Optimal Granularity for Portfolio Choice

Nicole Branger[1], Katarína Lučivjanská[2] and Alex Weissensteiner[3]

[1]Finance Center Muenster, University of Muenster, Germany
email: nicole.branger@wiwi.uni-muenster.de
[2]Faculty of Economics and Business Administration, VU University Amsterdam, Netherlands, k.lucivjanska@vu.nl
[3]School of Economics and Management, Free University of Bozen-Bolzano, Italy, email:alex.weissensteiner@unibz.it

July 14, 2016

Abstract

Many optimization-based portfolio rules fail to beat the simple 1/N rule out-of-sample because of parameter uncertainty. In this paper we suggest a grouping strategy in which we first form groups of equally weighted stocks and then optimize over the resulting groups only. In a simplified setting we show analytically how to optimize the trade-off between drawbacks from parameter uncertainty and drawbacks from deviating from the overall optimal asset allocation. We illustrate that the optimal group size depends on the volatility of the assets, on the number of observations and on how much the optimal asset allocation differs from 1/N. Out of sample back-tests confirm the validity of our grouping strategy empirically.

Keywords: Mean-variance optimization, 1/N rule, Parameter uncertainty, Optimal portfolio granularity.
JEL classification: G1, G11

1 Introduction

The classical theory of mean-variance analysis suggests that the optimal portfolio lies on the efficient frontier (Markowitz, 1952; Tobin, 1958). However, the estimation of the necessary input parameters is in practice so imprecise that the naive 1/N strategy, which does not use any historical data to optimize the composition of the portfolio of risky assets, often delivers a better performance (Duchin and Levy, 2009; DeMiguel et al., 2009). Other optimization based portfolio selection rules have problems to compete with the simple 1/N rule, too. DeMiguel et al. (2009) compare various portfolio selection rules which have been proposed to deal with estimation risk. The comparison is based on different data sets, and they show that none of these rules is able to beat 1/N consistently. The 1/N strategy has thus become a benchmark against which the performance of alternative portfolio strategies is assessed, and the challenge is to beat $1/N$.

In case of estimation risk, the performance of portfolio strategies depends on the trade-off between the benefits from optimization and the losses from estimation risk. The mean-variance portfolio realizes the maximal benefits from optimizing over the full asset menu, but suffers from large losses due to estimation risk. The $1/N$ strategy foregoes the benefits from optimizing the portfolio of risky assets, but largely reduces the losses from estimation risk. As DeMiguel et al. (2009) show, the $1/N$ strategy is superior in quite a lot of cases. The estimation window which is necessary to alleviate losses from parameter uncertainty is huge even for moderate numbers of assets. Thousands of monthly observations are needed for the mean-variance portfolio to outperform the naive 1/N strategy when considering a portfolio with just more than 10 assets.

In this paper, we suggest grouping strategies as an in-between option. Instead of buying the whole market without optimization (the naive 1/N rule) or basing the asset allocation decision on all available stocks, we collect single stocks in (equally weighted) groups and then optimize over these groups. The grouping strategy thus combines the benefits from optimizing the risk-return trade-off (over the groups) and the reduction of the losses from parameter uncertainty achieved by the naive 1/N rule (within the groups). We find that grouping strategies can often beat both the optimization over the whole asset menu and the 1/N-strategy if one chooses the group size optimally.

The obvious question is of course to which extent assets should be grouped. On the one hand, a larger group size implies less groups to optimize over and in this way mitigates parameter uncertainty. On the other hand, the equal weights of single stocks within each group can be far away from the optimal weights under known parameters, which favors a smaller group size. We rely on a simplified setup to derive the optimal reduction of the original asset menu. The optimal granularity depends on the differences between the assets, i.e. on their heterogeneity, and on the amount of estimation risk. The larger the heterogeneity of the assets (and the more the optimal portfolio deviates from the 1/N-strategy), the smaller the optimal group size. In line with intuition, the optimal group size also decreases in the length of the estimation window, i.e. in the precision of the estimated parameters.

We show that our analytical results carry over to realistic empirical setups. To do so, we apply the grouping strategy to industry portfolios, where we use data sets with 10, 12, and

30 industry portfolios.[1] We find that grouping can indeed help to increase the performance of the strategies, even if the ideal conditions assumed in the analytical analysis do not hold.

Our paper contributes to the literature that deals with the impact of estimation risk on optimal portfolio choice. High parameter uncertainty is mainly driven by the difficulty to estimate the expected returns (Merton, 1980). Therefore, a part of the literature focuses on exploiting only information in the covariance matrix, which is less vulnerable to estimation errors (see Jagannathan and Ma, 2003, among others). Another part of the literature suggests different portfolio selection rules that try to account for parameter uncertainty. First, one approach is to shrink the sample estimates of expected returns and/or of the covariance matrix towards some average values or parameters obtained from factor models (Chan et al., 1999; Green and Hollifield, 1992; Ledoit and Wolf, 2004a,b; Kourtis et al., 2012). Second, Michaud (1989) implements a technique called re-sampling to mitigate issues with parameter uncertainty. Third, short-sale constraints are imposed to avoid extreme positions (Frost and Savarino, 1988). Interestingly, Jagannathan and Ma (2003) show that these constraints have the same effect as shrinking the covariance matrix. Fourth, Garlappi et al. (2007) and Goldfarb and Iyengar (2003) suggest non-Bayesian robust strategies. Fifth, Chen and Yuan (2016) restrict the set of portfolios to the leading eigenvectors of the variance-covariance matrix and optimize over those portfolios only. Finally, Kan and Zhou (2007) suggest a three-fund rule which combines the tangency portfolio and the risk-free asset with a third portfolio as e.g. the minimum-variance portfolio. They show that estimation risk causes the optimal weight of the third portfolio to differ from zero. Tu and Zhou (2011) propose to combine well-known portfolio selection rules in an optimal way with 1/N.

Given that grouping reduces the minimum observation length required for a positive-definite covariance matrix, our paper also relates to the shrinkage technique proposed by Ledoit and Wolf (2004a,b). Optimizing over the constituents of the Standard & Poors 500 index, e.g., would require around 42 years of monthly data. Even if such data were available (different companies were only founded recently), it is doubtful whether very old datasets are still relevant for practical applications. Investors who want to base their portfolio decision on more recent data can rely on the grouping strategy we suggest to profit from optimization benefits even with less data.

The remainder of the paper is organized as follows. Section 2 reviews the mean-variance analysis under parameter uncertainty. Section 3 presents a simulation study to illustrate the benefits from grouping. In Section 4 we discuss the theoretical background of our suggested strategy and derive an analytical solution for the optimal granularity in a simplified setup. In Section 5 we implement our strategy with real data and discuss the out-of-sample performance. Section 6 concludes. Proofs and technical details for all our results in the main text are collected in the appendix.

[1] We rely on the data provided by Kenneth French: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

2 Optimal Asset Allocation Under Parameter Uncertainty

We consider the classical Markowitz/Tobin decision problem of an investor who chooses a portfolio from the set of N risky assets and a riskless asset. We assume that the excess returns of the risky assets r_t are independently and identically distributed (iid) and follow a multivariate normal distribution, i.e. $r_t \sim N(\mu, \Sigma)$. Given the portfolio weights w for the risky assets ($1 - w'\mathbf{1}$ is invested in the riskless asset), the expected excess return of the portfolio is equal to $w'\mu$ and the variance of the portfolio is given by $w'\Sigma w$.

The investor maximizes the mean-variance utility function $U(w)$:[2]

$$\max_{w} w'\mu - \frac{\gamma}{2} w'\Sigma w. \tag{1}$$

When the parameters are known, the optimal weights of the risky assets are given by

$$w^* = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{2}$$

and $1 - \mathbf{1}'w^*$ is invested in the risk-free asset. The resulting optimal utility is

$$U(w^*) = \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu. \tag{3}$$

In practice the true parameters of the return distribution are unknown and have to be estimated from past observations. The maximum likelihood estimators are given by

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{4}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{\mu})'(r_t - \hat{\mu}), \tag{5}$$

where T is the number of observations used to calculate sample moments. Then, $\hat{w} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}$ is the maximum likelihood estimator of w^* (see Kan and Zhou, 2007). This so-called "*plug-in approach*" treats sample parameter estimates as if they were the true ones. The utility

$$U(\hat{w}) = \hat{w}'\mu - \frac{\gamma}{2}\hat{w}'\Sigma\hat{w} \tag{6}$$

then becomes a random variable, and it holds that $U(\hat{w}) \leq U(w^*)$ for any realization of $\{r_t\}_{t=1}^{T}$ that the estimation is based on. Estimation risk thus results in a utility loss.

In case of unknown parameters and maximum likelihood estimators for the expected returns and the covariance matrix, the expected utility $E[U(\hat{w})]$ is

$$E\big[U(\hat{w})\big] = k_1\frac{1}{2\gamma}\mu'\Sigma^{-1}\mu - \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}, \tag{7}$$

[2] $U(w)$ can also be interpreted as a certainty equivalent return, since the investor is indifferent between holding the portfolio w and receiving a risk-free return equal to $r + U(w)$.

where

$$k_1 \quad = \quad \frac{T}{T-N-2}\left(2-\frac{T(T-2)}{(T-N-1)(T-N-4)}\right),$$

for a derivation see Kan and Zhou (2007). The loss $U(w^*) - E[U(\hat{w})]$ increases in the number of assets and decreases in the length of the time window used for the estimation of the asset moments. A larger number of assets implies a larger set of parameters to estimate, and negative consequences of parameter uncertainty boost. On the other hand, if more data are available for the estimation of the first two moments and the correlation parameter uncertainty is less severe.

Alternatively, an investor can rely on the *naive* 1/N rule. The weights of the risky assets are all equal, and the expectations and the covariance matrix of excess returns are only needed to determine the optimal allocation of funds to the riskless asset and to the portfolio of risky assets. The portfolio weights are

$$w^{1/N} = \frac{1}{\gamma}\frac{\mathbf{1}_N'\mu}{\mathbf{1}_N'\Sigma\mathbf{1}_N}\,\mathbf{1}_N. \tag{8}$$

If the parameters of the return distributions are known, the naive strategy is suboptimal and only delivers the utility

$$U(w^{1/N}) = \frac{1}{2\gamma}\frac{(\mathbf{1}_N'\mu)^2}{\mathbf{1}_N'\Sigma\mathbf{1}_N}. \tag{9}$$

where $U(w^{1/N}) \leq U(w^*)$. However, the naive strategy can be superior to the mean-variance strategy if parameters have to be estimated and losses from parameter uncertainty are high. On the one hand, an investor following this strategy discards the opportunity to learn something from the data when deciding on the portfolio of risky assets. On the other hand, the portfolio of risky assets is not subject to parameter uncertainty. If the investor has to estimate the mean and the variance of the 1/N-portfolio, its expected utility is

$$E\big[U(\hat{w}^{1/N})\big] = \frac{T}{T-3}\Big(2-\frac{T}{T-5}\Big)\frac{1}{2\gamma}\frac{(\mathbf{1}_N'\mu)^2}{\mathbf{1}_N'\Sigma\mathbf{1}_N} - \frac{T}{2\gamma(T-3)(T-5)}. \tag{10}$$

Even if the 1/N rule cannot completely eliminate any risk related to parameter uncertainty, it considerably alleviates the problem compared to (7). DeMiguel et al. (2009) show that for a portfolio with 50 assets (calibrated to US stock market data), thousands of monthly observations are necessary for the sample-based mean-variance strategy to beat the naive rule.

The main aim of the present paper is to introduce a strategy that mitigates the utility losses caused by estimation risk, but that still partly exploits the information embodied in the data. We propose to optimize the portfolio granularity. For a high level of parameter uncertainty the benefits of 1/N outweigh the drawbacks of not optimizing the composition of the risky portfolio. For reasonable observation periods, however, grouping of similar assets (see Bjerring et al., 2016) and optimizing over the groups might be superior. We illustrate our idea by a simulation study in the next subsection, and by theoretical results as well as empirical back-tests in the following sections.

3 Grouping Using a Simulation Approach

We consider 30 risky assets ($N = 30$) whose returns follow a multivariate normal distribution with (unknown) expected returns μ and (unknown) covariance matrix Σ. The investor has to estimate μ and Σ based on $T = 480$ observations. In order to relate insights of this simulation study to the analytical results in the next section, we make a number of simplifying assumptions which are needed to keep tractability in Section 4. First, in our setting all asset returns have the same variance σ^2 and are uncorrelated.[3] Second, assets differ according to their expected returns, and we assume that they are equally spaced within an interval. We calibrate the volatility parameter σ and the interval for the expected return μ to match empirical moments from the 30 Fama-French industry portfolios.[4] Our sample period comprises 480 months from 1973M8–2013M7. We set σ equal to the average volatility of the return series of the 30 single industries. The upper and lower bound of the interval for μ are given by the best and worst average excess return.

We analyze three different strategies, the "plug-in" mean-variance rule using the full asset menu, the "plug-in" mean-variance rule using a reduced (grouped) asset menu with N/g groups (the g assets are equally weighted within each group), and finally the naive 1/N rule. In the simulation study, we group assets according to their similarities, i.e. their true (unknown) expected returns.[5]

Figure 1 shows the utility for all three strategies without parameter uncertainty (vertical solid lines), the expected utilities under parameter uncertainty (vertical dashed lines), and the distributions of utilities with parameter uncertainty (histograms). The group size is $g = 10$. In case of no parameter uncertainty, the utility follows from Equations (3) and (9). If the true parameters are unknown, the investor considers sample estimates, i.e. the expected utilities are given by (7) and (10). Furthermore, we rely on a simulation study to illustrate frequencies (not known in closed-form) of the utilities. We generate 50,000 samples of $T = 480$ monthly observations with $N = 30$ assets. For each sample we estimate the parameters, determine the optimal portfolio weight $\hat{w}$, and calculate the corresponding utility $U(\hat{w})$.

If the parameters are known, grouping reduces the utility since optimization is only applied across the groups, but not within each group. Therefore, the utility for the grouped assets (blue solid line) is smaller than the utility for the full asset menu (black solid line), and the 1/N strategy (red solid line) results in the smallest utility. If the investor has to estimate the parameters, she follows a suboptimal strategy, which implies that the utility under parameter certainty (solid lines) represents an upper bound for the utilities under parameter uncertainty (histograms). We interpret the difference between the utility without parameter uncertainty and with parameter uncertainty as a utility loss due to estimation risk. It depends on the length of the observation period, the number of assets, and the true

[3]Later, we relax the assumption of orthogonal asset returns and allow for a non-zero pairwise correlation.

[4]Available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[5]The grouping rule, although not implementable, allows us to illustrate the main drivers of our strategy. We postpone a discussion of possible group assignments for a real-world implementation to Section 5.

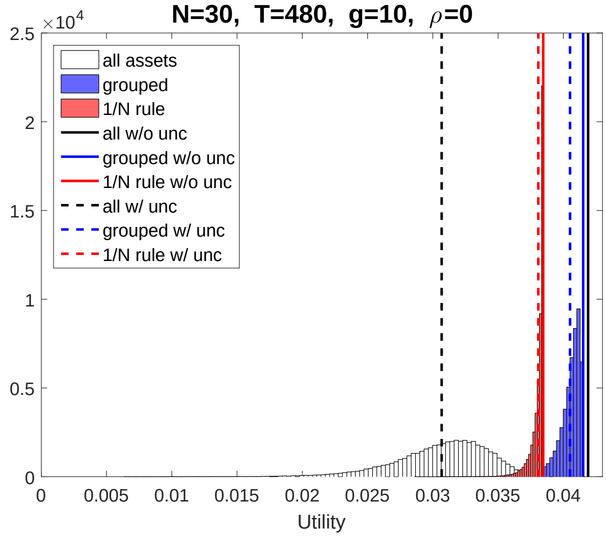

Figure 1: **Utility for a mean-variance investor** with the full asset portfolio (black), with grouping (blue) and using the 1/N rule (red). We show the utility for the strategy without parameter uncertainty (vertical solid lines), the expected utility under parameter uncertainty (vertical dashed lines), and the distributions of utilities with parameter uncertainty (histograms). The expected returns of the assets are equally-spaced. All assets have the same volatility, the returns are uncorrelated. The interval of expected returns and the average volatility are calibrated to monthly returns of the 30 Fama-French industry portfolios (1973M8-2013M7). The histograms are based on 50,000 simulations. Relative risk aversion is set to $\gamma = 5$.

parameters of the assets. In case of grouping, the utility loss (the difference between the blue solid line and a realization from the blue distribution) is much smaller than the loss in case of the full asset menu (the difference between the black solid line and a realization from the white distribution), and it is smallest for the 1/N strategy. In our example, the much smaller utility loss of the 1/N strategy from parameter uncertainty more than offsets its utility loss due to the foregone optimization, so that the expected utility is larger for the 1/N strategy than for optimization over the full asset menu. By a similar argument, the expected utility of the grouping strategy is also larger than that of the full strategy. Again, the huge reduction in the loss from estimation risk more than compensates the much smaller loss from suboptimality. The decisive point is the relative performance of the grouping strategy and the naive strategy. Here, the grouping strategy is superior to the naive strategy, since the slightly larger loss from parameter uncertainty is overcompensated by the benefits from optimizing over a larger asset menu.

Grouping can thus indeed increase the expected utility in case of parameter uncertainty. A natural question is then to find the optimal group size: how many stocks should be collected in a group? The investor has a large set of choices, going from $g = 1$, which corresponds to the optimization over the full asset menu, to $g = N$, which results in the 1/N strategy. In our example, the 1/N rule is able to beat the full asset menu strategy (in terms of the expected utility), but it is not superior to the grouping strategy with $g = 10$.

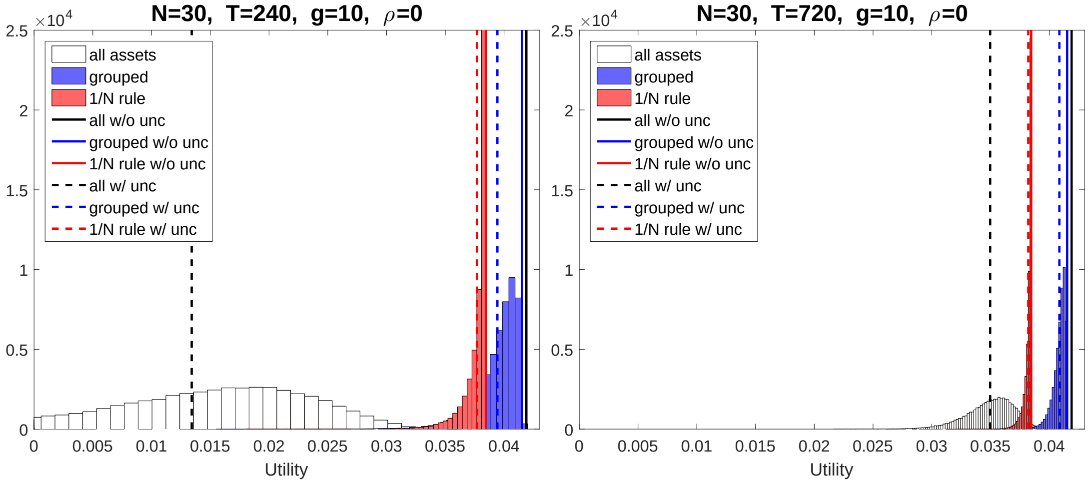

Figure 2: **Sensitivity with respect to the observation window T.** We assume 20 years of data (left) and 60 years of data (right). The graphs show the utility for a mean-variance investor with the full asset portfolio (black), with grouping (blue) and using the 1/N rule (red). We consider the utility for the strategy without parameter uncertainty (vertical solid lines), the expected utility under parameter uncertainty (vertical dashed lines), and the distributions of utilities with parameter uncertainty (histograms). The expected returns of the assets are equally-spaced. All assets have the same volatility, the returns are uncorrelated. The interval of expected returns and the average volatility are calibrated to returns of the 30 Fama-French industry portfolios (1973M8-2013M7). The histograms are based on 50,000 simulations. Relative risk aversion is set to $\gamma = 5$.

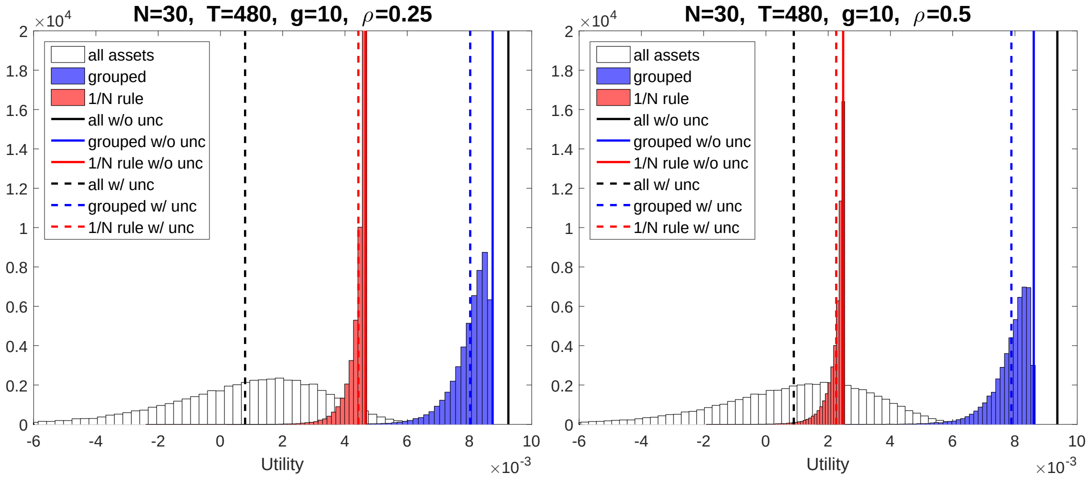

Figure 3: **Sensitivity with respect to the pairwise correlation coefficient ρ.** We assume a pairwise correlation $\rho = 0.25$ (left) and $\rho = 0.5$ (right). The graphs show the utility for a mean-variance investor with the full asset portfolio (black), with grouping (blue) and using the 1/N rule (red). We consider the utility for the strategy without parameter uncertainty (vertical solid lines), the expected utility under parameter uncertainty (vertical dashed lines), and the distributions of utilities with parameter uncertainty (histograms). The expected returns of the assets are equally-spaced. All assets have the same volatility. The interval of expected returns and the average volatility are calibrated to returns of the 30 Fama-French industry portfolios (1973M8-2013M7). The histograms are based on 50,000 simulations. Relative risk aversion is set to $\gamma = 5$.

In order to investigate the impact of different parameter choices on our results, we conduct a sensitivity analysis. First, to investigate the impact of different observation periods, Figure 2 shows utilities for 20 years (left) and 60 years of monthly data (right).[6] While the expected utilities without parameter uncertainty stay the same, the impact of estimation risk changes. For the shorter observation period there is more parameter uncertainty. Thus, strategies which try to optimize the composition of the risky asset portfolio suffer from higher utility losses and their performance worsens relative to the 1/N rule.[7] In terms of expected utility, the grouping strategy with $g = 10$ now performs more similarly to the 1/N strategy. With 60 years of data the moment estimates are much more reliable, so that the 1/N rule is clearly inferior to the grouping strategy in terms of expected utility. However, 60 years of data are still not enough for the full-asset menu strategy to outperform the 1/N rule (in line with the findings of DeMiguel et al., 2009).

Second, we relax the assumption of zero correlation and allow for a positive pairwise correlation, where we consider $\rho = 0.25$ (Figure 3, left) and $\rho = 0.5$ (Figure 3, right) while keeping all other parameters equal. Again, the grouping strategy generates the highest expected utility compared to the two other strategies. Without parameter risk (solid vertical lines), the losses from not optimizing over the full asset menu increase in

[6] For comparison reasons we use the same scaling of the axis in Figures 1, 2 and 4.

[7] In the left graph of Figure 2, the utility for optimization over the full asset menu can become negative. In these cases, the investor would be better off if he just invested in the risk-free asset.

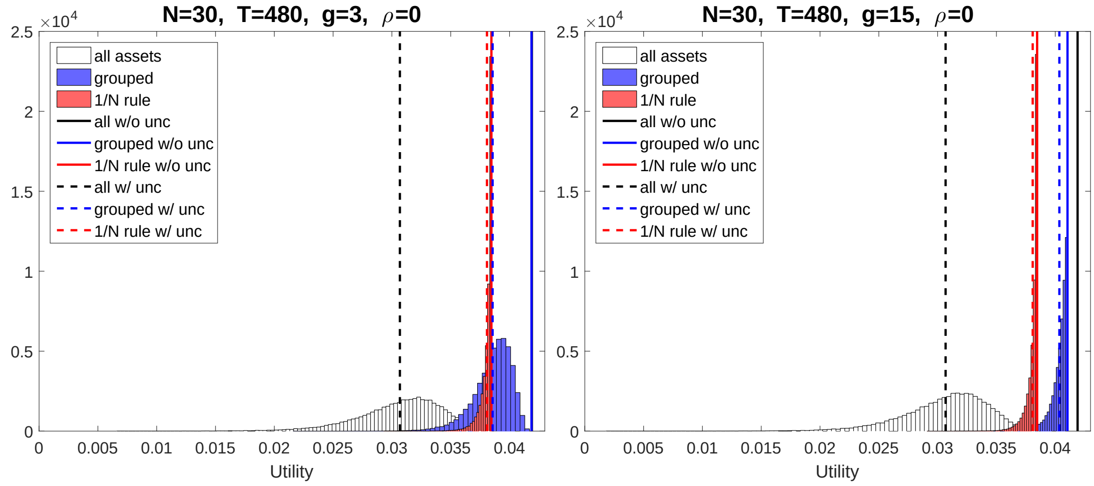

Figure 4: **Sensitivity with respect to the group size g.** We assume a group size $g = 3$ (left) and $g = 15$ (right). The graphs show the utility for a mean-variance investor with the full asset portfolio (black), with grouping (blue) and using the 1/N rule (red). We consider the utility for the strategy without parameter uncertainty (vertical solid lines), the expected utility under parameter uncertainty (vertical dashed lines), and the distributions of utilities with parameter uncertainty (histograms). The expected returns of the assets are equally-spaced. All assets have the same volatility. The interval of expected returns and the average volatility are calibrated to returns of the 30 Fama-French industry portfolios (1973M8-2013M7). The histograms are based on 50,000 simulations. Relative risk aversion is set to $\gamma = 5$.

the correlation. At the same time, the average losses from estimation risk stay more or less the same. In consequence, the difference between the plug-in rule based on the full asset menu and the 1/N rule declines for larger correlations. Ceteris paribus, a higher positive correlation makes assets more heterogeneous. Intuitively, if two assets differ only in expected returns, then the higher the correlation between the assets the more extreme the optimal portfolio weights (long in the asset with the higher and short in the asset with the lower expected return). Forcing weights to be equal to each other is thus more restrictive for large correlations.

Finally, we investigate different group sizes for the grouping strategy. Figure 4 illustrates the expected utility for smaller groups with $g = 3$ (left) and larger groups with $g = 15$ (right). For a group size of 3 assets, the strategy outperforms the full-asset-menu strategy, but is only marginally superior to the 1/N rule. The benefits from the optimization do not outweigh the losses caused by parameter uncertainty. In this case, a larger group size is more desirable. On the other hand, forming too big groups might not be optimal either. If two groups with $g = 15$ assets (right) are created, more potential optimization gains are lost. Although the grouping strategy still outperforms both the full-asset-menu strategy and the 1/N rule, the results are slightly worse than for $g = 10$ (see Figure 1). This suggests a hump-shaped pattern of the expected utility as a function of the group size, and thus an optimal size of the groups strictly between 1 and N.

The previous examples provide the basic intuition for the benefits and losses of optimization if there is estimation risk. Instead of relying on two extremes, either the mean-variance optimization based on the full asset menu or the 1/N rule, we can take a route in between and create groups of assets. The optimal portfolio granularity depends on the uncertainty in asset moments and the heterogeneity of single assets.

4 Analytical Results for the Optimal Granularity

In this section we derive the optimal granularity of the asset menu in a simplified setting. As illustrated in our simulation study, on the one hand it might not be optimal to solve the mean-variance problem over the full asset menu due to parameter estimation risk. On the other hand, using the 1/N rule that does not exploit any information from the data to decide on the composition of the portfolio of risky assets can be suboptimal, too. We argue that optimizing over groups composed of equally-weighted assets might improve the performance since it provides an optimal balance between the benefits from optimization (which increase in the number of assets) and the losses due to estimation risk (which also increase in the number of assets). In the following, we determine the optimal group size in closed-form for the case where only the expected returns are unknown.[8] The omission of more complicated cases is due to tractability.[9]

The expected utility with unknown μ and known Σ for a sample-based mean-variance investor is given by (see Kan and Zhou, 2007):

$$E\big[U(\hat{w})|\Sigma\big] = \frac{\theta^2}{2\gamma} - \frac{N}{2\gamma T}, \tag{11}$$

where $\theta^2 = \mu'\Sigma^{-1}\mu$ is the squared Sharpe ratio. Instead of optimizing over the whole asset menu, the investor can reduce the number of assets from N to $\bar{N}$, with $\bar{N} = N/g$.[10] In general, the reduction is beneficial in terms of expected utility if

$$\frac{\theta^2}{2\gamma} - \frac{N}{2\gamma T} \leq \frac{\bar{\theta}^2}{2\gamma} - \frac{\bar{N}}{2\gamma T}, \tag{12}$$

$$\theta^2 - \bar{\theta}^2 \leq \frac{N - \bar{N}}{T}. \tag{13}$$

Note that the relative performance of the two strategies is independent of the level of risk aversion γ. The formation of groups reduces the number of parameters to estimate and thus reduces the loss from estimation risk, see right-hand side of Equation (13). If this reduction is larger than the loss in the Sharpe ratio from giving up optimization potential, see left-hand side of Equation (13), grouping is superior to optimizing over the full asset menu.

First, we assume that all assets have the same expected return μ, the same volatility σ

[8] In line with Garlappi et al. (2007) we focus on parameter uncertainty in the expected returns of assets because as shown by Merton (1980) they are much harder to estimate than the variances and covariances. Moreover, Chopra and Ziemba (1993) shows that errors in estimating expected returns are over 10 times as costly as errors in estimating variances, and over 20 times as costly as errors in estimating covariances.

[9] We perform an empirical study in Section 5 to show that our grouping strategy also works in practice.

[10] We assume that the number of assets g in each group is the same, i.e. $N \bmod g = 0$.

and are orthogonal to each other, i.e. the covariance matrix Σ is diagonal. Then (13) can be rewritten as

$$N\frac{\mu^2}{\sigma^2} - \bar{N}g\frac{\mu^2}{\sigma^2} \leq \frac{N-\bar{N}}{T}. \tag{14}$$

Since $\bar{N} = N/g$, this gives

$$N\frac{\mu^2}{\sigma^2} - N\frac{\mu^2}{\sigma^2} \leq \frac{N-\bar{N}}{T}. \tag{15}$$

Given the chosen setting, grouping has no impact on the Sharpe ratio, but it reduces estimation risk. Thus, equal-weighted grouping is always beneficial, and the optimal solution is to choose the maximal group size $g = N$, that is the $1/N$ strategy. Here, the use of the $1/N$ strategy restricts the optimal risky portfolio in case of estimation risk to the optimal portfolio without estimation risk. Only for $T \to \infty$ the expected utility is the same for the grouping strategies and for optimization over the full asset menu.

Next, we consider the more interesting case of heterogeneous assets. Here, grouping imposes a suboptimal asset allocation due to equal weights within a group. This allows us to study the trade-off between a lower Sharpe ratio induced by grouping and benefits from less parameters to estimate.

To account for heterogeneity in assets and in line with our simulation study in Section 3, we assume that the expected returns are equally-spaced, $\{\mu_i; i = 1, \ldots, N, \mu_i = \mu + s(i-1)\}$. All other assumptions are the same as before. Of course, this extended setting nests the special case that all assets have equal expected returns (for $s = 0$). The higher s, the more heterogeneous the optimal asset allocation. Intuitively, with increasing s optimization becomes relatively better compared to grouping and the 1/N rule. In the following, assets are grouped with respect to their expected return.[11]

Proposition 1. *Assume that expected returns are equally spaced with $\mu_i = \mu + (i-1)s$ ($i = 1, \ldots, N$), all assets have the same volatility σ, and are uncorrelated. Expected returns are unknown, the covariance matrix is known. Then, the expected utility for a grouping strategy with group size g is given by*

$$E\big[\tilde{U}(\hat{w})|\Sigma\big] = \frac{\bar{\theta}_1^2}{2\gamma} - \frac{N}{2\gamma g T}, \tag{16}$$

where

$$\bar{\theta}_1^2 = \frac{N}{\sigma^2}\left[\left(\frac{2\mu + s(N-1)}{2}\right)^2 + s^2\frac{(N-g)(N+g)}{12}\right]. \tag{17}$$

The optimal group size is

$$g_1^* = \sqrt[3]{\frac{6\sigma^2}{s^2 T}}. \tag{18}$$

The expected utility increases in the optimal squared Sharpe ratio, $\bar{\theta}_1^2$, and decreases due the estimation risk caused by unknown expected returns. While the first term in the squared brackets of the squared Sharpe ratio (17) depends on the average expected return only, the second term captures the benefits from adjusting the portfolio to differences in

[11] The implicit assumption here is of course that the decision maker can order the assets according to their expected returns.

returns. Since we apply the same weight to all assets in one group, i.e. weights within a group are independent of differences in the asset moments, the second term becomes smaller with a larger group size g. The Sharpe ratio thus decreases in g due to foregone optimization potential. The decisive question is whether this loss is compensated by a decline in estimation risk. The loss from estimation risk decreases in the group size, so that a larger g leads to a higher expected utility. The optimal group size follows from the trade-off between a lower Sharpe ratio and less estimation risk. In line with economic intuition, the optimal group size g_1^* is decreasing in the number of observations T, decreasing in the differences of the expected returns s, and increasing in the asset volatility σ. Note that both a smaller s and a larger σ reduce the heterogeneity of the assets.[12] The optimal portfolio without estimation risk is thus closer to $1/N$, and the utility gains from optimizing decrease. The losses from a larger group size are thus smaller, while the reduction in the losses from estimation risk stays the same, which implies a larger optimal group size.

By rearranging (18) it follows that the 1/N rule is optimal for

$$\frac{6\sigma^2}{s^2N^3} \geq T,$$

while the use of the full universe is recommended if

$$\frac{6\sigma^2}{s^2} \leq T.$$

We illustrate these thresholds in Figure 5. For the full asset menu of 50 assets, the 1/N rule is optimal if we use less than 10 monthly observations for estimation, and the mean-variance optimization over the full asset menu is optimal for more than 100 000 monthly observations. Most important, for a wide range of reasonably long observation windows, the grouping strategy is superior to the other two strategies.

We illustrate the relevance of our grouping strategy by the two following figures. First, we compare the grouping strategy to optimization over the full asset menu. Figure 6 shows the benefits and losses from grouping as a function of the number of observations T. From Equations (16) and (17), we get that the loss from a lower squared Sharpe ratio equals

$$\theta^2 - \bar{\theta}_1^2 \quad = \quad \frac{N(g^2-1)s^2}{12\sigma^2} \tag{19}$$

and the benefit from less estimation risk is

$$\frac{N-\bar{N}}{T} \quad = \quad \frac{N(1-1/g)}{T}. \tag{20}$$

If the benefit from less parameter uncertainty is higher than the loss from the lower Sharpe ratio, then grouping is beneficial. The difference of the squared Sharpe ratios in (19), which corresponds to the benefits of optimization over the full set of assets relative to the set of grouped assets, does not vary with T. The benefits from less estimation risk (20) decrease

[12] Given the assumptions in Proposition 1, the optimal portfolio weights for optimization over the full asset menu are equally spaced. The difference between the smallest and the largest portfolio weight is $\frac{(N-1)s}{\gamma\sigma^2} = \frac{\mu_N-\mu_1}{\gamma\sigma^2}$.

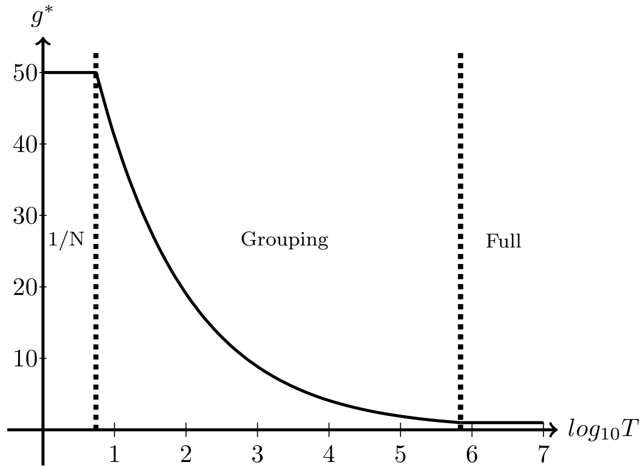

Figure 5: **Optimal group size.** The figure shows the optimal group size as a function of the sample length T (in months). The expected returns of the assets are equally-spaced. All assets have the same volatility, the returns are uncorrelated. We use $N = 50$, $\mu_N - \mu = 0.1/12$, and $\sigma^2 = 0.04/12$. Relative risk aversion is set to $\gamma = 5$. The x-axis is shown in a log scale.

with the number of observations as estimates get more precise. Therefore, in particular for a small number of observations T the reduction in estimation risk due to grouping can be substantial. The intersection of both lines gives the minimum number of observations for which optimization based on the full sample outperforms optimization based on the groups with size g. In Figure 6 we show asset menus with 50 and 100 assets, and compare the optimization using the full universe with the optimization using groups of assets with size $g = 5$ and $g = 10$. In all cases thousands of monthly observations are needed for the mean-variance strategy applied to the full menu of assets to beat the grouping strategy. If less observations are available, the benefits from grouping outweigh the losses from optimizing over the reduced asset menu only.

Second, Figure 7 reports the expected utility as a function of the group size when the number of assets is fixed. It illustrates the benefits of the grouping strategy as compared to the two extreme strategies: the mean-variance strategy over the full asset menu, which corresponds to $g = 1$, and the naive 1/N rule, which corresponds to $g = N$. For 50 assets (left graph), the optimal group size is around 14 for a sample length of $T = 240$ months and drops to around 8 for a sample length of $T = 1200$ months. For the small sample, the $1/N$-strategy is superior to the optimization over the full asset menu, but it is still beaten by the optimal grouping strategy. For the long sample, the mean-variance strategy is better than the naive strategy. Nevertheless, the grouping strategy is able to improve on the better of the two strategies again, even if the utility gain is smaller than for the small sample size. For $N = 100$ assets, the overall ranking of the strategies is basically the same. The optimal group size is now around 22 for the small sample and around 13 for the large sample. Doubling the number of assets (and keeping the maximal difference of expected

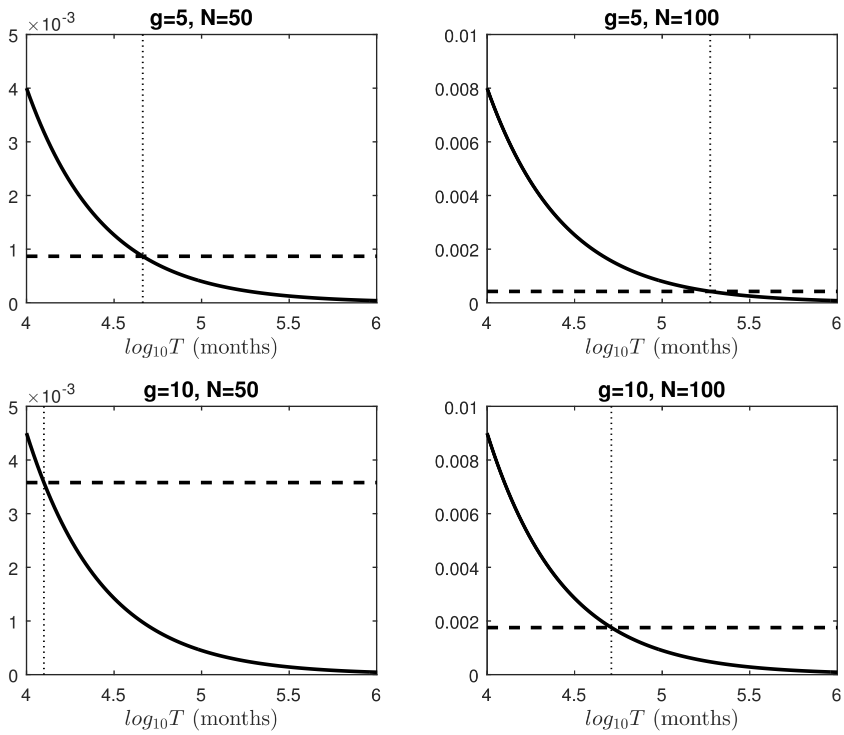

Figure 6: **Benefits and losses from the grouping strategy.** The solid line corresponds to the benefits from less estimation risk, while the horizontal dashed line gives the losses from a smaller Sharpe ratio. The intersection of both lines is the minimum number of observations for which optimization over the full asset menu outperforms optimization over the groups with size g. The expected returns of the assets are equally-spaced. All assets have the same volatility, the returns are uncorrelated. We set $\mu_N - \mu = 0.1/12$ and $\sigma^2 = 0.04/12$. The x-axes are shown in a log scale.

returns constant) thus less than doubles the optimal group size, and the number of groups increases.[13]

Finally, we relax the assumption of zero correlation. We assume that there is a constant correlation ρ for all pairs of assets (see e.g. Boyle et al., 2012). This non-zero correlation modifies the Sharpe ratio, but not the loss due to estimation risk. Therefore, it also influences the optimal group size. We summarize its effects in the following proposition.

Proposition 2. *Assume that expected returns are equally spaced with $\mu_i = \mu + (i-1)s$ ($i = 1, \ldots, N$), all assets have the same volatility σ, and the pairwise correlation of returns is ρ. Expected returns are unknown, the covariance matrix is known. Then, the expected*

[13]If we fix $\mu_N - \mu$ and set $s = \frac{\mu_N - \mu}{N-1}$, the optimal group size is

$$g_1^* = \sqrt[3]{\frac{6\sigma^2(N-1)^2}{(\mu_N-\mu)^2 T}}.$$

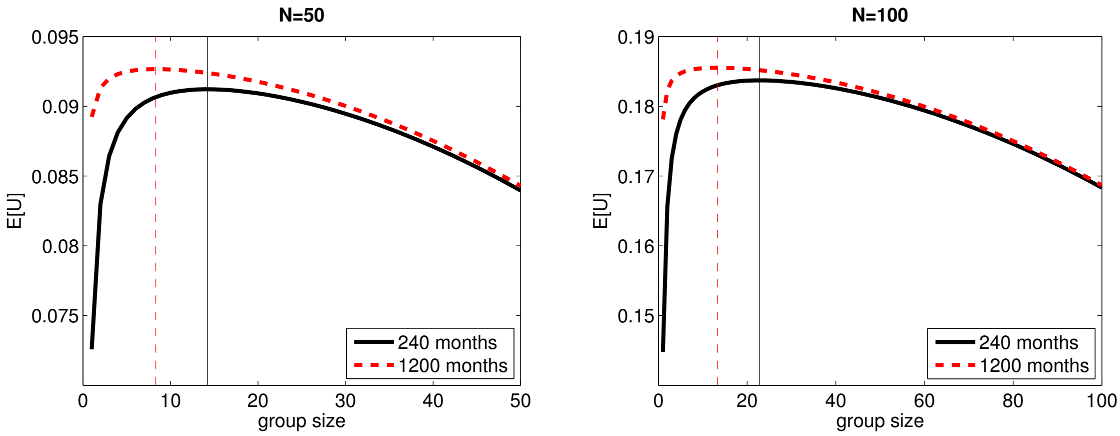

Figure 7: **Expected utility for different group sizes.** The graphs show the expected utility for $N = 50$ (left) and $N = 100$ (right) assets. The sample length is $T = 240$ months (black solid line) and $T = 1200$ months (red dashed line), respectively. The resulting optimal group size g^* is indicated by the vertical lines. The expected returns of the assets are equally-spaced. All assets have the same volatility, the returns are uncorrelated. We set $\mu_N - \mu = 0.1/12$, and $\sigma^2 = 0.04/12$. Relative risk aversion is set to $\gamma = 5$.

utility for a grouping strategy with group size g is given by

$$E\left[\tilde{U}(\hat{w})|\Sigma\right] = \frac{\bar{\theta}_2^2}{2\gamma} - \frac{N}{2\gamma g T}, \tag{21}$$

where

$$\bar{\theta}_2^2 = \frac{N}{\sigma^2}\left[\frac{1}{1+\rho(N-1)}\left(\frac{2\mu + s(N-1)}{2}\right)^2 + s^2\frac{(N-g)(N+g)}{12(1-\rho)}\right]. \tag{22}$$

The optimal group size is

$$g_2^* = \sqrt[3]{\frac{6\sigma^2(1-\rho)}{s^2 T}}. \tag{23}$$

As expected, a non-zero correlation of the assets does not change the utility loss due to estimation risk, i.e. the second term of (21). However, this is not true for the Sharpe ratio in (22). The first term in square brackets depends on the average expected return, and the impact of the correlation on this term is the same for all group sizes g. The impact of the heterogeneity of groups, indicated by the second term in square brackets, however, depends on both g and the correlation coefficient. The drop in Sharpe ratio from grouping is

$$\frac{Ns^2(g^2-1)}{12\sigma^2(1-\rho)},$$

and this drop increases in correlation ρ. The larger losses from foregone optimization and the constant benefits from less estimation risk lead to a smaller optimal group size g_2^*. It follows from (23) that the 1/N rule is optimal for

$$\frac{6(1-\rho)\sigma^2}{s^2N^3} \geq T,$$

while optimization over the full asset menu is optimal if

$$\frac{6(1-\rho)\sigma^2}{s^2} \leq T.$$

Both bounds decrease in correlation ρ.

Figure 8 illustrates the sensitivity of the optimal group size with respect to an increase in the correlation coefficient. We plot the expected utility for different group sizes and different levels of pairwise correlations. An increase in correlation from 0 (solid line) to 0.5 (dashed line) reduces the optimal group size by around 20% for both observation windows.

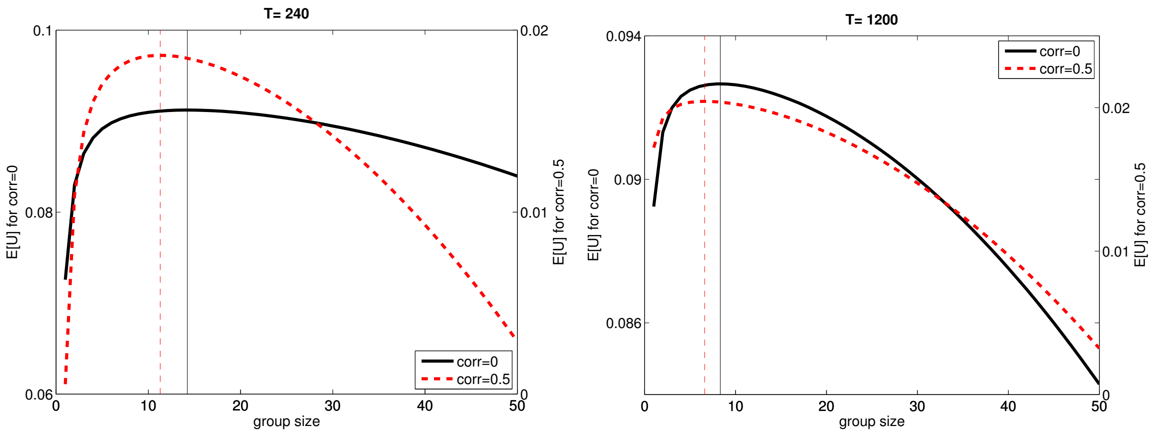

Figure 8: **Expected utility for different assets with equal pairwise correlation.** The graphs show the expected utility for a sample of $T = 240$ months (left) and $T = 1200$ months (right), respectively. The pairwise correlation of the returns is $\rho = 0$ (black solid line) and $\rho = 0.5$ (red dashed line), respectively. The resulting optimal group size g^* is indicated by the vertical lines. The expected returns of the assets are equally-spaced. All assets have the same volatility. We set $N = 50$, $\mu_N - \mu = 0.1/12$, and $\sigma^2 = 0.04/12$. Relative risk aversion is set to $\gamma = 5$.

In this section we have derived the benefits from optimal granularity within a simplified setup. Although quite strong assumptions have to be imposed to keep analytical tractability, they enable us to describe the main drivers in the trade-off between the optimization potential and estimation risk. In the next section we demonstrate that the optimal granularity strategy is also beneficial when applied to real data.

5 Out-of-sample Evaluation

The results in Section 3 and 4 are based on the assumptions that (a) parameters are constant over time, and (b) assets differ with respect to their expected returns only, which is therefore used as the discrimination criterion to group them. These assumptions will most likely not hold in real markets.

In this section we use different empirical data sets to analyze the impact of the group size and the length of the observation period on expected utility. We find that the optimal group size is indeed often in between the full asset universe ($g = 1$) and the naive 1/N rule ($g = N$). We also show that the group size which gives the highest out-of-sample expected utility in the empirical analysis is similar to the theoretically optimal group size implied by our simplified model.

First, we use the 30 industry portfolios of Fama/French as full asset universe, which provides a fair compromise between the length of available data (1926M7 – 2013M7) and the number of different equally sized groups which can be formed. As observation periods we consider 5, 10, 25, and 50 years of monthly data, i.e. $T \in \{60, 120, 300, 600\}$, and a relative risk aversion $\gamma = 5$. In order to ensure a level playing field when comparing different observation periods we consider out-of-sample returns from July 1976 onward.[14]

Given that the true expected return is unknown in an out-of-sample exercise, and given that assets might also differ with respect to variances and pairwise correlations, we use a one-factor approach to group the assets. We regress the return of each industry portfolio on the market return and group the portfolios according to their betas.[15] The assumption is that portfolios with a higher beta have a higher expected return. For each month in the out-of-sample period from 1976M7–2013M7, we first estimate the moments from the preceding T observations and calculate the portfolio weights for different choices of the group size g. Out-of-sample portfolio returns are then calculated with the realized assets returns one months later, when the exercise is repeated. The mean and variance of the time series of out-of-sample portfolio returns are then used to calculate the utility for the different observation period/granularity strategies.

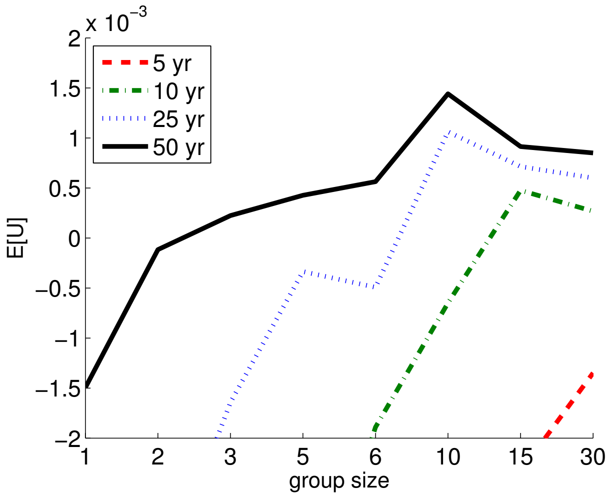

Figure 9: **Expected utility for different combinations of group size g and observation periods.** The full asset menu is given by the 30 Fama-French industry portfolios. The relative risk aversion is set to $\gamma = 5$.

[14]The observation period 1926M7–1976M6 is necessary to estimate the parameters for our longest observation period.

[15]The success of the grouping strategy will depend on the ability of the investor to group the assets by their (unknown) true optimal portfolio weights. We do not claim that our grouping strategy meets this criterion, but rather show that even a (maybe) suboptimal grouping rule can lead to utility gains.

Figure 9 illustrates our main results. First, losses due to estimation risk may be so severe that the expected utility turns negative, and the investor is better off if he just holds the risk-free asset. Second, in line with the model predictions in Section 4, the best results in terms of expected utility are achieved for the very long observation periods. This is consistent with the recommendation of Kritzman et al. (2010), who argue that most back-test studies in academic papers use too short time intervals. Third, even for very long observation periods (50 years) optimization over the full universe ($g = 1$) is significantly worse than the 1/N rule ($g = 30$) in terms of expected utility. Fourth, in line with our analytical results, the optimal group size decreases in the length of the observation period. While for very short periods the 1/N rule (i.e. $g = N$) is best, for 10 years the optimal group size of is 15, and for 25 years as well as 50 years the optimal group size is 10.

The graph also illustrates the expected utility losses from parameter uncertainty. These losses are the larger the shorter the sample period. For the 1/N strategy sample estimates are only used to decide on how to allocate funds to the risk-free asset and the equally weighted portfolio of risky assets, and utility losses are small. For smaller group sizes, sample estimates are also used to optimize the composition of the portfolio of risky assets. The impact of estimation risk increases, and utility losses become larger.

Similarly, in Figure 10 we consider the 10 (left) and 12 (right) Fama-French industry portfolios and give the expected utility for different group sizes and observation periods. Again, for observation periods longer than five years we find a hump-shaped pattern in expected utility. An intermediate group size is superior both to the naive 1/N strategy and to the mean-variance strategy based on the full asset menu.

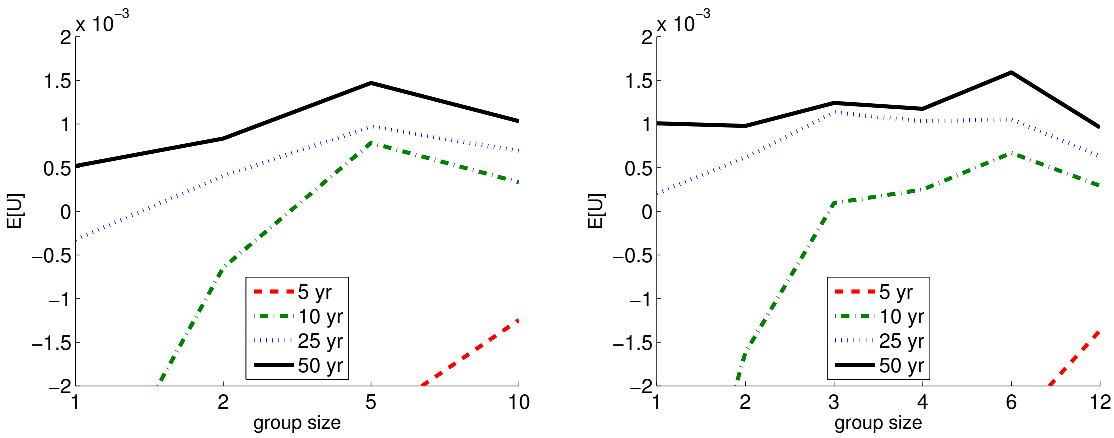

Figure 10: **Expected utility for different combinations of group size g and observation period.** The full asset menu is given by the 10 Fama-French industry portfolios (left graph) and the 12 Fama-French industry portfolios (right graph), respectively. The relative risk aversion is set to $\gamma = 5$.

We also compare the optimal group size in the empirical analysis to the theoretical predictions of our simplified model. Table 1 reports the theoretically optimal group size for all cases considered in Figures 9 and 10, where the parameters are calibrated to match μ

and Σ in the data.[16] The group size which gives the highest out-of-sample expected utility often coincides with the theoretical optimal values. In some cases, in particular for the short observation period of 5 years, our theoretical results underestimate the optimal group size. In reality parameters are time-varying, and volatilities as well as pairwise correlations differ among the assets. The omission of all these issues in our analytical results might cause "overconfidence" in the optimization approach, and therefore imply a too fine granularity. Thus, our analytical results might be considered as lower bounds for the optimal group size.

Table 1: Theoretical optimal group size

The table reports the optimal group size based on the theoretical model in Section 4. The interval for expected returns, the average volatility and correlation are calibrated based on the first 50 years of data (when the out-of-sample evaluation begins). The optimal group size is calculated with (23).

Observation window T	5 years	10 years	25 years	50 years
30 FF industry portfolios	15	10	10	6
12 FF industry portfolios	6	6	4	4
10 FF industry portfolios	5	5	5	5

Overall, our results demonstrate that there is a trade-off between losses due to parameter uncertainty and benefits from optimization, and that investors might benefit from grouping assets and then optimizing over the grouped portfolios. We show that the optimal grouping strategy is in many cases superior to the naive 1/N rule and to optimization based on the full universe.

6 Conclusions

We suggest a novel portfolio rule to deal with parameter uncertainty in the mean-variance framework which combines the benefits from both the naive 1/N rule and the optimization over the full asset menu. The 1/N strategy minimizes losses from parameter uncertainty, but at the same time foregoes benefits from optimizing the portfolio weights. Due to high parameter uncertainty, the 1/N rule can be superior to optimization over the full asset universe. We argue that the trade-off between losses due to estimation risk and optimization benefits leads to an optimal portfolio granularity. For the grouping strategy estimation risk is reduced by forming groups with an optimal size (implementing the 1/N rule within each group) while optimization benefits are still exploited over the groups of assets. We theoretically derive the optimal group size in a simplified setup and show that the grouping strategy can indeed be superior to optimization over the full asset menu and to the 1/N rule.

[16]We assume that expected returns are equally spaced and we set the lowest and highest return equal to the smallest and largest average return of the 30 industry portfolios. The volatility is set equal to the average volatility of all portfolios, the correlation is set equal to the average of all pairwise correlations. For more details see Section 3.

Our results are confirmed in an empirical study in which we apply the grouping strategy to different sets of Fama-French industry portfolios. The poor performance of the “plug-in” mean-variance portfolio rule can often be improved by switching to the grouping strategy, which in our examples is able to beat also 1/N. Put together, our approach optimally combines the benefits of the mean-variance portfolio rule and the naive 1/N in order to improve the out-of-sample performance under parameter uncertainty.

Appendix: Proofs of propositions

PROOF OF PROPOSITION 1. By grouping N assets with $\bar{N}g = N$, the expected returns of the equally weighted group i is given by

$$\bar{\mu}_i = \frac{\mu + sg(i-1) + \mu + s(ig-1)}{2} = \mu + \frac{s((2i-1)g-1)}{2}. \tag{24}$$

For the grouped asset menu we have

$${\theta_1}^2 = \frac{g}{\sigma^2}\left[\bar{\mu}_1^2 + \bar{\mu}_2^2 + \cdots + \bar{\mu}_{\bar{N}}^2\right]. \tag{25}$$

By recognizing that the sum of the first $\bar{N}$ odd squared numbers is

$$\sum_{i=1}^{\bar{N}}(2\bar{N}-1)^2 = \frac{\bar{N}(2\bar{N}-1)(2\bar{N}+1)}{3}$$

and that the sum of the first $\bar{N}$ odd numbers is

$$\sum_{i=1}^{\bar{N}}(2\bar{N}-1) = \bar{N}^2,$$

the squared Sharpe ratio $\bar{\theta}_1^2$ is given by

$$\bar{\theta}_1^2 = \frac{g}{\sigma^2}\left[\bar{N}\mu^2 + \mu s(\bar{N}g-1)\bar{N} + \frac{s^2}{4}\left(g^2\frac{\bar{N}(2\bar{N}-1)(2\bar{N}+1)}{3} - 2g\bar{N}^2 + \bar{N}\right)\right] \tag{26}$$

$$= \frac{g\bar{N}}{\sigma^2}\left[\mu^2 + \mu s(\bar{N}g-1) + \frac{s^2}{4}\left(g^2\frac{(2\bar{N}-1)(2\bar{N}+1)}{3} - 2g\bar{N} + 1\right)\right]. \tag{27}$$

Substituting $\bar{N}$ with N/g in (27) we get

$$\bar{\theta}_1^2 = \frac{N}{\sigma^2}\left[\mu^2 + \mu s(N-1) + \frac{s^2}{4}\left(\frac{g^2}{3}\left(\frac{2N}{g}-1\right)\left(\frac{2N}{g}+1\right) - 2N + 1\right)\right]. \tag{28}$$

Further simplifications give (17) in Proposition 1.

To find an optimal group size we take the first derivative of the expected utility

$$\frac{\partial E[U(\hat{w})|\Sigma]}{\partial g} = \frac{N}{2\gamma}\left(-\frac{s^2 g}{6\sigma^2} + \frac{1}{Tg^2}\right), \tag{29}$$

where the first term on the right-hand side indicates the marginal drawback of increasing the equally weighted assets groups (the disadvantage is proportional to the group size), while the second term quantifies the marginal benefit by increasing the group size. By setting the first derivative equal to zero and solving for g we get (for $g > 0$)

$$g^* = \sqrt[3]{\frac{6\sigma^2}{s^2 T}}. \tag{30}$$

□

PROOF OF PROPOSITION 2. First, we derive the value of $\bar{\theta}_2^2$ for the group size g and correlation ρ among individual assets. The expected return $\bar{\mu}_i$ for group i is the same as in the case of zero correlation and is given by (24). The covariance matrix consists of the variance $\bar{\sigma}^2$ and the same covariance $\bar{\sigma}_{ij}$ for each pair of grouped assets

$$\bar{\sigma}^2 = \frac{\sigma^2}{g}(1+(g-1)\rho), \tag{31}$$

$$\bar{\sigma}_{ij} = \rho\sigma^2. \tag{32}$$

Note that the correlation among groups of assets is thus

$$\bar{\rho}_{ij} = \frac{g\rho}{1+(g-1)\rho}. \tag{33}$$

The diagonal elements in the inverse of the covariance matrix are given by

$$\bar{a} = \frac{[1+\rho(N-g-1)]g}{(1-\rho)\sigma^2(1+\rho(N-1))} \tag{34}$$

and the off-diagonal elements

$$\bar{b} = \frac{-\rho g^2}{(1-\rho)\sigma^2(1+\rho(N-1))} \tag{35}$$

The squared Sharpe ratio $\bar{\theta}^2$ consists of two parts: an effect from the diagonal elements and an effect of the non-diagonal elements of the covariance matrix's inverse. The first part can be constructed in a similar way as in the previous example with non-zero correlation (27) and is equal to

$$\bar{a}D = \bar{a}\left[\bar{N}\mu^2 + \mu s(\bar{N}g-1)\bar{N} + \frac{s^2}{4}\left(g^2\frac{\bar{N}(2\bar{N}-1)(2\bar{N}+1)}{3} - 2g\bar{N}^2 + \bar{N}\right)\right]. \tag{36}$$

The second part, related to the off-diagonal elements is given by

$$\begin{aligned} \bar{b}O &= \bar{b}\Big[\bar{N}(\bar{N}-1)\mu^2 + [(\bar{N}^3-\bar{N}^2)g - \bar{N}(\bar{N}-1)]\mu s \\ &+ \frac{1}{12}[g^2\bar{N}(3\bar{N}^3-4\bar{N}^2+1) + 6g\bar{N}^2(1-\bar{N}) + 3\bar{N}(\bar{N}-1)]s^2\Big]. \end{aligned} \tag{37}$$

Finally, we can calculate the squared Sharpe ratio, for grouped assets with nonzero correlation

$$\begin{aligned} \bar{\theta}_2^2 &= \bar{a}D + \bar{b}O \\ &= N\frac{\mu^2+\mu(N-1)s}{(1+\rho(N-1))\sigma^2} \\ &+ N\frac{(-3+g^2+6N-4N^2+\rho(N-1)(-3+g^2-(N-3)N))s^2}{12(\rho-1)(1+\rho(N-1))\sigma^2}. \end{aligned} \tag{38}$$

Further simplification gives

$$\bar{\theta}_2^2 \quad = \quad \frac{N}{\sigma^2}\left[\frac{1}{1+\rho(N-1)}\left(\frac{\mu+\mu+s(N-1)}{2}\right)^2 + s^2\frac{(N-g)(N+g)}{12(1-\rho)}\right]. \tag{39}$$

To decide the optimal grouping size we calculate the first derivative of the expected utility given in (21)

$$\frac{\partial E[U(\hat{w})|\Sigma]}{\partial g} = \frac{N}{2\gamma}\left(\frac{-gs^2}{6(1-\rho)\sigma^2} + \frac{1}{Tg^2}\right), \tag{40}$$

where the first term on the right-hand side indicates the marginal drawback of increasing the equally weighted assets groups (the disadvantage is proportional to the group size), while the second term quantifies the marginal benefit by increasing the group size. By setting this first derivative equal to zero and solving for g we get (for $g > 0$)

$$g^* = \sqrt[3]{\frac{6\sigma^2(1-\rho)}{s^2 T}}. \tag{41}$$

□

References

Bjerring, T., Ross, O., and Weissensteiner, A. (2016). Feature selection for portfolio optimization. *Annals of Operations Research*. forthcoming.

Boyle, P., Garlappi, L., Uppal, R., and Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management Science*, 58(2):253–272.

Chan, L. K. C., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, 12(5):937–974.

Chen, J. and Yuan, M. (2016). Efficient portfolio selection in a large market. *Journal of Financial Econometrics*. forthcoming.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.

Duchin, R. and Levy, H. (2009). Markowitz versus the Talmudic portfolio diversification strategies. *The Journal of Portfolio Management*, 35(2):71–74.

Frost, P. A. and Savarino, J. E. (1988). For better performance. *The Journal of Portfolio Management*, 15(1):29–34.

Garlappi, L., Uppal, R., and Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81.

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.

Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, 47(5):1785–809.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Kourtis, A., Dotsis, G., and Markellos, R. (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking and Finance*, 36:2522–2531.

Kritzman, M., Page, S., and Turkington, D. (2010). In defense of optimization: The fallacy of 1/N. *Financial Analysts Journal*, 66:31–39.

Ledoit, O. and Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2004b). Honey, I shrunk the sample covariance matrix: Problems in mean-variance optimization. *The Journal of Portfolio Management*, 30:110–119.

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

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

Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.

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

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99:204–215.