AP 2013–03

Chair for Applied Stochastics and Risk Management

Helmut Schmidt Universität
Universität der Bundeswehr Hamburg
Faculty of Economics and Social Sciences
Department of Mathematics/Statistics

Working Paper

A Theoretical Foundation of Portfolio Resampling

Gabriel Frahm

May 27, 2014

A Theoretical Foundation of Portfolio Resampling

Gabriel Frahm

Helmut Schmidt University
Faculty of Economics and Social Sciences
Department of Mathematics/Statistics
Chair for Applied Stochastics and Risk Management
Holstenhofweg 85, D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2791
E-mail: frahm@hsu-hh.de

Working Paper
Please use only the latest version of the manuscript. Distribution is unlimited.

Supervised by: Prof. Dr. Gabriel Frahm
Chair for Applied Stochastics and Risk Management

URL: www.hsu-hh.de/stochastik

A Theoretical Foundation of Portfolio Resampling*

Gabriel Frahm†
Helmut Schmidt University
Department of Mathematics/Statistics
Chair for Applied Stochastics and
Risk Management

May 27, 2014

Abstract

A portfolio-resampling procedure invented by Richard and Robert Michaud is the subject of a highly controversial discussion and big scientific dispute. It has been evaluated in many empirical studies and Monte Carlo experiments. Apart from the contradictory findings, the Michaud approach still lacks a theoretical foundation. I prove that portfolio resampling has a strong foundation in the classic theory of rational behavior. Every noise trader could do better by applying the Michaud procedure. By contrast, a signal trader who has enough prediction power and risk-management skills should refrain from portfolio resampling. The key note is that in most simulation studies, investors are considered as noise traders. This explains why portfolio resampling performs well in simulation studies but could be mediocre in real life.

Keywords: Asset allocation, Mean-variance analysis, Noise trader, Out-of-sample performance, Portfolio resampling, Resampled efficiency, Signal trader.

JEL Subject Classification: Primary G11, Secondary D81.

*Many thanks belong to Uwe Jaekel, Bob Korkie, Christoph Memmel, Gert Mittring, Winfried Pohlmeier, Rainer Schüssler, and Niklas Wagner for their fruitful comments and suggestions. Moreover, I thank all participants of the workshop on portfolio optimization in Konstanz (2012) for their contributions.
†Phone: +49 40 6541-2791, e-mail: frahm@hsu-hh.de.

1. Motivation

This paper provides a theoretical foundation of portfolio resampling. Based on the classic theory of rational behavior (von Neumann and Morgenstern, 1944), I derive simple conditions under which portfolio resampling can be justified. The given results do not require any specific assumption on the investment strategy, probability distribution of asset returns, or on the utility function of the investor. I clarify the impact of uncertainty on capital allocation in a very general setting. The methodological framework allows me to obtain universal and reliable statements on this subject.

This work is inspired by a famous experiment conducted by Harry Markowitz and Nilufer Usmen (2003). They compared a Bayesian portfolio optimization method with a portfolio-resampling technique invented by Richard Michaud and Robert Michaud (Michaud, 1998).[1] Markowitz and Usmen expected the Bayesian strategy to do better than the resampling approach but, surprisingly, the Michaud strategy won the battle.

Markowitz and Usmen (2003) frankly state that,

> "the results represent something of a crisis for the theoretical foundations of portfolio theory."

At the end of their paper, Markowitz and Usmen raise some basic questions, which can be stated as follows:

Q1. *How does Michaud's procedure relate to the classic theory of rational behavior according to von Neumann and Morgenstern (1944)?*

Q2. *Why does portfolio resampling performs so good in the experiment?*

Q3. *How much does the Michaud procedure contribute to the risk-adjusted expected return in practice? This means after taking the true distribution of asset returns, transaction costs, etc., into account.*

Q4. *Are Savage's axioms of subjective probability (1954) violated by the Michaud player and if so, do we have to blame the Michaud player or the axioms?*

Q5. *Does the Bayesian strategy perform so bad, because the chosen priors are incorrect or because of a poor numerical approximation of the posteriors?*

First of all, these questions are highly relevant from an academic point of view but, nevertheless, Markowitz and Usmen (2003) emphasize that their experiment has

[1] The Michaud approach is a US patented procedure with patent number 6,003,018.

also practical implications for mean-variance analysis. It is still unclear why Bayesian analysis, which is a sophisticated method, apparently does not succeed over such a relatively simple method like portfolio resampling.

Michaud's portfolio resampling is a patented procedure. This means nobody can use or distribute the algorithm without license or permission. According to Kovaleski (2001), Michaud's firm *New Frontier Advisors* "has licensed the optimizer to about 10 financial institutions, including money managers, financial services companies and consultants." Hence, some practitioners might wish to know why, and under which circumstances, portfolio resampling could be useful in real-life applications.

The portfolio-resampling approach has been published in Michaud (1998) and only a few years later it has become the subject of a highly controversial discussion and big scientific dispute. In the following I will give an overview on the literature. It is easy to divide the list of authors into the proponents and opponents of the Michaud approach.

Fletcher and Hillier (2001) compare the out-of-sample performance of the Michaud procedure with traditional mean-variance analysis based on empirical data of equity indices between 1983 and 2000. They consider several estimation approaches for the mean excess returns: The historical sample mean, the James-Stein shrinkage estimator, a single-factor estimator, and a conditional model. Their study reveals that portfolio resampling does slightly better than traditional mean-variance analysis, although the results are not significant.

With regard to the Fletcher-Hillier study, Michaud and Michaud (2003) warn that "A casual reader might conclude that resampled-efficiency improvements are not of investment significance." Their opinion is that a backtest cannot provide reliable information, since a "good" strategy may perform poorly and a "bad" strategy may perform well in any given period. Moreover, Michaud and Michaud (2003) point out that, "This is why I used simulation methods to prove the superiority of resampled efficiency."

Scherer (2002) tries to explain the potential benefits and pitfalls of the Michaud procedure. He concludes that portfolio resampling can be useful to develop statistical tests for the difference between two portfolios but at the end of his paper he writes, "What is not clear, however, is why averaging over resampled portfolio weights should represent an optimal portfolio construction solution to deal with estimation error." Further, he argues that portfolio resampling should not be compared with the naive Markowitz approach, where sample means and sample (co-)variances are treated like true moments and thus estimation risk is completely neglected. He recommends to use Bayesian portfolio optimization strategies instead.

In fact, Markowitz and Usmen (2003) apply a Bayesian strategy in their experiment but, in contrast to Fletcher and Hillier (2001), they conduct a Monte Carlo simulation and not an empirical study. As already mentioned, they come to the conclusion that the Michaud procedure dominates the Bayesian procedure. At the same time, Scherer (2004) propagates Bayesian strategies as a better alternative to portfolio resampling and asserts that this is the "correct way to deal with estimation error." He concentrates on the drawbacks of portfolio resampling and concludes that the Michaud approach has "serious statistical and decision theoretic limitations."

By contrast, Kohli (2005) conducts an empirical study based on stock-market data from 2011 to 2013 and find "no conclusive advantage or disadvantage of using resampling as a technique to obtain better returns," but "resampled portfolios do seem to offer higher stability and lower transaction costs." Scherer (2006) runs a new Monte Carlo simulation, which reveals that the James-Stein shrinkage estimator outperforms portfolio resampling. Nevertheless, the Michaud procedure is able to outperform the naive Markowitz approach but "the exact mechanics remain unclear (and unformulated)." Similarly, Scherer (2007) states that portfolio resampling is a "practitioner-based heuristics with no rooting in decision theory." Finally, Scherer (2006) notes that it is impossible to compare portfolio resampling with Bayesian procedures: "For every distribution, a prior will be found that will outperform resampling (and vice versa). Equally, for every prior, a distribution will be found where resampling outperforms."

Herold and Maurer (2006) compare Bayesian strategies with heuristic approaches including portfolio resampling. In their study they distinguish between "conditional" and "unconditional" strategies. Unconditional strategies assume that asset returns are serially independent and identically distributed (i.i.d.). By contrast, the conditional strategies are based on the assumption that asset returns are predictable. Their study incorporates a long history of US stock-market data and a shorter history of European stock prices. It turns out that the Michaud procedure does not significantly improve the Sharpe ratio.

Wolf (2007) compares the Michaud approach with an active investment strategy based on a shrinkage estimator for the covariance matrix of asset returns. This is done on the basis of US stock-market data from 1983 to 2002. He finds that the shrinkage estimator systematically outperforms portfolio resampling.

Harvey et al. (2008) repeat the Markowitz-Usmen experiment after changing the priors and the numerical algorithm for calculating the posteriors. Their results are in direct contrast to Markowitz and Usmen (2003), i.e., Bayes beats Michaud. They state that "the Bayes player was handicapped because the algorithm that was used to evaluate the predictive distribution of the portfolio provided only a rough approximation."

Michaud and Michaud (2008) reply that Harvey et al. (2008) use an inefficient method for portfolio resampling and, moreover, the two inventors argue that the numerical implementation of the Michaud procedure was less accurate than the implementation of the Bayesian estimates.

Barros Fernandes et al. (2012) assert that "Several out-of-sample evaluations have shown results in favor of resampling methodology, using different sets of data." Becker et al. (2013) perform an empirical study and a Monte Carlo experiment to compare portfolio resampling with several optimization methods and find that Markowitz outperforms Michaud on average. The empirical study is based on US and European stock-market data from 1988 to 2007. However, the resampling method implemented by Becker et al. (2013) suffers from the same drawback as in Harvey et al. (2008).

Summing up, there is still no clear answer to the questions raised by Markowitz and Usmen (2003). Most firms and professional investors would never implement a rule-based investment strategy without a backtest. Hence, it is necessary to evaluate every investment rule ex post on the basis of historical price data, at least for demonstrating its potential benefit. Nevertheless, a serious empirical study that aims at *comparing* different strategies must control for non-normality, serial dependence, selection bias, multiple testing, etc. After taking all that issues properly into account, in general it is hard to find *significant* results in favor of any strategy (Frahm et al., 2012). This means the question whether portfolio resampling is better or worse than other investment strategies cannot be solved by statistical inference.

Unfortunately, the same holds for Monte Carlo simulation. As is pointed out by Scherer (2006), one can always create situations where a Bayesian strategy performs better than any other strategy and vice versa. Every strategy works good if the underlying assumptions are satisfied and so it is relatively easy to construct a Monte Carlo experiment where Strategy A outperforms Strategy B and to find another setting where the opposite is true. For this reason, Monte Carlo simulation can only be used to compare strategies under a *hypothetical* situation but does not explain how the strategies perform in real life.

In my opinion, the only way to provide satisfactory answers to the basic questions raised by Markowitz and Usmen is to derive precise *theoretical* arguments for or against portfolio resampling. This goal can be only achieved by analytical investigation and not by observation or simulation. I hope that the results presented in this paper provide new insights and a better understanding of the mechanisms governing individual investment decisions in a world of uncertainty.

2. Preliminary Assumptions and Definitions

In the following $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ represents an N-dimensional *column* vector. The equality "$X = Y$" for any two random vectors X and Y of the same dimension means that X and Y are almost surely equal. The symbol $\mathbf{0}$ denotes a vector of zeros and $\mathbf{1}$ is a vector of ones. The dimensions of $\mathbf{0}$ and $\mathbf{1}$ will be clear from the context. Let $R = (R_1, R_2, \ldots, R_N)$ be an N-dimensional random vector of risky asset returns. More precisely, R denotes the vector of returns in excess of the risk-free interest rate, i.e., a number $r \geq 0$ at the beginning of the investment period. The prefix "excess" will be dropped for convenience.

Let $w = (w_1, w_2, \ldots, w_N)$ be a vector of portfolio weights. The weight of the risk-free asset follows by $w_0 = 1 - w'\mathbf{1}$. This guarantees that the budget constraint $w_0 + w'\mathbf{1} = 1$ is satisfied for every portfolio $w \in \mathbb{R}^N$. Moreover, let $\mathcal{C} \subseteq \mathbb{R}^N$ be any convex set of constraints on the portfolio weights. It might contain any number of linear equality and inequality constraints like, e.g., short-selling constraints or nonlinear inequality constraints like, e.g., a norm constraint $\|w\| \leq \delta > 0$ (DeMiguel et al., 2009). It is worth emphasizing that $\mathcal{C}$ is required to be *convex*, which appears to be a realistic assumption in most practical situations.

Consider an investor or trader who is about to select a portfolio of risky assets.[2] Once the decision is made, the entire portfolio is held for some fixed period of time. Although the investment period is fixed, the length of the period can be very short, very long or anything in between. At the end of the investment period, the investor liquidates all positions and possibly re-allocates his capital. He aims at maximizing his expected utility by choosing the "best" portfolio weights at the beginning of each investment period. Whenever it is not necessary to distinguish specific points in time, I will omit time indices for notational convenience.

Let $\mathcal{I}$ be any information that is available to the investor at the beginning of the investment period. To avoid unnecessary complications, I will not give a formal, i.e., measure theoretic, definition of $\mathcal{I}$. The information $\mathcal{I}$ can be simply interpreted as "observation and experience" (Markowitz, 1952):

> "The process of selecting a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio."

[2] In the following I use both terms "investor" and "trader" synonymously.

Unlike the second stage, the first stage of capital allocation can be a complicated process rather than a simple calculus. The problem is that investors are faced with uncertainty and so it is not clear to them which kind of information to choose and how to combine this information to obtain reliable expectations. In the following discussion I use the terms *risk* and *uncertainty* according to the usual definition by Knight (1921).

For example, $\mathcal{I}$ could be a set of technical indicators like, e.g., moving averages, candlesticks, support and resistance lines, etc. In this case $\mathcal{I}$ is a function of the price history at the beginning of the investment period. This will be denoted by $\mathcal{H}$. Equivalently, we could write "$\mathcal{I} \subseteq \mathcal{H}$" or we could say that "$\mathcal{I}$ is implied by $\mathcal{H}$." In most practical situations we have that $\mathcal{I} \subseteq \mathcal{J}$, where $\mathcal{J}$ is any kind of information like, e.g., public or private information. In these cases it is usually assumed that $\mathcal{H} \subseteq \mathcal{J}$.

The information $\mathcal{I}$ is a random object. Let $\Im$ be the *support* of $\mathcal{I}$, i.e., the space of all information that can occur to the investor at the beginning of the investment period. It is assumed that the information $\mathcal{I}$ implies one and only one vector of portfolio weights, i.e., $\tilde{w}(\mathcal{I}) \in \mathcal{C} \subseteq \mathbb{R}^N$. This means there exists a map $\tilde{w} : \Im \to \mathbb{R}^N$ which assigns an element of $\mathcal{C}$ to each information $\mathcal{I} \in \Im$. This *map* is said to be the *strategy*, whereas the *output* of $\tilde{w}$, i.e., $\tilde{w}(\mathcal{I})$, is said to be a *portfolio*. In terms of decision theory, $\tilde{w}(\mathcal{I})$ represents an *act* or a *decision* based on a *state* or *event* $\mathcal{I}$. Since $\mathcal{I}$ is stochastic, $\tilde{w}$ in general represents a non-deterministic random vector and then it is said to be a *mixed* strategy. In the special case where $\tilde{w}$ is fixed, i.e., deterministic, it is said to be a *pure* strategy. The vector R of asset returns can be interpreted as the *response* of the nature. In our framework the nature is simply the financial market.

The dot product $\tilde{w}'R$ represents the excess return on the strategy $\tilde{w}$, whereas the actual return on the strategy amounts to $r + \tilde{w}'R$. The investor aims at finding an optimal strategy. Let $W > 0$ be the current wealth of the investor and consider any single-period utility function $u\colon [0,\infty[\, \to \mathbb{R}$. Hence, as is usual in the literature on portfolio optimization, I assume that the investor tries to maximize a single-period expected utility. More precisely, the investor treats each period as if it were the last one and so his decisions are *myopic* (Mossin, 1968).

3. Capital Allocation under Risk and Uncertainty

3.1. Allocation under Risk

The investor aims at maximizing his expected utility

$$\mathrm{E}\Big(u\big(W(1 + r + \tilde{w}'R)\big)\Big), \qquad \tilde{w} \in \mathcal{C}.$$

Under the circumstances of *risk* (Knight, 1921), it is simply possible to calculate the *optimal portfolio* under the constraint $\mathcal{C}$, i.e., to apply the *pure* strategy

$$w^* = \arg\max_{w \in \mathcal{C}} \mathrm{E}\Big(u\big(W\,(1 + r + w'R)\big)\Big).$$

So why should the investor take observation and experience into account? Let

$$\tilde{w}^*(\mathcal{I}) = \arg\max_{w \in \mathcal{C}} \mathrm{E}\Big(u\big(W\,(1 + r + w'R)\big)\,|\,\mathcal{I}\Big)$$

be the optimal portfolio conditional on the information $\mathcal{I}$. In contrast to w^*, $\tilde{w}^*$ is a random vector and thus it represents a *mixed* but not a pure strategy. It is clear that

$$\mathrm{E}\Big(u\big(W\,(1 + r + \tilde{w}^{*\prime}R)\big)\,|\,\mathcal{I}\Big) \geq \mathrm{E}\Big(u\big(W\,(1 + r + w^{*\prime}R)\big)\,|\,\mathcal{I}\Big)$$

and so, by the law of total expectation, we obtain

$$\mathrm{E}\Big(u\big(W\,(1 + r + \tilde{w}^{*\prime}R)\big)\Big) \geq \mathrm{E}\Big(u\big(W\,(1 + r + w^{*\prime}R)\big)\Big).$$

Hence, the optimal strategy is more favorable than the optimal portfolio. This means maximizing the expected utility *conditional* on $\mathcal{I}$ can only increase but never decrease the expected utility of the investor. To put it another way, it is always recommended to search for as much information as possible and to "act conditionally" on $\mathcal{I}$.[3]

A function $f : \Im \to \mathbb{R}$ is said to be *increasing* if and only if $f(\mathcal{I}) \leq f(\mathcal{J})$ for every information $\mathcal{I}, \mathcal{J} \in \Im$ such that $\mathcal{I} \subseteq \mathcal{J}$.

In a world of risk, the expected utility of an investor increases with information.

Hence, the investor should gather as much information as possible to make the most of his money. This means observation and experience are precious things and should never be disdained in a risky world. This obvious conclusion changes essentially when the world becomes uncertain, which is the topic of the following sections.

3.2. Allocation under Uncertainty

Definition 1 (Irrelevant information)**.** *Let $\mathcal{I}$ and $\mathcal{J}$ be some information such that $\mathcal{I} \subseteq \mathcal{J}$. The complementary information $\mathcal{J} \setminus \mathcal{I}$ is said to be* irrelevant given $\mathcal{I}$ *if and only if*

$$\mathbb{P}(R \leq x\,|\,\mathcal{J}) = \mathbb{P}(R \leq x\,|\,\mathcal{I})$$

for all $x \in \mathbb{R}^N$. In particular, $\mathcal{J}$ is simply called irrelevant *if and only if*

$$\mathbb{P}(R \leq x\,|\,\mathcal{J}) = \mathbb{P}(R \leq x)$$

for all $x \in \mathbb{R}^N$.

[3] Here it is implicitly assumed that the search costs for the information $\mathcal{I}$ are negligible.

Irrelevant information $\mathcal{I}$ cannot be used for predicting future asset returns, since R does not depend on $\mathcal{I}$. If $\mathcal{I}$ is not irrelevant, i.e., relevant, the investor could at least ignore all parts of $\mathcal{I}$ that are irrelevant given a subset of $\mathcal{I}$. The problem is that the real world is *uncertain* (Knight, 1921) and so the investor cannot see which parts of $\mathcal{I}$ are irrelevant. Moreover, since the investor does not know the distribution of R conditional on $\mathcal{I}$, he is not able to apply the optimal strategy $\tilde{w}^*$. Therefore, he applies another strategy $\tilde{w} \in \mathcal{C}$.

Suppose that $\tilde{w}$ is determined by any kind of relevant or irrelevant information $\mathcal{I}$. Then, once again by the law of total expectation, we obtain

$$\mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big) \leq \mathrm{E}\Big(u\big(W(1+r+\tilde{w}^{*\prime}R)\big)\Big).$$

Now, consider the *mean portfolio* $\bar{w} = \mathrm{E}(\tilde{w}) \in \mathcal{C}$ associated with $\tilde{w}$.[4] It is clear that

$$\mathrm{E}\Big(u\big(W(1+r+\bar{w}'R)\big)\Big) \leq \mathrm{E}\Big(u\big(W(1+r+\tilde{w}^{*\prime}R)\big)\Big).$$

This means the mean portfolio cannot beat the optimal strategy, but $\bar{w}$ could at least be better than the actual strategy $\tilde{w}$. In fact, this turns out to be *always* true under the following very general assumptions:

A1. The investor has an increasing and strictly concave utility function u.

A2. The strategy $\tilde{w}$ is stochastically independent of the vector R of asset returns.

A sufficient condition for **A2** is that the strategy $\tilde{w}$ of the investor is determined by some irrelevant information $\mathcal{I}$. Another possibility is that $\tilde{w}$ is a pure strategy.

Theorem 1. *Consider a mixed strategy $\tilde{w} \in \mathcal{C}$ and suppose that the assumptions **A1** and **A2** are satisfied. Then*

$$\mathrm{E}\Big(u\big(W(1+r+\bar{w}'R)\big)\Big) > \mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big),$$

where $\bar{w} = \mathrm{E}(\tilde{w}) \in \mathcal{C}$ is the mean portfolio associated with $\tilde{w}$.

Definition 2 (Noise trader). *An investor with a mixed strategy $\tilde{w}$ is called a* noise trader *if and only if $\tilde{w}$ is stochastically independent of R. Otherwise he is called a* signal trader.

Theorem 1 implies that every noise trader with increasing and strictly concave utility function should try to substitute his strategy by the associated mean portfolio. This fundamental result does not depend on the specific utility function.

[4]Here $\bar{w}$ denotes the *theoretical* expectation of $\tilde{w}$ but not some corresponding approximation or estimate.

Example 1 (The account manager)**.** Suppose that a noise trader acts as an account manager. Then every risk averse client, independent of his own utility function, would benefit if the trader substitutes his strategy $\tilde{w}$ by the mean portfolio $\bar{w}$. For the sake of simplicity we may assume that for each asset $i = 1, \ldots, N$, the trader either decides to go long ($\tilde{w}_i = 1$), short ($\tilde{w}_i = -1$), or flat ($\tilde{w}_i = 0$). It follows that $\bar{w} = p - q$, where the vectors $p = (p_1, p_2, \ldots, p_N)$ and $q = (q_1, q_2, \ldots, q_N)$ contain his probabilities to go long and short, respectively. In case $p = q$, i.e., if his trading strategy is symmetric, it turns out that $\bar{w} = \mathbf{0}$ and so it would be better for his clients to stop trading at all.

Example 2 (The blindfolded chimpanzee)**.** According to Malkiel (2003) "A blindfolded chimpanzee throwing darts at the Wall Street Journal could select a portfolio that would do as well as the experts." Suppose that $N = 2$ and consider some number $\delta > 0$. The set of constraints $\mathcal{C} = \{w \in \mathbb{R}^2 : \|w\|_2 \leq \delta\}$ represents the dartboard. (A shot is simply repeated if it is outside $\mathcal{C}$.) It is clear that the chimpanzee is a noise trader. Instead of letting him throw the next dart, it would be better to take the "mean shot" on the dartboard, i.e., $\bar{w} = \mathrm{E}(\tilde{w}) \in \mathcal{C}$ (which need not be $\mathbf{0} \in \mathbb{R}^2$).

Example 3 (The i.i.d. assumption)**.** Consider the process $\{R_t\}_{t \in \mathbb{Z}}$ of asset returns and suppose that $\tilde{w}_t$ is a function of the asset-return history $\mathcal{I}_t = (R_t, R_{t-1}, \ldots)$ at any time $t \in \mathbb{Z}$, i.e., at the beginning of some investment period. Every portfolio optimization method that aims at estimating the parameters of the conditional or unconditional asset-return distribution on the basis of past asset returns falls into this category. It is often assumed that asset returns are serially independent and identically distributed. Under this widespread "i.i.d. assumption," the history of asset returns is *irrelevant*. In this case every investor applying a simple strategy or even a *a highly sophisticated statistical method* based on past asset returns is just a noise trader! This means the mean portfolio $\bar{w}_t$ is more favorable than his actual strategy $\tilde{w}_t$.[5]

Example 4 (Random-walk hypothesis)**.** Fama's (1970) famous *random-walk model* states that

$$\mathbb{P}\big(R_{t+1} \leq x \,|\, \mathcal{J}_t\big) = \mathbb{P}\big(R_{t+1} \leq x\big) = F(x), \qquad \forall\, x \in \mathbb{R}^N,\ t \in \mathbb{Z},$$

where $\mathcal{J}_t$ represents any information at time t. It is usually assumed that $\mathcal{H}_t \subseteq \mathcal{J}_t$. This means $\mathcal{J}_t$ contains the history of asset prices and it is often regarded as public or private information manifested at time t. Hence, the random-walk model implies that the asset returns are serially independent and identically distributed. If the strategy $\tilde{w}_t$ is determined by $\mathcal{I}_t \subseteq \mathcal{J}_t$ and the random-walk hypothesis is true with respect to $\mathcal{J}_t$, the investor is a noise trader and once again $\bar{w}_t$ beats the actual strategy $\tilde{w}_t$.

[5]Here the mean portfolio $\bar{w}_t$ is a function of time, since the distribution of $\tilde{w}_t$ might depend on $t \in \mathbb{Z}$.

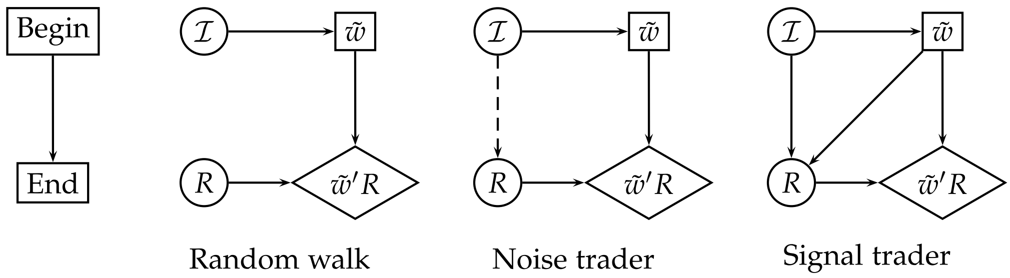

Figure 1: Influence diagrams.

Figure 1 contains three different influence diagrams, which illustrate the possible interconnections between $\mathcal{I}$, $\tilde{w}$, R, and $\tilde{w}'R$. The first diagram represents the random-walk model. If the strategy $\tilde{w}$ is a function of some information $\mathcal{I}$ and R does not depend on $\mathcal{I}$, it cannot depend on $\tilde{w}$ either. Hence, the return on the strategy, i.e., $\tilde{w}'R$, is the product of two independent random vectors.

A noise trader *might* use relevant information. This is illustrated by the dashed line between $\mathcal{I}$ and R in the second diagram of Figure 1. Nevertheless, he is not able to create a strategy $\tilde{w}$ such that R depends on $\tilde{w}$. This is not meant in the sense that the trader were able to *affect* market prices. Here "dependence" merely refers to a stochastic, but not a causal relation between $\tilde{w}$ and R.

The situation of a signal trader is depicted in the third diagram of Figure 1. A signal trader has access to relevant information *and* he can create a strategy $\tilde{w}$ such that $\tilde{w}$ and R depend on each other. For example, it could be that the trader has found an appropriate set of predictors for the vector R of asset returns and applies a conditional strategy based on his forecasts.

The fact that every statistical portfolio optimization method, that is based only on the history of asset returns, leads to a noise-trading strategy might be counterintuitive. The random-walk hypothesis implies that future asset returns do not depend on historical price data. Hence, past asset returns are only "irrelevant" in the sense that they cannot be used to *predict* future asset returns.[6] This is not to say that they are *useless* from a statistical point of view. In particular, without any history of asset returns it is impossible to *estimate* the parameters of the asset-return distribution, i.e., $x \mapsto F(x)$.

3.3. The Bias-Variance Trade-Off

Under quite general conditions regarding the utility function u and the distribution of the return on the strategy, the expected utility of the investor can be approximated (up

[6] In Section 4.2 I discuss the apparent link between the random-walk and the efficient-market hypothesis.

to an affine-linear transformation) by the "mean-variance" objective function

$$\mathrm{E}\left(\tilde{w}'R - \frac{\alpha}{2}\,(\tilde{w}'R)^2\right) \tag{1}$$

(Kroll et al., 1984, Markowitz, 2012, Pulley, 1981). Assumption **A1** guarantees that α, i.e., the *relative risk aversion* (Arrow, 1971, Pratt, 1964), is positive.

In the following I write $\mu = \mathrm{E}(R)$ and $\Sigma = \mathrm{E}(RR')$. Further, I assume that it is not possible to construct a risk-free portfolio with risky assets, i.e., Σ is positive definite. Suppose that the investor is a noise trader. Then (1) equals

$$\mathrm{E}\left(\tilde{w}'\mathrm{E}(R\,|\,\tilde{w}) - \frac{\alpha}{2}\,\tilde{w}'\mathrm{E}(RR'\,|\,\tilde{w})\,\tilde{w}\right) = \mathrm{E}\left(\tilde{w}'\mu - \frac{\alpha}{2}\,\tilde{w}'\Sigma\tilde{w}\right).$$

The quantity

$$\varphi_\alpha(\tilde{w}) = \tilde{w}'\mu - \frac{\alpha}{2}\,\tilde{w}'\Sigma\tilde{w}$$

is often referred to as the *out-of-sample performance* (Kan and Zhou, 2007) or *out-of-sample certainty equivalent* (Frahm et al., 2012) of $\tilde{w}$.[7]

The expected out-of-sample performance, i.e., $\mathrm{E}(\varphi_\alpha(\tilde{w}))$, is a popular means for evaluating investment strategies. In particular, it is used by Markowitz and Usmen (2003) in their experiment. Hence, a theoretical investigation of the potential impact of portfolio resampling on the expected out-of-sample performance is crucial. This shall be tackled in the subsequent analysis.

Nevertheless, as it can be seen from the derivation above, the expected value of $\varphi_\alpha(\tilde{w})$ in general can only be used in place of (1) if the investor is considered as a *noise trader*. More precisely, the first and second moments of R, conditional on the strategy $\tilde{w}$, must not depend on $\tilde{w}$. This means the expected out-of-sample performance is a meaningful measure if the investor is assumed to be a noise trader. By contrast, if he is a signal trader, $\mathrm{E}(\varphi_\alpha(\tilde{w}))$ becomes meaningless. This point turns out to be essential later on, when clarifying the potential benefits and pitfalls of portfolio resampling.

Consider the mean-variance optimal portfolio without any constraints on the portfolio weights, i.e.,

$$w^* = \arg\max_{w\in\mathbb{R}^N}\left(w'\mu - \frac{\alpha}{2}\,w'\Sigma w\right) = \frac{\Sigma^{-1}\mu}{\alpha}.$$

The *loss* of every noise-trading strategy $\tilde{w}$ is defined as $\varphi_\alpha(w^*) - \varphi_\alpha(\tilde{w})$ and

$$\mathcal{L}_\alpha(\tilde{w}) = \mathrm{E}\big(\varphi_\alpha(w^*) - \varphi_\alpha(\tilde{w})\big)$$

is said to be the *expected loss* of $\tilde{w}$. The following theorem guarantees that $\mathcal{L}_\alpha(\tilde{w}) \geq 0$.

[7]In most publications, Σ typically denotes the covariance matrix $\mathrm{Var}(R)$ but not the uncentered second-moment matrix of R. For short-term asset returns the difference between $\mathrm{Var}(R)$ and $\mathrm{E}(RR')$ is negligible.

Theorem 2. *Let $\tilde{w} \in \mathcal{C}$ be any strategy and suppose that the assumptions **A1** and **A2** are satisfied. The expected loss of $\tilde{w}$ amounts to*

$$\mathcal{L}_\alpha(\tilde{w}) = \frac{\alpha}{2}\,\mathrm{E}\Big((\tilde{w} - w^*)'\Sigma\,(\tilde{w} - w^*)\Big) = \mathcal{B}_\alpha(\tilde{w}) + \mathcal{V}_\alpha(\tilde{w}) \geq \mathcal{L}_\alpha(\bar{w})$$

where $w^* = \Sigma^{-1}\mu/\alpha$, $\bar{w} = \mathrm{E}(\tilde{w}) \in \mathcal{C}$,

$$\mathcal{B}_\alpha(\tilde{w}) = \frac{\alpha}{2}\,(\bar{w} - w^*)'\Sigma\,(\bar{w} - w^*)\,, \quad \textit{and} \quad \mathcal{V}_\alpha(\tilde{w}) = \frac{\alpha}{2}\,\mathrm{E}\Big((\tilde{w} - \bar{w})'\Sigma\,(\tilde{w} - \bar{w})\Big)\,.$$

The inequality $\mathcal{L}_\alpha(\tilde{w}) \geq \mathcal{L}_\alpha(\bar{w})$ is strict if and only if $\tilde{w}$ is a mixed strategy.

Hence, the expected loss of every noise-trading strategy $\tilde{w}$ can be divided into two parts: The *bias*, i.e., $\mathcal{B}_\alpha(\tilde{w})$, and the *variance*, i.e., $\mathcal{V}_\alpha(\tilde{w})$, of $\tilde{w}$. This illustrates the fundamental bias-variance trade-off, which has been often observed in the literature on portfolio optimization. In particular, Theorem 2 clarifies why it can be better to prefer a biased but robust strategy versus an unbiased strategy that is strongly affected by estimation errors. For an analytical investigation of the bias-variance trade-off and the impact of estimation errors under the i.i.d. assumption see, e.g., Frahm (2011), Frahm and Memmel (2010), Jobson and Korkie (1980) as well as Kan and Zhou (2007).

Hence, the more the strategy of a noise trader is affected by noise, the bigger his expected loss. Now, it is intuitively clear that the financial success of a noise or signal trader essentially depends on (i) the quality of information $\mathcal{I}$ and (ii) his capability to ignore the irrelevant part of $\mathcal{I}$, i.e., to separate signal from noise.

> *In a world of uncertainty, the expected utility of a risk averse investor tends to decrease, the more his strategy is exposed to irrelevant information.*

4. Does Portfolio Resampling Really Make Sense?

4.1. The Michaud Approach

The Michaud procedure (Michaud, 1998) works as follows:

1. Consider a sample of past asset returns with sample size n.
2. Estimate the unknown parameters μ and Σ by the empirical moments $\hat{\mu}$ and $\widehat{\Sigma}$.[8]
3. Generate a sample of n independent asset returns from $\mathcal{N}_N(\hat{\mu},\widehat{\Sigma})$.
4. Estimate $\hat{\mu}$ and $\widehat{\Sigma}$ by the empirical moments $\hat{\hat{\mu}}$ and $\widehat{\widehat{\Sigma}}$.

[8]Here Σ refers to the *covariance matrix* of R (Michaud, 1998).

5. Calculate the weights of 101 efficient portfolios $\hat{\tilde{w}}_0, \hat{\tilde{w}}_1, \ldots, \hat{\tilde{w}}_{100}$. The first one is the *minimum-variance portfolio*, i.e.,

$$\hat{\tilde{w}}_0 = \arg\min_{w \in \mathcal{C}} w' \hat{\hat{\Sigma}} w \,,$$

whereas the last one is the *maximum-mean portfolio*, i.e.,

$$\hat{\tilde{w}}_{100} = \arg\max_{w \in \mathcal{C}} w' \hat{\hat{\mu}} \,.$$

Here it is assumed that $\mathcal{C} \subseteq \{ w \in \mathbb{R}^N : w'\mathbf{1} = 1,\ w \geq \mathbf{0} \}$.[9] The other 99 portfolios maximize $w_i' \hat{\hat{\mu}}$ such that $w_i \in \mathcal{C}$ and

$$\sqrt{w_{i+1}' \hat{\hat{\Sigma}} w_{i+1}} - \sqrt{w_i' \hat{\hat{\Sigma}} w_i} = \sqrt{w_i' \hat{\hat{\Sigma}} w_i} - \sqrt{w_{i-1}' \hat{\hat{\Sigma}} w_{i-1}} \,, \qquad i = 1, \ldots, 99 \,.$$

6. Repeat Step 3 to Step 5 499 times.

7. Let $\hat{\tilde{w}}_{ij}$ be the portfolio $i = 0, 1, \ldots, 100$ of draw $j = 1, \ldots, 500$ and calculate the average holdings

$$\bar{\tilde{w}}_i = \frac{1}{500} \sum_{j=1}^{500} \hat{\tilde{w}}_{ij} \,, \qquad i = 0, 1, \ldots, 100 \,.$$

8. Search for the $\bar{\tilde{w}}_i$ that maximizes

$$\bar{\tilde{w}}_i' \hat{\mu} - \frac{\alpha}{2} \bar{\tilde{w}}_i' \widehat{\Sigma} \bar{\tilde{w}}_i \,.$$

In the previous sections it has been shown that every noise trader with increasing and strictly concave utility function should substitute his strategy $\tilde{w}$ by the associated mean portfolio $\bar{w}$. The problem is that $\bar{w}$ is not known in real life. As is shown above, the Michaud procedure aims at *approximating* 101 mean portfolios $\bar{w}_0, \bar{w}_1, \ldots, \bar{w}_{100}$. In fact, each variant of the aforementioned portfolio-resampling procedure turns out to be just a method for approximating one or more mean portfolios.

All simulation studies mentioned in the introduction have one thing in common: The investors are considered as noise traders. In this case Theorem 1 and Theorem 2 are applicable and so it is not surprising that the Michaud procedure does well in the simulation studies. However, the empirical studies discussed in the introduction show that portfolio resampling does not systematically outperform other approaches that take estimation risk likewise into account, e.g., constraining portfolio weights, applying Bayesian analysis, shrinkage estimation, robust portfolio optimization, etc. These approaches are always designed to find an optimal trade off between bias and

[9]Hence, $\mathcal{C}$ belongs to the N-dimensional Euclidean simplex.

variance and so there is no theoretical argument why portfolio resampling should dominate any of these strategies.

There is another important drawback of the Michaud approach.[10] Suppose that the desired portfolio $w \in \mathcal{C}$ is a function of an unknown parameter $\theta \in \Theta$, i.e., $w = f(\theta)$.[11] Since θ is unknown, it is typically approximated by some estimate $\hat{\theta} \in \Theta$. Let $G(\cdot;\theta)$ be the cumulative distribution function of $\hat{\theta}$ and assume that $\hat{\theta}$ is unbiased, i.e., $\mathrm{E}(\hat{\theta};\theta) = \theta$ for all $\theta \in \Theta$. Consider the *plug-in strategy* $\hat{w} = f(\hat{\theta})$ and suppose that the function f is linear.[12] Now, the Michaud player generates a large number of independent estimates $\hat{\hat{\theta}}_1, \hat{\hat{\theta}}_2, \ldots, \hat{\hat{\theta}}_m \sim G(\cdot;\hat{\theta})$ and calculates the corresponding portfolios $\hat{\hat{w}}_1, \hat{\hat{w}}_2, \ldots, \hat{\hat{w}}_m \sim \hat{\hat{w}}$ with $\hat{\hat{w}}_j = f(\hat{\hat{\theta}}_j)$ for $j = 1,2,\ldots,m$. Finally, he chooses the sample mean $\bar{\hat{w}}_m = \frac{1}{m}\sum_{j=1}^{m} \hat{\hat{w}}_j$. From the Law of Large Numbers it follows that

$$\bar{\hat{w}}_m \xrightarrow{\text{a.s.}} \mathrm{E}(\hat{\hat{w}};\hat{\theta}) = \mathrm{E}\Big(f(\hat{\hat{\theta}});\hat{\theta}\Big) = f\Big(\mathrm{E}(\hat{\hat{\theta}};\hat{\theta})\Big) = f(\hat{\theta}) = \hat{w}, \qquad m \longrightarrow \infty.$$

This means portfolio resampling can only make sense if f is nonlinear or the estimator for θ is biased, which can be achieved by using constraints on the portfolio weights or nonlinear estimation methods like, e.g., shrinkage estimation. Similar arguments can be found in Scherer (2004).

However, we must not throw out the baby with the bath water. Wolf (2007) notes that the Michaud procedure can be considered as a special case of a statistical method called "bagging," which has been invented by Breiman (1996).[13] Consider any method to predict a variable on the basis of some information $\mathcal{I}$. Since $\mathcal{I}$ is stochastic, the prediction suffers from estimation risk. Breiman (1996) proves that it would be better to choose the mean of all forecasts if the loss function is quadratic.

"Bagging" stands for *bootstrap aggregating* and means to aggregate the predictions based on a large number of bootstrap replications of $\mathcal{I}$. As is pointed out by Breiman (1996), bagging cannot *always* lead to an improvement, since the bootstrap replications come from an empirical but not the true distribution of $\mathcal{I}$. The potential improvement depends essentially on the prediction variability. More precisely, the more variability, the more room is left for improvement.

The same conclusion can be drawn by Theorem 2, regarding the question whether the Michaud procedure leads to an improvement or not. Consider any strategy $\tilde{w}$

[10]This point has been suggested by Christoph Memmel in a personal communication.

[11]In the random-walk model (see Example 4), θ is the parameter of F, i.e., the joint cumulative distribution function of the asset returns.

[12]For example, it is typically assumed that the random-walk model is true, Σ is known, and $\mathcal{C} = \mathbb{R}^N$. Then the desired portfolio corresponds to $w^* = \Sigma^{-1}\mu/\alpha$, which is a linear function of μ.

[13]I would like to thank Winfried Pohlmeier for this hint.

and let $\bar{\tilde{w}}$ be the resampled version of $\tilde{w}$. Moreover, suppose that **A1** and **A2** are satisfied. Theorem 2 implies that the expected out-of-sample performance of $\bar{\tilde{w}}$ exceeds the expected out-of-sample performance of the actual strategy $\tilde{w}$ if and only if

$$\mathcal{B}_\alpha(\bar{\tilde{w}}) - \mathcal{B}_\alpha(\tilde{w}) < \mathcal{V}_\alpha(\tilde{w}) - \mathcal{V}_\alpha(\bar{\tilde{w}})\,.$$

This means portfolio resampling improves a noise-trading strategy if and only if the variance reduction is larger than the bias increment. For example, the naive Markowitz approach does not take estimation errors into account and so this strategy suffers from a large variance. Hence, the Michaud procedure typically beats the naive Markowitz approach as long as $\mathcal{C} \neq \mathbb{R}^N$ (Scherer, 2002, 2004). However, as already mentioned, the Michaud strategy does not systematically outperform other investment strategies with lower variance. It seems more reasonable to apply portfolio resampling on these kind of strategies rather than to improve on the naive Markowitz approach.

In fact, Fletcher and Hillier (2001) compare several strategies in their "actual" and "resampled" version. It turns out that the *unconditional* strategies can be improved by resampling, whereas the *conditional* strategies often become worse after applying the resampling procedure! These empirical results perfectly agree with Assumption **A2**, which has been used for deriving the theorems above. This point will be clarified in more detail in the next section.

4.2. The Noise-Trader Assumption

The random-walk model (see Example 4) was *the* basic assumption in finance theory for many decades. It is often confused with the *efficient-market hypothesis* (Fama, 1970). The former is only an auxiliary model for the latter. Campbell et al. (1997), LeRoy (1973) and Lucas (1978) demonstrate that a random walk is neither necessary nor sufficient for a "fair game," i.e., for an efficient market according to Fama (1970).

Recent approaches to the efficient-market hypothesis only require that the market is *arbitrage free* (Jarrow and Larsson, 2012, Jarrow, 2012, Ross, 2005). Let $\{P_t\}$ be the price process of any asset. A financial market with a finite number of assets and a finite lifetime $T > 0$ is said to be *efficient* with respect to some information flow $\{\mathcal{F}_t\}$ if and only if $P_t = \mathrm{E}(\lambda_{t,T} P_T \,|\, \mathcal{F}_t)$ for all $0 \leq t \leq T$, where $\lambda_{t,T}$ represents a *stochastic discount factor* (Cochrane, 2005, Ferson, 2007).[14] Hence, market efficiency does not rule out the predictability of asset prices (Timmermann and Granger, 2004). Here $\{\mathcal{F}_t\}$ is a superfiltration of $\{\mathcal{H}_t\}$, i.e., the evolution of asset prices. If $\{\mathcal{F}_t\}$ represents a flow

[14] For a precise definition of $\lambda_{t,T}$ see Back (2010) and Frahm (2013).

of *relevant* information and $\{\mathcal{H}_t\}$ "fully reflects" $\{\mathcal{F}_t\}$ (Fama, 1970), it is obvious that $\{\mathcal{H}_t\}$ is relevant, too.[15]

What does this mean for us? Lo et al. (2002) state that, "Among some circles, technical analysis is known as 'vodoo finance'," but "several academic studies suggest that [...] technical analysis may well be an effective means for extracting useful information from market prices." Hence, in an efficient market, an investor might improve his expected utility by collecting price data, but he cannot expect to realize "abnormal profits," i.e., arbitrage, on the basis of historical asset prices.

Nevertheless, Black (1986) writes that, "people sometimes trade on noise as if it were information. If they expect to make profits from noise trading, they are incorrect." This is an essential point. If asset prices are predictable, why should people act as noise traders? More precisely, if R depends on some information $\mathcal{I}$, why is it not simply possible to create a strategy $\tilde{w}$, that is determined by $\mathcal{I}$, such that $\tilde{w}$ and R depend on each other? The answer is simple: Because we are living in a world of uncertainty. We would have to search for a suitable information $\mathcal{I}$ and calculate the true conditional probability distribution of the asset returns, i.e., $\mathbb{P}(R \leq x \,|\, \mathcal{I})$. These steps cannot be accomplished in a world of uncertainty.

Black's statement can be formalized as follows: The general objective function (1) can be written as $\mathrm{E}\big(\phi_\alpha(\tilde{w})\big)$ with

$$\phi_\alpha(\tilde{w}) = \tilde{w}'R - \frac{\alpha}{2}\,(\tilde{w}'R)^2\,.$$

In fact, the expected return on any strategy $\tilde{w}$ amounts to

$$\mathrm{E}(\tilde{w}'R) = \bar{w}'\mu + \sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i)\,.$$

Hence, a noise trader can only earn $\bar{w}'\mu$, i.e., the expected return on the mean portfolio. By contrast, a signal trader is able to produce the (additional) profit $\sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i)$. This can be done only by using *relevant* information and this is what all people try to achieve in a financial market.

Theorem 3. *For every strategy $\tilde{w}$ we have that*

$$\mathrm{E}\big(\phi_\alpha(\tilde{w})\big) = \bar{w}'\mu + \sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i) - \frac{\alpha}{2}\left(\mathrm{E}\big(\tilde{w}'\Sigma\,\tilde{w}\big) + \sum_{i,j=1}^{N} \mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j, R_iR_j\big)\right).$$

If the investor acts on irrelevant information or if he is simply unable to separate signal from noise, we obtain

$$\sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i) = \sum_{i,j=1}^{N} \mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j, R_iR_j\big) = 0\,.$$

[15] More details on this subject can be found in Frahm (2013).

By Jensen's inequality, it holds that $\mathrm{E}\big(\tilde{w}'\Sigma\tilde{w}\big) \geq \mathrm{E}\big(\bar{w}'\Sigma\bar{w}\big)$ and so $\bar{w}$ is more favorable than $\tilde{w}$. By contrast, a signal trader should *not* favor the mean portfolio if and only if

$$\sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i) - \frac{\alpha}{2} \sum_{i,j=1}^{N} \mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j, R_iR_j\big) > \mathcal{V}_\alpha(\bar{w}) \,. \tag{2}$$

Here $\sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i)$ quantifies his *prediction power*, whereas $\sum_{i,j=1}^{N} \mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j, R_iR_j\big)$ reflects the quality of his *risk management* (where a negative value is preferable).

Prediction: Whenever the trader expects that the asset return R_i is large, he will choose a large portfolio weight $\tilde{w}_i$ and vice versa. If his expectations are valid, there will be a positive correlation between $\tilde{w}_i$ and R_i. This means $\sum_{i=1}^{N} \mathrm{Cov}(\tilde{w}_i, R_i)$ measures his overall prediction power. The better his forecasts, the larger ceteris paribus the expected return on his strategy.

Risk management: By choosing a low exposure, i.e., $\tilde{w}_i^2$, in case R_i^2 is expected to be large and a high exposure whenever it is presumably small, the trader can try to reduce the overall variance of his strategy. Similarly, he could use each pair of assets i and j ($i \neq j$) for diversification by choosing a large cross exposure $\tilde{w}_i\tilde{w}_j$ whenever he expects that R_iR_j is small and vice versa. Hence, $\sum_{i,j=1}^{N} \mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j, R_iR_j\big)$ measures the investor's capability to control the variance of the return on the strategy by placing appropriate orders according to his own expectations. This is nothing else than risk management.

Hence, for a noise trader it can be better to renounce his strategy altogether and to approximate the mean portfolio by portfolio resampling. By contrast, the fluctuation of a strategy applied by a signal trader comes from *relevant information* and reflects his prediction power as well as his risk-management capabilities (see Theorem 3). This fluctuation possibly makes his strategy profitable and then portfolio resampling could be self-defeating.

> *Portfolio resampling is not suitable for a signal trader if his prediction power and risk-management skills are "good enough."*

The problem is that *unconditional* strategies are not intended to make use of relevant information. Thus even if the investor *has* relevant information, he should not expect to get a satisfactory result by applying an unconditional strategy. This is confirmed by the empirical studies mentioned in the introduction. In particular, portfolio resampling represents an unconditional strategy.

> *Portfolio resampling does not turn a noise trader into a signal trader but it can turn a signal trader into a noise trader.*

Suppose that it is possible to observe n realizations $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$ of some strategy $\tilde{w}$ and the n associated vectors $R_2, R_3, \ldots, R_{n+1}$ of asset returns, i.e., $(\tilde{w}_t, R_{t+1}) \sim (\tilde{w}, R)$ for $t = 1, 2, \ldots, n$. More precisely, it is assumed that $\{(\tilde{w}_t, R_{t+1})\}$ is an ergodic stationary process, but the portfolio weights and asset returns need not be serially independent. Now, it is reasonable to estimate the investor's prediction power, his risk-management capability, and the variance of his strategy by the *method of moments* (Hansen, 1982). These estimates might be used to judge whether the trader has enough power to go without portfolio resampling, i.e., whether Eq. 2 is satisfied.

4.3. The Sample-Mean Strategy

Let $\tilde{w}$ be a noise-trading strategy. It is clear that the mean portfolio is unknown and approximating $\bar{w}$ by portfolio resampling can be unsatisfactory, in particular if the sample size is small. Now, an interesting question is that of whether it simply suffices to substitute $\bar{w}$ by the *sample-mean strategy* associated with $\tilde{w}$, i.e.,

$$\hat{\tilde{w}} = \frac{1}{n} \sum_{t=1}^{n} \tilde{w}_t .$$

In contrast to $\bar{w}$, which is unknown to the investor, the sample-mean strategy is *feasible* in a world of uncertainty. The key note is that $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$ are *true* observations rather than bootstrap replications of $\tilde{w}$ and thus applying $\hat{\tilde{w}}$ could be always better than the actual strategy $\tilde{w}$. This is confirmed by the next theorem.

Theorem 4. *Let $\tilde{w} \in \mathcal{C}$ be a mixed strategy and suppose that the assumptions* ***A1*** *and* ***A2*** *are satisfied. Further, let $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$ be a sample of $n \geq 2$ irrelevant copies of $\tilde{w}$,*[16] *where $\tilde{w}_1'R, \tilde{w}_2'R, \ldots, \tilde{w}_n'R$ are not almost surely identical. Then we have that*

$$\mathrm{E}\Big(u\big(W(1 + r + \hat{\tilde{w}}'R)\big)\Big) > \mathrm{E}\Big(u\big(W(1 + r + \tilde{w}'R)\big)\Big),$$

where $\hat{\tilde{w}} = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t \in \mathcal{C}$ is the sample-mean strategy associated with $\tilde{w}$.

This means every noise-trading strategy $\tilde{w}$ becomes *inadmissible* if it is possible to obtain just a *finite number $n \geq 2$* of copies of $\tilde{w}$. This statement remains valid even if $\hat{\tilde{w}}$ is "far away" from the true expectation $\bar{w}$. Moreover, the observations $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$ need not be independent. The essential condition of Theorem 4 is that these observations are distributed like $\tilde{w}$.

As a typical example suppose that $\tilde{w}$ is some technical trading strategy determined by the past asset returns in a rolling window with fixed window length. Let the process $\{R_t\}$ of asset returns be stationary. Since each $\tilde{w}_t$ is a function of a constant

[16] This means R is stochastically independent of $\tilde{w}_t \sim \tilde{w}$ for $t = 1, 2, \ldots, n$.

number of asset returns up to time t, the process $\{\tilde{w}_t\}$ is stationary, too. Under the i.i.d. assumption, $\tilde{w}$ is a noise-trading strategy. Hence, due to Theorem 4 it is better to create a sample $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n \sim \tilde{w}$ by historical simulation and to apply the corresponding sample mean $\hat{\tilde{w}}$ of $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$ rather than the actual strategy $\tilde{w}$.

In most simulation studies, $\tilde{w}$ represents a statistical portfolio optimization method. It is often assumed that practitioners use a rolling-window approach for estimating the unknown parameters and in the vast majority of cases, the experiments are conducted under the i.i.d. assumption. Hence, the investors are noise traders by construction and $\{\tilde{w}_t\}$ is a stationary process. From Theorem 4 it follows that $\tilde{w}$ is inadmissible in that context. This means applying $\tilde{w}$ violates the principles of rational behavior.

In fact, it makes no sense to use a rolling window for an *unconditional* strategy or if it is clear to the investor that he cannot predict future asset returns, or any function thereof. Consider the blindfolded chimpanzee from Example 2. Why should we use only 10 shots on the dartboard for calculating the sample mean $\hat{\tilde{w}}$, although we know that the chimpanzee is a noise trader and it is simple to create more observations of $\tilde{w}$ so as to reduce the impact of noise?

5. The Markowitz-Usmen Enigma Revisited

The classic theory of rational behavior is due to von Neumann and Morgenstern (1944). It presumes that the decision maker lives in a world of risk. Of course, the problem is that virtually all economic decisions have to be made under uncertainty. There exist many alternative approaches that account for the fact that most real-world phenomena are not risky but uncertain in the sense of Knight (1921). A very robust approach is simply to refrain from probability distributions, e.g., by applying maximin or minimax rules. Another approach is based on the *Choquet integral* (Choquet, 1953), which relies on capacities rather than probabilities. Moreover, the Bayesian approach is based on subjective probabilities (Ramsey, 1926, Savage, 1954).[17] As Bradley (2007) points out,

> "Bayesian decision theories are formal theories of rational agency: they aim to tell us both what the properties of a rational state of mind are (the theory of pure rationality) and what action it is rational for an agent to perform, given the state of mind (the theory of choice)."

This means Bayesian decision theories can be decomposed into an *epistemic* and a *normative* part.[18] The same holds for every other sort of decision theory, independent

[17] Here I do not distinguish between theories where a rational individual act *as if* he maximizes the expected utility based on his subjective probabilities and other theories implying that he actually *does*.

[18] This is also reflected by Markowitz' two-stage portfolio optimization approach discussed in Section 2.

of whether it is Bayesian or not. The first goal, i.e., the epistemic part of the theory, is typically achieved by finding a set of axioms regarding the beliefs and preferences of an individual who is to be considered as "rational." Interestingly, these axioms usually do not say anything about the quality of beliefs and it is not required that the information is "true" in an objective sense, even if this information could be falsified by evidence. Hence, a decision that is based on wrong information can be rational and so an irrational individual can create decisions that are *systematically* better (from an objective viewpoint) than those of a rational individual.

Bayesian decision theories promote an *internal* point of view. This means the strategy of a rational decision maker can be always considered as optimal *given the subjective information he has*, whereas the strategies of other rational subjects typically appear to be suboptimal. Hence, it makes little sense to ask the individual himself whether his strategy is better or worse than the strategy of another individual. This question can be satisfactorily answered only by adopting an *external* point of view by applying objective criteria.

This means for evaluating an investment strategy it is not helpful to stick to the question whether the investor is rational or not. His strategy should be measured by the out-of-sample performance and not by the way he comes to his conclusions. It is no accident that in the Markowitz-Usmen experiment it is neither the rational decision maker nor the Michaud player, who evaluates the outcomes of the strategies, but the referee. The referee lives in a world of risk and not of uncertainty. Thus he has more information than the investors and therefore he is able to calculate the performance measures. This is simply done by approximating the expected utilities with Monte Carlo simulation. Of course, this procedure is inherently frequentistic. Nevertheless, in this context, expected utility is not wrongly considered as a normative theory for decisions under uncertainty. It is just a way to *evaluate* each investment strategy.

Now, I come back to the basic questions raised by Markowitz and Usmen (2003). **Q1**, **Q2**, and **Q3** can be readily answered by using the arguments in the previous sections. For **Q4** and **Q5** I rely on the previous arguments given in this section.

Q1. *How does Michaud's procedure relate to the classic theory of rational behavior according to von Neumann and Morgenstern (1944)?*
Theorem 1 shows that the Michaud procedure has a theoretical foundation in the classic theory of rational behavior.

Q2. *Why does portfolio resampling performs so good in the experiment?*
The Markowitz-Usmen experiment perfectly fits into the framework developed in this paper. The simulation study is constructed in such a way that the "Bayesian player"

and the "Michaud player" are noise traders. The strategies are compared to each other by the expected out-of-sample performance, i.e., the performance measure involved in Theorem 2. Moreover, there are short-selling constraints. The Bayesian approach can lead to boundary solutions, whereas the Michaud procedure takes the average of a large number of points in a convex set and thus leads to better diversified portfolios. This situation is aggravated by the fact that the prior chosen by Markowitz and Usmen is suboptimal. This point will be detailed below.

Q3. *How much does the Michaud procedure contribute to the risk-adjusted expected return in practice? This means after taking the true distribution of asset returns, transaction costs, etc., into account.*
Theorem 3 provides an immediate solution to this question. The expected value of $\phi_\alpha(\tilde{w})$ can be interpreted as a risk-adjusted expected return. A noise trader would gain at most

$$\mathcal{V}_\alpha(\tilde{w}) = \frac{\alpha}{2}\,\mathrm{E}\Big((\tilde{w} - \bar{w})'\Sigma\,(\tilde{w} - \bar{w})\Big)$$

by switching to the mean portfolio. Taking transaction costs into account makes only sense if we assume that $\bar{w}$ is repeatedly applied after each investment period, i.e., if $\bar{w}$ represents a so-called *constant rebalanced portfolio*. Relative to a noise-trading strategy, a constant rebalanced portfolio produces low turnover and thus keeps transaction costs small. Hence, things become even worse for a noise trader if the market is imperfect. By contrast, Theorem 3 shows that portfolio resampling can have a negative impact if the trader acts on relevant information. More precisely, a signal trader could in principle lose

$$\sum_{i=1}^{N}\mathrm{Cov}(\tilde{w}_i, R_i) - \frac{\alpha}{2}\sum_{i,j=1}^{N}\mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j, R_iR_j\big) - \mathcal{V}_\alpha(\tilde{w})$$

by moving from the trading strategy to the mean portfolio. Hence, how much a signal trader can gain or lose by portfolio resampling depends essentially on his prediction power and risk-management skills.

Q4. *Are Savage's axioms of subjective probability (1954) violated by the Michaud player and if so, do we have to blame the Michaud player or the axioms?*
In the Markowitz-Usmen experiment it is not the Michaud player, who evaluates the outcomes, but the referee. So the question is whether the *referee* violates the axioms. The experimental design reveals that he is a frequentist. Hence, the problem is that Markowitz and Usmen (2003) consider Savage's (1954) axioms as correct but evaluate the strategies in the original sense of von Neumann and Morgenstern (1944), i.e., in a frequentistic way. As pointed out before, internalist and externalist theories contradict each other, since a decision that is optimal from a "subjective" viewpoint need not be

optimal from an "objective" point of view and vice versa.

Q5. *Does the Bayesian strategy perform so bad, because the chosen priors are incorrect or because of a poor numerical approximation of the posteriors?*
Harvey et al. (2008) argue that the numerical algorithm used by Markowitz and Usmen (2003) for calculating the predictive distribution provides only a rough approximation. Another argument refers to the prior. Markowitz and Usmen assume that μ is uniformly distributed, but they indeed observe that this prior is not plausible from an economic viewpoint (Markowitz and Usmen, 2003). If the support of a prior is too large, it does not contribute much to the posterior and then Bayesian analysis becomes void. Since the decision essentially depends on the chosen prior, it should be *at least* valid from an economic point of view and its choice should not be (only) driven by pure mathematical reasons (Bade et al., 2008). Of course, this is my very personal, i.e., *subjective*, opinion.

6. Conclusion

Contrary to common belief, portfolio resampling has a strong theoretical foundation in the classic theory of rational behavior according to von Neumann and Morgenstern (1944). If the world was risky but not uncertain in the sense of Knight (1921), nobody would trade on irrelevant information. Without relevant information, the best choice of a trader is a pure strategy. Unfortunately, the world is uncertain and investors make their decisions on the basis of observation and experience. The more a strategy is exposed to irrelevant information, the worse the position of the investor.

Every noise trader could do better by choosing the mean portfolio and the Michaud procedure is nothing else than a computational method to approximate $\bar{w}$. Hence, portfolio resampling could be useful for a noise trader. By contrast, a signal trader with enough prediction power and risk-management skills should refrain from portfolio resampling. The reason is that the Michaud approach aims at eliminating the variance of an investment strategy, but it is precisely the variability that can make the strategy of a signal trader profitable and then portfolio resampling could be self-defeating.

Most empirical studies apply unconditional rather than conditional strategies to evaluate the impact of portfolio resampling. Unconditional strategies are not intended to make use of relevant information and so they can be often improved by portfolio resampling. However, the Michaud approach is not able to improve the performance of conditional strategies. This effect cannot be understood when conducting a Monte Carlo experiment under the widespread i.i.d. assumption. The i.i.d. assumption implies that people are noise traders and this is the reason why the Michaud approach does well in simulation studies.

The analytical results presented in this paper demonstrate that portfolio resampling in principle could be useful, but its potential benefit depends essentially on the quality of information that is used by the investor and the way he processes this information. In particular, the most nonchalant but yet unrealistic assumption in finance, i.e., the random-walk hypothesis, is the reason why portfolio resampling performs well in simulation studies but could be mediocre in real life.

Proofs

In the following it is implicitly assumed that the given expectations, variances, and covariances exist and are finite. Moreover, it is assumed that Σ is positive definite, so that the inverse Σ^{-1} exists.

Proposition 1. *Let f be an increasing and strictly concave function from $\mathbb{R}$ to $\mathbb{R}$. Further, let g be a concave function from $\mathbb{R}^N$ to $\mathbb{R}$. Then $f \circ g$ is strictly concave, too.*

Proof: Since f is strictly concave, it holds that

$$\pi f\big(g(x_1)\big) + (1-\pi) f\big(g(x_2)\big) < f\big(\pi g(x_1) + (1-\pi) g(x_2)\big)$$

for every $\pi \in]0,1[$. Further, since g is concave,

$$\pi g(x_1) + (1-\pi) g(x_2) \leq g\big(\pi x_1 + (1-\pi) x_2\big).$$

Since f is increasing, it turns out that

$$\pi f\big(g(x_1)\big) + (1-\pi) f\big(g(x_2)\big) < f\big(g(\pi x_1 + (1-\pi) x_2)\big)$$

for every $\pi \in]0,1[$, which means that $f \circ g$ is strictly concave.

Theorem 1

We have that

$$\mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big) = \int\int u\big(W(1+r+w'r)\big)\, p(w,r)\, dr\, dw$$

and, since $\tilde{w}$ is independent of R, it follows that $p(w,r) = p(w)p(r)$. This means

$$\begin{aligned}\mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big) &= \int\int u\big(W(1+r+w'r)\big)\, p(w)p(r)\, dr\, dw \\ &= \int\left[\int u\big(W(1+r+w'r)\big)\, p(w)\, dw\right] p(r)\, dr.\end{aligned}$$

Since u is increasing and strictly concave in $W(1+r+w'r)$, which is linear and thus concave in w, Proposition 1 implies that $u\big(W(1+r+w'r)\big)$ is strictly concave in w. Further, since $\tilde{w}$ is mixed, its probability distribution is not degenerated and thus

$$\int u\big(W(1+r+w'r)\big)\,p(w)\,dw < u\big(W(1+r+\bar{w}'r)\big)$$

by Jensen's inequality. Hence, it follows that

$$\mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big) < \int u\big(W(1+r+\bar{w}'r)\big)\,p(r)\,dr = \mathrm{E}\Big(u\big(W(1+r+\bar{w}'R)\big)\Big).$$

Since $\mathcal{C}$ is convex and $\tilde{w}\in\mathcal{C}$, we have that $\bar{w}=\mathrm{E}(\tilde{w})\in\mathcal{C}$.

Theorem 2

The out-of-sample performance of $w^*=\Sigma^{-1}\mu/\alpha$ amounts to

$$\varphi_\alpha(w^*) = \frac{\mu'\Sigma^{-1}\mu}{2\alpha}.$$

Further, we have that

$$(\tilde{w}-w^*)'\Sigma(\tilde{w}-w^*) = \tilde{w}'\Sigma\tilde{w} - 2\tilde{w}'\Sigma w^* + w^{*\prime}\Sigma w^* = \tilde{w}'\Sigma\tilde{w} - \frac{2\tilde{w}'\mu}{\alpha} + \frac{\mu'\Sigma^{-1}\mu}{\alpha^2},$$

so that

$$\frac{\alpha}{2}(\tilde{w}-w^*)'\Sigma(\tilde{w}-w^*) = \frac{\alpha}{2}\tilde{w}'\Sigma\tilde{w} - \tilde{w}'\mu + \frac{\mu'\Sigma^{-1}\mu}{2\alpha} = \varphi_\alpha(w^*) - \varphi_\alpha(\tilde{w}).$$

Hence,

$$\frac{\alpha}{2}\mathrm{E}\Big((\tilde{w}-w^*)'\Sigma(\tilde{w}-w^*)\Big) = \mathrm{E}\Big(\varphi_\alpha(w^*)-\varphi_\alpha(\tilde{w})\Big) = \mathcal{L}_\alpha(\tilde{w}).$$

The decomposition of $\mathcal{L}_\alpha(\tilde{w})$ into $\mathcal{B}_\alpha(\tilde{w})$ and $\mathcal{V}_\alpha(\tilde{w})$ follows immediately from

$$\begin{aligned}\mathcal{L}_\alpha(\tilde{w}) &= \frac{\alpha}{2}\mathrm{E}\Big(\big[(\tilde{w}-\bar{w})+(\bar{w}-w^*)\big]'\Sigma\big[(\tilde{w}-\bar{w})+(\bar{w}-w^*)\big]\Big)\\ &= \frac{\alpha}{2}\mathrm{E}\Big((\tilde{w}-\bar{w})'\Sigma(\tilde{w}-\bar{w})\Big) + \frac{\alpha}{2}(\bar{w}-w^*)'\Sigma(\bar{w}-w^*) = \mathcal{V}_\alpha(\tilde{w}) + \mathcal{B}_\alpha(\tilde{w}).\end{aligned}$$

Since Σ is positive definite we have that

$$\mathcal{V}_\alpha(\tilde{w}) = \frac{\alpha}{2}\mathrm{E}\Big((\tilde{w}-\bar{w})'\Sigma(\tilde{w}-\bar{w})\Big) \geq 0$$

and from

$$\mathcal{B}_\alpha(\tilde{w}) = \frac{\alpha}{2}(\bar{w}-w^*)'\Sigma(\bar{w}-w^*) = \mathcal{L}_\alpha(\bar{w})$$

it follows that $\mathcal{L}_\alpha(\tilde{w}) \geq \mathcal{L}_\alpha(\bar{w})$. Moreover, since $\tilde{w}\in\mathcal{C}$ and $\mathcal{C}$ is a convex set, it holds that $\bar{w}=\mathrm{E}(\tilde{w})\in\mathcal{C}$. If $\tilde{w}$ is a mixed strategy, Jensen's inequality implies that $\mathcal{V}_\alpha(\tilde{w})>0$, i.e., the inequality is strict. Vice versa, if the inequality is strict, it must hold that $\mathcal{V}_\alpha(\tilde{w})>0$. This is only possible if $\tilde{w}$ is mixed.

Theorem 3

First of all, we have that

$$\mathrm{E}\big(\phi_\alpha(\tilde{w})\big) = \mathrm{E}(\tilde{w}'R) - \frac{\alpha}{2}\,\mathrm{E}\Big((\tilde{w}'R)^2\Big)\,,$$

where

$$\mathrm{E}(\tilde{w}'R) = \sum_{i=1}^{N}\mathrm{E}(\tilde{w}_iR_i) = \sum_{i=1}^{N}\bar{w}_i\mu_i + \sum_{i=1}^{N}\mathrm{Cov}(\tilde{w}_i,R_i) = \bar{w}'\mu + \sum_{i=1}^{N}\mathrm{Cov}(\tilde{w}_i,R_i)\,.$$

Moreover, it follows that

$$\mathrm{E}\Big((\tilde{w}'R)^2\Big) = \mathrm{E}\big(\tilde{w}'RR'\tilde{w}\big) = \mathrm{E}\big(\tilde{w}'\Sigma\tilde{w}\big) + \mathrm{E}\Big(\tilde{w}'\big(RR'-\Sigma\big)\tilde{w}\Big)\,,$$

where

$$\mathrm{E}\Big(\tilde{w}'\big(RR'-\Sigma\big)\tilde{w}\Big) = \sum_{i,j=1}^{N}\mathrm{E}\Big(\tilde{w}_i\tilde{w}_j\big(R_iR_j - \mathrm{E}(R_iR_j)\big)\Big) = \sum_{i,j=1}^{N}\mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j,R_iR_j\big)\,.$$

This means

$$\mathrm{E}\big(\phi_\alpha(\tilde{w})\big) = \bar{w}'\mu + \sum_{i=1}^{N}\mathrm{Cov}(\tilde{w}_i,R_i) - \frac{\alpha}{2}\left(\mathrm{E}\big(\tilde{w}'\Sigma\tilde{w}\big) + \sum_{i,j=1}^{N}\mathrm{Cov}\big(\tilde{w}_i\tilde{w}_j,R_iR_j\big)\right).$$

Theorem 4

Since $\tilde{w}$ and its copies $\tilde{w}_1,\tilde{w}_2,\ldots,\tilde{w}_n$ are irrelevant, it follows that

$$p(w_t,r) = p(w_t)p(r) = p(w)p(r) = p(w,r)$$

and thus

$$\begin{aligned}
\mathrm{E}\Big(u\big(W(1+r+\tilde{w}_t'R)\big)\Big) &= \int\!\!\int u\big(W(1+r+w_t'r)\big)\,p(w_t,r)\,dr\,dw_t \\
&= \int\!\!\int u\big(W(1+r+w'r)\big)\,p(w,r)\,dr\,dw \\
&= \mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big)\,, \qquad t=1,2,\ldots,n\,,
\end{aligned}$$

i.e.,

$$\mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big) = \mathrm{E}\left(\frac{1}{n}\sum_{t=1}^{n}u\big(W(1+r+\tilde{w}_t'R)\big)\right).$$

Since $u\big(W(1+r+x)\big)$ is strictly concave in x, we have that

$$\frac{1}{n}\sum_{t=1}^{n}u\big(W(1+r+\tilde{w}_t'R)\big) \leq u\left(W\left(1+r+\frac{1}{n}\sum_{t=1}^{n}\tilde{w}_t'R\right)\right) = u\Big(W\big(1+r+\hat{\tilde{w}}'R\big)\Big)$$

with probability one, but since $\tilde{w}_1'R, \tilde{w}_2'R, \ldots, \tilde{w}_n'R$ are not almost surely identical, it follows that

$$\frac{1}{n}\sum_{t=1}^{n} u\big(W(1+r+\tilde{w}_t'R)\big) < u\Big(W(1+r+\hat{\tilde{w}}'R)\Big)$$

with positive probability. This means

$$\mathrm{E}\Big(u\big(W(1+r+\tilde{w}'R)\big)\Big) < \mathrm{E}\Big(u\big(W(1+r+\hat{\tilde{w}}'R)\big)\Big).$$

Finally, since $\mathcal{C}$ is convex and $\tilde{w}_t \in \mathcal{C}$ for $t = 1, 2, \ldots, n$, we have that $\hat{\tilde{w}} = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t \in \mathcal{C}$.

References

Arrow, K.J., "The theory of risk aversion," in Yrjo Jahnssonin Saatio, ed., *Aspects of the Theory of Risk Bearing*, Markham, 1971, pp. 90–109.

Back, K., "Martingale pricing," *Annual Review of Financial Economics*, 2010, *2*, 235–250.

Bade, A., G. Frahm, and U. Jaekel, "A general approach to Bayesian portfolio optimization," *Mathematical Methods of Operations Research*, 2008, *70*, 337–356.

Becker, F., M. Gürtler, and M. Hibbeln, "Markowitz versus Michaud: portfolio optimization strategies reconsidered," *The European Journal of Finance*, 2013. DOI: 10.1080/1351847X.2013.830138.

Black, F., "Noise," *Journal of Finance*, 1986, *41*, 529–543.

Bradley, R., "A unified Bayesian decision theory," *Theory and decision*, 2007, *63*, 233–263.

Breiman, L., "Bagging predictors," *Machine Learning*, 1996, *24*, 123–140.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press, 1997.

Choquet, G., "Theory of capacities," *Annales de l'Institut Fourier*, 1953, *5*, 131–295.

Cochrane, J.H., *Asset Pricing*, Princeton University Press, 2005.

DeMiguel, V., L. Garlappi, F.J. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Management Science*, 2009, *55*, 798–812.

Fama, E.F., "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, 1970, *25*, 383–417.

Fernandes, J.L. Barros, J.R. Haas Ornelas, and O.A. Martínez Cusicanqui, "Combining equilibrium, resampling, and analyst's views in portfolio optimization," *Journal of Banking and Finance*, 2012, *36*, 1354–1361.

Ferson, W., "Market efficiency and forecasting," in "Forecasting Expected Returns in the Financial Markets," Academic Press, 2007.

Fletcher, J. and J. Hillier, "An examination of resampled portfolio efficiency," *Financial Analysts Journal*, 2001, *57*, 66–74.

Frahm, G., "The determinants of the risk functions of estimators for expected asset returns," Technical Report, University of Cologne 2011.

_ , "The fundamental theorem of asset pricing for liquid financial markets," Working paper AP 2013–1, Helmut Schmidt University 2013.

_ and C. Memmel, "Dominating estimators for minimum variance portfolios," *Journal of Econometrics*, 2010, *159*, 289–302.

_ , T. Wickern, and C. Wiechers, "Multiple tests for the performance of different investment strategies," *Advances in Statistical Analysis*, 2012, *96*, 343–383.

Hansen, L.P., "Large sample properties of generalized method of moments estimators," *Econometrica*, 1982, *50*, 1029–1054.

Harvey, C.R., J.C. Liechty, and M.W. Liechty, "Bayes vs. resampling: a rematch," *Journal of Investment Management*, 2008, *6*, 1–17.

Herold, U. and R. Maurer, "Portfolio choice and estimation risk. A comparison of Bayesian to heuristic approaches," *ASTIN Bulletin*, 2006, *36*, 135–160.

Jarrow, R., "The Third Fundamental Theorem of Asset Pricing," *Annals of Financial Economics*, 2012, *7*. DOI: 10.1142/S2010495212500078.

Jarrow, R.A. and M. Larsson, "The meaning of market efficiency," *Mathematical Finance*, 2012, *22*, 1–30.

Jobson, J.D. and B. Korkie, "Estimation for Markowitz efficient portfolios," *Journal of the American Statistical Association*, 1980, *75*, 544–554.

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

Knight, F.H., *Risk, Uncertainty and Profit*, Hart, Schaffner and Marx, 1921.

Kohli, J., "An Empirical Analysis of Resampled Efficiency," Master's thesis, Worcester Polytechnic Institute 2005.

Kovaleski, D., "Northfield to utilize New Frontier's portfolio optimization process," *Pensions & Investments*, 2001, *29*, 39.

Kroll, Y., H. Levy, and H.M. Markowitz, "Mean-variance versus direct utility maximization," *Journal of Finance*, 1984, *39*, 47–61.

LeRoy, S.F., "Risk aversion and the martingale property of stock prices," *International Economic Review*, 1973, *14*, 436–446.

Lo, A.W., H. Mamaysky, and J. Wang, "Foundations of technical analysis: computational algorithms, statistical inference and empirical implementation," in E. Acar and S. Satchell, eds., *Advanced Trading Rules*, Butterworth-Heinemann, 2002, chapter 2, pp. 42–111.

Lucas, R.E., "Asset prices in an exchange economy," *Econometrica*, 1978, *46*, 1429–1445.

Malkiel, B., "The Efficient Market Hypothesis and its critics," *Journal of Economic Perspectives*, 2003, *17*, 59–82.

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

_ , "The "great confusion" concerning MPT," *AESTIMATIO, The IEB International Journal of Finance*, 2012, *4*, 8–27.

_ **and N. Usmen**, "Resampled frontiers versus diffuse Bayes," *Journal of Investment Management*, 2003, *1*, 1–17.

Michaud, R.O., *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*, McGraw-Hill, 1998.

_ **and R. Michaud**, ""An examination of resampled portfolio efficiency": comment and response," *Financial Analysts Journal*, 2003, *59*, 15–16.

_ **and** _ , "Defense of Markowitz-Usmen," Technical Report, New Frontier Advisors 2008.

Mossin, J., "Optimal multiperiod portfolio policies," *The Journal of Business*, 1968, *41*, 215–229.

Pratt, J.W., "Risk aversion in the small and in the large," *Econometrica*, 1964, *32*, 122–136.

Pulley, L.B., "A general mean-variance approximation to expected utility for short holding periods," *Journal of Financial and Quantitative Analysis*, 1981, *16*, 361–373.

Ramsey, F.P., "Truth and probability," in "Philosophical Papers," Cambridge University Press, 1926.

Ross, S.A., *Neoclassical Finance*, Princeton University Press, 2005.

Savage, L.J., *The Foundations of Statistics*, Wiley, 1954.

Scherer, B., "Portfolio resampling: review and critique," *Financial Analysts Journal*, 2002, *58*, 98–109.

_ , "Resampled efficiency and portfolio choice," *Financial Markets and Portfolio Management*, 2004, *18*, 382–398.

_ , "A note on the out-of-sample performance of resampled efficiency," *Journal of Asset Management*, 2006, *7*, 170–178.

_ , "Can robust portfolio optimisation help to build better portfolios?," *Journal of Asset Management*, 2007, *7*, 374–387.

Timmermann, A. and C.W.J. Granger, "Efficient market hypothesis and forecasting," *International Journal of Forecasting*, 2004, *20*, 15–27.

von Neumann, J. and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

Wolf, M., "Resampling vs. shrinkage for benchmarked managers," *Wilmott Magazine*, 2007, *January 2007*, 76–81.