Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Double bachelor Bsc^2 in Econometrics and Economics

Combining Combination Rules: Improving Portfolio Performance Even Further?

Isabelle Pelt (535620)

Supervisor:	dr. (Alberto) A. Quaini
Second assessor:	dr. (Rasmus) SOR Lönn
Date final version:	1st July 2023

Abstract

We are interested in analyzing possible portfolio performance improvements when incorporating combinations of the naive with one of the four sophisticated diversification rules, based on the research of Tu & Zhou (2011). Our first research goal is to identify whether combining rules is beneficial for portfolio performance. We find this to increase portfolio performance. As an extension, we combine combination rules, to investigate whether performance is possibly improved even further. We find that it indefinitely improves portfolio performance.

1 Introduction

It is a well-known fact that diversification of capital increases the risk-adjusted returns of a portfolio. Distribution across (amongst others) various sectors and industries, asset classes, and firms causes a decrease in risk exposure and therefore in losses as unsystematic risk is diversified away, given that each react distinctly to similar determining events. As investors aim to maximize wealth, diversifying capital successfully is of extreme importance.

The most straightforward diversification strategy is the naive rule. This rule claims that wealth should be distributed equally amongst all the assets considered, causing the portfolio weights to be of identical magnitude per asset. Hence, this rule is neither theory- nor data-based and does not require any estimation of variables of interest. Markowitz (1952), Nobel-Prize winner and the founding father of Modern Portfolio Theory, introduces the *optimal* Markowitz rule on diversification, stating that investors should always aim to be on the efficient frontier to maximize their utility, by distributing their wealth over various assets which are not perfectly correlated.

In their research, DeMiguel et al. (2009) compare portfolio performance when distributing wealth using the naive rule, the Markowitz rule, and additional sophisticated rules. The additional sophisticated rules are designed to reduce the impact of errors in estimation, which arise as we do not know the true values of the parameters in the model. This issue is known to be a notorious problem for optimal diversification rules. Amongst others, DeMiguel et al. (2009) and Tu & Zhou (2011) find that the impact of estimation errors often outweighs the benefits of the optimal rules, causing the naive rule to outperform the sophisticated ones. Therefore, we replicate the research by Tu & Zhou (2011) and consider *combining* the sophisticated rules with the naive rule. The naive rule is given to be neither theory- nor data-based, thus it is biased and will never converge to the true rule, the portfolio strategy whenever the parameters of the models are known. The sophisticated rules on the other hand are theory- and data-based, often unbiased whenever the sample size takes on large values and known to have a substantial variance, especially whenever the sample size is small. Thus, the introduction of combining rules suggests a trade-off between the bias of the naive rule and the variance of the sophisticated rule. Accessing this trade-off leads to possible better portfolio performance, in which we are interested. To examine whether combining the naive with a sophisticated diversification rule leads to significant portfolio performance improvement compared to their single counterparts, similar to Tu & Zhou (2011), we distinguish between a simulated and empirical application. For the simulated application, we draw excess realized returns under certain assumptions, whilst, for the empirical application, these are retrieved from real datasets. The reason for making a distinction between the two applications is that only the simulated application allows for the insightful investigation of the impact of the presence of mispricing in the factor model, together with the influence of the risk-averse behavior of an MV investor on portfolio performance. This is not achievable for the empirical application, as possible mispricing is already incorporated into the series of returns, and the risk-averse behavior of the investor is unknown.

To extend the research by Tu & Zhou (2011), we consider combining the combination rules, thus combining the sophisticated rules combined with the naive rule, which we refer to as "double combining", for ease of understanding. These rules are given to outperform their sole counterparts (Tu & Zhou (2011) and displayed in Section 6), even under the presence of estimation errors. We allow all rules to have an equal share of the combined portfolio weights. This is because this diversification rule is straightforward to implement, as it does not require additional derivations to obtain the combination coefficients (described in Section 5). Next, additional errors in estimating the combination coefficients will not arise, preventing the gain in performance by implementing the double combined rule to be offset by the presence of estimation risk. Double combining rules gives another trade-off between the bias of the naive rule and the variance of the sophisticated rule, but not exactly similar to when the sole rules are combined. This is as now, we obtain a 'trade-off of trade-offs': every combined rule has an established optimal trade-off between bias and variance, which we implement partially. Therefore, double combining rules is regarded as shrinking the trade-off of the double combined rule towards the trade-offs of the combination rules. Shrinking is given to be a beneficial and effective tool to handle estimation errors in sophisticated rules (DeMiguel et al. (2009), Jorion (1986)). Thus, our objective is to investigate whether incorporating shrinking twice (once to obtain the combined rules, again to obtain the double combined rule) whilst integrating a combination coefficient for the extension of the research does not yield additional errors in estimation, potentially generating an increase in portfolio performance. Hence, our research question is:

> Will equally combining the combination rules lead to significant portfolio performance improvement compared to all diversification rules considered, and possible outperformance of the naive rule?

If the double combined rule outperforms the naive rule, our extension additionally preserves Markowitz's financial economical theory. To answer the research question, we first investigate, similar to Tu & Zhou (2011), whether combining sophisticated rules with the naive rule increases portfolio performance and possibly outperforms the naive rule. Then, we regard double combining the combination rules that perform satisfactorily, this being the Markowitz (1952), Kan & Zhou (2007) and Jorion (1986) rules (validated by the results depicted in Section 6), and examine whether the double combined rule outperforms the combination rules, the sophisticated rules, and the naive rule. The reason for choosing to not combine several combined rules with the naive rule anew is that the naive rule is given to not converge to the true optimal rule as the sample size increases. As we desire to have portfolio weights close to the true rule, we choose to combine sophisticated combination rules solely.

We find, similar to Tu & Zhou (2011), that combining sophisticated rules with the naive rule generally increases performance, potentially leading to the combined rule outperforming the naive rule where the sole sophisticated rule could not, even at small sample sized or when the naive rule is hard to beat.

For our extension, we find it to consistently outperform the naive rule and the single sophisticated rules, whilst we do not encounter our extension too often convincingly outperform the combination rules.

The paper is organized as follows: Section 2 summarizes the relevant literature. Section 3 explains the applications incorporated and how these obtain their datasets. Section 4 clarifies combining rules, and describes the combination rules we consider and our extension. Section 5 illustrates the way the performance of a rule is examined. Section 6 discusses the obtained results. Section 7 summarizes our findings and suggests topics for further investigation. In the Appendix, we motivate our datasets considered in Section A. Section B further clarifies the methodology behind our combination rules. We consider some additional plots and tables of our results in Section D. Lastly, Section E describes limitations encountered during our research.

2 Literature

Besides the research executed by DeMiguel et al. (2009), Michaud (1989) and Tu & Zhou (2011) learn that the optimal Markowitz rule is often outperformed by the naive distribution strategy. Black & Litterman (1992) and Kan & Zhou (2007) show that this is due to its errors in estimating the parameters of the model, whereas the naive strategy does not require any estimation. This finding is again verified by the research of Platanakis et al. (2018), as they put their focus on the distinct diversification strategies for portfolios in the cryptocurrency market, and find no significant outperformance of the optimal strategies. They find that this is again because the gain of implementing the sophisticated strategies is canceled out by the error in parameter estimation. Ackermann et al. (2017) address portfolio performance for the distinct strategies in the currency market, and (indirectly) show the impact of error in parameter estimation on the performance of advanced diversification strategies as well. They now discover that the sophisticated strategies outperform, and devote this to the fact that the input variable, the interest rate, predicts future returns not subject to errors in estimation.

The implementation of sophisticated strategies to obtain better performance of portfolios is therefore regarded as questionable: the severe influence of the errors in parameter estimation often does not lead to better performance compared to the naive diversification strategy, which is problematic. The sophisticated strategies are elements of the major fundamental for investing strategies and thus financial economics, namely the Modern Portfolio Theory introduced by Markowitz (1952), the idea that whenever investors attain a fully-diversified allocation of wealth, higher expected returns are only obtained by taking on more extreme allocations and hence more risk (Brandt (2010)). This theory is still of great importance in economic theories, more than fifty years after its introduction. Diversifying wealth following Markowitz's theory is executed by many distinct individuals and organizations. Individual investors can choose to diversify wealth according to his theory, that of retrieving a credible risk-return trade-off, by, amongst others, passive investing (e.g. Ben-David et al. (2017)). Passive investing or exchange-traded funds (ETFs) are more demanded than ever, causing Markowitz's theory to grow in relevance as well. The goal of such an investor could either be to expand the capital for short-term reasons or to save for his retirement. Prosperous investors invest according to Markowitz's theory as well, whilst they have the additional possibility to diversify away risk by investing in additional assets such as real estate or private equity. In terms of organizations, we could think of those which have generated profit in the past and have the desire to expand it, and therefore choose to invest in similar assets as the wealthy individuals, such as real estate, stocks, and many more. Additionally, institutional investors such as pension funds and foundations, which are known to possess great capital, often choose to diversify their capital to protect it or to achieve long-term financial goals. Multiple extensions of Markowitz's theory are introduced, such as the Post-Modern Portfolio Theory, which is a more robust model that additionally focuses on minimizing downside risk and allows to conquer some of the basic issues of the theory, such as assuming the returns to be normally distributed (Rom & Ferguson (1994)). The reason for Markowitz's theory being a prevalent theory for investing strategies, is partly due to the comprehensive, clear, and understandable explanation in the Markowitz (1952) paper containing Markowitz's theory on portfolio choice, which is given in his book on portfolio selection (Markowitz (1959)), intentionally causing it to be accessible for less quantitative investors to understand its working.

All together, confirming the credibility of Markowitz's theory is of extreme importance. Tu & Zhou (2011) let this be one of their objectives in their research, and verify its usefulness (which is reaffirmed by the replication results obtained in Section 6) by introducing the idea of combining sophisticated strategies with the naive rule, which leads to a significant increase in portfolio performance. They economically substantiate that risk-averse investors, the only type of investors we consider for this research, would prefer a combined portfolio over their sole counterparts chosen at random, as it is given that they favor the mean performance of a well- and worse-performing portfolio over a random selection of the two. For our research, we as well aim at reaffirming the reliability of the theory. We however do this in a slightly distinct manner, as we not only show that combining rules outperforms the naive rule, we furthermore display that the double combined rule outperforms it as well. Therefore, we are to verify his investment theory twice.

Combining the naive with the sophisticated rules is prosperous in improving portfolio performance as estimation errors are handled more rigidly, according to Tu & Zhou (2011)). They highlight that the naive rule is known to act well whenever the sample size, T, is small, whilst the sophisticated rule generates substantial variance. Utilizing the naive rule for shrinkage allows the estimate of the mean of the multivariate normal distribution to be enhanced. For our extension, we as well find an improvement in portfolio performance due to estimation errors treated more rigorously. This is as we combine rules which deal with such errors beforehand (explained in Section 1), whilst choosing not to generate additional errors in estimation, by predetermining the value of the combination coefficient.

Our research, together with that of Tu & Zhou (2011), is not the first to combine portfolio strategies. Much research on combining has been executed over the years, substantiating its relevance and usefulness of it. Examples of previously conducted research on combining allocation strategies and hence its credibility are displayed in the work of Pástor & Stambaugh (2000), Bonaccolto & Paterlini (2020), DeMiguel et al. (2009), Jorion (1986), Kan & Zhou (2007) and Brandt et al. (2009). These researches however distinct to Tu & Zhou (2011) and our extension, but are important to highlight to state that combining is an efficient and suited tool to generate better portfolio performance of sophisticated rules, under the presence of estimation errors. First, we consider combining in the manner of Pástor & Stambaugh (2000). They work with Bayesian model averaging: an estimate for a parameter is obtained by taking the average of the obtained estimates for the parameter using different regarded models, whilst letting the weights be equal. This approach can be interpreted as combining diversification strategies (Tu & Zhou (2011)). Pástor & Stambaugh (2000) find that whenever errors due to estimation are present, combining is beneficial. Additionally, Brandt et al. (2009) introduces the idea of letting the portfolio weight per asset be determined by the characteristic of the asset considered, and calculating the coefficient of this function by maximizing the investor's average utility gained from a certain portfolio strategy. Therefore again, this research does not introduce the idea of combining the naive with a sophisticated rule. Next, Jorion (1986) focuses on shrinking the mean of the excess returns to a certain target, hence focus is laid on combining this parameter with its 'original' estimated value and target, and not directly the portfolio rules. Another distinct example of combining is given by Kan & Zhou (2007). They illustrate that the introduction of a third fund, a non-risky asset, allows for a reduction in estimation risk and thus utility loss, as the solution is now closer to optimal portfolio choice. Combining with the naive rule is supported by Tu & Zhou (2011), as they theoretically substantiate that not doing so gives worse portfolio performance. DeMiguel et al. (2009) consider this only with the minimum-variance portfolio strategy. They motivate their choice by stating that since the estimates of the mean are harder to obtain than the covariance matrix of returns, it is preferred to estimate the second one only. When executing their research, they find this combination to not statistically significantly outperform the sole naive diversification strategy (DeMiguel et al. (2009)). This combination is closely related, but not exactly similar to the ones we are evaluating. We take into account the results obtained by DeMiguel et al. (2009), and therefore choose not to combine the naive rule with the minimum-variance strategy. Lastly, Bonaccolto & Paterlini (2020) empirically analyze the performance of N distinct diversification strategies which are optimally combined, utilizing a data-driven manner whilst having no requirements on the beliefs of the implemented data, and find that combining outperforms sole strategies, specifically in the presence of estimation error when optimally adjusting to differences in market conditions. Thus, taking into account the existence of the researches executed as just mentioned, combining in such a manner to improve portfolio performance has not been analyzed, giving possibly unique, useful, and interesting results.

To obtain an outperformance of the sophisticated strategies compared to the naive rule under the presence of estimation error and therefore to replicate the work by Tu & Zhou (2011), we first combine the naive rule with the following sophisticated rules: the Markowitz (1952) rule, the Kan & Zhou (2007), MacKinlay & Pástor (2000), and Jorion (1986) rule.

The first sophisticated rule considered is the Markowitz (1952) rule. To recall, this rule depicts investors aiming to be on the efficient frontier to obtain a credible risk-return (or mean-variance (MV)) trade-off, relying on their willingness to undergo risk. Choosing to incorporate the Markowitz rule as one of the sophisticated rules considered for combining is rather straightforward: our objective is to confirm the credibility of Markowitz's investing theory, therefore inclusion of this theory is regarded as required.

The second rule considered is demonstrated by Kan & Zhou (2007), also known as the three-fund rule. This rule works with a non-Bayesian approach and is already a combination of diversification strategies. Distinct to Tu & Zhou (2011), two sophisticated strategies are now combined, instead of combining the naive strategy with a sophisticated strategy. Kan & Zhou (2007) choose to combine the Markowitz rule with the popular strategy containing a moment restriction, namely the minimum-variance portfolio. They state that the impact of estimation errors is minimized even further whenever investors are given the possibility to invest in another risky portfolio, additional to the tangency portfolio and risk-free asset (Kan & Zhou (2007), DeMiguel et al. (2009)). Kan & Zhou (2007) find this rule to enhance models utilizing the Jorion (1986) shrinkage method. Therefore, it is interesting to evaluate whether the Kan & Zhou (2007) rule attains better portfolio performance.

The third rule, the MacKinlay & Pástor (2000) rule, extends the CAPM model and is the first rule considered that implements a Bayesian approach to estimation error. The prior is determined by the belief an investor has on a certain asset pricing model, in this research either the CAPM or Fama-French three-factor model. DeMiguel et al. (2009) explain that they, hence, let $\bar{\mu}$ rely on the prior conviction, where the size of shrinkage depends on the information available in the dataset, and the flexibility in the conviction relative to it. MacKinlay & Pástor (2000) motivate the introduction of their rule by stating that the incorporation of an unobserved factor in an APM of returns whilst the factor model is exactly defined leads to mispricing in the covariance matrix of the residuals. When allowing for manipulation of mispricing, MacKinlay & Pástor (2000) find improvement in portfolio choice as expected returns estimates are found more stable and precise. We implement this rule as the third sophisticated rule to be combined with the naive rule, as it is a relatively well-performing rule with a Bayesian approach to estimation errors (DeMiguel et al. (2009)).

The last rule, the Jorion (1986) rule, is nicely clarified by Kan & Zhou (2007). They explain that it considers the Bayes-Stein shrinkage estimators, and thus is the second rule considered to utilize a Bayesian approach to estimation errors. It is given to frequently attain better portfolio performance compared to estimators obtained using a Bayesian approach with a diffuse prior (clarified by, amongst others, Barry (1974)), according to findings of Jorion (1986). Shrinking, first introduced and further clarified by Stein (1956) and James & Stein (1992), refers to letting μ 'diminish' towards the general "grand mean" (Jorion, 1986, p.285) $\bar{\mu}$, whereas Jorion (1986) lets this be the mean of the excess returns of the minimum-variance portfolio, μ_{min} (denoted as μ_g in this paper). Jorion (1986) finds that such shrinkage leads to improvement in performance. Kan & Zhou (2007) additionally show that Jorion (1986) can be interpreted as a three-fund rule. We let this rule be the last sophisticated rule considered, as it as well performs relatively when considering sophisticated rules with a Bayesian approach to estimation errors (DeMiguel et al. (2009)).

Research similar to our extension, that of double combining rules, has not as yet been executed. Tu & Zhou (2011) theoretically evaluate the benefits of combining more than two rules, and those of choosing to combine two sophisticated rules, therefore excluding a combination with the naive rule. For the first instance, they mention that obtaining the estimated optimal combination coefficient generates additional estimation errors, hence not necessarily leading to an improvement in performance. We coincide with their theoretical evaluation, and, as explained in Section 4.6, we choose to double combine 'naively', thus we allow each combination rule considered to attain the same weight in our extension.[1] This translates to not (necessarily) optimally double combining, which prevents estimation errors to increase further. We choose to not do so as it would be computationally extremely expensive to derive expressions for the combination coefficients, if not impossible. Few of the combination rules considered for our extension, the CKZ and CPJ rules, possess estimated combination coefficients as their analytical expressions are challenging to obtain. Thus, deriving an expression for the combination coefficients for our extension whilst utilizing biased combination coefficients does not guarantee a performance improvement. Even if we were to use the analytical expression for the CPJ rule, encountering this issue is still likely as one of the three combination coefficients is estimated and thus biased. Therefore, we believe that naively double combining is appropriate to do so.

3 Data

For both applications, accurately replicating the results of Tu & Zhou (2011) is accessible. For the simulated analysis, this is as we follow the same assumptions as the models. For the empirical analysis, this is because the datasets considered are publicly available. As we incorporate the same datasets, the results should be certainly similar. Slight deviations in results for the simulated application are possibly caused by the fact that we are randomly drawing 10,000 times the excess returns for the N assets from a multivariate normal distribution with a given mean and covariance matrix, together with drawing the diagonal values of the residual covariance matrix from a uniform distribution. The differences in results are however not that great, due to a large number of repetitions we execute to obtain the results.

3.1 Simulated Application: Initializing the Models

Before clarifying the exact initialization of the models for the simulated setting, it is important to highlight that the model is retrieved 10,000 times at random. We obtain the performance criteria of interest by taking the average of the 10,000 values obtained. Having the number of repetitions equal to 10,000 causes the resulting variables of interest to hardly differ, thus the estimated criteria to be rather reliable.

We choose to work with both the one-factor and the three-factor model. We motivate these models in Section A.1 of the Appendix.

[1] It is however interesting to evaluate the performance of a portfolio whenever we allow the double combining to be executed optimally, in the case when we find equally double combining to not significantly increase performance even further. We motivate this further in Section 7.

First, we employ the CAPM model, similar to DeMiguel et al. (2009) and Tu & Zhou (2011). We have for the market factor the annual mean excess returns being equal to 8%, whilst its standard deviation being 16%. We allow the factor loadings to be equally spread between the values 0.5 and 1.5. For the covariance matrix of the residuals, we have that it is a diagonal matrix with its values drawn from U[0.10,0.30]. We work with the same extensions as Tu & Zhou (2011), that of a varying relative risk-averse behavior of the investor, therefore we allow the parameter γ to equal either 1 or 3. Having this parameter to be larger than zero is equivalent to being risk-averse as an investor, whilst having a larger value corresponds to an investor being more risk-averse. Another varying input parameter is the α's of mispricing. To determine the influence of mispricing on portfolio performance, we allow the [Nx1] vector of mispricing to be either equal to zero for every risky asset or to be equally spread, such that $\alpha \in [-2\%, 2\%]$. Incorporating both extensions is of interest as we want to investigate whether a difference in behavior has an impact on the performance of combining, whilst for mispricing, this could well be present as we are considering a (quite unrealistic) one-factor model. Throughout the whole simulated application, we let the number of risky assets, N, equal 25. With this amount, we diversify away idiosyncratic risk (e.g. Elton & Gruber (1997) display this), whilst preventing the incorporation of too many assets leads to a great challenge in outperforming the market. Next, we let the sample size T equal 120, 240, 480, 960, 3000, and 6000, to obtain the sample mean and covariance matrix of the series of the excess returns of length T, from whereon we obtain the various diversification rules. We know that the sophisticated rules converge to the true rule as the sample size increases whilst it having a substantial variance when T is small (Tu & Zhou (2011)), therefore it is of interest to consider these various sizes.

To decrease the impact of mispricing and to further investigate the profitability of combining rules for portfolio theory, we consider the Fama-French three-factor model (Fama & French (1993)) additionally, similar to Tu & Zhou (2011). We utilize the same, monthly sample period as Tu & Zhou (2011), that of July 1963 to August 2007 (thus 530 observations) for the factor portfolios from Kenneth French's data-site. We must be aware that the implementation of the one-factor model immediately gives the annualized excess returns time series, whereas, for the three-factor model, we additionally have to annualize the monthly returns. Having the annualized returns allows for drawing the mean and covariance matrix of returns, which are regarded as the true parameters. Therefore, we obtain the annualized returns for asset n at time t as

$$R_{n,t} = \left(\left(1 + \frac{\tilde{R}_{n,t}}{100} \right)^{\frac{1}{12}} - 1 \right) * 12. \tag{1}$$

We divide the rate of return of asset n at time t, $\tilde{R}_{n,t}$, by 100 as we want these to be displayed in decimals. Adding it to one allows us to take this value to the power $\frac{1}{12}$ as we are considering compounded returns, exposing the time that it covers which is given to be a month, as we consider monthly datasets only. We then have to subtract one by this number to obtain the monthly return, which we lastly multiply by 12 to get the annualized excess returns, such that we match the annualized α's of mispricing.

For these factor portfolios, we obtain their mean and covariance matrices of the three factors, recognized as the true parameters μ and Σ. We allow the factor loadings of the model to be evenly spread on various ranges, depending on the factor considered: [0.9,1.2] for the Mkt, [-0.3,1.4] for the SMB, and [-0.5,0.9] for the HML factor, whereas we randomly pair these. With the three-factor model initialized, we again obtain $\hat{\mu}$ and $\hat{\Sigma}$ from the excess returns series (further clarified in Section B.2), from whereon we calculate the various diversification rules.

3.2 Empirical Application: Real Datasets Considered

All seven, real datasets considered are given in Table 7 in the Appendix in Section A.2, where the first five refer to some used by DeMiguel et al. (2009), and the last two are additionally introduced datasets, similar to Tu & Zhou (2011). Six out of seven datasets considered are retrieved from Kenneth French's dependable data-site, whereas the last is obtained from the reliable MSCI data-site. Note that, distinct to DeMiguel et al. (2009), we choose to not include the dataset containing 20 portfolios constructed based on the SMB and HML factors, with Fama-French's three-factor portfolios. This is substantiated by their finding, as they show that the results regarding portfolio performance do not significantly differ from the distinct portfolio strategies considered and therefore are not engaging to include. We motivate these datasets in Section A.2 in the Appendix.

For some datasets, in addition to Fama and French's three factors, the Momentum (UMD) factor is included. The motivation for including this factor is given in Section A.2 in the Appendix.

The first, third, fourth, and fifth datasets contain 497 observations, the second contains 379 observations, and the last two contain 530 observations of monthly data per asset considered.

4 Description of the Combination Rules Considered

As explained, we focus on the combination of the naive diversification rule with a more sophisticated combination rule: either the Markowitz (1952), Kan & Zhou (2007), MacKinlay & Pástor (2000) or Jorion (1986) rule. First, we clarify the general working of combining two rules, followed by an illustration of the combination with the distinct sophisticated rules, the additional methods implemented for the empirical application together with the performance criteria, and lastly the extension. For this section, explanations and terminology coincide largely with the work of Tu & Zhou (2011).

4.1 Combination of Two Rules

Calculating the vector of portfolio weights when combining the naive with the sophisticated rules is done in the following manner, similar to Tu & Zhou (2011):

$$\hat{w}_c = (1-\delta)w_e + \delta\tilde{w}, \tag{2}$$

where we let w_e be the vector of portfolio weights when implementing the naive rule, thus this being $w_e = \frac{\vec{1}_N}{N}$, where N is the number of risky assets considered.[2] We have for the data-based

[2]Similar to Tu & Zhou (2011), we choose to not incorporate the riskless asset, which would give $w_e = \frac{\vec{1}_N}{N+1}$ for the naive rule, as it does not significantly alter the results.

rule the excess returns on the risky assets, R_{T+1} at time T+1, a vector of size N, which are given to be multivariate normal and IID distributed, with mean μ and covariance matrix Σ, therefore, the portfolio return of the combination rule $\hat{w}_c$ at time T+1 is $R_{pT+1} = r_{fT+1} + \hat{w}_c' R_{T+1}$. For our application, we choose the risk-free rate, r_f, equal zero at any given time T, as we consider the *relative* performance of the various diversification rules. Thus, as these are identical for every portfolio strategy, the risk-free rates cancel each other out when regarding the distinct performances. $\tilde{w}$ refers to the vector of portfolio weights when implementing the sophisticated rule, which is estimated and data-based.

We let $\delta \in [0,1]$ be the combination coefficient, displaying the magnitude of the vector of combined portfolio weights determined by the sophisticated rule, whereas 1 - δ tells the amount of the combined weights determined by the naive rule. This coefficient has to be estimated as well as it is unknown. When estimating δ, estimation errors arise. For the various combinations, these errors are typically however not too grand (Tu & Zhou (2011)), as the coefficient is a sole parameter. Therefore, an improvement in performance compared to the single diversification rules regarded is rather probable.

Similar to Tu & Zhou (2011), our objective is to maximize the utility of an MV investor. We do not know μ and Σ for the sophisticated rules. In this case, this objective function is the most straightforward to maximize, as, choosing to maximize the Sharpe ratio of a strategy is analytically challenging, as this function is "highly nonlinear" (Tu & Zhou, 2011, p.208). Thus, we minimize the following expected loss function, given by Tu & Zhou (2011) which is a well-known function considered in ordinary statistical decision theory (implemented by Brown (1976), DeMiguel et al. (2009), and many more), to obtain the estimated value of the combination coefficient:

$$L(w*, \hat{w}_c) = U(w^*) - \mathbb{E}[U(\hat{w}_c)], \tag{3}$$

where $U(w^*)$ refers to the expected utility of the "true optimal portfolio rule" (Tu & Zhou, 2011, p.206). In this case, we know the mean and covariance matrix of the excess returns of the N risky assets considered. We take the expectation of the utility of the sophisticated (combination) rule, $U(\hat{w}_c)$ which is given to be a random variable (Tu & Zhou (2011)), so we acquire $\mathbb{E}[U(\hat{w}_c)]$. Consequentially, the out-of-sample performance $U(\hat{w}_c)$ is given to be a random variable. Therefore, we choose to calculate the expected out-of-sample performance, $\mathbb{E}[U(\hat{w}_c)]$, to investigate the performance of a certain rule.

Thus, assuming $R_{T+1} \overset{\text{IID}}{\sim} \mathcal{N}(\mu, \Sigma)$, we obtain the expected utility of the sophisticated (combination) rule, $U(\hat{w}_c)$:

$$U(\hat{w}_c) = r_{fT+1} + \mu' \hat{w}_c - \frac{\gamma}{2} \hat{w}_c' \Sigma \hat{w}_c \tag{4}$$

where γ tells the relative risk aversion of a particular investor. As clarified, we let r_f equal zero for any time t. Hence, the expected utility function of the sophisticated (combined) strategies can be interpreted as subtracting the mean of the excess returns of the (combined) portfolio $\mu_p = \mu_t' \hat{w}_c$ by its variance, $\sigma_p^2 = \hat{w}_c' \Sigma \hat{w}_c$, whilst incorporating the risk-averse behavior of the investor (Kan & Zhou (2007)). When optimizing the expected utility function of the true rule, we diversify our wealth through a combination of the tangency portfolio and the risk-free asset

(Kan & Zhou (2007)), giving the optimal portfolio weights $w^* = \frac{\Sigma^{-1}\mu}{\gamma}$ (Tu & Zhou (2011)).

For every sophisticated rule, thus the diversification rules not being the true or naive rule, we do not know the true μ and Σ. Therefore, we utilize the sample mean and covariance matrix of the excess returns, denoted as $\hat{\mu}$ and $\hat{\Sigma}$ respectively, to obtain the rules. Once these are retrieved, we utilize the true μ and Σ to calculate the performance criteria considered. The motivation to do this is that we are interested in comparing the relative performance of the (combined) portfolios, thus making use of the true parameters allows for this comparison to be more apparent and efficient.

4.2 Combination with the Markowitz (1952) Rule

The first sophisticated rule discussed by Tu & Zhou (2011) we likewise consider to combine with the naive rule is the Markowitz (1952) rule. This rule, also known as the maximum likelihood (ML) rule, aims to optimally distribute wealth across the assets considered, whilst solely caring about μ and Σ (DeMiguel et al. (2009)). The motivation to utilize this rule is given in Section 2.

We employ the (estimated) scaled ML rule, $\bar{w} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu}$, and use Equation (2) to obtain the combined ML rule (or the CML rule):

$$\hat{w}^{cml} = (1 - \delta_{ml})w_e + \delta_{ml}\bar{w}, \tag{5}$$

In this certain case, our objective is to minimize the loss function obtained by Tu & Zhou (2011), substantiated with proofs given in their Appendix, in Sections A.1 and A.2. Thus, we have for the loss function:

$$L(w^*, \hat{w}_c) = \frac{\gamma}{2}[(1 - \delta_{ml})^2\pi_1 + \delta_{ml}^2\pi_2], \tag{6}$$

with $\pi_1 = (w_e - w^*)'\Sigma(w_e - w^*)$ and $\pi_2 = E[(\bar{w} - w^*)'\Sigma(\bar{w} - w^*)]$. Tu & Zhou (2011) clarify that π_1 tells the impact from the bias of the naive rule, whereas π_2 denotes the impact from the variance of the scaled ML rule. We allow the optimal combination coefficient δ_{ml} to represent the trade-off proposed between the bias and the variance, therefore we have:

$$\delta_{ml}^* = \frac{\pi_1}{\pi_1 + \pi_2}. \tag{7}$$

Tu & Zhou (2011) show, in Proposition 1 in their research, that if $\delta_{ml}^* \in (0, 1)$, combining the two rules outperforms the single diversification rules in terms of expected loss. They find that this always holds, as the naive rule is never equal to the true optimal rule as it is given to never converge to it, which is a condition needing to be satisfied to make this claim. If the CML does not outperform the naive or ML rule, we assign this to large errors in estimating δ_{ml}^*. Its estimate is given in Equation (27) in Section B.2, together with the motivation for employing the scaled ML rule, and the derivation of the estimated CML rule, $\hat{w}^{CML}$.

4.3 Combination with the Kan & Zhou (2007) Rule

The second sophisticated rule regarded is the KZ rule, which is founded by Kan & Zhou (2007). They introduce the idea of allowing investors to invest in another risky portfolio to minimize the impact of estimation errors on portfolio performance. Further motivation for incorporation of this rule is given in Section 2.

We have for the (estimated) Kan & Zhou (2007) rule

$$\hat{w}^{KZ} = \frac{c_3}{\gamma}[(\frac{\hat{\eta}}{\hat{\eta}+\frac{N}{T}})\hat{\Sigma}^{-1}\hat{\mu} + (\frac{\frac{N}{T}}{\hat{\eta}+\frac{N}{T}})\hat{\mu}_g\hat{\Sigma}^{-1}\vec{1}_N], \tag{8}$$

with $c_3 = \frac{(T-N-1)(T-N-4)}{T(T-2)}$. When minimizing the expected loss function, we obtain the optimal combination rule:

$$\hat{w}^{ckz} = (1-\delta^*_{kz})w_e + \delta^*_{kz}\hat{w}^{KZ}, \tag{9}$$

with

$$\delta^*_{kz} = \frac{\pi_1 - \pi_{13}}{\pi_1 - 2\pi_{13} + \pi_3}. \tag{10}$$

We again estimate the combination coefficient for this rule, giving us the estimated combined KZ rule, $\hat{w}^{CKZ}$. According to Proposition 3 by Tu & Zhou (2011), the purpose of this combination coefficient is that the CKZ rule should outperform the naive rule if the estimation errors arising with the estimation of the coefficient are small, together with the naive rule not being exactly equal to the true rule. We find this to often hold when evaluating the results in Section 6. All derivations on these estimations are given in Section B.3.

4.4 Combination with the MacKinlay & Pástor (2000) Rule

For the MP rule, MacKinlay & Pástor (2000) extend the CAPM (the capital asset pricing model). MacKinlay & Pástor (2000) motivate the introduction of their rule by stating that the incorporation of an unobserved factor in an APM of returns whilst the factor model is exactly defined leads to mispricing in the covariance matrix of the residuals. Further reasoning on the inclusion of this rule is given in Section 2.

We obtain the (estimated) MP rule:

$$\hat{w}^{MP} = \frac{(\hat{\Sigma}^{MP})^{-1}\hat{\mu}^{MP}}{\gamma}, \tag{11}$$

where $\hat{\mu}^{MP}$ and $\hat{\Sigma}^{MP}$ are the ML estimators of the parameters in their latent model factor, clarified further in Section B.4.

Similar to Tu & Zhou (2011), we minimize the expected loss function and obtain the optimal combination coefficient given as

$$\delta^*_{mp} = \frac{\pi_1 - (w_e - w^*)'\Sigma\,\mathbb{E}[\hat{w}^{MP} - w^*]}{\pi_1 - 2(w_e - w^*)'\Sigma\,\mathbb{E}[\hat{w}^{MP} - w^*] + \mathbb{E}[(\hat{w}^{MP} - w^*)'\Sigma(\hat{w}^{MP} - w^*)]}. \tag{12}$$

We obtain the optimal combination rule as:

$$\hat{w}^{cmp} = (1 - \delta_{mp})w_e + \delta_{mp}\hat{w}^{MP}. \tag{13}$$

Section B.4 provides the derivation of the estimated combination coefficient, together with the derivation of the estimated combined MP rule, $\hat{w}^{CMP}$.

4.5 Combination with the Jorion (1986) Rule

We denote the Jorion (1986) rule as $\hat{w}^{PJ}$ (as Jorion's first name is Philippe). The PJ rule considers the Bayes-Stein shrinkage estimators, given to frequently attain better portfolio performance compared to estimators obtained using a Bayesian approach with a diffuse prior (Jorion (1986)). Further motivation for including this rule in our research is given in Section 2.

We obtain the (estimated) PJ rule:

$$\hat{w}^{PJ} = \frac{1}{\gamma}(\hat{\Sigma}^{BS})^{-1}\hat{\mu}^{BS}. \tag{14}$$

To calculate this rule, we must calculate the Bayes-Stein estimators. Clarification of these is given in Section B.5.

We calculate δ^*_{pj} in the following manner, similar to Tu & Zhou (2011) when minimizing the expected loss for the investor implementing the given combination rule:

$$\delta^*_{pj} = \frac{\pi_1 - (w_e - w^*)'\Sigma\,\mathbb{E}[\hat{w}^{PJ} - w^*]}{\pi_1 - 2(w_e - w^*)'\Sigma\,\mathbb{E}[\hat{w}^{PJ} - w^*] + \mathbb{E}[(\hat{w}^{PJ} - w^*)'\Sigma(\hat{w}^{PJ} - w^*)]}, \tag{15}$$

under the assumption of T>N+4.
We obtain the optimal combination rule as:

$$\hat{w}^{cpj} = (1 - \delta_{pj})w_e + \delta_{pj}\hat{w}^{PJ}. \tag{16}$$

Explanation of the estimation of the combination coefficient and the estimated combined PJ rule, $\hat{w}^{CPJ}$, are given in Section B.5.

4.6 Extension

To extend the paper of Tu & Zhou (2011), we introduce a diversification rule that takes a combination of the combined rules. We choose to combine solely the combination rules that significantly improve the performance of their sole sophisticated counterparts and (often) outperform the naive rule in the empirical application. As confirmed by our results, Tu & Zhou (2011) show that these criteria do not hold for the MacKinlay & Pástor (2000) rule. Thus, we regard the combined Markowitz (1952), Kan & Zhou (2007), and Jorion (1986) rules. We obtain the combined portfolio weights for the sophisticated rules with the naive rule as explained in Sections 4.2, 4.3, and 4.5. We choose the vector of combination coefficients, $\vec{\delta} = (\delta_1, \delta_2, \delta_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, thus attaining equal shares to every combination rule. It is possible to carry out this combination, as we assume the returns to be multivariate normal and IID distributed (amongst others highlighted in Section 4.1). Therefore, we obtain the combination of the combined rules, whilst assuming T>N+4:

$$\begin{aligned} \hat{w}^{Ext.} &= \delta_1 \hat{w}^{CML} + \delta_2 \hat{w}^{CKZ} + \delta_3 \hat{w}^{CPJ} \\ &= \frac{\hat{w}^{CML} + \hat{w}^{CKZ} + \hat{w}^{CPJ}}{3} \end{aligned} \tag{17}$$

We again distinguish between a simulated and empirical application. We incorporate the same performance criteria per application for a given model to evaluate portfolio performance by employing a comparison to the true and naive rule, the sole sophisticated rules, and their combined counterparts.

5 Methodology for Examining Performance

5.1 Empirical Application: Rolling-Window Approach and the In-Sample ML Rule

For the empirical application, we calculate the portfolio returns at time $T+1$, denoted as R_{T+1}, utilizing a rolling-window approach, similar to Tu & Zhou (2011): given a real dataset considered with monthly data points of sample size T, we let the window be either of length $M = 120$ or $M = 240$. By employing the data points of the latest M months, the various optimal portfolio rules are estimated at time t, which are then used to calculate the next month's investment strategy. We obtain $T - M$ observations for the realized excess returns for a certain diversification rule z, $r_{z,t+1} = w'_{z,t} r_{t+1}$, where r_{t+1} denotes the realized excess return on the risky assets in month $t + 1$.

Besides the rolling-window approach, we approximate the value for the true optimal rule, as its value is unknown. We do so using the similar maximum likelihood approach to Tu & Zhou (2011) for the full sample period and obtain an ML estimator. We are aware that the full dataset is not available beforehand, and therefore we coincide with Tu & Zhou (2011) and allow the ML estimator to solely serve as a benchmark, to explore the impact of estimation errors on out-of-sample results, as this approach considers the ML rule without estimation risk (DeMiguel et al. (2009)).

5.2 Performance Criteria

5.2.1 Simulated Application

To evaluate the performance of portfolios when combining the naive with sophisticated diversification rule, we introduce several performance criteria. Similar to Tu & Zhou (2011), we investigate for the simulated application the utilities as explained in Section 2 (and their standard errors) of an MV investor per portfolio strategy, and allow for the potential presence of mispricing for the returns, together with variation in the risk-averse parameter γ. Additionally, we examine the Sharpe ratios (as well as their standard errors) of the strategies. The formula and further motivation of this ratio is given in Section C.1.1, together with the formula and motivation of the standard errors of these criteria.

5.2.2 Empirical Application

For the empirical application, we more extensively evaluate the performance of a dataset by reporting additional criteria besides the Certainty-Equivalent-Return (CER) as evaluated by Tu & Zhou (2011). We give the motivation for evaluating this criterion and its derivation in Section C.1.2 in the Appendix.

The additional criteria considered are the Sharpe ratio and drawdown (DD). Recall that we now apply the rolling-window approach for the real datasets as explained in Section 5.1. The Sharpe ratio gives the risk-adjusted returns and is clarified further in Section 5.2.1, and Section C.1.1 in the Appendix. The drawdown denotes the utmost decline in the portfolio returns in regards to the maximum attained in the past (Chekhlov et al. (2004)). For the drawdown, we regard the $T - M$ realized excess returns, ER obtained per dataset for a given rule c:

$$DD_c = \left(\frac{ER_{c,max} - ER_{c,min}}{ER_{c,max}} \right). \tag{18}$$

The drawdown is related to the well-known risk function 'Value-at-Risk' (VaR), introduced by the JPMorgan Group in 1994, which displays the potential loss an investor can endure at a predetermined level of risk over a given horizon, often chosen to be 5%. We choose to regard the drawdown instead of the VaR, as the VaR additionally requires the standard deviation of excess returns, with its true value unknown to us. The drawdown only considers the observed realized excess returns and therefore no additional estimation errors arise, thus from an accuracy perspective, we choose to work with this criterion. It allows for a better understanding of rules maximizing the risk-adjusted returns, an investor's objective. We care about the relative riskiness of a strategy as the largest share of individuals involved in portfolio management is risk-averse, and hence prefers strategies that undergo less risk. We carry out this assumption for our research as we evaluate performance primarily assuming $\gamma = 3$, instead of $\gamma = 1$. We again calculate this criterion implementing the in-sample ML rule as the true rule for a given dataset. This rule however is not too insightful, as we choose the objective of our research to be maximizing the utility of an investor, which is equivalent to maximizing the Sharpe ratio of a portfolio strategy (Tu & Zhou (2011)). Thus now, our true rule will not necessarily give the minimum drawdown, and therefore cannot be used as a benchmark now. Hence, in Table 10, we report the findings in italics and choose to not regard them when interpreting the table. We still however find the criterion useful, as we can evaluate the *relative* performance of the rules, and thus investigate whether (double) combining rules increases portfolio performance, given to be the main objective of this paper.

The combination of the three criteria considered allows to verify the strategies which attain the relative best risk-reward trade-off, given to be the main objective of Markowitz's fundamental financial theory.

6 Results

In regards to replicating the research of Tu & Zhou (2011), we first consider utilities of the one-factor model without mispricing under varying values of γ given in Table 1, after which we are interested in the utilities of an investor in the case where mispricing is present, the investor being more risk-averse thus γ equal to 3 for both the one-factor and the three-factor model, given in Table 2. Then, we calculate the Sharpe ratios for an one-factor and three-factor model, with mispricing either present or not, depicted in Table 3. We then provide the standard errors of the utilities and Sharpe ratios for the three-factor model in Table 4 under the presence of mispricing, and γ equal to 3. We report the estimates of the combination coefficients and their standard errors in Table 5, under the same assumptions as in Table 4. Lastly, we calculate the CERs, Sharpe ratios, and drawdowns in Tables 6, 9 and 10, respectively, for the various datasets for the empirical application. The last two tables can be found in the Appendix, in Section D.5.

To evaluate the performance of our extension, we incorporate the exact environments as just explained. These results are displayed in the tables mentioned.

6.1 Table 1: Utilities in One-Factor with varying γ

Table 1: Utilities in one-factor model without mispricing, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, and these rules whenever they are combined with the naive rule. In this one-factor model we do not consider mispricing, so $\alpha = 0$. We let γ equal either 3 (given in Panel A) or 1 (given in Panel B). We have for the number of risky assets $N = 25$ and the number of repetitions being 10,000, for all the values of the sample size T considered.

	T					
Rule	**120**	**240**	**480**	**960**	**3000**	**6000**
Panel A: γ = 3						
True	3.30	3.30	3.30	3.30	3.30	3.30
1/N	2.96	2.96	2.96	2.96	2.96	2.96
ML	-2.01	0.96	2.20	2.77	3.13	3.22
KZ	1.29	2.13	2.61	2.89	3.15	3.22
MP	2.85	3.07	3.13	3.17	3.20	3.21
PJ	0.98	2.07	2.59	2.89	3.15	3.22
CML	2.83	2.93	2.99	3.07	3.18	3.23
CKZ	2.88	2.92	2.96	3.03	3.16	3.22
CMP	3.06	3.09	3.09	3.09	3.07	3.07
CPJ	2.80	2.89	2.94	3.03	3.16	3.22
Extension	2.87	2.94	2.99	3.06	3.18	3.23
Panel B: γ = 1						
True	9.91	9.91	9.91	9.91	9.91	9.91
1/N	6.32	6.32	6.32	6.32	6.32	6.32
ML	-6.03	2.86	6.58	8.29	9.40	9.65
KZ	3.91	6.39	7.82	8.68	9.44	9.66
MP	9.63	9.19	9.40	9.53	9.61	9.63
PJ	6.08	6.22	7.78	8.67	9.45	9.66
CML	6.46	7.21	8.02	8.73	9.45	9.67
CKZ	6.46	6.80	7.60	7.83	7.72	7.51
CMP	8.26	8.44	8.60	8.79	8.98	9.14
CPJ	6.69	6.73	7.57	7.82	7.70	7.47
Extension	6.49	7.02	7.89	8.34	8.66	8.67

Table 1 displays the utilities of an investor in an one-factor model with a varying risk-averse coefficient γ, as we let its value be equal to either 3 (Panel A) or 1 (Panel B). These utilities are plotted in Figures 1 and 2 in the Appendix, Section D.1. The goal of exhibiting these results is to understand the impact of variation in risk-averse behavior, and to understand in which setting (double) combining is the most beneficial. The decisions on the initialization of the one-factor model are further clarified in Section 3.1. We do not consider mispricing to be present, therefore $\alpha = 0$.

First, we consider an investor who is more risk-averse, $\gamma = 3$, thus an investor who receives a higher utility when he undergoes less portfolio risk. Panel A shows that none of the uncombined sophisticated rules outperform the naive rule in terms of utility generated when the sample size T is small, namely equal to 120. When combining however, thus when regarding the utility of CML, CKZ, CMP and CPJ, we find that the CMP rule is the sole rule to outperform the naive rule at a T this small. In general, thus for any value of T, the performance of the MP rule combined with the naive rule is rather satisfactory. We even find that for $T \geq 240$, the sole MP rule outperforms the naive rule, whilst the other sophisticated rules do not outperform the naive rule, unless $T \geq 3000$. This finding coincides with that of Tu & Zhou (2011), as they also discover that the MP rule performs well relatively at small sample sizes, as it greatly outperforms other sophisticated rules. Additionally, the findings by DeMiguel et al. (2009) confirm the relatively good performance of the MP rule, whilst not necessarily outperforming the naive rule. As neither of the two clarify why the MP rule performs as well, we assume the model proposed by MacKinlay & Pástor (2000) is simply suited in this setting. Note that combining with the naive rule for the MP rule, thus the CMP rule compared to the MP rule, in general, does not increase performance (we have this for $T \geq 480$), whilst combining the naive with the other sophisticated rules solely increases performance. For the CMP rule, we do not have an analytical expression for our estimated combination coefficient, and therefore choose to implement the Jackknife approach to obtain the estimate. Whenever we do possess an analytical expression (as for the CML and CKZ rule), Tu & Zhou (2011) state that combining rules will lead to an increase in portfolio performance if the errors in estimating the combination coefficients are not too grand, and if the naive rule is not equal to the true rule (first mentioned in Sections 4.2 and 4.3). For the CMP rule, we cannot make such a statement. It is not evident that combining leads to outperformance of the sophisticated rule. A decrease in the performance of the CMP rule in contrast to the MP rule is most likely due to the fact that the implementation of the Jackknife approach generates sizeable estimation errors. Tu & Zhou (2011) however find that combining the MP rule generally increases, instead of decreases, performance. We devote this difference in results to the way we implement the Jackknife approach, which is further clarified in our limitations section, Section E in the Appendix. The combination coefficient for the CMP rule does not optimize the trade-off between the bias generated by the naive rule and the variance generated by the sophisticated rule as clarified in Section 1. As our objective is to improve portfolio performance utilizing combining rules, we it is of interest to regard the performance of the other sophisticated rules and their combinations.

The true rule does not require the estimation of the mean and covariance matrix of the returns, as it utilizes its theoretical values, hence it does not generate any estimation errors. Logically, the true rule always outperforms other sophisticated diversification rules, which are obtained based on their sample counterparts and therefore endure the impact of errors in estimation of the matrices. For the naive rule, we incorporate the theoretical parameters as well. The rule is not theory- nor data-based (Tu & Zhou (2011)) thus its performance does not depend on the sample size. Hence, the bias in the rule causes the rule to never converge to the true rule. Table 1 together with Figure 1 confirm that the sophisticated rules are asymptotically unbiased (Tu & Zhou (2011)) as we find the utilities to converge to the true value as the sample size T increases, thus the sophisticated rules converging to the true rule.[3] This is as the errors generated when estimating the sample parameters decrease in magnitude whenever T increases, thus the impact of the generated variance decreases as well. Table 1 additionally shows that optimally combining the sophisticated rules accelerates the combination rules converging towards the true rule, as whenever we let T increase, we see that all the combined rules start outperforming the naive rule from $T = 960$ onwards, whilst their uncombined counterparts do not as yet: this occurs from $T = 3000$ onwards. The reason that the sophisticated rules and their combinations require rather large sample sizes is because the naive rule is given to perform very well, thus being hard to beat. Tu & Zhou (2011) claim that the way the one-factor model is initialized happens to be beneficial for the performance of the naive rule. They state that now, choosing to diversify naively is almost equal to choosing to diversify following the true rule. They give two explanations for this finding. First, they clarify that the initialization of the factor loadings in particular causes the naive rule to be close to the true rule. Additionally, they show that as mispricing is not allowed, the factor model is on the efficient frontier (Tu & Zhou (2011)). We know that the true rule is proportional to the factor portfolio by the given value of γ. Having $\gamma = 3$ now, we actually find that the naive rule is close to the true rule.

Whenever T is small, we note that combining is the most beneficial for the ML rule. The sole ML rule performs not too sufficiently in these cases, whereas combining greatly increases its performance. Therefore, for T being small, we can state that the ML rule suffers mostly from estimation errors. This coincides with theory, as the sophisticated rules are designed to better handle estimation errors (DeMiguel et al. (2009)), which matches the higher obtained utilities for such rules compared to the ML rule. This often holds for any setting considered in the remaining sections, thus we choose to not repeat this finding.

We now consider a less risk-averse investor, thus $\gamma = 1$ and therefore we have that the utility received is higher whenever we undergo more risk in comparison to an investor who has $\gamma = 3$. We find that the utilities obtained are of greater magnitude than when we set $\gamma = 3$ (be aware that this also holds for the uncombined ML rule whenever $T = 120$, as we now obtain a more negative utility in this setting). For the CMP rule, we find it to outperform the naive rule for any T, whilst it performs worse in contrast to its uncombined counterpart, the MP rule. Similarly as when we set $\gamma = 3$, we state that the errors in estimating the combination coefficient offset the gain of combining. We again devote the well performance of the (combined) MP rule to the actual model implemented being to its advantage. As our objective remains to increase portfolio performance by means of combining rules, we anew are not too interested in the MP rule and its combination.

We now find that the other combination rules, the CML, CKZ and CPJ rule, outperform the naive rule whenever T is small, namely from $T = 120$ onwards, whilst to recall, this is the case from $T = 960$ onwards whenever the investor is more risk-averse, as we found the naive rule to perform fairly well when $\gamma = 3$. Having a lower value for γ now causes the naive rule to not be as similar to the true rule as when we set $\gamma = 3$, which is confirmed by the larger difference in utility between the two rules. Thus now, the naive rule is outperformed more easily because it

[3] Whenever we endure potential exceptions for the rest of our research, we solely discuss these in their corresponding sections.

performs worse in this setting.

Tu & Zhou (2011) find that for $\gamma = 1$, combining the KZ, the MP and PJ[4] rule with the naive rule does not necessarily lead to an increase in performance, especially at larger sample sizes. Our results coincide with theirs, as we notice a similar pattern in portfolio performance when combining these rules. We contribute this to sizeable errors in estimating the combination coefficients, leading to the combination coefficients not optimally trading off the bias generated by the naive rule and variance generated by the sophisticated rule. We additionally find the CKZ and CPJ rule to *decrease* in performance as T increases. Therefore, these rules do not seem to converge to the true rule. Figure 2 displays how these rules actually converge to the naive rule. We suggest this, especially as the sole sophisticated rules do converge towards the true rule, to rely on the fact that the combination coefficients of these rules strongly shrink the sophisticated towards the naive rule.

Another note to make is that, for $\gamma \in \{1,3\}$, the CKZ rule outperforms the CPJ rule for any value of T. We devote this to the fact that both combination rules utilize the same expression for the estimated optimal combination coefficient, which is designed for the CKZ rule. Hence, it is understandable that the CKZ rule outperforms.

We find our extension to perform very well. We find that in general, double combining rules increases portfolio performance even further for the one-factor model without mispricing, for $\gamma \in \{1,3\}$. It performs best when we set $\gamma = 3$. Coinciding with theory, it converts to the true rule as T increases.

Whenever we set $\gamma = 3$, it outperforms the successful naive rule already at a small sample size, namely from $T = 480$ onwards. When we do not regard the (combined) MP rule, we find the extension to always outperforms the single sophisticated rules, thus for any T. For every T, our extension outperforms all the combination rules, except in the cases when $T = 120$, the CKZ rule then performs slightly better, and when $T = 960$, where the CML rule outperforms slightly. Therefore, we can state that in general equally double combining rules increases portfolio performance even further. We find that by definition not (necessarily) setting the combination coefficient to an optimal value, hence not generating any additional estimation errors, suffices to improve portfolio performance.

If $\gamma = 1$, our extension outperforms the naive rule at any T. We now however do not experience the extension to outperform all the combination rules as when $\gamma = 3$ for various values of T, as we find the CML rule to outperform the extension for all values of T, except when $T = 120$. We can contribute this to the finding that the estimation errors of the combination coefficient of the CKZ and CPJ rule are sizeable, and consequentially offset the benefit of combining their single sophisticated counterparts with the naive rule, which appears to not be an issue for the (combined) ML rule. As our extension is given to be an equal combination of these three rules, and as we have highlighted the fact that the CKZ and CPJ rule converge to the naive and not the true rule as T increases, it is not shocking that our extension, in this case, does not converge towards the true rule at a high rate, and is outperformed by one of its single counterparts

[4] Tu & Zhou (2011) do implement an analytical expression for the combination coefficient of the CPJ rule. Therefore, it is interesting to note that they still find the errors in estimating this coefficient are sizeable enough to offset the gains of combining the PJ rule with the naive rule.

which does not endure conversion towards the naive rule. It is not guaranteed that optimally double combining rules will eliminate this issue as this additionally requires estimation of the combination coefficient (theoretically stated by Tu & Zhou (2011)). However, investigating this issue remains interesting for further research.

Table 1 and Figures 1 and 2 have allowed us to evaluate the influence of the risk-averse behavior of an MV investor. As we are interested in increasing portfolio performance when combining rules, we set $\gamma = 3$ for the rest of our research.

6.2 Table 2: Utilities in One-Factor and Three-Factor Model

Table 2: Utilities in one-factor and three-factor model with mispricing, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, and these rules whenever they are combined with the naive rule. Utilities in the one-factor model are given in Panel A, for the three-factor model these can be found in Panel B. Mispricing is considered, so $\alpha \in$ [-2%,2%]. We let γ equal 3. We have for the number of risky assets $N = 25$ and the number of repetitions being 10,000, for all the values of the sample size T considered.

	T					
Rule	**120**	**240**	**480**	**960**	**3000**	**6000**
Panel A: One-Factor Model						
True	3.88	3.88	3.88	3.88	3.88	3.88
1/N	2.96	2.96	2.96	2.96	2.96	2.96
ML	-1.61	1.47	2.75	3.33	3.71	3.80
KZ	1.46	2.42	3.03	3.41	3.72	3.80
MP	3.17	3.37	3.45	3.48	3.51	3.52
PJ	1.10	2.35	3.02	3.40	3.72	3.80
CML	2.93	3.11	3.32	3.51	3.73	3.80
CKZ	3.01	3.05	3.11	3.16	3.02	2.97
CMP	3.22	3.33	3.40	3.43	3.46	3.46
CPJ	2.91	3.01	3.10	3.17	3.02	2.98
Extension	2.98	3.09	3.23	3.37	3.42	3.43
Panel B: Three-Factor Model						
True	0.38	0.38	0.38	0.38	0.38	0.38
1/N	-1.02	-1.02	-1.02	-1.02	-1.02	-1.02
ML	-4.12	-1.56	-0.54	-0.08	0.23	0.31
KZ	-0.78	-0.29	0.00	0.15	0.30	0.37
MP	-0.32	-0.31	0.00	0.00	0.00	-0.20
PJ	-1.00	-0.30	-0.01	0.17	0.30	0.33
CML	-0.79	-0.47	-0.20	0.02	0.24	0.32
CKZ	-0.36	-0.26	0.00	0.15	0.30	0.38
CMP	-0.45	0.00	0.00	0.00	0.00	-0.37
CPJ	-0.50	-0.30	-0.02	0.17	0.30	0.33
Extension	-0.54	-0.28	-0.05	0.15	0.26	0.36

Table 2 displays the utilities of an investor in an one-factor model (Panel A) and a three-factor model (Panel B), under the presence of mispricing. These utilities are plotted accordingly in Figures 3 and 4 in the Appendix, Section D.2. The goal of exhibiting these results is to understand the impact of mispricing per factor model and to understand in which setting (double) combining is the most beneficial. To recall, we set $\gamma = 3$. The decisions on the initialization of the one-factor model are further clarified in Section 3.1.

When we allow mispricing to be present, in the one-factor model, we find that the utility of the true rule increases, whereas it stays the same for the naive rule. Therefore, the naive rule is now less close to the true rule. Tu & Zhou (2011) explain that this is because the factor model is not positioned on the efficient frontier anymore, causing the naive rule to not solely garner "the proportion, but also the composition of the optimal portfolio incorrect" (Tu & Zhou (2011), p.209). Consequentially, outperforming the naive rule is more accessible. Hence, we find that the naive rule is already outperformed by the sophisticated rules when $T \geq 960$ (recall that the threshold we encountered for outperformance when mispricing is not present (Table 1) is $T \geq$

3000). We find all the combination rules to outperform at a smaller sample size as well, namely when $T = 480$ instead of $T = 960$ (the threshold when regarding the one-factor model without mispricing 1). The sole sophisticated rules outperform the naive rule when we have $T \geq 960$.

For the (combined) MP rule, we again find it to perform well at any T. We find both rules to converge (slowly) towards the true rule, matching the findings of Tu & Zhou (2011). Combining mostly decreases the performance of the rule due to sizeable errors in estimating the combination coefficient, causing the bias-variance trade-off again to not be optimized. As we wish to increase portfolio performance by combining rules, we are solely interested in the performance of the other sophisticated rules and their combinations.

Similar to Tu & Zhou (2011), we find that combining the KZ and PJ rule with the naive rule does not necessarily lead to an increase in performance, especially at larger sample sizes. We contribute this to sizeable errors in estimating the combination coefficients, leading to the combination coefficients not optimally trading off the bias generated by the naive rule and variance generated by the sophisticated rule. We additionally find the CKZ and CPJ rule to decrease in performance as T increases. Therefore, these rules do not seem to converge to the true rule. Figure 1 displays how these rules actually converge to the naive rule. We again suggest this to rely on the fact that the combination coefficients of these rules strongly shrink the sophisticated towards the naive rule.

Besides the one-factor model, we consider the more realistic three-factor model with the presence of mispricing, where its initialization is further clarified in Section 3.1. Recall that with the presence of mispricing, the naive rule performs poorly as it "gets not only the proportion, but also the composition of the optimal portfolio incorrect" (Tu & Zhou (2011), p.209). Hence, we find the threshold for the sample size such that the naive rule is outperformed by the sophisticated rules becomes even smaller, as it is now at $T \geq 480$. We find the combined sophisticated rules to outperform the naive rule for any value of T.

For the three-factor model, similar to Tu & Zhou (2011), we find that combining rules is beneficial in general for portfolio performance for any given value of T, and that combination rules converge towards to true rule whenever T increases.

For the MP rule, we again find that combining does not increase performance, which we anew devote to the sizeable errors in estimating the combination coefficient, due to the implementation of the Jackknife approach. Additional to note, is that both the MP and the CMP rule do not converge to the true rule in this setting. We suggest this to rely on the initialization of the three-factor model: in Section E in the Appendix, we clarify that it is mandatory to draw certain assumptions to obtain the maximum likelihood estimator of the rule. These assumptions however are tremendous for the results obtained with the rule. Therefore, as highlighted in this section, the results for the (combined) MP rule are not too insightful in this setting. We however find the decline in performance to be stronger for the CMP rule, hence, we contribute a share of this behavior to the estimation of the combination coefficient.

For our extension, we find in both factor models that it outperforms the naive rule, for any value of T. It outperforms the sophisticated rules whenever T is not too large in both models.

When including our findings in Section 6.1, we find a slight improvement in portfolio performance when we do not allow mispricing to be present when double combining rules. We however find this improvement to not be too dominant, and thus do not claim that setting $\alpha = 0$ in the model increases performance significantly when diversifying according to our extension. Double combining rules increases performance when regarding the naive and sophisticated rules, but not when regarding all its counterparts, the single combination rules considered.

In the one-factor model, it in general outperforms the CKZ and CPJ rules. Recall that these rules appear to converge to the naive rule as T increases in this setting. It is slightly outperformed by the CML rule whenever we have $T \geq 240$. Additionally, we find that our extension outperforms the sophisticated rules for smaller values of the sample size, thus $T \leq 480$ in the one-factor model. Thus, equally double combining suffices to outperform its single counterparts, the CKZ and CPJ rule, but not the CML rule. Hence, we could investigate whether optimally double combining these rules leads to an improvement in portfolio performance for any T. Be aware that this is not a guarantee, as optimally double combining requires estimation of the combination coefficient (theoretically stated by Tu & Zhou (2011)). It is still interesting to investigate and therefore left for further research.

In the three-factor model, we do not find the CKZ and CPJ rules to converge to the naive rule. Now, our extension is slightly outperformed by these rules for $T \geq 480$. It outperforms the sophisticated rules for smaller values of T, namely when we have $T \leq 240$. Thus in the three-factor model with mispricing, equally double combining does not suffice to outperform its single counterparts, the CKZ and CPJ rule. It is not guaranteed that optimally double combining rules will eliminate this issue as this additionally requires estimation of the combination coefficient. However, investigating this issue remains interesting for further research.

6.3 Tables 3 and 8: Sharpe Ratios in One-Factor and Three-Factor Model, with and without Mispricing

Table 3 displays the Sharpe ratios in an one-factor model without (Panel A) and with (Panel B) the presence of mispricing. These Sharpe ratios are plotted in Figures 5 and 6 in the Appendix, Section D.3. The goal of exhibiting these results is to understand the impact of mispricing in the one-factor model and to understand in which setting (double) combining is the most beneficial. The decisions on the initialization of the one-factor model are further clarified in Section 3.1.

Additional to Tu & Zhou (2011), we choose to report the Sharpe ratios in a three-factor model without (Panel A) and with (Panel B) the presence of mispricing, given in Table 8, in the Appendix in Section D.3. We do so as we found our extension to perform quite well in the latter setting, therefore we are interested in evaluating the corresponding Sharpe ratios. These Sharpe ratios are interpreted and plotted in Figures 7 and 8 in the Appendix, Section D.3. The goal of exhibiting these results is to understand the impact of mispricing in the three-factor model and to understand in which setting (double) combining is the most beneficial. The decisions on the initialization of the three-factor model are further clarified in Section 3.1. We find that our extension performs best when we allow mispricing to be present. We find that in both cases, combining improves portfolio performance. The naive rule is barely outperformed when we allow mispricing to not be present, neither by the sophisticated rules, the combination rules, and our extension. Double combining does not guarantee an increase in portfolio performance, independent of the assumptions on α. We do note that an increase in performance by implementing double combining is more likely when we allow mispricing to be present.

Table 3: Sharpe ratios in one-factor model, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, and these rules whenever they are combined with the naive rule. In the one-factor model, we let the alphas of mispricing be either zero (given in Panel A) or evenly spread between [-0.02,0.02] (given in Panel B), thus in the second case allowing it to be present. We have for the number of risky assets $N = 25$ and the number of repetitions being 10,000, for all the values of the sample size T considered.

	T					
Rule	**120**	**240**	**480**	**960**	**3000**	**6000**
Panel A: $\alpha = 0$						
True	44.51	44.51	44.51	44.51	44.51	44.51
1/N	43.65	43.65	43.65	43.65	43.65	43.65
ML	27.79	34.25	38.70	41.39	43.45	43.98
KZ	31.71	36.93	39.95	41.85	43.51	43.99
MP	43.53	43.76	43.85	43.89	43.93	43.95
PJ	31.97	37.03	39.99	41.85	43.52	43.99
CML	40.29	40.66	41.29	42.24	43.56	44.01
CKZ	39.01	38.96	40.78	42.51	43.56	43.75
CMP	43.66	43.74	43.76	43.78	43.77	43.79
CPJ	38.65	39.01	40.80	42.48	43.57	43.75
Extension	42.54	42.83	42.98	43.24	43.79	44.08
Panel B: $\alpha \in$ [-0.02,0.02]						
True	48.25	48.25	48.25	48.25	48.25	48.25
1/N	43.65	43.65	43.65	43.65	43.65	43.65
ML	31.32	38.15	42.65	45.27	47.26	47.75
KZ	33.20	39.19	43.02	45.37	47.29	47.75
MP	45.71	45.87	45.95	46.01	46.03	46.05
PJ	33.29	39.22	43.06	45.39	47.27	47.75
CML	43.59	44.33	45.26	46.22	47.40	47.79
CKZ	43.04	43.32	43.59	44.38	43.95	43.72
CMP	45.05	45.38	45.56	45.64	45.64	45.62
CPJ	42.79	43.26	43.60	44.39	43.94	43.72
Extension	43.24	43.75	44.41	45.37	45.74	45.81

First, we consider the one-factor model without mispricing. Recall that the naive rule performs very well, therefore it being hard to beat in performance. This is as now the factor model is on the efficient frontier. Panel A shows this, as outperformance only occurs for all sophisticated rules whenever we have $T \geq 6000$, whilst, for the combined rules, this threshold is the same. Additionally, we find that combining a sophisticated rule with the naive rule, in general, increases portfolio performance in terms of Sharpe ratios. The only sophisticated rule for which combining is not beneficial is the MP rule. We again devote this to the sizeable errors in estimating its combination coefficient. We find that all sophisticated rules and their combinations converge to the true rule as T increases.

Whenever we consider mispricing, we are aware that the naive rule performs worse, as we have that the factor portfolio is not on the efficient frontier anymore. We can confirm this by regarding the increase in the difference between the Sharpe ratio of the true and the naive rule. Outperforming the naive rule is now more accessible, which is confirmed by the fact that the sophisticated rules do so whenever we have $T \geq 3000$, and $T \geq 960$ for the combination rules, therefore our thresholds being lower compared to the case where we did not allow mispricing to be present. We find that for small values of T, combining rules leads to an increase in portfolio performance. We however again find the CKZ and CPJ rules to converge to the naive, instead of the true rule, and thus decline in performance whenever T increases, whilst their uncombined counterparts do converge to the true rule. We explain, when interpreting the previous results in Table 2, that this could be (partly) clarified by the way the combination coefficient is derived.

For our extension in the one-factor model, we find it to increase portfolio performance most whenever we do not allow mispricing to be present. It then outperforms the naive rule most of the time. Independent of the assumptions on α, we find it to in general increase performance even further, in contrast to the combination rules.

In the one-factor model without the presence of mispricing, we find that it outperforms all the combination rules (when disregarding the CMP rule) whenever $T \leq 3000$, together with it outperforming all the sophisticated rules for any value of T. Our extension outperforms the naive rule whenever $T \geq 3000$. This value is rather large, as, again, the naive rule performs well whenever we do not consider mispricing to be present.

When considering mispricing, we find our extension to outperform the CKZ and CPJ rule for any value of T, whilst it is always outperformed by the CML rule. Our extension is given to be an equal combination of these three rules, and as we have highlighted the fact that the CKZ and CPJ rule converge to the naive and not the true rule as T increases, it is not surprising that our extension now does not converge towards the true rule at a high rate, and is outperformed by one of its single counterparts which does not endure conversion towards the naive rule. For further research, it is interesting to evaluate whether optimally double combining resolves this issue. Therefore, without the presence of mispricing, our extension, in general, increases portfolio performance even further, whilst, under the presence of mispricing, this is solely the case for the CKZ and CPJ rule.

6.4 Table 4: Standard Errors of Utilities and Sharpe Ratios in Three-Factor Model

Table 4: Standard errors of utilities and Sharpe ratios in three-factor model with mispricing, in percentage points. We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, and these rules whenever they are combined with the naive rule. Standard errors of the utilities are given in Panel A, whilst the standard errors for the Sharpe ratios in the three-factor model can be found in Panel B. We let the alphas of mispricing be evenly spread between [-0.02,0.02], therefore allowing mispricing to be present. We let γ equal 3. We have for the number of risky assets $N = 25$ and the number of repetitions being 10,000, for all the values of the sample size T considered.

	T					
Rule	**120**	**240**	**480**	**960**	**3000**	**6000**
Panel A: Standard Errors of Utilities						
True	0.00	0.00	0.00	0.00	0.00	0.00
1/N	0.00	0.00	0.00	0.00	0.00	0.00
ML	0.18	0.04	0.01	0.00	0.00	0.00
KZ	0.06	0.02	0.01	0.00	0.00	0.00
MP	0.00	0.00	0.00	0.00	0.00	0.00
PJ	0.08	0.02	0.01	0.00	0.00	0.00
CML	0.04	0.02	0.01	0.00	0.00	0.00
CKZ	0.03	0.02	0.01	0.00	0.00	0.00
CMP	0.00	0.00	0.00	0.00	0.00	0.00
CPJ	0.04	0.02	0.01	0.00	0.00	0.00
Extension	0.04	0.02	0.01	0.00	0.00	0.00
Panel B: Standard Errors of Sharpe Ratios						
True	0.00	0.00	0.00	0.00	0.00	0.00
1/N	0.00	0.00	0.00	0.00	0.00	0.00
ML	0.26	0.17	0.10	0.05	0.02	0.01
KZ	0.27	0.17	0.10	0.05	0.02	0.01
MP	0.00	0.00	0.00	0.00	0.00	0.00
PJ	0.27	0.17	0.10	0.05	0.02	0.01
CML	0.15	0.11	0.07	0.04	0.02	0.01
CKZ	0.21	0.16	0.09	0.05	0.02	0.01
CMP	0.00	0.00	0.00	0.00	0.00	0.00
CPJ	0.22	0.16	0.10	0.05	0.02	0.01
Extension	0.18	0.13	0.08	0.05	0.02	0.01

Tables 1, 2 and 3 have displayed that, in general, combining rules leads to a (significant) increase in portfolio performance. Tu & Zhou (2011) highlight that this could be well due to the presence of sizeable standard errors of the performance criteria considered. Therefore, Table 4 displays the standard errors of the utilities (Panel A) and Sharpe ratios (Panel B) in a three-factor model with mispricing being present, and again $\gamma = 3$, to verify or invalidate this. The decisions on the initialization of the three-factor model are further clarified in Section 3.1. We find similar values for the standard errors of the utilities and Sharpe ratios in the one-factor model, allowing us to draw similar conclusions as in the three-factor model. We hence choose to not additionally report these values.

We find that combining sophisticated rules with the naive rule often leads to a decrease (whenever we have $T = 120$ for the utilities, $T \leq 240$ for the Sharpe ratios) in the magnitude of the standard errors of both the utilities and the Sharpe ratios, for any value of T. For our extension, however, we find its errors to be smaller than those obtained with the CKZ and CPJ rule, whilst in contrast, the errors of the CML rule are slightly smaller. Equally double combining hence not necessarily leads to a decrease in the magnitude of the standard errors. Note that the standard errors of the true and naive rule are always equal to zero for any value of T, as these rules are data-independent (Tu & Zhou (2011)). Therefore, we state that the significant increase in performance obtained due to combining is not due to large standard errors, but the actual methodology of optimally combining rules itself.

6.5 Table 5: Combination Coefficients in Three-Factor Model

Table 5 displays the combination coefficients in the three-factor model for the various combination rules. Panel A considers the CML rule, Panel B the CKZ rule, Panel C the CMP rule, Panel D the CPJ rule, and Panel E our extension. Per rule, we display the true combination coefficients as given by Tu & Zhou (2011) in Table 5 of their paper (δ, δ_k, δ_m and δ_j), our true combination coefficient (δ_{ml}, δ_{kz}, δ_{mp}, δ_{pj} and $\delta_{Ext.}$) and lastly the estimated combination coefficient ($\hat{\delta}_{ml}$, $\hat{\delta}_{kz}$, $\hat{\delta}_{mp}$, $\hat{\delta}_{pj}$ and $\hat{\delta}_{Ext.}$) and its standard error, given underneath in parentheses. We have plotted the according values in Figure 9, in the Appendix, Section D.4. The goal of exhibiting these results is to become acquainted with the share of the naive rule that contributed to the combination rule, and possibly clarify striking results. We obtain our own calculated true values for the combination coefficient by replacing the sample mean and covariance matrix of returns with their theoretical values.

Similar to the findings of Tu & Zhou (2011), we find that for any combination rule, the contribution of the naive rule, recognized as $1 - \delta$, decreases as the sample size T increases. This matches theory, as the sophisticated rules and their combinations converge to the true rule as T increases (Tu & Zhou (2011)). Additionally, the bias in the estimated combination coefficient decreases in magnitude as the sample size increases. This is clarified by the fact that the impact of estimation errors decreases for larger sample sizes (DeMiguel et al. (2009)).

For the CML rule, we find our estimated combination coefficient to be biased upwards for any value of T. We obtain for the CKZ and CPJ rule almost similar values for any T, which

Table 5: Combination coefficients in three-factor model with mispricing, in percentage points.[a]
We consider as the diversification strategies the naive rule combined with the Markowitz (1952) rule (Panel A), with the Kan & Zhou (2007) rule (Panel B), with the MacKinlay & Pástor (2000) rule (Panel C), with the Jorion (1986) rule (Panel D) and our extension (Panel E). For this three-factor model, we let the alphas of mispricing be evenly spread between [-0.02,0.02], therefore allowing mispricing to be present. We set γ equal to 3. We denote per rule the true combination coefficient as given by Tu & Zhou (2011) in Table 5 of their paper (δ, δ_k, δ_m and δ_j), our true combination coefficient (δ_{ml}, δ_{kz}, δ_{mp}, δ_{pj} and $\delta_{Ext.}$) and lastly the estimated combination coefficient ($\hat{\delta}_{ml}$, $\hat{\delta}_{kz}$, $\hat{\delta}_{mp}$, $\hat{\delta}_{pj}$ and $\hat{\delta}_{Ext.}$) and its standard error, given underneath in parenthesis. We have for the number of risky assets $N = 25$ and the number of repetitions being 10,000, for all the values of the sample size T considered.

					T		
Parameters		**120**	**240**	**480**	**960**	**3000**	**6000**
Panel A: CML							
True	δ	15.74	29.93	47.57	65.12	85.60	92.27
	δ_{ml}	18.90	34.95	53.54	70.93	89.64	95.04
Estimated	$\hat{\delta}_{ml}$	28.93	43.20	59.78	75.09	90.41	95.22
		(1.01)	(0.74)	(0.39)	(0.15)	(0.02)	(0.01)
Panel B: CKZ							
True	δ_k	53.78	68.09	79.87	88.35	95.81	97.84
	δ_{kz}	89.30	92.69	95.16	98.69	100.00	100.00
Estimated	$\hat{\delta}_{kz}$	66.33	92.41	98.64	99.73	99.99	100.00
		(1.20)	(0.76)	(0.28)	(0.14)	(0.02)	(0.00)
Panel C: CMP							
True	δ_m	28.50	42.54	56.06	66.62	76.25	78.93
	δ_{mp}	78.77	81.34	78.05	75.70	74.93	74.84
Estimated	$\hat{\delta}_{mp}$	93.08	75.89	71.16	76.09	74.18	75.17
		(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Panel D: CPJ							
True	δ_j	27.56	46.65	65.29	80.32	93.90	97.17
	δ_{pj}	89.30	92.50	95.16	98.69	100.00	100.00
Estimated	$\hat{\delta}_{pj}$	66.33	91.84	98.64	99.73	99.99	100.00
		(1.20)	(0.74)	(0.28)	(0.14)	(0.02)	(0.00)
Panel E: Extension							
True	$\delta_{Ext.}$	65.50	72.41	81.15	90.85	96.76	98.33
Estimated	$\hat{\delta}_{Ext.}$	48.55	76.75	85.59	92.51	96.96	98.21
		(1.10)	(0.72)	(0.27)	(0.11)	(0.01)	(0.00)

[a] We demonstrate the combination coefficients of our extension, but be aware that these are not too straightforward to interpret. It namely is given that our extension puts an equal weight on every combination rule, therefore, the combination coefficient of our extension is regarded as the average of the combination coefficients of the three combination rules. The value for the combination coefficient solely depicts the weight the extension puts on the naive rule, this being $(1 - \delta_{Ext.})$. $\delta_{Ext.}$ namely does not display the way the weight per rule is distributed. Obtaining this distribution is still feasible, as we know the combination coefficients per combined rule, from which we take $\frac{1}{C}$, where C denotes the number of combination rules considered, thus $C = 3$.

is supported by the fact that we choose to estimate the combination coefficient of the CPJ rule with the expression gained for the CKZ rule, as Tu & Zhou (2011) point out that deriving an analytical expression for the combination coefficient for the CPJ rule is rather challenging. We advocate this decision more extensively in Section 4.5. For these combination rules, we discover our combination coefficient to be biased downwards for smaller values of T, $T \leq 240$, whilst it being biased upwards for larger values, as then the estimated coefficients mostly exceed the true coefficients in magnitude. The bias is quite severe whenever T is small. The CMP endures the largest bias in its combination coefficient when comparing its estimated value to the true value of Tu & Zhou (2011). Additionally to the issue we share with Tu & Zhou (2011), that of the CMP rule not having an analytical expression for the combination coefficient and therefore estimating its value by means of a Jackknife approach, we extensively specify the shortcomings of the rule in our setting in Section E in the Appendix. Tu & Zhou (2011) highlight that having such a large bias translates to a minimal improvement (or even a decrease) in the performance of the MP rule whenever we combine it with the naive rule, which is confirmed by the findings in Table 2, Panel B, which displays the utilities per rule when assuming the same model. For the CMP rule, distinct to Tu & Zhou (2011), we find its combination coefficient to not converge to 100 as T increases. As a possible explanation, we devote this finding to the way the coefficient is calculated: we interpret the Jackknife approach slightly distinctly to Tu & Zhou (2011), which is further clarified in Section E in the Appendix. Similar to Tu & Zhou (2011), we obtain the smallest standard errors for the CMP rule. This is as both the MP and CMP rules do not increase in performance extensively as T increases. Having a rather constant performance explains the small standard errors. We do however find the combination coefficients of the CMP rule to converge to their true values as T increases. We again devote this to the assumptions drawn to obtain the required maximum likelihood, as further clarified in Section E in the Appendix. We find our estimated combination coefficient to be more similar to the true values of the MP rule given by Tu & Zhou (2011), than the estimates of the coefficients of Tu & Zhou (2011) themselves. Again, we suggest this being as we estimated the combination coefficients in a distinct manner. Having an estimated combination coefficient that converges to its true value allows the performance of the CMP rule to be more accurate. Consequentially, more verifiable conclusions are drawn.

We discover the naive rule to still contribute to our extension when $T = 6000$, whilst this is not the case for their single counterparts, namely the CKZ and CPJ rule. Tu & Zhou (2011) however as well find the naive rule to contribute at such a large sample size, therefore, it is interesting to see that our extension reintroduces the naive rule contributing at $T = 6000$. Important to note that, as given in the remark of Table 5, the combination coefficient of our extension only allows for the interpretation of the share contributed by the naive rule for the extension. Additionally, the bias in our estimated combination coefficient is relatively small for any T, allowing for a (potential) increase in performance when implementing our extension.

6.6 Table 6: CER per Real Dataset

Table 6: CER per portfolio strategy, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, and these rules whenever they are combined with the naive rule. The CERs of an MV investor for the distinct portfolio strategies, when we set the rolling window M = 120, are given in Panel A, whilst for M = 240 these can be found in Panel B. For the true or the in-sample ML rule, we utilize the complete dataset, therefore its result not depending on the length of the window. The datasets considered are clarified in Section 3.2. We set $\gamma = 3$. The number of repetitions is 10,000.

Rule	Industry N = 10+1	International N = 8+1	Mkt/ SMB/HML *N = 3*	FF20 1-Factor *N = 20+1*	FF20 4-Factor *N = 20+4*	FF25 3-Factor *N = 25+3*	Indu49 3-Factor *N = 49+3*
Panel A: M = 120							
In-Sample ML	12.84	0.78	0.86	7.04	21.64	23.53	17.89
1/N	0.61	0.10	0.32	0.74	0.76	0.75	0.61
ML	-102.02	-0.66	0.05	-7.06	-29.32	-70.49	-508.92
KZ	-102.91	-0.25	0.22	-6.69	-33.20	-78.71	-610.86
MP	0.07	0.03	-0.28	0.04	0.04	0.02	0.05
PJ	-99.54	-0.17	0.29	-5.28	-26.94	-67.09	-491.85
CML	-99.00	-0.28	0.11	-5.30	-26.39	-66.58	-489.11
CKZ	-1.97	0.10	0.39	-4.74	-7.04	8.98	-118.87
CMP	0.07	0.03	-0.17	0.04	0.06	0.02	0.06
CPJ	2.74	-0.17	0.30	-3.43	2.15	15.19	-76.73
Extension	-24.84	0.04	0.42	-4.49	-10.98	-7.74	-204.03
Panel B: M = 240							
In-Sample ML	12.84	0.78	0.86	7.04	21.64	23.53	17.89
1/N	0.67	0.07	0.26	0.60	0.61	0.65	0.59
ML	-29.42	-0.42	0.46	0.60	-2.87	-14.69	-68.55
KZ	-29.16	-0.09	0.58	1.18	-4.36	-16.79	-81.16
MP	0.07	-0.01	-0.45	0.05	0.03	0.06	0.08
PJ	-27.80	0.07	0.57	1.66	-1.67	-12.73	-61.47
CML	-27.71	-0.10	0.51	1.42	-1.01	-12.52	-61.06
CKZ	-3.81	-0.11	0.85	1.18	4.44	11.13	-12.45
CMP	0.07	-0.01	-0.04	0.05	0.05	0.06	0.09
CPJ	-1.64	0.07	0.57	1.66	8.50	13.38	-3.89
Extension	-9.46	-0.01	0.86	1.34	3.31	5.91	-23.09

Table 6 displays the Certainty-Equivalent-Return (CER) per diversification rule for the various real datasets considered, as discussed in Section 3.2. The goal of exhibiting these results is to examine which rules give valuable portfolio performance, which is translated to a high CER, and to understand when it benefits most to (double) combine rules. We implement a rolling-window approach as clarified in Section 5.1 with either a length of $M = 120$ (Panel A) or $M = 240$ (Panel B). Recall that the true, or in-sample ML rule, is obtained utilizing the complete dataset (explanation is given in Section 5.1).

DeMiguel et al. (2009) state in their research that in order for the sophisticated diversification rule to outperform the naive rule, the window must have a length of about 3000 months. When regarding both Panels A and B from Table 6, we in general coincide with their statement, as the naive rule is barely outperformed by the sophisticated rules, for any dataset, for any M. Similar to Tu & Zhou (2011), we find that the portfolios implementing the sophisticated rules increase in performance when the length of the window M is larger, as the magnitude of the estimation errors is decreased. Coinciding with their findings, we have that the ML rule performs worst, and the KZ rule performs best (when disregarding the MP rule), for any value of M, for any dataset considered.

The combination rules however do outperform the naive rule more often, even when $M = 120$. When we allow $M = 240$, we see that for the larger share of datasets considered, the combination rules outperform the naive rule. In particular, similar to Tu & Zhou (2011), the CKZ and CPJ rules outperform the naive rule most often. We as well find that the worst-performing rule is the CML rule, as it outperforms the naive rule less often and is less convincing. Not only causes combining in general increase the CER, we as well occasionally find outperformance of combination rules in cases when their single counterparts did not outperform the naive rule. Hence, we again state that combining rules is beneficial for portfolio performance.

Concerning the datasets considered, for the sophisticated rules and their combinations, we find similar to Tu & Zhou (2011) that the industry portfolios, thus "Industry" and "Indu49 3-Factor" generate the lowest CERs. We as well find that the inclusion of the Momentum factor, thus when comparing the performance of the "FF20 1-Factor" to that of the "FF20 4-Factor" dataset, increases portfolio performance, whereas, for the latter dataset, we also find combining to be more prosperous.

For our extension, we find it to increase in performance whenever we utilize the larger rolling window, thus $M = 240$, and see it then to outperform the naive rule for the largest share of real datasets considered. We discover it to outperform in general the single sophisticated rules for any dataset, for any M, and for $M = 240$ specifically, diversification implementing our extension even generates money (as we have a positive CER (Tu & Zhou (2011))) on a risk-adjusted base where these single sophisticated rules lose money. Whether our extension increases portfolio performance in terms of CER compared to the combination rules varies per dataset considered. Therefore, we cannot straightforwardly state that double combining increases performance even further.

6.7 Tables 9 and 10: Sharpe Ratio and Drawdown per Real Dataset

For their empirical application, Tu & Zhou (2011) solely regard the CER as a performance criterion. We additionally display per dataset the Sharpe ratio in Table 9 and the drawdown in Table 10 in Section D.5 in the Appendix. The goal of exhibiting these results is to examine which rules perform well in the empirical application, as we wish to be as conclusive as possible. The motivation for including these two criteria is given in Section 5.2.2.

When regarding both the Sharpe ratios and drawdown, we find that combining the sophisticated rules with the naive rule in general increases performance. We however do not discover such dominating results for our extension.

Similar to Tu & Zhou (2011), we find that for the empirical application, the (combined) MP rule undergoes a significant drop in performance. We as well find these rules to often be outperformed by the naive rule, almost independent of the length we assume for the window. Hence, we state this rule to perform well in the simulated, but simply not in the empirical application.

For the empirical application, it is interesting to examine whether optimally double combine rules *does* generate a (more) dominant performance of our extension, as for none of the performance criteria evaluated (the CER, Sharpe ratio, and drawdown), we found this to be the case. We leave this for further research.

7 Conclusions

Countless investors pursue the Modern Portfolio Theory introduced by Markowitz (1952) to diversify wealth. DeMiguel et al. (2009) however learn that this diversification technique and its multiple extensions are repeatedly outperformed by the naive rule in terms of various performance criteria. Tu & Zhou (2011) introduce the idea of *combining* the sophisticated rules with the naive rule, to investigate whether an increase in performance is attained. They let this be the objective of their research, and find that if this increase occurs, the credibility of Markowitz's fundamental is reaffirmed. To extend their research, we introduce the idea of double combining: we equally combine combination rules considered by Tu & Zhou (2011).

To answer our research question, we investigate portfolio performance for a simulated and empirical application.

For the simulated application, we initialize our one-factor and three-factor model similar to Tu & Zhou (2011) and DeMiguel et al. (2009), and allow the risk-averse behavior to vary, together with allowing mispricing to be present or not. We find that the naive rule is harder to beat whenever the investor is more risk averse, together with when we allow mispricing to not be present. We find that combining the sophisticated rules with the naive rule frequently improves performance, even at small sample sizes. Therefore, combining rules decreases the impact of generated errors when estimating the rules. In cases where the sophisticated rules did not outperform the naive rule, combining the rule with the naive rule consistently allows for outperformance. In this application, the presence of mispricing does not seem to affect our extension greatly, whereas it performs better if the investor is more risk-averse. It consistently outperforms the naive rule and the single sophisticated rules, whilst we do not encounter our extension to often convincingly outperform the combination rules.

For the empirical application, we consider the same real datasets as Tu & Zhou (2011). Combining a sophisticated rule, which does not outperform the naive rule by itself, with the naive rule increases performance, and potentially leads to an outperformance of the naive rule.

In this application, we find slightly less dominating results for our extension. It consistently outperforms the naive rule and the single sophisticated rules, whilst outperformance in regards to the combination rules is not as evident.

Similar to Tu & Zhou (2011), we find that for the empirical application, the (combined) MP rule undergoes a significant drop in performance in comparison to the simulated application. We as well find that these rules are often outperformed by the naive rule, almost independent of the length we assume for the window. Besides this, we do not observe, corresponding Tu & Zhou (2011), purely dominating results for the (combined) KZ rule in terms of portfolio performance in contrast to the (combined) PJ rule, even though the KZ rule is designed to enhance models utilizing the shrinkage technique of Jorion (1986).

For our extension, we predetermined the combination coefficients. Consequentially, a potential bias could arise, therefore possibly preventing the extension to converge to its true optimal rule. Tu & Zhou (2011) theoretically state that optimally double combining generates additional errors in estimating the combination coefficients which potentially offset the gains of combining. Nevertheless, we are interested in examining whether we achieve more dominant results in terms of a further increase in performance of the double combined rule, in contrast to its single counterparts, the combination rules, if we double combine optimally. The reason for this is substantiated by the research of DeMiguel et al. (2013), as they observe that the degree of shrinkage is important for the performance of a portfolio, which can be interpreted as the values of the combination coefficients. We leave this for further research.

Additionally, it is interesting to examine whether double combining different combination rules does give more dominant results in terms of an increase in performance. DeMiguel et al. (2009) evaluate the performance of 10 distinct sophisticated rules (those not being the Markowitz (1952) or the naive rule), which can be considered for double combining.

A Motivation for the Datasets Considered

A.1 Motivation for the Simulated Application

The one-factor model, which is the CAPM introduced by Sharpe (1964) and Lintner (1965), is known to be the fundamental equilibrium model for finance, due to its clear definition (Ross (1978)). The factor utilized for the model is observable and testable. With the implementation of some assumptions, we obtain results that are well used to obtain an intuition on the assets' returns. Additional to the one-factor model, we consider the three-factor model. As denoted by Tu & Zhou (2011), the one-factor model is quite unrealistic, leading to its performance most probably being fairly impacted by the presence of mispricing. Fama & French (1993) additionally introduce two factors with explanatory power in asset returns, giving their three-factor model. To recall, this model contains the US equity market (Mkt), size (SMB), and value (HML) factor. Mkt denotes the value-weighted excess return on the market of all CRSP firms considered, thus those listed on the NASDAQ, NYSE, or AMEX. The SMB or size factor, short for Small-Minus-Big, is the second factor introduced by Fama & French (1993). This factor is calculated on a monthly basis as the average of three portfolios constructed with small firms, subtracted by the average of three portfolios constructed with large firms. According to Fama & French (1993), smaller-cap firms tend to outperform large-cap firms in the long run, thus a portfolio containing more small-cap firms should be profitable in comparison to the market. The third factor is the HML or book-to-market factor, short for High-Minus-Low, which tells whether a firm is a value firm (those with a high book-to-market ratio) or whether it is a growth firm (those with a low ratio). According to Fama & French (1993), growth firms tend to outperform value firms in the long run, thus a portfolio containing more growth firms should be profitable in comparison to the market. They find that the errors in estimating expected returns for the CAPM are about three to five times the size of the errors in the three-factor model. The three-factor model however has its own shortcomings. Fama & French (1996) are unable to give a fair explanation of the persistence of temporary returns, known as the Momentum. Therefore, we consider both models for our simulated setting.

A.2 Motivation for the Empirical Application

We choose to work with the seven datasets depicted in Table 7 as we believe that this amount allows for an extensive interpretation of the possible benefits of combining diversification rules. The international dataset obtained using the MSCI data-site allows us to understand the performance of combining rules for markets outside the U.S., besides its performance across sections as in the several industry portfolios, where we distinguish between a 'general' and more detailed industry portfolios. Combining the several portfolios based on the size and book-to-market factor with either solely the market factor, all three Fama-French factors and the three Fama-French factors with the Momentum factor allows for an increase in understanding of the usefulness of combining when including the distinct number of factors to the portfolios, as additional risky assets. Hence, we consider as one of the datasets Fama-French's three-factor portfolios, to exclusively investigate their performance when combining rules.

Besides, Fama and French's Mkt, SMB, and HML factors, we consider the Momentum

(UMD) factor. As explained in Section 3.1, the three-factor model is unable to explain the momentum in returns. Therefore, introducing the Momentum factor is of interest as we wish to investigate whether capturing this issue leads to an increase in portfolio performance. The UMD is calculated, again on a monthly basis, by taking the average of the two highest previous portfolio returns, with the lowest. In this way, the UMD factor allows for investing in 'previous winners' and 'past losers'. Amongst the seven datasets considered, we have exactly one that is additional to the three factors, including the UMD factor (FF20 4-factor).

The first and last datasets given in Table 7 contain industry portfolios. For the first one, we consider amongst others the consumer (non-)durables (e.g. food and tobacco (non-durables), or cars and furniture (durables)), manufacturing, and energy industry. For the last dataset, we regard 49 industries, hence the ten 'main' industries given in the first dataset are further unfolded. We now allow the industries to be for example 'entertainment' or 'machinery'. For the dataset regarding 49 industries, we discovered some missing values, denoted by -99.99. For accuracy reasons regarding the results, we have replaced these values with 0. We executed this replacement 96 times, as we had to replace 72 values for 'Hlth', thus health, and 24 times for 'Softw', thus software.

For the second dataset, we are considering the large and mid-cap equity performance of eight distinct countries obtained from the Morgan Stanley Capital International data-site (MSCI). We choose these to be the same as DeMiguel et al. (2009), therefore being the US, Canada, Japan, the UK, France, Germany, Italy, and Switzerland. These annualized returns are calculated using the month-end US-dollar value of the country equity index (DeMiguel et al. (2009)): we take the current month's index, subtract the previous month's index and divide by the current month's index in order to obtain a rate of return. As the data is available from December 1969 onwards, our first observation for a country's index is from January 1970 onwards.

After the transformation of the data of the international dataset, we solely have datasets containing the monthly rate of returns of the various risky assets. We incorporate the formula given in Equation 1 and obtain the annualized excess returns of the N risky assets.

Table 7: Datasets utilized for empirical application

	Dataset	Abbreviation	N	Sample Period	Source
DeMiguel, Garlappi And Uppal (2009)	Ten Industry Portfolios With The Mkt Portfolio	Industry	10 + 1	07/1963 - 11/2004	Kenneth French's Data-site
	Eight Country Indexes With The World Index	International	8 + 1	01/1970 - 07/2001	MSCI
	Fama-French's Three-Factor Portfolios	Mkt/SMB/HML	3	07/1963 - 11/2004	Kenneth French's Data-site
	20 Portfolios Based On The SMB And HML Factors, With The Mkt Portfolio	FF20 1-Factor	20 + 1	07/1963 - 11/2004	Kenneth French's Data-site
	20 Portfolios Based On The SMB and HML Factors, With Fama-French's Three-Factor Portfolios and the Momentum Factor Portfolio	FF20 4-Factor	20 + 4	07/1963 - 11/2004	Kenneth French's Data-site
Additional	25 Portfolios Based On The SMB And HML Factors, With Fama-French's Three-Factor Portfolios	FF25 3-Factor	25 + 3	07/1963 - 08/2007	Kenneth French's Data-site
	49 Industry Portfolios Based On The SMB And HML Factors, With Fama-French's Three-Factor Portfolios	Indu49 3-Factor	49 + 3	07/1963 - 08/2007	Kenneth French's Data-site

Note: The monthly returns of the portfolios are constructed at the end of June each year. The number of assets, N, in a dataset is given as a sum, the first integer denoting the number of risky assets considered, whereas the second integer tells the number of factor portfolios included in the dataset. The portfolio returns in the datasets obtained from Kenneth French's data-site are retrieved from the database of the CRSP (Center for Research in Security Prices). The 1-month T-Bill return is obtained from the database of Ibbotson and Associates, Inc. (again from Kenneth French's data-site). The 20 portfolios based on the SMB and HML factors are obtained following the reasoning of Wang (2005); we choose to not include the portfolios containing the largest firms. The reason for this is that Fama-French's three factors are about a linear combination of all 25 portfolios, constructed based on the SMB and HML factors (Wang (2005)).

B Methodology for the Combination Rules

B.1 Sample Mean and Covariance Matrix

We assume $R_{T+1} \overset{\text{IID}}{\sim} \mathcal{N}(\mu, \Sigma)$ and obtain for the sample mean and covariance matrix of the excess returns (Kan & Zhou (2007)):

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t \tag{19}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})' \tag{20}$$

B.2 The Markowitz (1952) Rule

B.2.1 Scaled ML Rule

The Markowitz (1952) rule is known as $\hat{w}^{ML} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{\gamma}$. The derivations of $\hat{\mu}$ and $\hat{\Sigma}$ are given in Equations (19) and (20) in Section B.1. We scale this rule for bias, denoted as $\bar{w}$, in order to ensure it is unbiased and better performing (Tu & Zhou (2011)). $\bar{w}$ is given as:

$$\bar{w} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu}, \tag{21}$$

with $\tilde{\Sigma} = \frac{T}{T-N-2}\hat{\Sigma}$ (Kan & Zhou (2007)).

B.2.2 Estimated Combination Coefficient and Estimated Combination Rule

To obtain an estimate for δ^*_{ml}, we need estimates for π_1 and π_2 (given in Equations (24) and (25), which require an estimate for θ^2, given as $\tilde{\theta}^2_a = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$ (Kan & Zhou (2007)). This estimate however is given to be biased, according to Kan & Zhou (2007). They provide the following, unbiased estimator:

$$\tilde{\theta}^2_a = \frac{(T-N-2)\tilde{\theta}^2 - N}{T} + \frac{2(\tilde{\theta}^2)^{\frac{N}{2}}(1+\tilde{\theta}^2)^{\frac{-(T+2)}{2}}}{TB_{\frac{\tilde{\theta}^2}{1+\tilde{\theta}^2}}(\frac{N}{2}, \frac{T-N}{2})} \tag{22}$$

where we have that

$$B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}dy \tag{23}$$

which is given to be the incomplete beta distribution, with parameters a and b.

We obtain for the estimates of π_1 and π_2:

$$\hat{\pi}_1 = w_e'\hat{\Sigma}w_e - \frac{2}{\gamma}w_e'\hat{\mu} + \frac{1}{\gamma^2}\tilde{\theta}^2_a, \tag{24}$$

$$\hat{\pi}_2 = \frac{1}{\gamma^2}(c_1 - 1)\tilde{\theta}^2_a + \frac{c_1}{\gamma^2}\frac{N}{T}, \tag{25}$$

with $c_1 = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}$. We need T>N+4 in order to guarantee $\hat{\Sigma}^{-1}$ to exist (Tu & Zhou (2011)). If this condition is satisfied, we obtain the estimated optimal combination rule (with the ML rule as the sophisticated rule) as

$$\hat{w}^{CML} = (1 - \hat{\delta}_{ml})w_e + \hat{\delta}_{ml}\bar{w}, \tag{26}$$

with the optimal estimated combination coefficient being

$$\hat{\delta}_{ml} = \frac{\hat{\pi}_1}{\hat{\pi}_1 + \hat{\pi}_2}. \tag{27}$$

We show that when estimating δ^*_{ml}, we obtain minor errors, as mentioned in Section 2, which consequently do not ensure an outperformance of the $\hat{w}^{CML}$ rule compared to $\bar{w}$ or the naive rule. As $\hat{w}^{CML}$ however converges to the true optimal rule as T goes to infinity, combining the rules is still beneficial. We additionally find that $\hat{w}^{CML}$ outperforms $\bar{w}$ for various values of T, whilst it often does so for the naive rule; when this is not the case, it attains results close to those when incorporating the sole naive rule.

B.3 The Kan & Zhou (2007) Rule

B.3.1 Estimated Combination Coefficient and Estimated Combination Rule

For the estimated values of the π's, we have $\hat{\pi}_1$ calculated as in Equation (24), whereas $\hat{\pi}_{13}$ and $\hat{\pi}_3$ calculated similar to Tu & Zhou (2011) using

$$\hat{\pi}_{13} = \frac{1}{\gamma^2}\tilde{\theta}_a^2 - \frac{1}{\gamma}w_e'\hat{\mu} + \frac{1}{\gamma c_1}([\hat{\eta}w_e'\hat{\mu} + (1-\hat{\eta})\hat{\mu}_g w_e'\vec{1}_N] - \frac{1}{\gamma}[\hat{\eta}\hat{\mu}'\tilde{\Sigma}^{-1}\hat{\mu} + (1-\hat{\eta})\hat{\mu}_g\hat{\mu}'\tilde{\Sigma}^{-1}\vec{1}_N]) \tag{28}$$

$$\hat{\pi}_3 = \frac{1}{\gamma^2}\tilde{\theta}_a^2 - \frac{1}{\gamma^2 c_1}(\tilde{\theta}_a^2 - \frac{N}{T}\hat{\eta}), \tag{29}$$

where $\tilde{\theta}_a^2$ is the estimator of θ^2, clarified in Section B.2. We adopt the derivation of $\hat{\mu}_g$, the estimator of the expected excess return of the sample global minimum-variance portfolio, and $\hat{\eta}$, the estimator of the squared slope of the asymptote of the minimum-variance frontier, from Kan & Zhou (2007), thus we obtain

$$\hat{\mu}_g = \frac{\vec{1}'_N\hat{\Sigma}^{-1}\hat{\mu}}{\vec{1}'_N\hat{\Sigma}^{-1}\vec{1}_N}, \tag{30}$$

$$\hat{\eta} = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} - \frac{(\hat{\mu}'\hat{\Sigma}^{-1}\vec{1}_N)^2}{\vec{1}'_N\hat{\Sigma}^{-1}\vec{1}_N} = (\hat{\mu} - \hat{\mu}_g\vec{1}_N)'\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\vec{1}_N). \tag{31}$$

We therefore acquire the estimated combination coefficient of the KZ rule

$$\hat{\delta}_{kz} = \frac{\hat{\pi}_1 - \hat{\pi}_{13}}{\hat{\pi}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3}, \tag{32}$$

with proof of this derivation given in the Appendix of Tu & Zhou (2011), in Section A.3. Thus, we obtain the following estimated combination rule, whilst assuming T>N+4:

$$\hat{w}^{CKZ} = (1 - \hat{\delta}_{kz})w_e + \hat{\delta}_{kz}\hat{w}^{KZ}. \tag{33}$$

B.4 The MacKinlay & Pástor (2000) Rule

The MacKinlay & Pástor (2000) (MP) rule extends the CAPM model, given as:

$$R_t = \alpha + \beta f_t + \epsilon_t, \tag{34}$$

with α referring to mispricing in the excess returns on the N risky assets, and f_t being a latent factor.

B.4.1 Maximum Likelihood Estimators

The MP rule requires $\hat{\mu}^{MP}$ and $\hat{\Sigma}^{MP}$, which are the maximum likelihood (ML) estimators of the parameters in their latent model factor. These are obtained by minimizing the next function:

$$f(\mu, a, \sigma^2) = (N-1)ln(\sigma^2) + ln(\sigma^2 + a\mu'\mu) + \frac{1}{\sigma^2}[tr(\hat{U}) + \frac{\sigma^2(\mu'\mu - 2\hat{\mu}'\mu) - a\mu'\hat{U}\mu}{\sigma^2 + a\mu'\mu}], \tag{35}$$

where $\hat{U} = \hat{\Sigma} + \hat{\mu}\hat{\mu}'$. As we make rational decisions for the parameters implemented into the function, we make assumptions similar to Tu & Zhou (2011): in the one-factor model, we let the latent factor be normally distributed with $\mathbb{E}(f_t) = 0$, $\mathbb{V}(f_t) = \sigma_f^2$, $\mathbb{E}(f_t, \epsilon_t) = 0_N$, and the covariance matrix of the residuals being $\sigma^2 I_N$, I_N referring to the identity matrix with dimension [NxN]. We allow it to be exactly related, therefore we have that $\mu = \beta\gamma_f$ with γ_f referring to the factor risk premium, calculated as $\gamma_f = (\beta'\beta)^{-1}\beta'\mu_R$, with $\mu_R = \mathbb{E}(R_{T+1}) = \hat{\mu}$. Consequentially, we find $\Sigma = \sigma^2 I_N + a\mu\mu'$, with $a = \frac{\sigma_f^2}{\gamma_f^2}$. We obtain the ML estimator $\hat{\mu}^{MP}$ directly from minimizing the function given in Equation (35) with respect to the assessed parameters μ, a, σ^2, whilst for the covariance matrix for the MP rule, we have to perform the additional calculation $\hat{\Sigma}^{MP} = \hat{\sigma}^{2,ML} I_N + \hat{a}^{ML}\hat{\mu}^{MP}\hat{\mu}^{MP'}$.

Tu & Zhou (2011) do not specify their decision on the input parameters in the three-factor model. Hence, we choose the input parameters of the objective function we minimize (given in Equation (35)) to be similar to those in the one-factor model. We do so as MacKinlay & Pástor (2000) extend the CAPM and hence the one-factor model for their rule, thus the assumptions on the input parameters are originated based on this model. Therefore, we again let the latent factor be the market factor with its factor loadings as given in Section 3.1, which we simulate similar to the way it is simulated in the one-factor model. We let $\hat{\mu}$ and $\hat{\Sigma}$ refer to the sample moments of the returns generated assuming a model with the market factor as its sole factor. We choose not to assume the latent factor to be the three factors of the three-factor model as we then encounter issues with dimensions of the input parameters (e.g., our factor risk premium would be a vector of three parameter, our σ_f^2 a [3x3] covariance matrix). Obtaining the input parameters according to the one-factor model does not take into account the relation between the three-factors (as covariances are disregarded), leading to a possible bias in the ML estimators acquired after minimizing the function, and therefore potential unreliability of the results. We further discuss this issue in Section E in the Appendix.

For the empirical application, we calculate the input parameters using the observed returns. We re-estimate their values per window considered, as we re-estimate the rule per window, similar to Tu & Zhou (2011). The factor loadings are required as well, which are not given beforehand as for the simulated application. We therefore choose these to be equally distributed amongst the N risky assets considered, and thus let these be $\vec{1}_N$. The choice of the factor loadings impacts the performance of the rule. Hence, we mention this assumption as well as a limitation in our research for the MP rule in Section E in the Appendix.

B.4.2 Estimated Combination Coefficient and Estimated Combination Rule

To acquire the (estimated) combination coefficient, we need the expected value of the estimated MP rule, $\mathbb{E}(\hat{w}^{MP})$, which is hard to calculate analytically. Tu & Zhou (2011) implement a Jackknife approach, mentioned in (amongst others) Shao & Tu (2012), which is developed by Maurice Quenouille in 1949 (Quenouille (1949)) and refined in 1956 (Quenouille (1956)). The Jackknife approach allows for a reduction in the bias of the coefficient and obtains an estimate of its standard error, therefore is suited to be implemented here. To retrieve the estimator of this coefficient, denoted as $\hat{\delta}_{mp}$, we make use of the following derivations:

$$\mathbb{E}[\hat{w}^{MP} - w^*] \approx T(\hat{w}^{MP} - w^*) - \frac{T-1}{T}\sum_{t=1}^{T}(\hat{w}^{MP}_{-t} - w^*), \tag{36}$$

$$\begin{aligned}\mathbb{E}[(\hat{w}^{MP} - w^*)'\Sigma(\hat{w}^{MP} - w^*)] \approx\ & T[(\hat{w}^{MP} - w^*)'\tilde{\Sigma}(\hat{w}^{MP} - w^*)] \\ & - \frac{T-1}{T}\sum_{t=1}^{T}[(\hat{w}^{MP}_{-t} - w^*)'\tilde{\Sigma}(\hat{w}^{MP}_{-t} - w^*)]\end{aligned} \tag{37}$$

$\hat{w}^{MP}_{-t}$ refers to the (estimation of the) MP rule, after removing the t-th observation from the dataset, t $\in$ {1,2,...,T}, and $\tilde{\Sigma}$ to the scaled sample covariance matrix, as given in Section 4.2. Removing an observation leads to (slightly) different values for $\hat{\mu}$ and $\hat{\Sigma}$, distinct results for the minimization function (Equation (35)), $\hat{\mu}^{MP}$ and $\hat{\Sigma}^{MP}$, and finally for the weights of the MP rule, $\hat{w}^{MP}_{-t}$.

Resultingly, we find the estimated value of the combination coefficient by plugging in the derivations of $\mathbb{E}[\hat{w}^{MP} - w^*]$ and $\mathbb{E}[(\hat{w}^{MP} - w^*)'\Sigma(\hat{w}^{MP} - w^*)]$, together with $\hat{\pi}_1$ and $\hat{\pi}_2$ as in Equations (24) and (25). Therefore, we gain the following estimated combination rule:

$$\hat{w}^{CMP} = (1 - \hat{\delta}_{mp})w_e + \hat{\delta}_{mp}\hat{w}^{MP}. \tag{38}$$

B.5 The Jorion (1986) Rule

B.5.1 Bayes-Stein Estimators

To obtain the (estimated) PJ rule, we first obtain the (estimated) Bayes-Stein estimators. We assume T>N+4, and find the Bayes-Stein estimator of the mean of the excess return as

$$\hat{\mu}^{BS} = (1-v)\hat{\mu} + v\hat{\mu}_g \vec{1}_N, \tag{39}$$

where $\hat{\mu}_g$ denotes the estimator of the mean excess return on the sample global minimum-variance portfolio (Kan & Zhou (2007)), whilst v indicates the weight given to the target. These are given by

$$\hat{\mu}_g = \frac{\vec{1}'_N \tilde{\Sigma}^{-1}\hat{\mu}}{\vec{1}'_N \tilde{\Sigma}^{-1}\vec{1}_N} = \frac{\vec{1}'_N \hat{\Sigma}^{-1}\hat{\mu}}{\vec{1}'_N \hat{\Sigma}^{-1}\vec{1}_N}, \tag{40}$$

$$v = \frac{N+2}{(N+2) + T(\hat{\mu} - \hat{\mu}_g\vec{1}_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\vec{1}_N)}, \tag{41}$$

with the scaled $\tilde{\Sigma}$ as given in Section 4.2. Using the derivations provided by Kan & Zhou (2007), we obtain the covariance matrix of the excess returns as

$$\begin{aligned}\hat{\Sigma}^{BS} &= (1+\frac{1}{1+\hat{\lambda}})\hat{\Sigma} + \frac{\hat{\lambda}}{T(T+1+\hat{\lambda})\frac{\vec{1}_N\vec{1}'_N}{\vec{1}'_N\hat{\Sigma}^{-1}\vec{1}_N}} \\ &= (1+\frac{1}{1+\hat{\lambda}})\tilde{\Sigma} + \frac{\hat{\lambda}}{T(T+1+\hat{\lambda})\frac{\vec{1}_N\vec{1}'_N}{\vec{1}'_N\tilde{\Sigma}^{-1}\vec{1}_N}},\end{aligned} \tag{42}$$

with

$$\hat{\lambda} = \frac{N+2}{(\hat{\mu} - \hat{\mu}_g\vec{1}_N)'\tilde{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g\vec{1}_N)}. \tag{43}$$

To obtain $(\hat{\Sigma}^{BS})^{-1}$, we incorporate the expression given in Equation (27), in Appendix B of Tu & Zhou (2011). Equation (26) from Tu & Zhou (2011), given as

$$\hat{\Sigma}^{BS} = d\tilde{\Sigma} + DD', \tag{44}$$

tells that d and DD' are defined as

$$d = (1 + \frac{1}{1+\hat{\lambda}}), \tag{45}$$

$$DD' = \frac{\hat{\lambda}}{T(T+1+\hat{\lambda})\frac{\vec{1}_N\vec{1}'_N}{\vec{1}'_N\tilde{\Sigma}^{-1}\vec{1}_N}}, \tag{46}$$

when regarding the expression for $\hat{\Sigma}^{BS}$ given in Equation (42). This gives for $(\hat{\Sigma}^{BS})^{-1}$

$$(\hat{\Sigma}^{BS})^{-1} = \frac{\tilde{\Sigma}^{-1}}{d} - \tilde{\Sigma}^{-1}DD'\frac{\tilde{\Sigma}^{-1}}{d^2 + dD'\tilde{\Sigma}^{-1}D}. \tag{47}$$

In order to find D, we split the expression of DD' as given in Equation (46) into a matrix and a scalar. We require DD' to be a positive semi-definite matrix, and we let $\frac{\vec{1}_N \vec{1}'_N}{\vec{1}'_N \hat{\Sigma}^{-1} \vec{1}_N}$ be the scalar, and the rest of the expression be the matrix. From thereon, we can take the square root of the matrix divided by the scalar, and obtain D. Straightforwardly taking the square root of DD' will not suffice, as then we obtain this matrix solely by executing a simple element multiplication of the two matrices D, instead of an actual matrix multiplication.

B.5.2 Estimated Combination Coefficient and Estimated Combination Rule

Tu & Zhou (2011) imply that it is challenging to obtain the actual value for δ^*_{pj} as the moments of $\hat{w}^{PJ}$ are hard to obtain analytically. Hence, they derive the phrase of the moments in their Appendix, Section B. Due to a certain unclarity in their derivations together with the final expression δ^*_{pj} being rather extensive, we have decided to evaluate the portfolio performance when incorporating the CPJ rule, having the combination coefficient derived in a manner similar to that of the CMP and the CKZ rule. Using either expression is possible as both rules undergo similar issues regarding the analytical assessment of the moments of the estimated portfolio weights. Additionally, the implementation of either replacement procedure is suited as the errors in estimating the combination coefficient are given to be rather insignificant and therefore not too impactful for the results, due to the coefficient being a sole parameter (explained in Section 4.1). Hence, Proposition 3 given by Tu & Zhou (2011) should hold as well for the CPJ rule, that of the CPJ rule outperforming the naive rule whenever these estimation errors are found to be small, together with the naive rule not being exactly equal to the true rule.

The CKZ rule utilizes an expression for $\mathbb{E}[\hat{w}^{KZ} - w^*]$ and $\mathbb{E}[(\hat{w}^{KZ} - w^*)'\Sigma(\hat{w}^{KZ} - w^*)]$, denoted as $\hat{\pi}_{13}$ and $\hat{\pi}_3$ and given in Equations (28) and (29), respectively, leading to the estimated combination coefficient given in Equation (48). For the CPJ rule, we obtain similar estimates of combination coefficients as for the CKZ rule as the input parameters do not change. We must be aware that the KZ rule is designed to decrease the influence of estimation errors by suggesting the possibility to invest in a third fund, whereas the PJ rule introduces Bayes-Stein shrinkage estimators to handle this issue. Hence, incorporation of the derivation of the estimated combination coefficient for the CPJ as for the CKZ rule, therefore, leads to a bias in $\hat{\delta}_{pj}$, which we take into account when interpreting our results, and thus mention in our limitations section, Section E in the Appendix. We however find this bias to be less problematic than if we were to obtain the estimated combination coefficient for the CPJ rule using the Jackknife approach similar to the CMP rule. When unraveling the issue regarding the analytical expression of $\mathbb{E}[\hat{w}^{PJ} - w^*]$ and $\mathbb{E}[(\hat{w}^{PJ} - w^*)'\Sigma(\hat{w}^{PJ} - w^*)]$ similar to the CMP rule, we allow the moments to be obtained re-introducing the Jackknife approach, whereas now we evaluate the portfolio weights obtained for the PJ rule, not the MP rule and its maximum likelihood estimators. Similar to Tu & Zhou (2011), we find that the portfolio, when combining the sophisticated rule with the naive rule whilst implementing the Jackknife approach as for the MacKinlay & Pástor (2000) rule, does not (significantly) increase in performance compared with their uncombined counterpart, together that for in the empirical application, it never outperforms the naive diversification strategy. For our extension, explained in further detail in Section 4.6, we choose to incorporate the sophisticated combination rules solely which do not show these characteristics. Therefore, it would be contradicting to calculate the PJ rule using the Jackknife approach as in the MP rule, as this approach results in such characteristics. Reasons for this remarkable behavior when estimating the given moments using the Jackknife are given in our limitations section, Section E in the Appendix. Thus, for our research, we obtain the estimates for the combination coefficients similar to the KZ rule, thus using the expression for the estimated combination coefficient given in Equation (48):

$$\hat{\delta}_{pj} = \frac{\hat{\pi}_1 - \hat{\pi}_{13}}{\hat{\pi}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3}. \tag{48}$$

We obtain the following estimated combination rule, whilst assuming T>N+4:

$$\hat{w}^{CPJ} = (1 - \hat{\delta}_{pj})w_e + \hat{\delta}_{pj}\hat{w}^{PJ}. \tag{49}$$

C Methodology for Examining Performance

C.1 Performance Criteria

C.1.1 Simulated Application

We obtain for a certain rule c the following expression for the Sharpe ratio, similar to DeMiguel et al. (2009):

$$\widehat{SR}_c = \frac{\mu' \hat{w}_c}{\sqrt{\hat{w}_c' \Sigma \hat{w}_c}}, \tag{50}$$

where we utilize true mean and covariance matrix of returns to obtain the Sharpe ratios and therefore to compare the relative performance of a rule, as explained in Section 4.1.

We consider the Sharpe ratio, also known as the "reward-to-variability ratio" (Sharpe, 1966, p.123), to evaluate portfolio performance, since it is an incredibly popular customary and thus trustworthy technique. For either the utility or the Sharpe ratio of a certain rule c, we have for the estimated standard errors $\widehat{SE}_c = \frac{\hat{\sigma}}{\sqrt{T}}$. The standard errors aid in depicting whether the values obtained, for either the utility or the Sharpe ratio of the rule, are representative in order to draw verifiable conclusions. As clarified at the beginning of this section, we obtain 10,000 values for the utility and Sharpe ratio of a certain rule, from which we obtain the estimated standard deviation $\hat{\sigma}$. We, therefore, have for the standard errors that $\hat{\sigma}$ refers to the estimated standard deviation of the values, and T is the sample size considered. We prefer the rules that give portfolio strategies with the highest utilities and Sharpe ratios.

C.1.2 Empirical Application

The Certainty-Equivalent-Return (CER) is a popular performance-assessing instrument and tells the best portfolio strategy, as a higher CER translates to a better performance of the rule. The CER namely gives the minimum risk-free rate an investor is willing to accept in order to choose to not invest in a certain portfolio strategy. We have for a given rule c:

$$CER_c = \hat{\mu}_c - \frac{\gamma}{2}\hat{\sigma}_c^2, \tag{51}$$

with $\hat{\mu}_c$ calculated as the average of the $T - M$ realized excess returns of a portfolio, and its standard deviation being $\hat{\sigma}_c$. Similar to Tu & Zhou (2011) and mentioned in Section 2, we do not incorporate the value of the average realized risk-free rate, as this is common across all portfolio strategies, and therefore is canceled out at comparison.

D Tables and Figures

D.1 Figures Corresponding with Table 1

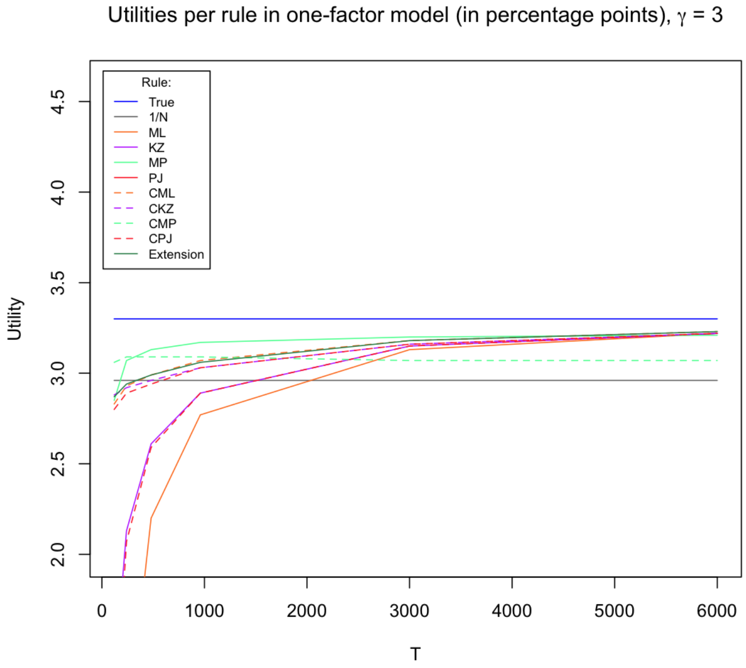

Figure 1: Utilities per rule in a one-factor model when $\gamma = 3$ without mispricing, given in percentage points

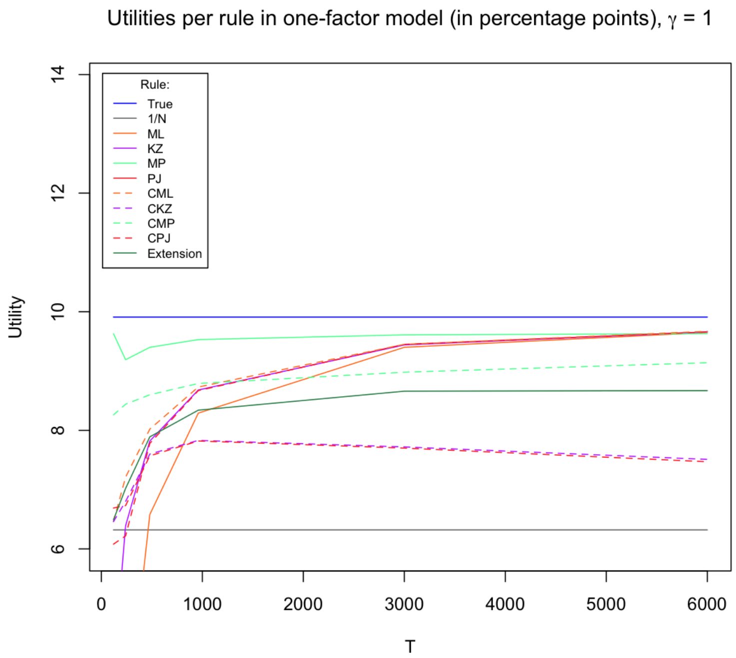

Figure 2: Utilities per rule in a one-factor model when $\gamma = 1$ without mispricing, given in percentage points

D.2 Figures Corresponding with Table 2

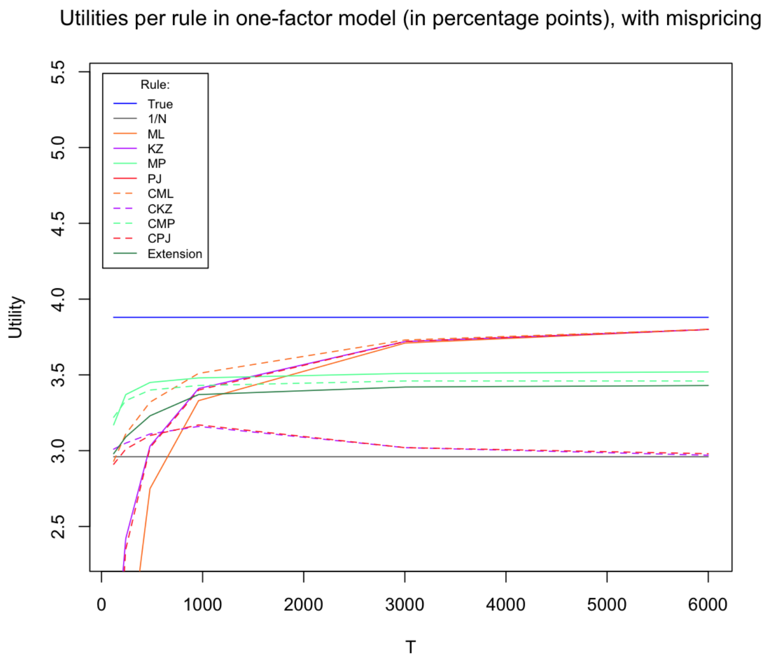

Figure 3: Utilities per rule in a one-factor model with mispricing and $\gamma = 3$, given in percentage points

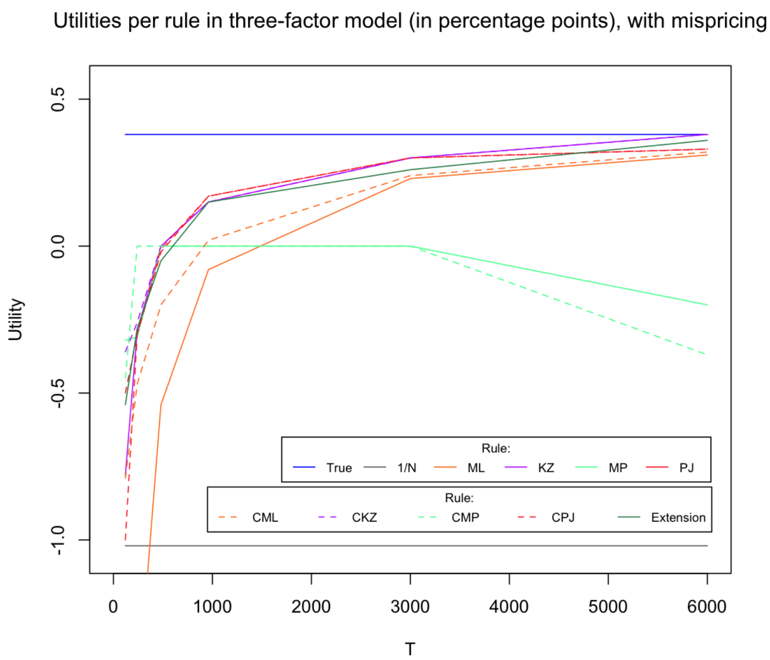

Figure 4: Utilities per rule in a three-factor model with mispricing and $\gamma = 3$, given in percentage points

D.3 Figures Corresponding with Table 3, Table 8 and Figures Corresponding with Table 8

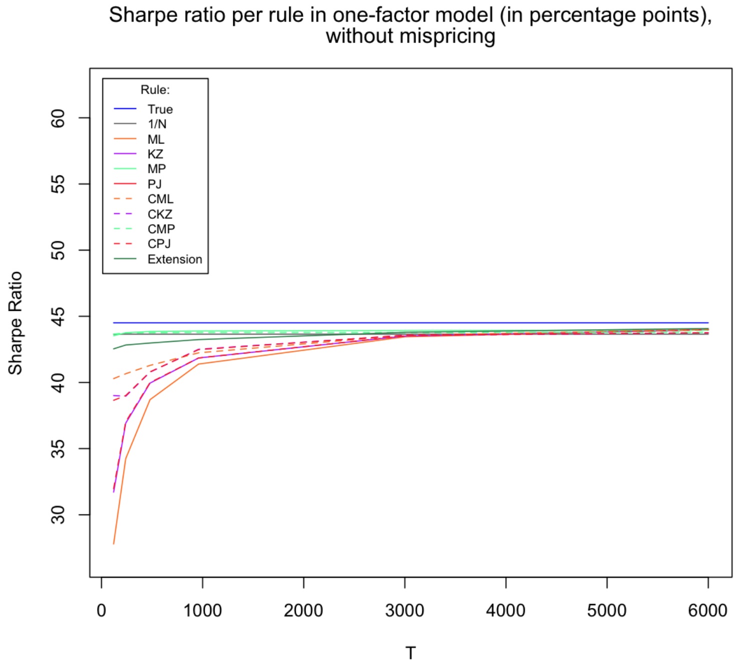

Figure 5: Sharpe ratios per rule in a one-factor model without mispricing and $\gamma = 3$, given in percentage points

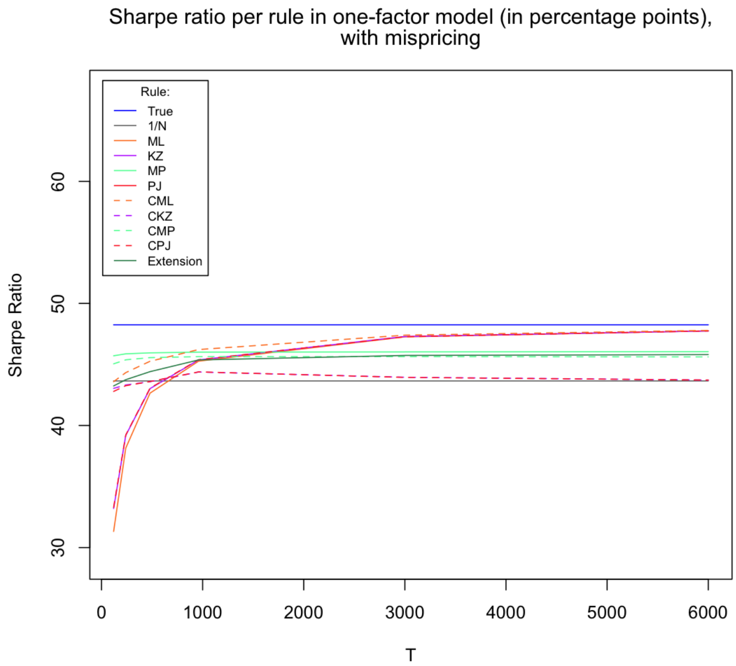

Figure 6: Sharpe ratios per rule in a one-factor model with mispricing and $\gamma = 3$, given in percentage points

Table 8: Sharpe ratios in three-factor model, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, and these rules whenever they are combined with the naive rule. In the three-factor model, we let the alphas of mispricing be either zero (given in Panel A) or evenly spread between [-0.02,0.02] (given in Panel B), thus in the second case allowing it to be present. We have for the number of risky assets $N = 25$ and the number of repetitions being 10,000, for all the values of the sample size T considered.

	T					
Rule	**120**	**240**	**480**	**960**	**3000**	**6000**
Panel A: $\alpha = 0$						
True	6.30	6.30	6.30	6.30	6.30	6.30
1/N	5.72	5.72	5.72	5.72	5.72	5.72
ML	0.76	1.18	1.67	2.27	3.63	4.45
KZ	1.18	1.81	2.61	3.46	4.83	5.46
MP	5.13	5.41	5.39	5.31	5.61	5.51
PJ	1.21	1.85	2.65	3.53	4.91	5.43
CML	4.76	4.67	4.55	4.44	4.64	4.95
CKZ	3.43	2.48	2.80	3.57	4.88	5.48
CMP	5.27	5.52	5.50	5.45	5.67	5.60
CPJ	3.31	2.49	2.84	3.63	4.96	5.46
Extension	3.55	4.04	3.68	4.00	4.82	5.26
Panel B: $\alpha \in$ [-0.02,0.02]						
True	15.00	15.00	15.00	15.00	15.00	15.00
1/N	5.72	5.72	5.72	5.72	5.72	5.72
ML	4.19	5.99	8.13	10.16	12.91	13.84
KZ	4.46	6.53	8.41	10.69	14.21	13.88
MP	8.98	9.36	9.28	9.29	9.24	9.04
PJ	4.62	6.25	8.74	10.48	13.54	14.94
CML	6.56	7.53	8.95	10.55	12.98	13.86
CKZ	6.00	6.74	8.40	10.67	14.21	13.88
CMP	8.44	8.53	8.35	8.30	8.24	8.10
CPJ	6.04	6.48	8.73	10.47	13.54	14.94
Extension	6.35	7.03	9.46	11.43	12.08	14.88

We additionally report the Sharpe ratios in a three-factor model without (Panel A) and with (Panel B) the presence of mispricing, given in Table 8. We do so as we found our extension to perform quite well in the latter setting, therefore we are interested in evaluating the corresponding Sharpe ratios. These are plotted in Figures 7 and 8. The goal of exhibiting these results is to understand the impact of mispricing in the three-factor model and to understand in which setting (double) combining is most beneficial. The decisions on the initialization of the three-factor model are further clarified in Section 3.1.

We find when $\alpha = 0$, the sophisticated rules outperform the naive rule solely when $T = 6000$. When $\alpha \in$ [-0.02,0.02], this is the case when $T \geq 960$. We affirm the findings by Tu & Zhou (2011), as we find that combining a sophisticated rule with the naive rule improves portfolio performance in terms of Sharpe ratios, for any T, whether mispricing is present or not. We devote the exception, that of a decrease in performance when combining the MP with the naive rule when we allow mispricing to be present, to the sizeable errors arising when estimating the combination coefficient in this setting.

Our extension outperforms the naive rule when mispricing is present, whilst it does not when this is not the case. It generally improves performance in contrast to the sophisticated rules, whilst it does frequently not in contrast to the combination rules. We find that an increase in performance by implementing double combining is more likely when we allow mispricing to be present.

Whenever we allow mispricing to be present, it outperforms the naive and sophisticated rules for any T. The extension performs better than the CKZ and CPJ rule for any T, whilst it is outperformed by the CML rule when we have $T \leq 240$, which is in contrast to its performance in the one-factor model with mispricing, as displayed in Table 2. Thus, we find that equally double combining generally suffices to outperform its single counterparts.

When we set $\alpha = 0$, Panel A shows that the naive rule is never outperformed by our extension. In Section 6.3, we clarify that without mispricing, the factor portfolio is on the efficient frontier, leading to the naive rule being able to attain both a correct composition and proportion of the portfolio (Tu & Zhou (2011)). Additionally, as clarified in Section 3.1, the three-factor model is a more realistic model. Consequentially, the naive rule performs even better compared to its performance in the one-factor model, causing it to be challenging to beat. We find that in this setting, combining the sophisticated rules with the naive rule always improves portfolio performance, whilst our extension is always outperformed by the CML rule, whilst it is outperformed by the CKZ and CPJ if $T \geq 3000$, therefore performing slightly worse than in the one-factor model without mispricing. These findings highlight that equally double combining rules does not suffice to improve portfolio performance even further when we allow mispricing to not be present.

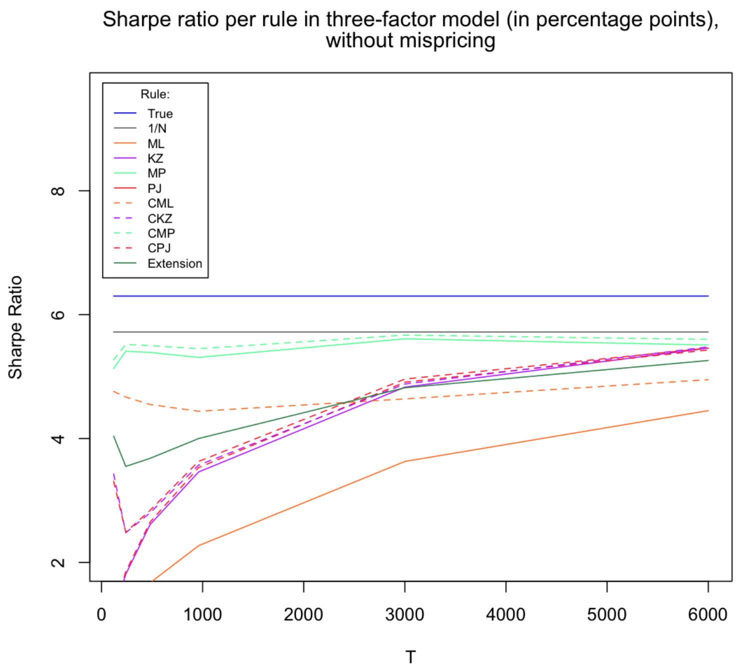

Figure 7: Sharpe ratios per rule in a three-factor model without mispricing and $\gamma = 3$, given in percentage points

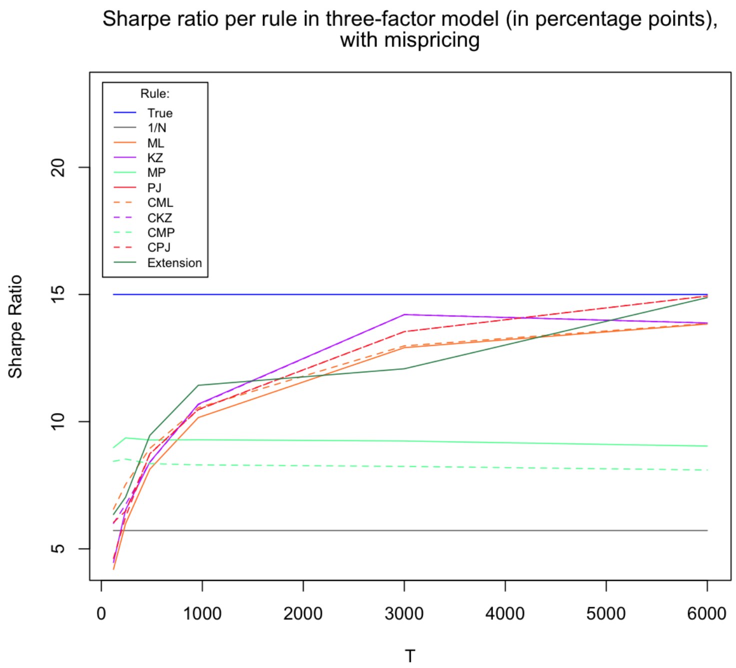

Figure 8: Sharpe ratios per rule in a three-factor model with mispricing and $\gamma = 3$, given in percentage points

D.4 Figure Corresponding with Table 5

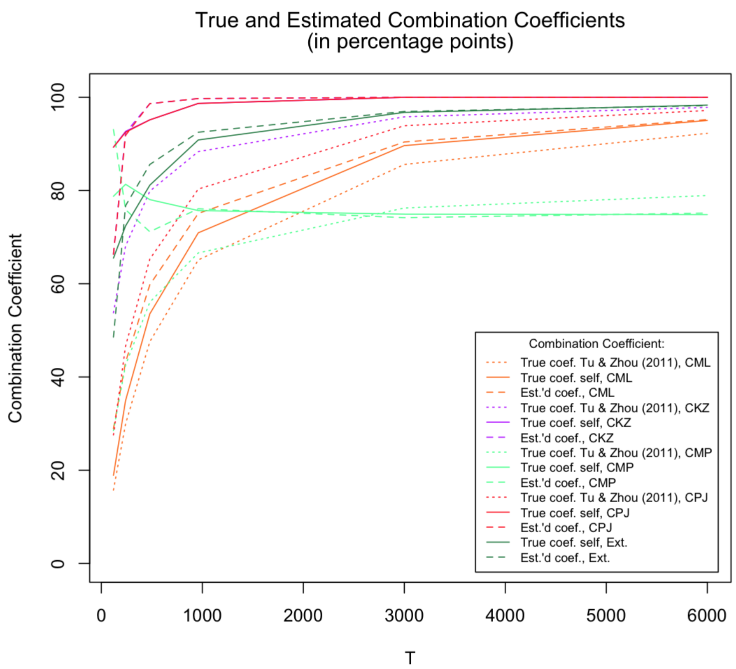

Figure 9: True and estimated combination coefficient for the various combination rules, given in percentage points

D.5 Tables 9 and 10: Sharpe Ratio and Drawdown per Real Dataset

Table 9: Sharpe ratio per portfolio strategy, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, these rules whenever they are combined with the naive rule (CML, CKZ, CMP and CPJ) and our extension. The Sharpe ratios (and their standard errors given in parentheses below) of an MV investor for the distinct portfolio strategies when we set the rolling window M = 120 are given in Panel A, whilst Panel B displays these values when setting M = 240. To calculate the true or the in-sample ML rule, we use the complete dataset. The datasets considered are clarified in Section 3.2. We set γ = 3. The number of repetitions is 10,000.

Rule	**Industry** **N = 10+1**	**Interna-tional** **N = 8+1**	**Mkt/ SMB/HML** *N = 3*	**FF20 1-Factor** *N = 20+1*	**FF20 4-Factor** *N = 20+4*	**FF25 3-Factor** *N = 25+3*	**Indu49 3-Factor** *N = 49+3*
Panel A: M = 120							
In-Sample ML	87.63 (0.00)	21.76 (0.00)	22.64 (0.00)	64.67 (0.00)	113.72 (0.00)	118.72 (0.00)	103.59 (0.00)
1/N	19.91 (0.00)	18.36 (0.00)	17.57 (0.00)	19.59 (0.00)	20.65 (0.00)	20.83 (0.00)	17.99 (0.00)
ML	65.06 (0.17)	16.64 (0.13)	18.20 (0.12)	47.74 (0.38)	83.27 (0.27)	81.71 (0.34)	58.49 (0.33)
KZ	65.07 (0.17)	18.23 (0.12)	19.28 (0.11)	47.57 (0.38)	83.08 (0.26)	81.70 (0.34)	58.49 (0.33)
MP	18.75 (0.19)	16.83 (0.24)	−0.25 (0.27)	19.52 (0.32)	20.14 (0.40)	19.88 (0.42)	13.40 (0.30)
PJ	65.07 (0.17)	18.46 (0.12)	19.69 (0.10)	47.55 (0.38)	83.06 (0.26)	81.70 (0.34)	58.49 (0.33)
CML	65.10 (0.17)	16.96 (0.12)	18.77 (0.11)	48.00 (0.36)	83.34 (0.27)	81.72 (0.34)	58.53 (0.33)
CKZ	62.73 (0.51)	18.84 (0.08)	19.32 (0.09)	47.63 (0.38)	83.17 (0.26)	81.86 (0.35)	58.66 (0.33)
CMP	18.75 (0.19)	16.83 (0.24)	7.87 (0.43)	19.56 (0.32)	21.60 (0.21)	21.87 (0.27)	15.17 (0.26)
CPJ	65.31 (0.17)	18.89 (0.08)	19.65 (0.08)	47.57 (0.37)	80.96 (0.58)	81.74 (0.35)	58.71 (0.32)
Extension	64.98 (0.18)	18.12 (0.11)	20.05 (0.08)	47.74 (0.38)	83.22 (0.26)	81.79 (0.35)	58.59 (0.33)
Panel B: M = 240							
In-Sample ML	87.63 (0.00)	21.76 (0.00)	22.64 (0.00)	64.67 (0.00)	113.72 (0.00)	118.72 (0.00)	103.59 (0.00)
1/N	19.91 (0.00)	18.36 (0.00)	17.57 (0.00)	19.59 (0.00)	20.65 (0.00)	20.83 (0.00)	17.99 (0.00)
ML	68.97 (0.32)	20.02 (0.05)	21.20 (0.05)	56.20 (0.10)	91.90 (0.23)	91.57 (0.31)	72.33 (0.30)
KZ	68.99 (0.32)	20.61 (0.02)	21.87 (0.03)	56.18 (0.10)	91.71 (0.25)	91.57 (0.31)	72.32 (0.30)
MP	19.14 (0.19)	18.13 (0.02)	−0.78 (0.26)	21.06 (0.22)	23.71 (0.37)	24.83 (0.29)	14.63 (0.26)
PJ	68.99 (0.32)	20.60 (0.02)	22.01 (0.03)	56.17 (0.10)	91.70 (0.25)	91.57 (0.32)	72.32 (0.30)
CML	69.02 (0.32)	20.10 (0.05)	21.43 (0.05)	56.37 (0.09)	91.99 (0.23)	91.57 (0.31)	72.34 (0.30)
CKZ	69.34 (0.31)	20.61 (0.02)	21.72 (0.05)	56.18 (0.10)	91.77 (0.24)	91.71 (0.29)	72.39 (0.29)
CMP	19.14 (0.19)	18.13 (0.02)	6.71 (0.42)	21.06 (0.22)	23.20 (0.26)	24.53 (0.27)	15.46 (0.24)
CPJ	68.79 (0.35)	20.61 (0.02)	21.89 (0.05)	56.30 (0.08)	81.88 (1.12)	91.28 (0.26)	72.19 (0.29)
Extension	69.18 (0.32)	20.61 (0.03)	21.86 (0.03)	56.25 (0.10)	91.84 (0.24)	91.64 (0.30)	72.37 (0.30)

Table 10 displays the Sharpe ratio per diversification rule for the various real datasets considered. The goal of exhibiting these results is to understand whether (double) combining rules allows for an increase in the risk-adjusted returns. We implement a rolling-window approach as clarified in Section 5.1 with either a length of $M = 120$ (Panel A) or $M = 240$ (Panel B). Recall that the true, or in-sample ML rule, is obtained utilizing the complete dataset (explanation is given in Section 5.1).
All sophisticated rules, (double) combined or not, outperform the naive rule in terms of Sharpe ratios, for any M. This is because the naive rule is not close to the true rule, causing it to not be hard to beat.

We find that for $M = 120$, combining the sophisticated rules with the naive rule in general increases portfolio performance, for any dataset considered. Double combining occasionally improves performance even further, depending on the combination rule and dataset considered. The standard errors in general decrease whenever we (double) combine rules. For $M = 240$, we again in general find an increase in performance. Standard errors both in- and decrease when combining. When we double combine rules, occasionally, portfolio performance is increased compared to the CML, CKZ and CPJ rules. Therefore, stating whether our extension further improves performance is ambiguous. We do find it to outperform all the combination rules when we consider the "International" dataset. For both values of M, the ML and CML rules perform the worst (when disregarding the MP and CMP rules) compared to the other sophisticated and combination rules, respectively.

In contrast to the findings for Table 6, we find the "Mkt/SMB/HML" strategy to perform the worst in terms of Sharpe ratios (where in Table 6, it performed decently), and that the industry portfolios ("Industry" and "Indu49 3-Factor") perform decently now (instead of the worst). We again find the inclusion of the Momentum factor to significantly increase the portfolio performance (when we compare the Sharpe ratios of "FF20 1-Factor" with "FF20 3-Factor").

Table 10: Drawdown per portfolio strategy, in percentage points.
We consider as the diversification strategies the true rule, the naive ($1/N$) rule, and the sophisticated rules, given to be the Markowitz (1952) rule, the Kan & Zhou (2007) rule, the MacKinlay & Pástor (2000) and Jorion (1986) rule, these rules whenever they are combined with the naive rule (CML, CKZ, CMP and CPJ) and our extension. The drawdown of an MV investor for the distinct portfolio strategies when we set the rolling window M = 120 is given in Panel A, whilst Panel B displays this value when setting M = 240. To calculate the true or the in-sample ML rule, we use the complete dataset. The datasets considered are clarified in Section 3.2. We set $\gamma = 3$. The number of repetitions is 10,000.

Rule	Industry N = 10+1	International N = 8+1	Mkt/SMB/HML *N = 3*	FF20 1-Factor *N = 20+1*	FF20 4-Factor *N = 20+4*	FF25 3-Factor *N = 25+3*	Indu49 3-Factor *N = 49+3*
Panel A: M = 120							
In-Sample ML	*207.21*	*232.91*	*156.44*	*167.65*	*124.43*	*133.90*	*139.60*
1/N	266.25	255.54	197.81	242.71	244.27	241.21	273.85
ML	135.16	224.46	290.53	181.36	179.13	139.71	140.87
KZ	134.64	263.88	273.14	186.13	178.51	139.56	140.40
MP	341.63	264.18	207.92	387.83	299.81	259.64	222.73
PJ	134.53	270.74	269.11	186.79	178.44	139.54	140.34
CML	134.49	235.33	290.35	184.46	179.91	139.64	140.40
CKZ	156.53	263.88	265.55	174.83	169.41	143.56	142.74
CMP	341.63	264.18	243.56	387.83	305.98	259.64	217.45
CPJ	153.56	270.74	267.81	164.79	148.24	141.28	140.46
Extension	145.97	257.00	271.78	178.07	177.08	141.69	141.59
Panel B: M = 240							
In-Sample ML	*207.21*	*232.91*	*156.44*	*167.65*	*124.43*	*133.90*	*139.60*
1/N	297.92	247.33	352.90	364.06	343.82	375.98	334.89
ML	197.30	197.94	218.30	143.00	147.74	178.41	156.56
KZ	196.21	263.89	219.43	143.28	147.89	178.80	156.47
MP	310.47	382.91	138.62	391.28	157.13	296.49	324.94
PJ	195.97	274.65	220.04	143.32	147.91	178.85	156.47
CML	195.88	216.95	218.74	142.33	147.64	179.39	156.52
CKZ	248.39	249.37	219.43	143.28	146.11	175.88	156.46
CMP	310.47	378.28	255.01	391.28	158.31	296.49	323.05
CPJ	232.58	272.27	220.04	143.32	146.77	178.68	155.97
Extension	224.67	242.71	219.44	142.98	146.67	177.33	156.49

Table 10 displays the drawdown per diversification rule for the various real datasets considered. The goal of exhibiting these results is to understand the riskiness of a certain portfolio rule, and whether combining allows for a decrease in it. Additionally, these results describe which portfolio strategy is considered the riskiest, namely those with a higher drawdown per rule. We implement a rolling-window approach as clarified in Section 5.1 with either a length of $M = 120$ (Panel A) or $M = 240$ (Panel B). Recall that the true, or in-sample ML rule, is obtained utilizing the complete dataset (explanation is given in Section 5.1). We explain in Section 5.2.2 that this rule now however is not too informative, as we choose the objective of our research to be maximizing the utility of an investor, which is equivalent to maximizing the Sharpe ratio of a portfolio strategy (Tu & Zhou (2011)). Thus now, our true rule will not necessarily give the minimum drawdown, and thus cannot be used as a benchmark now. We do not regard these findings when interpreting the table.

For any M, the naive rule is not necessarily outperformed by the sophisticated rules in terms of drawdown. We find that for $M = 120$, combining the sophisticated rules often decreases the drawdown, which can be interpreted as removing some of the risks a dataset undergoes, as a lower drawdown translates to a smaller absolute drop in returns since the maximum return is achieved.

Our extension does not dominate in terms of drawdown when comparing with the combination rules, together with the sole sophisticated rules and the naive rule. For $M = 240$, combining again seems to increase portfolio performance. Our extension does not necessarily decrease drawdown. We find that our extension generates the lowest drawdown for the "Industry", "FF25 3-Factor" and "Indu49 3-Factor" strategies when $M = 120$, whilst these being the "FF20 1-Factor", "FF20 3-Factor" and "Indu49 3-Factor" strategies when we set $M = 240$.

When regarding the drawdown for $M = 240$ for the extension, we find the strategies with the smallest value, the "FF20 1-Factor", "FF20 3-Factor" and "Indu49 3-Factor" strategies, to be as well the strategies with the highest Sharpe ratios as given in Table 9.

We find for both performance criteria considered in this section that the (combined) MP rule performs poorly. These rules are often outperformed by the naive rule, almost independent of the length we assume for the window.

E Limitations

When conducting our research, we come across some limitations. The values of our results happen to not exactly equal to those of Tu & Zhou (2011). As we have replicated their research exactly similar to our knowledge, we are not able to assign a direct underlying reason for this occurrence. The differences in actual values however in general do not affect our conclusions regarding our objective, that of investigating whether combining rules improves portfolio performance. We find our rules to still perform similarly when comparing with Tu & Zhou (2011). Hence, the conclusions drawn by us and therefore the credibility of our research remains predominantly strong.

For the simulated application, we infrequently find the combination rules, CKZ and CPJ, to converge to the naive rule, instead of the true rule by definition. We suggest this, especially as the sole sophisticated rules do converge towards the true rule, to rely on the fact that the combination coefficients of these rules strongly shrink the sophisticated towards the naive rule.

Besides this, we find, similar to Tu & Zhou (2011), that combining the MP rule with the naive rule decreases performance. We however additionally find the CMP rule to decrease in performance as T increases, thus it does not converge to the true rule. We devote this to the errors arising when estimating the combination coefficient. Similar to Tu & Zhou (2011), we incorporate the Jackknife approach. We however choose to not re-estimate the MP rule when deleting the t-th observation of the dataset, with $t \in \{1,2,...,T\}$ as this is computationally very expensive, especially if we wish to simulate the rule 10,000 times. Therefore, we choose to solely regard the MP when excluding the T-th observation, and thus calculate this rule once. This difference in obtaining the estimate of the combination coefficient could help explain why we find the CMP rule to decrease in value as T increases, instead of it converging to the true rule. Additionally to highlight is that for the true combination coefficient of the CMP rule, we do not re-estimate the MP rule when disregarding the T-th observation of the dataset. This is as the difference in portfolio weights obtained is not grand, and re-estimating the MP rule when deleting the T-th observation using the true parameters is computationally costly. This could explain why we find our estimated combination coefficients for the CMP rule to be close to the true values, closer to the true values than the estimates found by Tu & Zhou (2011). Furthermore, we notice the MP rule to be extremely dependent on the input parameters of the minimization function required to obtain the maximum likelihood estimators of the rule. We choose these parameters using reasoning, as we follow the assumptions made by Tu & Zhou (2011) on the CAPM. We however can be critical on these values, as, for example, Tu & Zhou (2011) assume that the diagonal values of the covariance matrix of the residuals are equal per asset considered, whilst we draw these values from a uniform distribution (following the assumptions of Tu & Zhou (2011) on the initialization of the factor models) and therefore not necessarily having this assumption to hold. Besides this, we are required to predetermine the values of the factor loadings for the three-factor model, which impacts results once more. Lastly, when minimizing the function required to obtain the maximum likelihood estimators, we occasionally do not generate output, as the log function in the objective function does not allow its input variable, σ^2 to be any value. This will affect the results regarding the (combined) MP rule again. Altogether, the MP and CMP do not necessarily perform reliably.

F Programming Code

We provide the programming code per diversification rule considered, for both the simulated and empirical application, by handing it in via Sin-Online. We have programmed all rules in RStudio. For the simulated application, per rule, we deliver two codes: one generates the performance of the rule in the one-factor model, whilst the other calculates the performance of the rule in the three-factor model. We extensively clarify the code by making use of comments, additional to the theory given in the paper. We additionally provide the code that delivers the plots, given in Section D in the Appendix.

References

Ackermann, F., Pohl, W. & Schmedders, K. (2017). Optimal and naive diversification in currency markets. *Management Science*, *63*(10), 3347–3360.

Barry, C. B. (1974). Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, *29*(2), 515–522.

Ben-David, I., Franzoni, F. & Moussawi, R. (2017). Exchange-traded funds. *Annual Review of Financial Economics*, *9*, 169–189.

Black, F. & Litterman, R. (1992). Global portfolio optimization. *Financial analysts journal*, *48*(5), 28–43.

Bonaccolto, G. & Paterlini, S. (2020). Developing new portfolio strategies by aggregation. *Annals of Operations Research*, *292*(2), 933–971.

Brandt, M. W. (2010). Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques* (pp. 269–336). Elsevier.

Brandt, M. W., Santa-Clara, P. & Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, *22*(9), 3411–3447.

Brown, S. J. (1976). *Optimal portfolio choice under uncertainty: a bayesian approach.* The University of Chicago.

Chekhlov, A., Uryasev, S. & Zabarankin, M. (2004). Portfolio optimization with drawdown constraints. In *Supply chain and finance* (pp. 209–228). World Scientific.

DeMiguel, V., Garlappi, L. & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, *22*(5), 1915–1953.

DeMiguel, V., Martin-Utrera, A. & Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, *37*(8), 3018–3034.

Elton, E. J. & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of banking & finance*, *21*(11-12), 1743–1759.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Fama, E. F. & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The journal of finance*, *51*(1), 55–84.

James, W. & Stein, C. (1992). Estimation with quadratic loss. *Breakthroughs in statistics: Foundations and basic theory*, 443–460.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis*, *21*(3), 279–292.

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

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance*, *20*(4), 587–615.

MacKinlay, A. C. & Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of financial studies*, *13*(4), 883–916.

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

Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments.* Yale University Press.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(1), 31–42.

Pástor, L. & Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, *56*(3), 335–381.

Platanakis, E., Sutcliffe, C. & Urquhart, A. (2018). Optimal vs naïve diversification in cryptocurrencies. *Economics Letters*, *171*, 93–96.

Quenouille, M. H. (1949). Problems in plane sampling. *The Annals of Mathematical Statistics*, 355–375.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, *43*(3/4), 353–360.

Rom, B. M. & Ferguson, K. W. (1994). Post-modern portfolio theory comes of age. *Journal of investing*, *3*(3), 11–17.

Ross, S. A. (1978). The current status of the capital asset pricing model (capm). *The Journal of Finance*, *33*(3), 885–901.

Shao, J. & Tu, D. (2012). *The jackknife and bootstrap.* Springer Science & Business Media.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425–442.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, *39*(1), 119–138.

Stein, C. (1956). *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution* (Tech. Rep.). Stanford University Stanford United States.

Tu, J. & Zhou, G. (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, *99*(1), 204–215.

Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, 673–705.