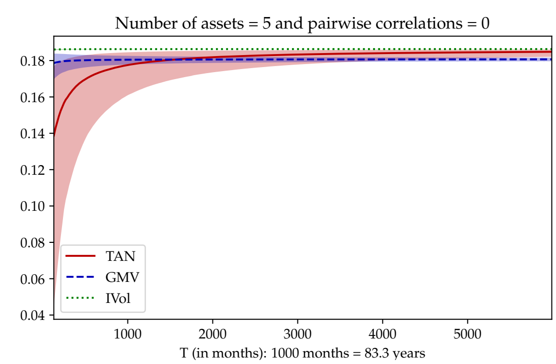
[28] DeMiguel et al. (2009b) perform a similar simulation exercise. We differentiate from them because we compare the performance of the tangency (TAN) and global minimum variance (GMV) portfolios to that of the inverse volatility portfolio (IVol), which is not discussed in DeMiguel et al. (2009b).

[29] In Internet Appendix A, we show the necessary and sufficient conditions for the IVol portfolio to be equivalent to the tangency portfolio in the absence of estimation error.

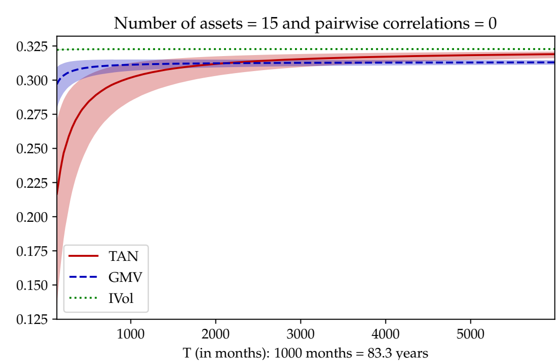
Panel C shows the simulations when there are 5 highly correlated assets (constant pairwise correlation of 0.7). If one knew the true parameters the Sharpe ratio of the TAN and IVol portfolios would be approximately 9.56% while for the GMV portfolio the Sharpe ratio would be 7.12%. Different from the results in Panel A and Panel B, Panel C shows that if investors knew the true parameters, the GMV portfolio would deliver a considerably lower Sharpe ratio compared to the TAN portfolio. This is not the case for the IVol portfolio which exhibits a Sharpe ratio that is very close to that of the TAN portfolio. In Internet Appendix A, we show that, in the absence of estimation error, the weights of the IVol portfolio are close to the weights of the TAN portfolio when assets have the same Sharpe ratios and equal pairwise correlations. Internet Appendix A shows that the assumption that assets have the same Sharpe ratios is not restrictive but rather supported by empirical evidence. We do not find the same support for the assumption of equal pairwise correlations which implies that the IVol strategy is not optimal from a theoretical point of view. There is therefore a tradeoff between theoretical optimality and estimation error: the tangency portfolio (optimal from a mean-variance standpoint) has a larger estimation error than the IVol portfolio (which is suboptimal theoretical) because it requires the estimation of expected returns and correlations while the IVol portfolio relies solely on the estimation of volatilities. This section highlights the existence of this trade-off between optimality and estimation error; however, which portfolio performs in the real world is ultimately an empirical question that we address in Section 4.

Figure C.1. The effect of the estimation window. This figure shows the Sharpe ratio performance of three different allocation rules as a function of the estimation window T. The three allocation rules plotted are: the optimal mean-variance tangency portfolio (TAN), the global minimum variance portfolio (GMV), and the inverse volatility portfolio (IVol). The y-axis displays the monthly Sharpe ratio, and the x-axis shows the estimation window T. For a fixed estimation window T, we simulate 10,000 portfolios with optimal weights calculated using the estimated mean return and variance-covariance matrix, $\widehat{\mu}$ and $\widehat{\Sigma}$. In each panel, the legend reports the name of the series. The red solid line shows the median Sharpe ratio of the simulated tangency portfolios (TAN), the blue dashed line displays the median Sharpe ratio for the GMV portfolios and the green dotted line depicts the median Sharpe ratio for the IVol portfolios. The shaded areas contain the 5^{th} and 95^{th} percentiles from the 10,000 simulations. Panel A and Panel B show the simulations when there are 5 and 15 uncorrelated assets, respectively. Panel C and Panel D show the simulations when there are 5 and 15 highly correlated assets (constant pairwise correlation of 0.7), respectively. In all panels, assets are assumed to have the same Sharpe ratios.

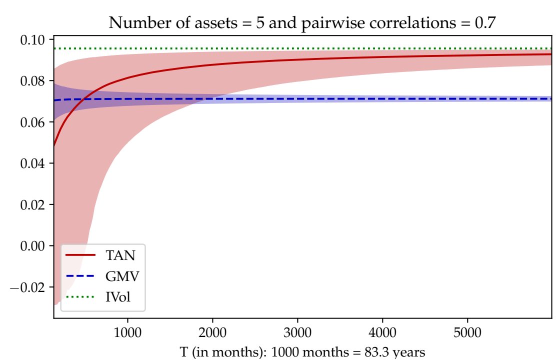
Panel A



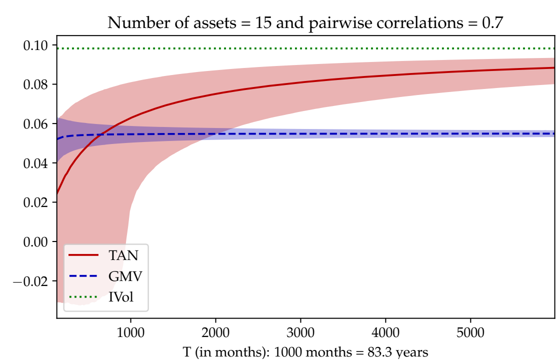
Panel B



Panel C



Panel D



Internet Appendix D Description of the anomalies used in this paper

Table D.1. This table lists the 76 anomalies used in this study. The column "Reference Paper" refers to the paper that contains the methodology used to build the anomaly. When authors make the data available until the end of 2019, we use their data. If not, we build the anomalies ourselves.

#	Anomaly	Reference Paper	Name
1	ABR1	Chan et al. (1996)	Cumulative abnormal returns around earnings announcement dates. Holding period 1 month.
2	ABR6	Chan et al. (1996)	Cumulative abnormal returns around earnings announcement dates. Holding period 6 months.
3	ACI	Hou et al. (2018)	Abnormal Corporate Investment.
4	ADM	Chan et al. (2001b)	Advertising expense-to-market.
5	BAB	Frazzini and Pedersen (2014)	Betting-against-beta.
6	BM	Fama and French (1993)	Sort by book-to-market equity.
7	CEI	Hou et al. (2018)	Composite Equity Issuance.
8	CLA	Ball et al. (2016)	Cash-based operating profits-to-lagged assets using yearly Compustat data.
9	CLAQ1	Ball et al. (2016)	Cash-based operating profits-to-lagged assets using quarterly Compustat data and holding period of 1 month.
10	CMA	Fama and French (2015)	Conservative minus Aggressive.
11	COP	Ball et al. (2016)	Cash-based operating profitability
12	DA	Hou et al. (2018)	Changes in in short-term investments.
13	DFIN	Hou et al. (2018)	Changes in net financial assets
14	DLTI	Hou et al. (2018)	Changes in in short-term investments.
15	DNCA	Hou et al. (2018)	Changes in non-current operating assets.
16	DNCO	Hou et al. (2018)	Changes in net non-current operating assets
17	DNOA	Hou et al. (2018)	Changes in net operating assets.
18	DPIA	Hou et al. (2018)	Changes in PPE and inventory-to-assets.

(Continued on next page)

Table D.1 – continued from previous page

#	Anomaly	Reference Paper	Name
19	DROE1	Hou et al. (2018)	4-quarter change in return on equity. Holding period of 1 month.
20	DROE12	Hou et al. (2018)	4-quarter change in return on equity. Holding period of 12 months.
21	DROE6	Hou et al. (2018)	4-quarter change in return on equity. Holding period of 6 months.
22	DWC	Hou et al. (2018)	Changes in net noncash working capital.
23	EM	Loughran and Wellman (2011)	Enterprise Multiple
24	HML	Fama and French (1993)	High minus Low.
25	HMLD	Asness and Frazzini (2013)	The devil in HML's details
26	IA	Hou et al. (2018)	Investment-to-assets.
27	IG	Hou et al. (2018)	Investment Growth, 1 year.
28	IG2y	Hou et al. (2018)	Investment Growth, 2 years.
29	IVC	Hou et al. (2018)	Inventory Changes.
30	IVG	Hou et al. (2018)	Inventory Growth.
31	MKT-RF		Excess Market Return.
32	NEI1	Barth et al. (1999)	The number of quarters with consecutive earnings increase.
33	NOA	Hirshleifer et al. (2004)	Net operating assets.
34	NOP	Boudoukh et al. (2007)	Net payout yield.
35	NSI	Pontiff and Woodgate (2008)	Net stock issues.
36	OA	Sloan (1996)	Operating Accruals.
37	OCA	Eisfeldt and Papanikolaou (2013)	Industry-adjusted organizational capital-to-assets.
38	OCP	Desai et al. (2004)	Operating cash-flow to price.
39	OP	Hou et al. (2018)	Payout Yield
40	OPA	Ball et al. (2016)	Operating profits to assets.
41	POA	Sloan (1996)	Percent operating accruals.
42	PTA	Sloan (1996)	Percent total accruals.
43	QMJ	Asness et al. (2019)	Quality minus Junk
44	R_EG	Hou et al. (2019)	Expected Growth Factor.
45	R_IA	Hou et al. (2015)	Investment Factor
46	R_ROE	Hou et al. (2015)	ROE Factor
47	R111	Jegadeesh and Titman (1993)	Price momentum, prior 11-month returns, holding period 1 month
48	R1112	Jegadeesh and Titman (1993)	Price momentum, prior 11-month returns, holding period 12 months
49	R1115A	Heston and Sadka (2008)	Years 11–15 lagged returns, annual

(Continued on next page)

Table D.1 – continued from previous page

#	Anomaly	Reference Paper	Name
50	R1115n	Heston and Sadka (2008)	Years 11–15 lagged returns, nonannual
51	R116	Jegadeesh and Titman (1993)	Price momentum, prior 11-month returns, holding period 6 months
52	R15A	Heston and Sadka (2008)	Years 1–5 lagged returns, annual
53	R1620A	Heston and Sadka (2008)	Years 16–20 lagged returns, annual
54	R1A	Heston and Sadka (2008)	Year 1-lagged return, annual
55	R1N	Heston and Sadka (2008)	Year 1-lagged return, nonannual
56	R25A	Heston and Sadka (2008)	Years 2–5 lagged returns, annual
57	R61	Jegadeesh and Titman (1993)	Price momentum, prior 6-month returns, holding period 1 months
58	R610A	Heston and Sadka (2008)	Years 6–10 lagged returns, annual
59	R610n	Heston and Sadka (2008)	Years 6–10 lagged returns, nonannual
60	R612	Jegadeesh and Titman (1993)	Price momentum, prior 6-month returns, holding period 12 months
61	R66	Jegadeesh and Titman (1993)	Price momentum, prior 6-month returns, holding period 6 months
62	RDM	Chan et al. (2001b)	R&D expense-to-market using Compustat yearly.
63	RE_1	Chan et al. (2001a)	Revisions in analysts' earnings forecasts - 1 month holding period
64	RE_6	Chan et al. (2001a)	Revisions in analysts' earnings forecasts - 6 months holding period
65	RER	Tuzel (2010)	Industry-adjusted real estate ratio
66	RESID11_1	Blitz et al. (2011)	11-month residual momentum, 1-month holding period
67	RESID11_12	Blitz et al. (2011)	11-month residual momentum, 12-month holding period
68	RESID11_6	Blitz et al. (2011)	11-month residual momentum, 6-month holding period
69	RESID6_12	Blitz et al. (2011)	6-month residual momentum, 12-month holding period
70	RESID6_6	Blitz et al. (2011)	6-month residual momentum, 6-month holding period
71	RMW	Fama and French (2015)	Robust minus weak factor
72	ROE1	Hou et al. (2015)	Return on Equity with holding period of 1 month.
73	ROE6	Hou et al. (2015)	Return on Equity with holding period of 6 months.
74	SP	Barbee Jr et al. (1996)	Sales-to-price ratio

(Continued on next page)

Table D.1 – continued from previous page

#	Anomaly	Reference Paper	Name
75	SUE1	Foster et al. (1984)	Standardized unexpected earnings. Holding period of 1 month.
76	SUE6	Foster et al. (1984)	Standardized unexpected earnings. Holding period of 6 months.

Internet Appendix E   Performance measures

We provide a performance comparison between IVolγ rules, tangency portfolio, 1/N rule, and a benchmark portfolio that uses only the covariance matrix: the 1-norm-constrained portfolio developed in DeMiguel, Garlappi, Nogales, and Uppal (2009a), which we label DGNU1. Before providing the results, we describe the performance measures that we use in our analysis. Following DeMiguel et al. (2009b), we consider 3 different measures: Sharpe ratio, Certainty Equivalent Returns (CEQ) and Turnover. All 3 measures are evaluated out-of-sample.

Sharpe ratio. Given the time-series of excess returns of the portfolio constructed using allocation rule i, the out-of-sample Sharpe ratio is defined as the sample average excess return of such rule ($\widehat{\mu}_i$) divided by its sample standard deviation ($\widehat{\sigma}_i$):

$$\widehat{SR}_i = \frac{\widehat{\mu}_i}{\widehat{\sigma}_i} \tag{E.1}$$

We check whether the difference in the Sharpe ratio of two different allocation rules is statistically different from zero using the test described in Internet Appendix B.

Certainty Equivalent Return (CEQ). The CEQ of allocation rule i is defined as the certain (i.e. risk-free) return that makes an investor indifferent between investing in i or receiving such certain return. Formally, the CEQ of allocation rule i is computed as

$$\widehat{CEQ}_i = \widehat{\mu}_i - \frac{\gamma}{2}\widehat{\sigma}_i^2 \tag{E.2}$$

where $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ are the mean and variance for average excess returns of the portfolio built using allocation rule i. The parameter γ is the risk aversion coefficient, which we set equal to 5.

We check whether the difference between the CEQs of the two allocation rules is statistically different from zero using the test for CEQ returns described in DeMiguel et al. (2009b).

Turnover. The Turnover of allocation rule i provides information on the number of trades required to implement such a rule. The higher the turnover, the more trades a strategy requires to be implemented, and the higher the transaction costs will be. Assuming that there are N assets and we have the time series of portfolio weights for T periods, the turnover of allocation rule i is calculated as

$$Turnover = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{N}\left(|\widehat{w}_{i,s,t+1} - \widehat{w}_{i,s,t}^{+}|\right) \tag{E.3}$$

where $\widehat{w}_{i,s,t+1}$ is the optimal weight of asset s according to allocation rule i at time $t+1$, and $\widehat{w}_{i,s,t}^{+}$ is the weight that asset s has before rebalancing at time $t+1$.

Return-gain. For each allocation rule, we compute the return-gain with respect to the 1/N rule. The return-gain is defined as the additional *annualized* return needed for the 1/N rule to perform as well as the allocation rule i in terms of the Sharpe ratio. We calculate the evolution of wealth for allocation rule i at time t ($\mathbf{W}_{i,t+1}$) as

$$\mathbf{W}_{i,t+1} = \mathbf{W}_{i,t}\left(1 + r_{i,P}\right)\left(1 - c_t \times \sum_{j=1}^{N} |\widehat{w}_{i,j,t+1} - \widehat{w}_{i,j,t^+}|\right)$$

where $r_{i,P}$ is the return of the portfolio before transaction costs using strategy i, c_t is the time-varying proportional transaction cost parameter, $\widehat{w}_{i,j,t+1}$ is the optimal weight of asset j according to allocation rule i at time $t+1$, and $\widehat{w}_{i,j,t^+}$ is the weight that asset j has before rebalancing at time $t+1$. The term $c_t \times \sum_{j=1}^{N} |\widehat{w}_{i,j,t+1} - \widehat{w}_{i,j,t^+}|$ represents the total transaction costs for rebalancing the portfolio at time $t+1$. We follow Brandt et al. (2009) and Hand and Green (2011) and assume that c_t decreases over time. We model transaction costs according to the following formula: $c_t = k_t \times 0.35\%$, where k_t is a multiplier that decreases linearly from 3.3 in January 1980 to 1.0 in January 2002 as in DeMiguel, Martin-Utrera, Nogales, and Uppal (2020). Also, k_t remains equal to 3.3 before 1980 and equal to 1.0 after January 2002. This parametrization gives use transaction costs that linearly decrease from 1.15% to 0.35% and they capture the decrease in transaction costs that have been observed over time thanks to the increased efficiency of financial markets.

The out-of-sample portfolio returns net of transaction costs according to allocation rule i are equal to the percentage change in wealth $\mathbf{W}_{i,t+1}/\mathbf{W}_{i,t} - 1$. Let μ_i and σ_i be the monthly average and standard deviation of the monthly portfolio returns net of transaction costs according to allocation rule i. The monthly return-gain of strategy i is defined as

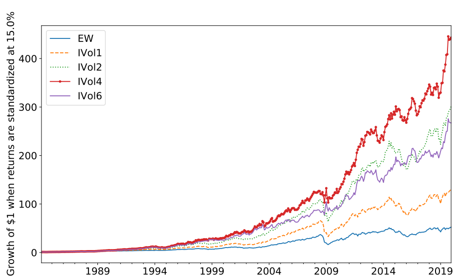
$$\text{return-gain}_i = \frac{\mu_i \times \sigma_{1/N}}{\sigma_i} - \mu_{1/N} \tag{E.4}$$

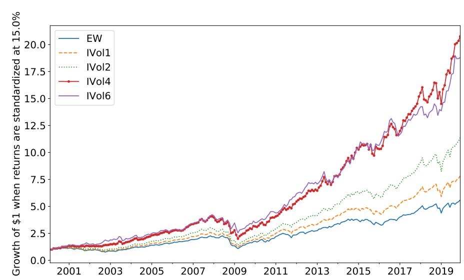
Internet Appendix F Growth of $1 for a given volatility

Figure F.1. Growth of $1 when returns are standardized at 15%. This figure shows the growth of $1 invested according to various allocation rules. The out-of-sample returns of the various rules are the ones described in Table 3 and we standardize their volatility to 15% (annualized) to make the strategies comparable. The different panels contain the results for different datasets. Panel A uses the returns from "Equity, Fixed Income, Commodities", Panel B uses the returns from "Equity, Fixed Income, Commodities, and 10 S&P Industries", Panel C uses the returns from "Equity and Fixed Income", Panel D uses the returns from "76 Anomalies", Panel E uses the returns from "10 Industries", Panel F uses the returns from "17 Industries". All datasets are described in Section 3.2.

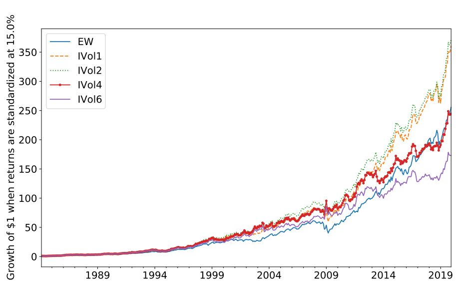
Panel B - Equity, Fixed Income, Commodities

Panel A - Equity, Fixed Income, Commoditiesand 10 S&P Industries

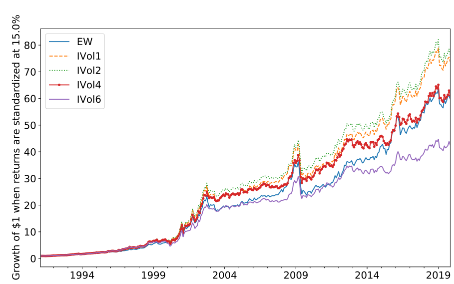




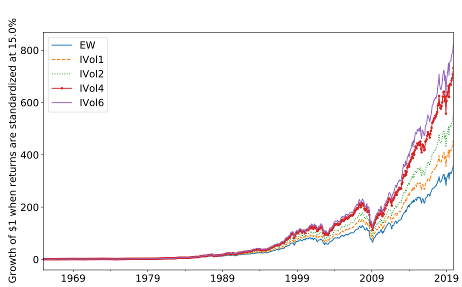
Panel C - Equity and Fixed Income



Panel D - 76 Anomalies



Panel E - 10 Industries



Panel F - 17 Industries