Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

A growth maximizing contrarian trading strategy

Author: **Marek Janča**

Supervisor: **PhDr. Ladislav Krištoufek**

Academic Year: **2010/2011**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 30, 2011

Signature

Acknowledgments

I owe many thanks to PhDr. Ladislav Krištoufek for his valuable comments and suggestions. The responsibility for all errors is mine.

Abstract

The purpose of our thesis is to build a contrarian trading strategy that would maximize growth rate of our wealth. In the first part, we derive the strategy by algebraic means. In particular, we exploit that growth maximization is equivalent to period by period maximization of log wealth. We approximate the log optimal portfolio by a mean-variance efficient portfolio and specify the first and second conditional moment by a dynamic econometric model. In the second part, we discuss deficiencies of our strategy and use Monte Carlo simulations to create a modification that should perform better. In the final part, we demonstrate viability of the strategy on historical data. Assuming unlimited leverage and mild transaction costs, the strategy was able to generate annual geometric mean return close to 24%.

JEL Classification	G11
Keywords	Portfolio Choice; Investment Decisions

Author's e-mail	jancamarek87@gmail.com
Supervisor's e-mail	ladislav.kristoufek@email.cz

Abstrakt

Účelem práce je vytvořit obchodní strategii, která by využívala jevu "contrarian profitability". První část práce se věnuje samotnému odvozování strategie. Nejprve využijeme faktu, že strategie maximalizuje růst, právě pokud v každé periodě maximalizuje logaritmus hodnoty našeho bohatství. Poté log-optimální portfolio aproximujeme portfoliem, které leží na efektivní hranici (termín z oblasti moderní teorie portfolia). První a druhé podmíněné momenty specifikujeme pomocí dynamického ekonometrického modelu. V druhé části prodiskutujeme nedostatky naší strategie a pomocí Monte Carlo simulací ji modifikujeme. V závěrečné části demonstrujeme životaschopnost strategie na historických datech. Za předpokladu neomezené páky a rozumných transakčních nákladů jsme byli schopni dosáhnout průměrného ročního zhodnocení kolem 24%.

Klasifikace JEL	G11
Klíčová slova	Výber portfolia; Investiční rozhodování

E-mail autora	jancamarek87@gmail.com
E-mail vedoucího práce	ladislav.kristoufek@email.cz

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since the early 1980's, a number of financial institutions have engaged in quantitative trading strategies that became later known as contrarian trading. The defining feature of these strategies was simultaneously opening long positions in stocks that had performed poorly in the past and shorting stocks that had performed well (selling past winners and buying past losers). Portfolios constructed in such a manner were rebalanced in short time intervals, usually ranging from minutes to days. Moreover, contrarian strategies were almost invariably market neutral. According to first-hand accounts, those strategies were able to generate excess risk adjusted returns (see for example Pole 2007; Thorp 2003). For obvious reasons, a detailed description of strategies used by practitioners was never revealed. The topic of contrarian profitability received some limited academic attention (for example Lehmann 1990; Lo & MacKinlay 1990a; Avellaneda & Lee 2008). The existing literature follows a bottom-up approach; i.e., haphazard trading rules are presented without any assertion of optimality for purely demonstrative purposes.

The purpose of this Bachelor Thesis is to build a contrarian trading strategy that would be able to replicate the windfall generated by the contrarian strategies in the past three decades. In comparison to the existing literature, we have chosen a top-down approach. We started by setting an objective measure of optimality - maximization of the growth rate of wealth - and derived the trading rules from there. For convenience, we have worked within a discrete time framework. While our effort to build the strategy has met with success, we consider the insight gained in the process to be much more valuable than the resulting strategy itself. In particular, we would like to highlight the discussion of problems linked to using a mean-variance efficient portfolio as a proxy for the growth optimal portfolio.

The rest of the thesis is structured as follows: In Chapter 2, We provide the motivation behind growth maximization and derive a trading strategy that is approximately optimal under some idealized assumptions. In Chapter 3, we comment on the practical limitation of our strategy and perform Monte Carlo simulation as a means of finding a modification that might perform better. In Chapter 4, we demonstrate our strategy on the historical returns of selected stocks. Chapter 5 concludes the thesis by summing up the results and indicating possible ways of improving our strategy. In order to allow the reader to fully focus, we have decided to keep the thesis as short as possible. For this reason, we have moved all non-essential material into the Appendix. Among other things, the Appendix contains derivations of formulae, proofs of various claims and detailed reasoning behind the selection of stocks used in Chapter 4. Source code of the simulations described in the thesis will be made available on request.

Chapter 2

The objective and approximate solution

We will begin by explaining why we have chosen growth maximization as our objective. We will proceed by introducing a general framework, notations and assumptions about institutional setting of the market as well as assumptions about return process. Finally, we will derive a strategy that (approximately) fulfills our objective.

2.1 Objective

Economic theory dictates that the relevant criterion that defines the best strategy is the maximization of a subjective utility function of wealth. While useful in theory, utility function is an elusive concept that is of little practical help when faced with investment decisions. To quote (Roy 1952):

> "In calling in a utility function to our aid, an appearance of generality is achieved at the cost of a loss of practical significance, and applicability in our results. A man who seeks advice about his actions will not be grateful for the suggestion that he maximize expected utility."

In the late 1950's Kelly (1956) and Latane (1959) proposed an alternative approach that later developed into the so called growth optimal portfolio theory. The cornerstone of this approach is not maximization of the subjective utility function but rather finding a strategy that maximizes the growth rate of wealth in the sense that, in the long run, it *almost surely* leads to more wealth than any other strategy.

This approach is especially appealing for short-term speculative strategies. To see why, realize that a short rebalancing interval implies a large number of periods in which the strategy will be applied. With a large number of periods, we can expect the limiting property to translate into actual results. Since developing a short-run speculative strategy is our exact goal, we have opted to take growth maximization as our objective.

In the majority of cases, the contrarian trading strategies used by practitioners were market neutral (Pole 2007). In other words, portfolios based on the strategies were ex ante designed to be uncorrelated with the market. In the process of building the strategy, we will retain this requirement. Doing so allows us to isolate gains stemming from contrarian profitability from market returns. For simplicity, we will use dollar neutrality (size of long positions is equal to size of short positions) as a proxy for market neutrality.

2.2 Framework

2.2.1 Time setting

As was pointed out in the introduction, the derivation will be conducted within a discrete time framework. We denote the number of periods in which we will be carrying out the strategy by T. For now, the length of a period is left unspecified. Decision point τ represents the end of period τ. ψ_τ will denote the information set available at decision point τ. In our case, it will contain the realized return from before and in the period τ.

2.2.2 Stocks, general characteristics

We denote the number of stocks of interest by n. We will only consider cases $n \geq 2$. In line with Hakansson & Ziemba (1995), we will assume perfect markets. This assumption embodies the following:

- There are no trading costs or taxes. (This assumption will be partially dropped in the last chapter.)
- The investor has no price impact. (This assumption will be partially dropped in the last chapter.)
- Any stock could be sold short.

- The investor has full use of funds obtained via short selling. (This assumption will be partially dropped in the last chapter.)
- Stocks are infinitely divisible; i.e., the invested amount can be any real number.

In order to simplify the analysis, we will also assume that the riskless rate of interest is zero.

2.2.3 Stocks, return process

Let $r_{\tau,i}$ be the simple rate of return of the stock i from time $\tau - 1$ to time τ. Formally:

$$r_{\tau,i} = \frac{P_{\tau,i} - P_{\tau-1,i}}{P_{\tau-1,i}},$$

where $P_{\tau,i}$ is the price of stock i at time τ. The use of simple returns (in contrast to continuously compounded log returns) is required by the upcoming sections. The vector of returns at period τ is defined as: $\boldsymbol{r_\tau} = (r_{\tau,1}, r_{\tau,2}, ..., r_{\tau,n})'$. In the subsequent text the term *rate of return* will be replaced by the shorter term *return*.

For convenience we will assume that the return process started in the infinite past and that the conditional second moments are finite in every period. Additional assumptions about the return process will be introduced later in this chapter.

To simplify the notation later on, we will denote the conditional expected return and variance of an individual stock i by:

$$\mu_{\tau,i} = \mathbb{E}[r_{\tau,i}|\psi_{t-1}],$$
$$\sigma^2_{\tau,i} = \text{var}[r_{\tau,1}|\psi_{t-1}].$$

Furthermore, we will denote the conditional variance-covariance matrix and vector of conditional expected returns by

$$\boldsymbol{\mu}_\tau = \mathbb{E}[\boldsymbol{r}_\tau|\psi_{t-1}],$$
$$\boldsymbol{\Sigma}_\tau = \text{var}[\boldsymbol{r}_\tau|\psi_{t-1}].$$

Next, we need a few assumptions that will guarantee the existence and uniqueness of the solution of optimization problems we encounter later. In particular we will assume:

- There are no arbitrage opportunities. Formally, in every period there is no $\boldsymbol{w_\tau}$ such that $\boldsymbol{w_\tau}'\boldsymbol{\iota} = 0$ and $\mathbb{P}[\boldsymbol{w}_\tau'\boldsymbol{r}_\tau > 1|\psi_\tau] = 1$. This assumption prevents us from generating riskless profit without any investment.

- $\boldsymbol{\Sigma}_\tau$ has rank n. With the first assumption in place, this assumption only excludes existence of stocks that has same return *almost surely*.

- The conditional expected return of at least one stock differs from the others, formally $\forall k \in \mathbb{R} : \mathbb{P}[\mu_\tau = k\boldsymbol{\iota}] = 0$.

2.2.4 Portfolio, wealth

$w_{\tau,i}$ will denote the net position in stock i held in the period t. $w_{\tau,i} < 0$ corresponds to selling short. For example, $w_{\tau,i} = 100$ means that our holding of stock i during the period τ are worth 100. Let $\boldsymbol{w}_\tau = (w_{\tau,1}, w_{\tau,2}, ..., w_{\tau,n})'$. Since a portfolio is uniquely defined by a vector of position sizes, we will be using the two terms interchangeably.

We will denote total conditional expected portfolio return and total conditional portfolio variance by

$$\begin{aligned}\mu_{\tau,p} &= \mathbb{E}[\boldsymbol{w}_t'\boldsymbol{r}_\tau|\psi_{t-1}],\\ \sigma^2_{\tau,p} &= \mathrm{Var}[\boldsymbol{w}_\tau'\boldsymbol{r}_t|\psi_{t-1}].\end{aligned}$$

Keep in mind that both $\mu_{\tau,p}$ and $\sigma^2_{\tau,p}$ are a function of w_τ.

V_τ will denote our wealth (or capital - the two words will be used synonymously) at the decision point t. V_0 will represent our starting wealth. Therefore, the formula that relates portfolio weights, returns and capital at any period is $V_\tau = V_{\tau-1} + \boldsymbol{w}_\tau'\boldsymbol{\tau}_t$. Intuitively, our wealth at the end of period τ is the sum of the wealth of the previous period and the rate of return of all stocks under consideration multiplied by our position in the particular stock.

2.3 Solution

To help the reader anticipate the solution to the problem outlined earlier, we provide a short summary of steps taken. Each step listed below corresponds to a separate subsection.

1. We will restate the problem of growth maximization as a problem of period-by-period maximization of log wealth.
2. We will derive an approximation of our objective function and show that under this approximation, the portfolio of our interest must necessary lie on an efficient frontier as understood by the modern portfolio theory.
3. We will formalize the problem of finding the efficient frontier and derive a closed form solution under the dollar neutrality constraint.
4. We will use the previous two results and identify the point at the efficiency frontier, where the approximation for our objective function is maximized.
5. We will use our belief about the behavior of the return process to specify functional forms of conditional moments used in the previous steps. At this point, we will also comment on estimation of the parameters entering the functional forms.

2.3.1 Growth rate maximization corresponds to period-by-period maximization of log wealth

Our goal is to find a strategy that optimizes the growth rate of our wealth in the sense that it *almost surely* leads to more capital than any other strategy in the long run. Because the formalization of the previous requirement would require us to introduce a disproportionate amount of new concepts, we have decided to omit it and proceed directly to the first important observation: According to Algoet & Cover (1988), a strategy maximizes the growth rate of wealth if and only if it sequentially maximizes the expected logarithmic wealth conditional on information set available at the relevant decision point. Formally, we are required to solve:

Maximize:

$$\mathbb{E}[\log(V_\tau)|\psi_{\tau-1}]. \tag{2.1}$$

Subject to:

$$\boldsymbol{w}_{\tau}'\boldsymbol{\iota} = 0,$$

sequentially at each decision point $\tau - 1$ for all $\tau \in \{1, ..., T\}$. ($\boldsymbol{\iota}$ denotes a unit vector.)

Because $V_{\tau-1}$ is already determined at the decision point $\tau - 1$, the problem can be equivalently stated using the objective function:

$$\boxed{\mathbb{E}[\log(\frac{V_{\tau}}{V_{\tau-1}})|\psi_{\tau-1}].} \tag{2.2}$$

The following paragraph does not constitute a part derivation. Rather it aims to provide some intuiting behind the first result. By the properties of the logarithmic function, we have

$$\log(\frac{V_{\tau}}{V_{\tau-1}}) \approx \frac{V_{\tau} - V_{\tau-1}}{V_{\tau-1}}$$

for small changes of V_{τ}. Note that the right side is a commonly used expression of the growth rate. In light of this realization the first result is intuitive; maximization of long term growth is equivalent to maximization of the growth rate period by period.

2.3.2 Approximation of the objective function

Maximizing the objective function (2.2) would require us to specify the distribution of $\boldsymbol{r}_{\tau}$ conditional on $\psi_{\tau-1}$. Even if we specify the conditional distribution, the optimization would be difficult because it would involve an evaluation of some non-trivial multiple integral. Consequently, we are forced to find a suitable approximation of (2.2). Doing so is the purpose of this section.

We will resort to the so called approximation by the first two moments. The approximation is partially based on the work Thorp (1997) and Pulley (1983). The idea behind the approximation is the following: Consider a portfolio $\boldsymbol{w}_{\tau}$ with an expected return $\mathbb{E}[\boldsymbol{w}_{\tau}\boldsymbol{r}_{\tau}|\psi_{\tau-1}]$ and variance $\text{var}[\boldsymbol{w}_{\tau}\boldsymbol{r}_{\tau}|\psi_{\tau-1}]$. Next, consider a finite sequence of bets on results of tosses with a biased coin such that the expected profit of the sequence of such bets is equal to the expected profit of our portfolio and the variance of profit resulting from such bets is equal to the portfolio variance. In our approximation, we will replace the original portfolio with a sequence of such bets and compute the value of the objective function. Finally, we observe what happens if we let the lengths of the sequence $M \to \infty$ while keeping the overall return and variance constant. (Heuristically, we consider longer and longer sequences of smaller and smaller bets.)

We will now carry out the procedure step by step. Firstly we define a finite sequence of independent random variables $y_1, ..., y_M$ in the following way:

$$\mathbb{P}[y_m = \frac{\mu_{\tau,p}}{M} + \frac{\sigma^2_{\tau,p}}{\sqrt{M}}] = \frac{1}{2},$$
$$\mathbb{P}[y_m = \frac{\mu_{\tau,p}}{M} - \frac{\sigma^2_{\tau,p}}{\sqrt{M}}] = \frac{1}{2}.$$

In light of the heuristic description provided above, y_m represents the result of a bet on the mth coin toss. Note that

$$\text{var}[\sum_{m=1}^{M} y_m] = \sigma^2_{\tau,p},$$
$$\mathbb{E}[\sum_{m=1}^{M} y_m] = \mu_{\tau,p}.$$

In other words, both the variance and expected return of the sum of the variables $y_1, ...y_M$ correspond to the conditional moments of the portfolio. We now replace the portfolio return with the return of the series of the bets on the coin tosses:

$$\begin{aligned}\frac{V_\tau}{V_{\tau-1}}|\psi_{\tau-1} &= \frac{V_{\tau-1} + \boldsymbol{w}'_\tau \boldsymbol{r}_\tau}{V_{\tau-1}}|\psi_{\tau-1} \\ &\approx \frac{V_{\tau-1} + \sum_{m=1}^{M} y_m}{V_{\tau-1}} \\ &= 1 + \sum_{m=1}^{M} \frac{y_m}{V_{\tau-1}}.\end{aligned}$$

If we assume that both values $y_1, ...y_M$ can attain are sufficiently small, we can approximate the above expression:

$$1 + \sum_{m=1}^{M} \frac{y_m}{V_{\tau-1}} \approx \prod_{m=1}^{M}(1 + \frac{y_m}{V_{\tau-1}}).$$

Our original interest lies in:

$$\mathbb{E}[\log(\frac{V_\tau}{V_{\tau-1}}|\psi_{\tau-1})].$$

With the help of the previous approximation and some elementary algebra, we obtain:

$$\begin{aligned}\mathbb{E}[\log(\frac{V_\tau}{V_{\tau-1}}|\psi_{\tau-1})] &\approx \mathbb{E}[\log(\prod_{m=1}^{M}(1+\frac{y_m}{V_{\tau-1}}))]\\ &= \mathbb{E}[\sum_{m=1}^{M}\log(1+\frac{y_m}{V_{\tau-1}})],\\ &= \sum_{m=1}^{M}\mathbb{E}[\log(1+\frac{y_m}{V_{\tau-1}})].\end{aligned}$$

From the definition of y_m:

$$\begin{aligned}\sum_{m=1}^{M}\mathbb{E}[\log(1+\frac{y_m}{V_{\tau-1}})] = \sum_{m=1}^{M}\Bigg(&\frac{1}{2}\log(1+\frac{\mu_{\tau,p}}{V_{\tau-1}M}+\frac{\sigma_{\tau,p}}{V_{\tau-1}\sqrt{M}})+\\ &+\frac{1}{2}\log(1+\frac{\mu_{\tau,p}}{V_{\tau-1}M}-\frac{\sigma_{\tau,p}}{V_{\tau-1}\sqrt{M}})\Bigg).\end{aligned}$$

Now, let us define $g_M(\mu_{\tau,p},\sigma_{\tau,p})$ as:

$$\begin{aligned}g_M(\mu,\sigma) = \sum_{m=1}^{M}\Bigg(&\frac{1}{2}\log(1+\frac{\mu_{\tau,p}}{V_{\tau-1}M}+\frac{\sigma_{\tau,p}}{V_{\tau-1}\sqrt{M}})+\\ &+\frac{1}{2}\log(1+\frac{\mu_{\tau,p}}{V_{\tau-1}M}-\frac{\sigma_{\tau,p}}{V_{\tau-1}\sqrt{M}})\Bigg).\end{aligned}$$

Expanding $g_M(\mu,\sigma)$ into a bivariate Taylor series around $(0,0)$ yields:

$$g_M(\mu_{\tau,p},\sigma_{\tau,p}) = \frac{\mu_{\tau,p}}{V_{\tau-1}} - \frac{\sigma^2_{\tau,p}}{2V^2_{\tau-1}} + o(\frac{1}{\sqrt{M}}).$$

For details, please refer to Section A.1. Finally, by letting the number of subperiod $M\to\infty$ we obtain:

$$\boxed{\mathbb{E}[\log(\frac{V_\tau}{V_{\tau-1}}|\psi_{\tau-1}) \approx g_\infty(\mu_{\tau,p},\sigma_{\tau,p}) = \frac{\mu_{\tau,p}}{V_{\tau-1}} - \frac{\sigma^2_{\tau,p}}{2V^2_{\tau-1}}.}$$

A simple examination of the above result reveals that $g_M(\mu_{\tau,p}, \sigma_{\tau,p})$ is increasing in expected portfolio return $\mu_{\tau,p}$ and decreasing in the portfolio variance $\sigma_{\tau,p}$. This result implies that the portfolio that maximizes $g_M(\mu_{\tau,p}, \sigma_{\tau,p})$ must have maximum expected return for a given amount of variance.

2.3.3 Identifying the efficient frontier

In the previous section, we have shown that the sought portfolio must have maximum expected return for a given amount of variance. Finding the set of such portfolios is the central topic of this subsection. As per our objective, we will require those portfolios to be dollar neutral.

The problem under consideration, yet with slightly different constraints and parameterization, was first introduced by Markowitz (1952) in an attempt to develop both a normative and positive strategy of portfolio selection. Markowitz named the set of all such portfolios *the efficient frontier* and the process of finding the set *mean-variance optimization*. We will be using this nomenclature (with a slight modification) throughout the thesis.

For a given positive level of portfolio conditional variance σ^2, the problem of finding a portfolio with maximum expected return can be formalized as follows:

Maximize:

$$\boldsymbol{w}_\tau' \boldsymbol{\mu}_\tau.$$

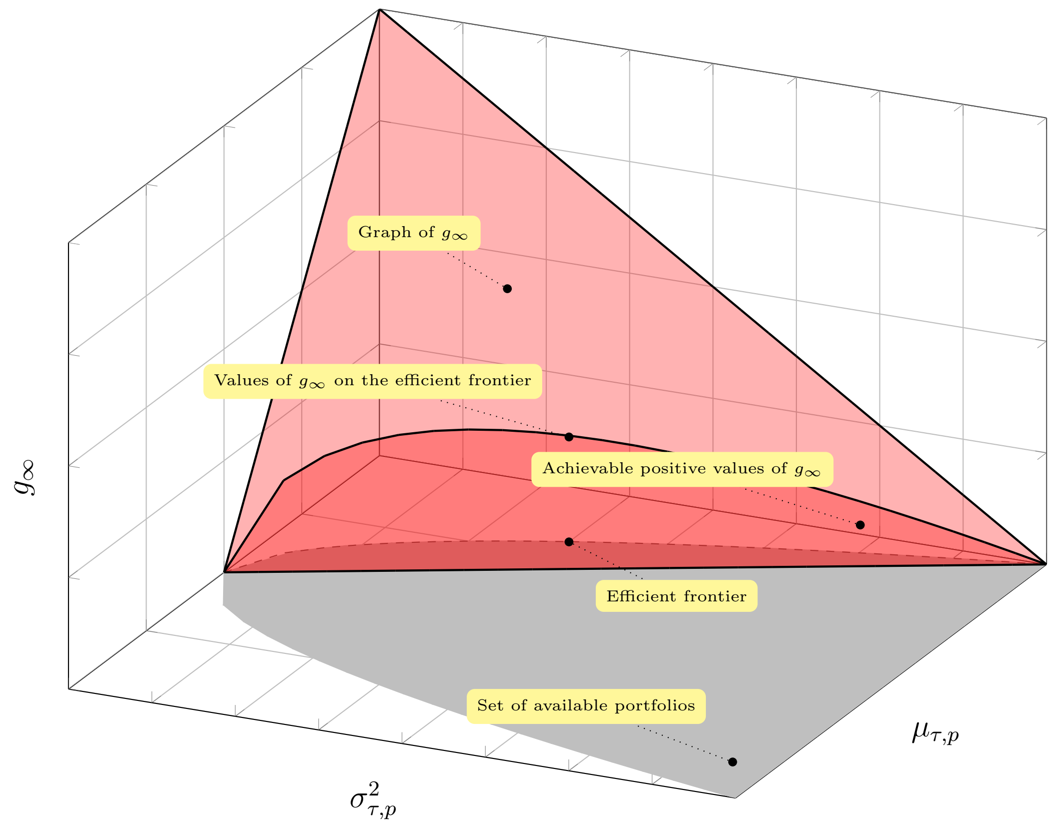
Subject to:

$$\boldsymbol{w}_\tau' \boldsymbol{\Sigma}_\tau \boldsymbol{w}_\tau = \sigma^2,$$
$$\boldsymbol{w}_\tau' \boldsymbol{\iota} = 0.$$

Let us first discuss the existence of the solution. For any given σ^2, the set of portfolios (when speaking about a portfolio, we are speaking about some $\boldsymbol{w}_\tau \in \mathbb{R}^n$) that satisfy all constraints is nonempty, closed and bounded. While the first two properties are obvious, the third one is a bit trickier to prove (the proof is provided in Section A.2). The objective function is a scalar product and as such is continuous. Therefore, the existence of a solution follows from theorem 4.16 in Rudin (1976).

With this knowledge in mind, we can rigorously define the efficient frontier as the set of solutions to the above problem for all nonnegative σ^2. (We realize that our definition slightly differs from how the efficient frontier is understood in the modern portfolio theory literature.) In light of this definition, we can restate the conclusion of the previous section: The portfolio that maximizes $g_M(\mu_{\tau,p}, \sigma_{\tau,p})$ must lie on the efficient frontier. The situation is depicted in Figure 2.1.

Figure 2.1: Efficient frontier and g_∞



Source: author's computations.

The next step is to a derive closed form solution of the above optimization problem. The problem can be solved by means of Lagrange multipliers (for details, please refer to Section A.3). The solution parameterized by σ is given by

$$\boxed{\boldsymbol{w}_\tau = \frac{\sigma}{\sqrt{K}} \boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}),} \tag{2.3}$$

where L and K are given by

$$L = (\boldsymbol{\iota}'\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{\iota})^{-1}(\boldsymbol{\iota}'\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{\mu}_\tau),$$
$$K = (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota})'\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}).$$

An important observation is that the optimal positions $\boldsymbol{w}_\tau$ of the mean-variance efficient portfolios for σ^2 are linear in σ. Consequently, the relative positions sizes in portfolio on the efficient frontier do not depend on the selected level of portfolio variance. In other words, movement alongside the efficient frontier corresponds to changes in total (absolute) size of the investment, not the relative position sizes.

Before we proceed, consider the expected return of the mean-variance efficient portfolio. The expected return is given by

$$\boxed{\boldsymbol{w}_\tau'\boldsymbol{\mu}_\tau = \frac{\sigma}{\sqrt{K}}\boldsymbol{\mu}_\tau'\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}).} \tag{2.4}$$

It is easy to see that the expected conditional portfolio return is linear in σ as well.

2.3.4 Which portfolio from the efficient frontier is the best?

We have previously derived an approximation of our objective function and found that the portfolio that maximizes that approximation lies on the efficient frontier. Furthermore, we have derived a closed form expression for the efficient frontier as a function of σ. Our current task is to select the best portfolio (in terms of our approximation) from the efficient frontier. Because we have parameterized the efficiency frontier by σ, our task comes down to choosing the right value of σ. Heuristically, we have already determined relative position sizes and what remains to be done is to determine absolute position sizes (the size of total absolute investment).

First step is to express the approximation $g_M(\mu_{\tau,p}, \sigma_{\tau,p})$ as a function of σ only. Our original formula was:

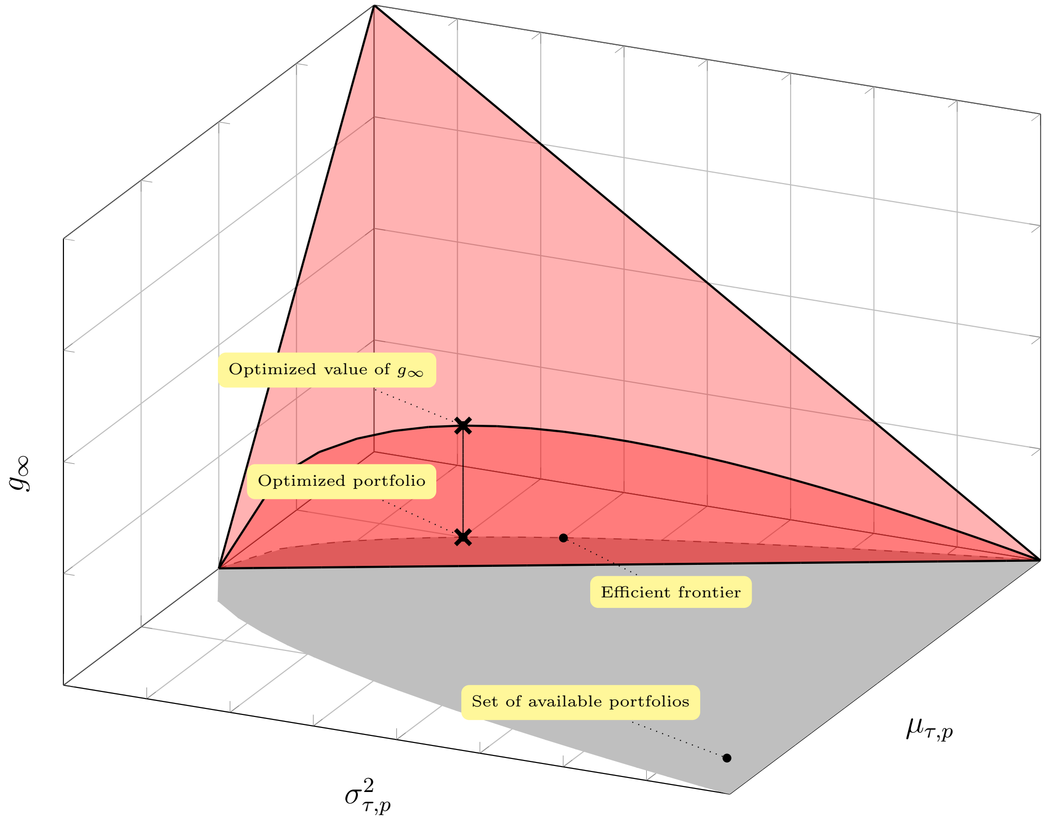
$$g_\infty(\mu_{\tau,p}, \sigma_{\tau,p}) = \frac{\mu_{\tau,p}}{V_{\tau-1}} - \frac{\sigma_{\tau,p}^2}{2V_{\tau-1}^2}.$$

By substituting (2.4) for the portfolio expected return $\mu_{\tau,p}$ and σ^2 for portfolio variance $\sigma_{\tau,p}^2$, we arrive at

$$g_\infty(\sigma) = \frac{\sigma}{V_{\tau_{-1}}\sqrt{K}}\boldsymbol{\mu}_\tau'\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}) - \frac{\sigma^2}{2V_{\tau-1}^2}.$$

The next step is to maximize the above expression with respect to σ. Because the expression is a negative quadratic function in σ, both the necessary and the sufficient condition for global maximum can be obtained by taking the

Figure 2.2: Optimal portfolio



Source: author's computations.

first derivative and setting it equal to 0:

$$\frac{\partial g_\infty(\sigma)}{\partial \sigma} = \frac{1}{V_{\tau-1}\sqrt{K}} \boldsymbol{\mu}_\tau' \boldsymbol{\Sigma}_\tau^{-1} (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}) - \frac{\sigma}{V_{\tau-1}^2} \stackrel{!}{=} 0.$$

Solving with respect to σ in turn yields:

$$\sigma = \frac{V_{\tau-1}}{\sqrt{K}} \boldsymbol{\mu}_\tau' \boldsymbol{\Sigma}_\tau^{-1} (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}).$$

Finally, we can insert the expression for variance into our expression for weights (2.3) to obtain:

$$\boxed{\boldsymbol{w_\tau} = \frac{V_{\tau-1}}{K} (\boldsymbol{\mu}_\tau' \boldsymbol{\Sigma}_\tau^{-1} (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota})) \boldsymbol{\Sigma}_\tau^{-1} (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}),} \tag{2.5}$$

which concludes the search for (approximately) optimal position sizes. The result is illustrated in Figure 2.2

2.3.5 Specifying the functional form of conditional expected return and variance

Until now, we have been operating with the conditional expected return $\boldsymbol{\mu}_\tau$ and conditional variance-covariance matrix $\boldsymbol{\Sigma}_\tau$ without any further specification. What remains to be done is to use our knowledge of contrarian profitability (buying losers and selling winners generates excess profits) to find a suitable functional form of the conditional expected return and to use well known facts about the behavior volatility to find a functional form of the conditional variance-covariance matrix. After we have specified the functional forms, we will discuss the estimation of the parameters that enter it.

Note that the specification of the conditional expected return captures our edge over the market and consequently is crucial for our success. Without a viable specification, all other work would be done in vain. The only reason why it is buried deep within the chapter rather than having a prominent place is our desire to present topics in a top-down manner.

In searching for the functional forms of the conditional expected return and conditional variance-covariance matrix, we will take advantage of well-known econometric models (or perhaps we can say that our model of the return process is stated in econometric terms). In particular, the conditional expected return will be specified by a simple linear regression model while the conditional variance-covariance matrix will be specified by a multivariate GARCH model.

Modeling the vector of conditional mean returns

Our objective is to find a suitable functional form of the conditional expected return. As we have already revealed, the conditional expected return will be specified by a linear regression model. Our choice of regressors is motivated by Lo & MacKinlay (1990a). Lo & MacKinlay (1990a) proposed a contrarian strategy which dictated position sizes:

$$w_{\tau,i} = \left(\frac{1}{n}\sum_{j=1}^{n} r_{\tau-1,j}\right) - r_{\tau-1,i}.$$

Expressed in words, positions in stock i is determined by the difference between the average return of all stocks and the return of stock i in the previous period. As documented by Khandani & Lo (2007), this simple strategy is able to generate excess risk adjusted profits. If we assume that the profitability was not a fluke, we can conclude that the deviation from average return is able to predict future returns (For detailed treatment, please see A.4).

Having provided the motivation behind the choice of regressors, we can now present the regression equations for individual stocks:

$$\begin{aligned} r_{\tau,1} &= \beta_1 d_{\tau-1,1} + \epsilon_{\tau,1}, \\ r_{\tau,2} &= \beta_2 d_{\tau-1,2} + \epsilon_{\tau,2}, \\ &\vdots \\ r_{\tau,n} &= \beta_n d_{\tau-1,n} + \epsilon_{\tau,n}, \end{aligned}$$

where

$$d_{\tau-1,i} = r_{\tau-1,i} - \frac{1}{n}\sum_{j=1}^{n} r_{\tau-1,j}.$$

Of course, we assume that the errors are uncorrelated with the regressors. Other assumptions about the behavior of the errors will be given in the next subsection.

The corresponding functional form of conditional expected return is then:

$$\begin{aligned} \mu_{\tau,1} &= \beta_1 d_{\tau-1,1}, \\ \mu_{\tau,2} &= \beta_2 d_{\tau-1,2}, \\ &\vdots \\ \mu_{t,n} &= \beta_n d_{\tau-1,n}. \end{aligned} \tag{2.6}$$

We have decided to exclude the intercept from the model. The reason is the following: If the intercept was included, the strategy would have systematically assigned more positive weight to the stocks with a high unconditional expected return. Consequently, the long-term return of those stocks and our strategy would have been correlated. Assuming that stocks with a high unconditional expected return are also stocks with high betas in the sense of Capital Asset Pricing Model, this would defy our goal of separating the long-term return of the market from the long-term return of our strategy.

Modeling the conditional variance - covariance matrix

The next task is to find a functional form of conditional variance-covariance matrix. This corresponds to specifying the process that generates disturbances in the previous model. In formal terms, if we put

$$\boldsymbol{\epsilon}_\tau = \begin{pmatrix} \epsilon_{\tau,1} \\ \epsilon_{\tau,2} \\ \vdots \\ \epsilon_{\tau,n} \end{pmatrix}$$

then we can write,

$$\boldsymbol{\epsilon_\tau} = \boldsymbol{\Sigma}_\tau^{1/2} \boldsymbol{v}_\tau, \tag{2.7}$$

where $v_{t,i}$ is white noise with variance equal to unity and $\boldsymbol{\Sigma}_\tau^{1/2}$ is the lower triangular component of the Cholesky decomposition of $\boldsymbol{\Sigma}_\tau$.

Stock returns are known to display volatility clustering (Cont 2001). Consequently, we will require the specification of the conditional variance-covariance matrix to capture this phenomenon. If we were dealing with a univariate time series, our task would be simple as we would resort to a basic model from GARCH class. Unfortunately there is no straightforward generalization of GARCH into the realm of multivariate time series and hence, further discussion is called for. We will restrict ourselves to models that only contain a single lag as those models are deemed to be sufficient (Engle 2001).

There are three well known problems inherent to general multivariate GARCH models (Silvennoinen & Terasvirta 2008). The first of these is an excessive number of parameters, the majority of which lack an economic interpretation. Excessive number of parameters results in overfitting. The second problem, partially linked to the first one, is that the resulting variance-covariance matrix is not by construction positive definite. Consequently, we need to impose restrictions. Third problem, linked to the first two, is that with a larger number of parameters and restricted parameter space, estimation becomes difficult and computationally demanding. Our aim is to find a model that does not suffer from these three shortcomings.

Our preferred approach (discussed for example in Silvennoinen & Terasvirta 2008) is to decompose the conditional variance-covariance matrix $\mathrm{Var}[\boldsymbol{r}_\tau|\psi_{\tau-1}]$ into a conditional correlation matrix and conditional variances. The two can then be modeled separately. Modeling the correlation matrix and variances separately makes it extremely easy to make the resulting variance-covariance matrix positive definite. Moreover, the parameters in this class of models usually have an intuitive interpretation. Finally, as will be shown later in more detail, we can estimate parameters of the model equation by equation.

We opted to use the most basic model of this class called Constant Conditional Correlation (CCC) (introduced by Bollerslev 1990). As the name implies, the cornerstone of the model is the assumption that the correlation matrix is constant over time. Hence the conditional variance-covariance matrix $\mathrm{Var}[\boldsymbol{r}_\tau|\psi_{\tau-1}]$ can be decomposed as follows:

$$\boldsymbol{\Sigma}_\tau = \boldsymbol{D_\tau P D_\tau},$$

where $\boldsymbol{P}$ is the time-invariant correlation matrix

$$\boldsymbol{P} = \begin{pmatrix} 1 & \rho_{1,2} & \dots & \rho_{1,n} \\ \rho_{2,1} & 1 & \dots & \rho_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \dots & 1 \end{pmatrix}$$

and $\boldsymbol{D_\tau}$ is a diagonal matrix of conditional standard deviations of returns of individual stocks:

$$\boldsymbol{D_\tau} = \begin{pmatrix} \sigma_{\tau,1} & 0 & \dots & 0 \\ 0 & \sigma_{\tau,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\tau,n} \end{pmatrix}.$$

Individual conditional variances are then modeled by a GARCH(1,1) as

$$\begin{aligned} \sigma^2_{\tau,1} &= \alpha_{1,0} + \alpha_{1,1}\epsilon^2_{\tau-1,1} + \gamma_1\sigma^2_{\tau-1,1}, \\ \sigma^2_{\tau,2} &= \alpha_{2,0} + \alpha_{1,2}\epsilon^2_{\tau-1,2} + \gamma_1\sigma^2_{\tau-1,2}, \\ &\vdots \\ \sigma^2_{\tau,n} &= \alpha_{n,0} + \alpha_{1,n}\epsilon^2_{\tau-1,n} + \gamma_1\sigma^2_{\tau-1,n}. \end{aligned}$$

The model has very appealing properties. The number of parameters grows at rate n^2. All parameters have straightforward interpretation and $\boldsymbol{\Sigma}_\tau$ is positive definite by construction. This concludes our search for a model of conditional variance.

Estimation

We have already specified the functional form of the conditional expected return and conditional variance-covariance matrix. What remains to be done is to estimate parameters that enter the specifications. In particular, we need to estimate:

- The matrix of time-invariant correlations $\boldsymbol{P}$
- The regression equations coefficients β_i for all $i \in \{1, ...n\}$
- The GARCH coefficients $\alpha_{i,0}, \alpha_{i,1}, \gamma_i$ for all $i \in \{1, ...n\}$

If the (simple) returns of our stocks were (conditionally) jointly normal, we could efficiently estimate all the parameters by means of maximum likelihood (Bollerslev 1986). Unfortunately, as suggested by Bera & Higgins (1993), the assumption of conditional normality is hard to justify in practice. Moreover, there is no consensus on the actual (conditional) distribution of stock returns and hence any attempt to construct a likelihood function for such a distribution would probably be misguided.

Our strategy will be to assume that the conditional distribution actually is conditionally normal and conduct the estimation under this assumption. As described by Bera & Higgins (1993), under relatively mild additional assumptions, the estimation procedure will be consistent even if the conditional distribution is misspecified (is not normal). (We have decided to omit those assumptions from the text as we believe they do not add any value to the discussion). Such an estimator is called the quasi-maximum likelihood.

If we denote the number of observations by S, $s \in \{1, ..., S\}$ The maximum likelihood function takes the form:

$$L = \prod_{s=1}^{S} (2\pi)^{-n/2} |\boldsymbol{\Sigma}_s|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon'}_s \boldsymbol{\Sigma}_s \boldsymbol{\epsilon}_s\right).$$

Because of the recurrent nature of the maximum likelihood function, numeric methods are needed in order to obtain the parameter estimates. Given the large number of parameters and the iterative nature of numeric methods, estimation becomes computationally demanding. This is inconvenient, because we plan to devote the next chapter to Monte Carlo experiments. Consequently, we are forced to tweak the estimation method. In particular, we will exploit the fact that CCC specification allow us - at the cost of a small efficiency loss -

to estimate the GARCH and regression parameters equation by equation, and estimate the matrix of time-invariant correlations by its sample equivalent.

Starting with the estimation of GARCH and regression parameters, the likelihood for stock i takes the form:

$$L = \prod_{s=1}^{S} (2\pi\sigma_{s,i}^2)^{-1/2} \exp\left(-\frac{\epsilon_{s,i}^2}{2\sigma_{s,i}^2}\right).$$

The estimation of the time-invariant correlation matrix is straightforward. The formula for sample correlation coefficient between stocks i and j is given by

$$\rho_{i,j} = \frac{\frac{1}{S-1}\sum_{s=1}^{S}(r_{s,i} - \bar{r}_{s,i})(r_{s,j} - \bar{r}_j)}{\sqrt{\frac{1}{S-1}\sum_{s=1}^{S}(r_{s,j} - \bar{r}_j)^2}\sqrt{\frac{1}{S-1}\sum_{s=1}^{S}(r_{s,i} - \bar{r}_i)^2}},$$

where

$$\bar{r}_i = \frac{1}{S}\sum_{s=1}^{S} r_{s,i}.$$

As was already mentioned, solving the maximum likelihood requires usage of numeric methods. Without further inquiring into the topic, we have opted to use the Quasi-Newton method as it is the default option in the software of our choice (MATLAB).

Chapter 3

Deficiencies of our solution and their partial remedy

In the previous section we laid out a trading strategy that aims to maximize growth. The optimality (in the sense of expected growth maximizing property) of the proposed strategy relied on assumptions about the properties of the return process, the accuracy of our approximations and prior knowledge of the parameters of our econometric model. Because none of the above is likely to be met in practice, the strategy will suffer from deficiencies.

Those deficiencies form the content of this chapter. We begin by providing a general discussion of the deficiencies and their interaction. We then perform Monte Carlo simulations on artificial time series in order to find a modification of our strategy that might perform better. The reader can consider this chapter to be a descent from the realm of slick theory to the realm of crude reality.

3.1 Deficiencies

In this section we will provide a general discussion of the deficiencies. We will start by explaining why our forecast of conditional expected return and variance are likely to be incorrect. Next we will discuss how the inaccuracy in the forecasts interacts with mean-variance optimization and how it influences our portfolio selection. Finally, we will discuss the effect of approximations used in Subsection 2.3.2.

3.1.1 Incorrect forecasts

First, the forecast conditional expected return and variance-covariance matrix might be incorrect, because the proposed functional form of the conditional moments is (very) likely to be misspecified (the true functional form of the conditional moments is likely to differ from the one we used). (Of course, all probabilistic models are to some extent misspecified. From the theoretical standpoint, the issue of misspecification can by bypassed by working in the linear projection framework (Hansen 2000). We have decided not to do so in order to keep the discussion simple.)

Second, even if we are somehow able to guess the correct functional form of conditional moments, we still need to estimate the parameters that enter it. Because the parameter estimates will be invariably laden with estimation errors, the accuracy of our forecasts will be compromised. In reality, we can expect to experience both problem - we will be using a misspecified model whose (projection) coefficients will be laden with estimation errors.

Finally, the return process might not be stationary, that is both the functional form of the conditional moments and the parameters might change over time. In the realm of financial markets, the non-stationarity might be attributed to institutional changes of the marketplace (regulations), technological changes and adaptive behavior of market participants (Lo 2005). Again, it is probable that we will experience non-stationarity along with the two initial problems. In the subsequent text, we will resort to the rolling window estimation. At this point we reveal that we do so to deal with the non-stationarity of the return process.

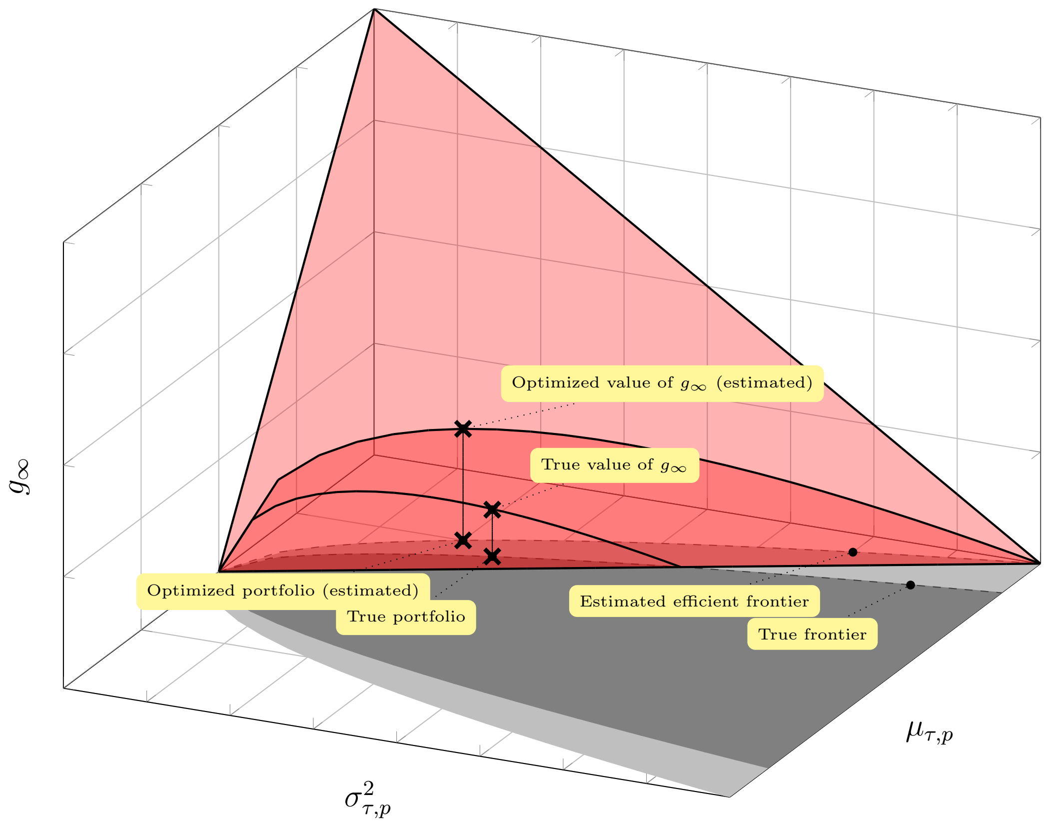
3.1.2 Mean-variance optimization bias

We have just shown that the forecast conditional moments are likely to be incorrect; i.e., for some assets, the forecast conditional expected return (or variance) becomes larger than the true (unobservable) conditional expected return while for others it becomes smaller. As was noted by Michaud (1989), the forecast errors interact with the mean-variance optimization. In particular, mean-variance optimization cannot distinguish between a situation where high forecast expected return reflects the true expected return and situations where a high forecast expected return is due to error. (The same holds for forecast variances.) Consequently, the mean-variance optimization routine tends to favor assets with an upward error in forecast expected return and a downward error in forecast variance. This feature of mean-variance optimization is known as error maximization property. An obvious consequence of error maximization is that estimated efficient frontier will be biased upwards in the sense that that the portfolios on the estimated efficient frontier will have overestimated expected return and underestimated variance (Broadie 1993; Ceria & Stubbs 2006).

Error maximization adversely affects our investment decision both at the level of the relative representation of the assets in the portfolio and at the level of the size of total (absolute) investment. The effect on relative representation is apparent from the previous paragraph and does not warrant further discussion. The effect on total (absolute) value of the portfolio is a bit trickier and will be discussed in more detail. Recall the size of total investment was chosen to maximize the expression g_∞ of the form $(k_\tau \sigma_{\tau,p})/V_{\tau-1} - \sigma^2_{\tau,p}/2V^2_{\tau-1}$, where k_τ represented the **estimated** trade-off between the conditional expected portfolio return and the conditional portfolio variance of the portfolio on the efficient frontier. As indicated in the previous paragraph, the estimated value of k_τ is destined to suffer from upward bias. Turning to the routine presented in Subsection 2.3.4, overestimated k_τ in turn implies that optimized value of σ will be too large, which in turn implies that the suggested position sizes will be too large in absolute value as well (overbetting).

The situation is depicted in Figure 3.1. The curve *Estimated efficient frontier* represents the set of portfolios obtained by using the forecast conditional moments and optimized position sizes (position sizes obtained by plugging forecast conditional moments into the mean-variance optimization routine). I.e., estimated efficient frontier represents the set of portfolio we think we obtain by mean variance optimization. *True frontier* (unobservable) is obtained by using optimized position sizes (position sizes obtained by plugging forecast conditional moments into the mean-variance optimization routine) and true conditional moment. I.e., the true frontier is the set of portfolios we really obtain. Similarly, *Optimized portfolio (estimated)* represents the portfolio we think we obtain by using the routine described in Chapter 2, while *True portfolio* represents the portfolio we really obtain. The explanation of *Optimized value of g_∞ (estimated)* and *True value of g_∞* would go along the same lines.

Figure 3.1: Effect of error maximization



Source: author's computations.

3.1.3 Effect of the approximations

In certain situations, the approximation by the first two moments presented in Chapter 2 might perform poorly. In particular, the approximation might fails when the return process displays high kurtosis. A detailed explanation follows. High kurtosis implies a high (higher than under normal distribution) probability of large gains and loses. Because the original objective function 2.2 treats losses and gains asymmetrically (is much more concerned with losses), it would dictate a reduction of position sizes. When the approximation is used, the high kurtosis is ignored and the reduction does not materialize. Hence, as in the case of error maximization, we end up overbetting. For a more elaborate treatment, please refer to Hakansson (1971).

A second case where our approximation might fail is when the conditional moments of the portfolio return become too large relative to our capital. To see why, remember that at one point, we have used $\sum_{m=1}^{M} \frac{y_m}{V_{\tau-1}} \approx \prod_{m=1}^{M}(1 + \frac{y_m}{V_{\tau-1}})$, which corresponded to neglecting higher order terms of the product (second expression). When the first two moments become too large, the higher order terms of the product stop being negligible and the approximation breaks down. At this point, the direction of the resulting bias in the position sizes cannot be determined.

3.2 Partial remedy

In the previous section we conducted a qualitative analysis of the possible deficiencies in our strategy with a focus on incorrect forecast of moments and error maximization property of mean-variance optimization. We concluded that the original strategy will be overbetting in most situations. The purpose of this section is to find a modification of the original strategy that counteracts the overbetting problem.

The modification will be found by means of a Monte Carlo simulation. To provide a brief overview, our plan is to create several modifications of the original strategy, evaluate their performance on an artificial time series generated in accord with our original model. The procedure will be repeated for several combinations of parameter values of the original model (scenarios). The concept of modification will be introduced first, next we will motivate our selection of scenarios and describe the simulation procedure. Finally, simulation results accompanied by a short commentary will be presented.

3.2.1 Modifications

Our original strategy dictates that in period τ position sizes are determined by:

$$\boldsymbol{w_\tau} = \frac{V_{\tau-1}}{K}(\boldsymbol{\mu}_\tau' \boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}))\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}). \tag{3.1}$$

By a modification, we will understand a strategy that dictates position sizes:

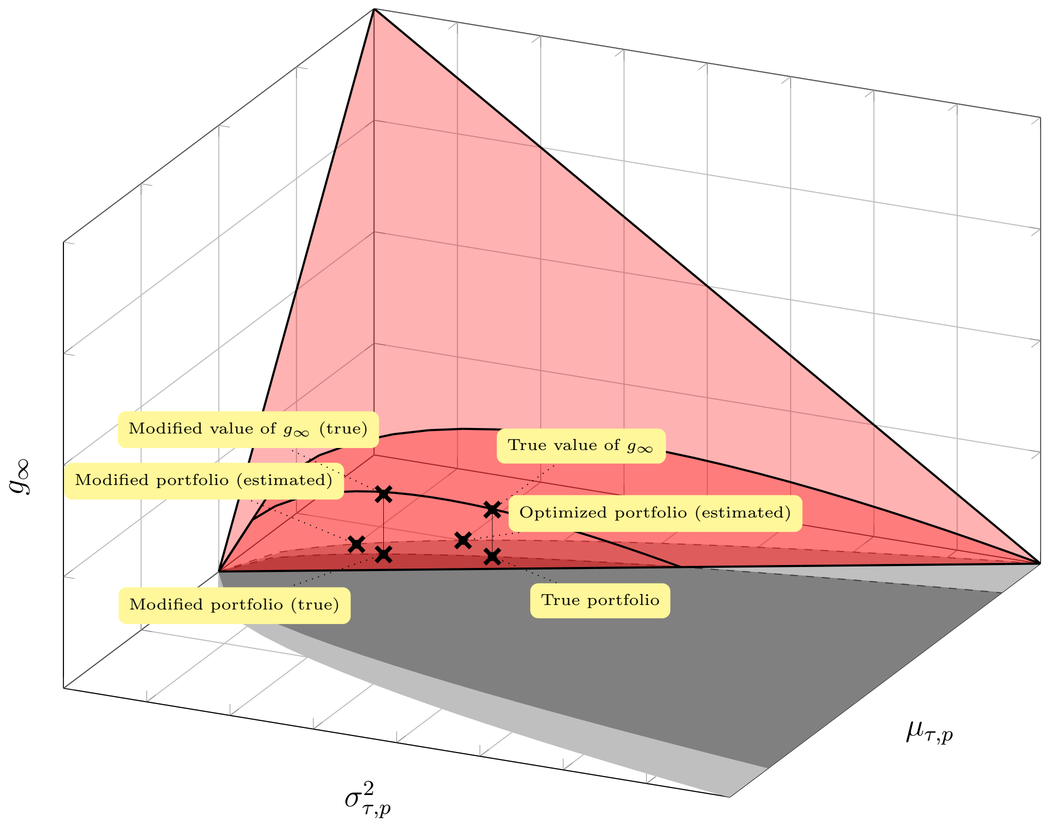
$$\boldsymbol{w_\tau^{(c)}} = c \cdot \frac{V_{\tau-1}}{K}(\boldsymbol{\mu}_\tau' \boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}))\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}), \tag{3.2}$$

where the c is a scalar constant that is time invariant. In the upcoming simulations, we will consider the following values of c

$$c \in \{0.05, 0.100.15, ..., 1.9, 1.95, 2\}.$$

This gives 40 strategies in total (the original strategy and 39 modifications).

Figure 3.2: Possible benefit of modifications



Source: author's computations.

Our motivation behind introducing this sort of modification is following: In the previous section, we have conjectured that the strategy will be overbetting in most scenarios. If we are right, the modifications with $c < 1$ are likely to outperform the original strategy. (The wealth would grow faster if we bet some constant fraction of what is implied by the original strategy.)

Figure 3.2 depicts how our modification can partially counter the overbetting problem stemming from the error maximization. *Optimized portfolio (estimated)*, *True portfolio* and *True value of* g_∞ have the same interpretation as in Figure 3.1. *Modified portfolio (estimated)* and *Modified portfolio (true)* were obtained from *Optimized portfolio (estimated)* and *True portfolio* by multiplying position sizes by some $c < 1$. *Modified value of* g_∞ *(true)* is the value of function g_∞ that corresponds to *Modified portfolio (true)*. Note that *Modified value of* g_∞ *(true)* is larger than the original *True value of* g_∞.

3.2.2 Scenarios

As we have explained, our goal is to find a modification that would counteract the overbetting problem. The optimal modification will of course depend on the degree to which the original strategy is overbetting, which is in turn likely to depend on parameters of the asset return model (the number of assets n, constant correlation matrix $\boldsymbol{P}$ and $\alpha_{i,0}$, $\alpha_{i,1}$, β_i and γ_i for each of the assets $1, ..., n$). Hence, we are forced to conduct simulations for different combinations of parameter values. This subsection explains our selection of combinations of parameter values. To avoid verbosity, we will call a selection of parameter value a *scenario*.

Naturally, we would like to cover as wide range of scenarios as possible. Unfortunately, the simulations are demanding in terms of computational power, which severely limits our options. To cope with this problem, we decided to only consider the case of three assets ($n = 3$) and fix all parameters aside from parameters that enter the specification of conditional mean (β_1, ..., β_n). The second step is motivated by the findings of Chopra & Ziemba (1993), which imply that the variation of parameters that enter the specification of conditional variances and covariances will have a much smaller effect on the degree to which the original strategy is overbetting compared with the variation of parameters that enter the specification of the conditional mean.

The following considerations were taken into account when determining the values of fixed parameters entering the specification of the conditional variance-covariance matrix:

- Estimates of GARCH parameter estimates are very similar across exchanges and time frames (Akgiray 1989; Lamoreux & Lastrapes 1990; Franses & VanDijk 1996). We will set:

$$\forall i \in \{1, ..., n\} : \alpha_{i,1} = 0.2, \gamma_i = 0.7$$

- Our choice of correlations will reflect the average daily correlation presented in Joshua M. Pollet (2010). Formally, we will set the time-invariant correlation matrix to:

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{pmatrix}$$

- For the unconditional component of variance, we will put

$$\forall i \in \{1, ..., n\} : \alpha_{i,0} = 0.0001$$

The choice of unconditional component of variance does not matter. To see why, realize that if we multiply $\alpha_{i,0}$ by some constant k for all i, the set of available portfolios will not change.

This concludes our discussion of fixed parameters.

As we explained earlier, the only parameters which will vary between scenarios are the parameters entering the conditional expected return specification. We will now comment on our choices of parameters β_i. Our first step is to fix the β_i between all stocks. That is, we fix:

$$\beta_1 = \beta_2 = ... = \beta_n = \beta.$$

Again, this step is motivated by a lack of computational power and a lack of guidelines according to which to choose relative values of $\beta_1, \beta_2, ..., \beta_n$.

Next, realize that there is a relationship between values of parameter β and a fraction of the explained variance (FEV) of return of any individual stock (the fraction of the explained variance is the population equivalent of the R^2 statistic). A detailed derivation of the relationship is provided in Section A.5. Because considering various values of FEV is much more illustrative, we will frame our choice of β_1, ..., β_n in terms choosing values of FEV.

The values of FEV were determined as follows. Initially, we considered the interval [0.001, 0.1] and its partition given by points 0.001, 0.005, 0.01, 0.05 and 0.1. We conducted the simulation for the inner points and found out that the constants corresponding to the best modifications lay far apart. We solved the situation by refining the partition until the absolute difference between the constants was no larger than 0.1. The process will be clear once we present the table with the result.

Under our selection of parameters, the return process of all the assets is weak-stationary (Detailed proof is provided in Section A.6). Stationarity implies that both the unconditional expected return and unconditional variance remain stable, which corresponds to the behavior of the return processes observed in practice (We have deliberately avoided stating that real return processes are stationary).

Finally, we need to specify the distribution of the white noise in Equation 2.7. We opt for Gaussian white noise. While this choice is arbitrary, so would be any other. The only advantage of this choice is that it makes sure that the conditions required for consistent estimation mentioned in the Section 2.3.5 are met.

3.2.3 Simulation Procedure

We will now describe the simulation procedure in detail. Because of the hierarchical structure of the simulation, we believe that the simulation is best described in the form of a nested list. We will perform 100 simulations for each scenario under consideration. A single simulation is performed as follows:

1. Generate a sample of length $1000 + 250$ from the return process described by our initial model.
2. Fix the starting wealth. (For example by setting it equal to 1)
3. Evaluate the performance of the original strategy. To do so, perform the following steps starting at $t = 251$
   (a) Forecast the conditional expected return and variance-covariance. The estimation of parameters is performed on a rolling window of past 250 observations.
   (b) Determine the position sizes according to the rule derived in the second chapter.
   (c) Evaluate the portfolio return and update our wealth.
   (d) Increase t by one and repeat the last three steps until $t = 1250$ is reached.
4. Evaluate the performance of all modified strategies. This is done in the same manner as evaluating the performance of the original strategy with the obvious difference in the step where position sizes are set.
5. Compare the level of end wealth of all 40 strategies (the original strategy and 39 modifications). Record the strategy that yields the highest wealth in the terminal period.

The number of periods (1000) and the number of simulations (100) reflect our limited computational power. The length of the estimation window reflects the number of trading days in a year (if one period corresponds to a trading day) or the approximate length of a business cycle (if one period corresponds to a week).

3.2.4 Identifying the best modification

This subsection describes how we identify (more precisely estimate) the best modification (in terms of growth) corresponding to any particular scenario. Because we fear that treating the matter in a formal (probabilistic) way would cause confusion, we have decided to keep the discussion on a predominantly intuitive level.

We have already explained that the output of a single simulation is a modification (represented by a number from $\{0.05, 0.10\ 0.15, ..., 1.9, 1.95, 2\}$) that yielded the highest terminal wealth (after 1000 periods). If we assume that 1000 periods are enough for limiting properties to kick in, we can conclude that the modification that yielded the highest terminal wealth is also the one that asymptotically maximizes growth in the scenario under consideration.

For every scenario, we are, however, not conducting a single simulation but rather a series of 100 simulations. Because the 100 simulations are based on time series with a random element, the modifications yielding the highest terminal wealth might (and will) differ between the 100 simulations. Hence, our task is to aggregate the modifications yielding the highest terminal wealth corresponding to those 100 simulations in a meaningful way. We have decided to do that by picking the modification that yielded the highest terminal weight the most often. (In probabilistic terms this would correspond to estimating the best modification by the modus of our sample).

The reader might ask why we conducted 100 simulations (each of which contained 1000 periods) instead of conducting one long simulation that was 100 000 periods long. We opted for 100 individual simulations, because we wanted to keep track of the dispersion of individual results.

3.2.5 Results

Table 3.1 and Figure 3.3 summarize the estimates of the best modification of the scenarios corresponding to various values FEV. The values of β matching with the values of FEV are also presented.

Before we proceed, let us shortly comment on the results. First of all, in all the scenarios considered, the coefficient c corresponding to the optimal modification is lower than one. This corresponds to our earlier conjecture that

Table 3.1: Best modification for given fraction of explained variance

FEV (R^2 equivalent)	Value of β	Best modification (c)
0.1	0.41	0.45
0.05	0.3	0.55
0.03	0.23	0.6
0.01	0.13	0.7
0.007	0.11	0.6
0.005	0.096	0.5
0.003	0.074	0.4
0.002	0.061	0.35
0.00175	0.057	0.25
0.0015	0.053	0.15
0.001	0.043	0.05

Source: author's computations.

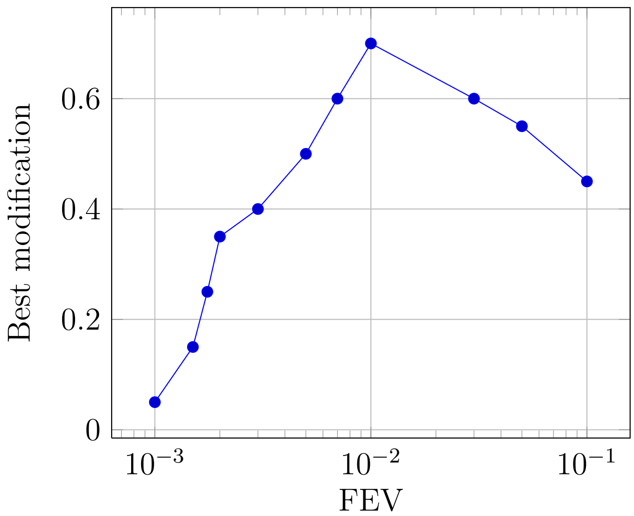
the original strategy will in general be overbetting. Moreover, the overbetting problem seems rather severe since in most scenarios the original strategy would be more than twice the amount suggested by the modification identified on the basis of a simulation.

Secondly, for small values of FEV, the coefficient corresponding to the optimal modification is extremely small. That implies that the original strategy breaks down. We believe the phenomenon can be attributed to error maximization. In particular, we believe that the strategy breaks down because the errors in the forecasts of the conditional expected returns become too large relative to the true values of the conditional expected returns.

Thirdly, note that as the FEV starts to increase, the coefficient corresponding to the optimal modification starts getting closer to one, which implies that the best modification starts to converges towards the original strategy. Again, we believe this can be attributed to error maximization. In this case, the errors in the forecasts of the conditional expected returns starts getting smaller relative to the true values of the conditional expected returns as the FEV begins to increase.

Finally, at some point ($FEV = 0.01$) the coefficient corresponding to the best modification starts to decline again. We believe that the decline is linked to the poor functioning of the second approximations discussed in the Subsection 3.2.1.

Figure 3.3: Simulation results



Source: author's computations.

Chapter 4

Practical demonstration

In the previous two chapters we have derived and tweaked a growth maximizing contrarian trading strategy. The purpose of this chapter is to demonstrate its viability by applying it to the historical returns of three selected stocks. We will begin by explaining how the tweaked strategy will be applied. Next, we will comment on our stock selection. Finally, we will present and discuss the results.

4.1 Application

In Chapter 3 we determined modification of the original strategy suitable for some idealized scenarios. The purpose of this subsection is to explain how we use the modifications to determine the position sizes in practice. We believe that the procedure is best presented as a list. At every decision point, position sizes will be determined as follows:

1. Use the rolling window of the past 250 observations to estimate the model parameters by the method described in Chapter 2.

2. Forecast the conditional expected return and conditional variance-covariance matrix.

3. Compute the estimates of FEV (R^2) for each stock. Take the median value. Determine the best modification using the median value and spline of the results obtained in Chapter 3.

4. Use the previously determined modification and the estimates of the conditional expected return and variance-covariance to determine the position sizes.

4.2 Stocks and period length selection

Let us now comment on the stock and period length selection. We have selected weekly returns of the stocks of the following three companies from the oil and gas mining equipment and affiliated services sector traded on the US stock exchanges: Schlumberger N.V, Baker Hughes Incorporated, Halliburton Company. The stock were selected on the basis of economic reasoning rather than data mining. Most importantly, the choice reflects our desire to avoid spurious results due to non-synchronous trading. Detailed reasoning can be found in Appendix B. Returns were computed using adjusted (for splits and dividends) closing prices from the `finance.yahoo.com` database. Our data sample ranges from April 1987 to April 2010. We have set the last data point so that it corresponds to the date when first research regarding the thesis was done, as it seemed to be the least arbitrary choice.

4.3 Results

Finally, we will present the results. As promised, we will partially drop the assumptions about trading costs, market impact and about the availability of proceeds shorts sales are available. That is, we will present both results obtained under those assumptions and the results obtained after discarding them. Drawing from Avellaneda & Lee (2008), we will accommodate the trading costs and market impact by introducing a fee of 0.1% of the traded volume. To accommodate regulations regarding proceeds from short sales, we will cap maximum leverage by two.

Table 4.1 contains annualized geometric mean returns when full proceeds from short sales are assumed. Aside from the return achieved by applying the strategy in the manner described in Section 4.1 (dynamic), we will also list the returns of the original unmodified strategy and the returns achieved by betting a constant fractions of what is suggested by the original strategy. The symbol ↯ implies that the strategy has lost all available capital.

The most striking observation is that if we applied the strategy in the manner described in Section 4.1 and if we were able to bypass regulations regarding proceed from short sales, we would achieve an outstanding annual geometric mean return of 24% even after accounting for trading costs. To put the number in a different perspective, if we started with an initial capital of \$20,000, our wealth would have exceeded \$1,000,000 by the end of April 2010.

Table 4.1: Annualized geometric mean return in %, unconstrained leverage

Strategy Application	No trading costs	Trading costs 0.1
Dynamic	34.6	24
Original	$\lightning$	$\lightning$
Constant fraction (0.25)	23.4	20
Constant fraction (0.5)	29.4	23
Constant fraction (0.75)	9.8	$\lightning$

Source: author's computations.

On the other hand, is we applied the strategy in the original unmodified form, we would have ended up bankrupt even if we had faced no transaction costs. This result illustrates the deficiencies discussed in Chapter 3 and highlights the usefulness of our modifications.

Table 4.2 contains annualized geometric mean returns when leverage is capped by two. Again, we will also list the return achieved by the strategy applied in the original form as well as returns the obtained by betting constant fractions.

Table 4.2: Annualized geometric mean return in %, constrained leverage

Strategy	No trading costs	Trading costs 0.1%
Dynamic	10.4	6.6
Original	5.2	-0.9
Constant fraction (0.75)	6.1	0
Constant fraction (0.5)	6.7	1.4
Constant fraction (0.25)	7.3	2.5

Source: author's computations.

Under the cap on leverage, the application of the strategy in the manner described in Section 4.1 still beats any other manners of application, but the total return is much lower (a meager 6.6%). Unlike when leverage was unconstrained, we can now make a meaningful comparison with the market portfolio. If, at the starting date, we leveraged our initial capital to the maximum extent (2x leverage) to purchase a market portfolio (as represented by the S&P 500 index), we would have achieved an annualized geometric mean return of about 9%. Hence, our strategy underperformed the market on average by about 2.4%. At this point we must emphasize that our strategy was not optimized for trading costs and constrained leverage. Hence, better results can be achieved by doing so. Moreover, the strategy was designed as to be uncorrelated with the market and therefore offers some diversification potential. For these two reasons, we consider the results to be a success rather than a disappointment.

Chapter 5

Conclusion

Our goal was to develop a contrarian speculative trading strategy. We have chosen growth maximization as opposed to maximization of a subjective utility function as our guiding principle. Furthermore, we have required our strategy to be market neutral (dollar neutral was used as a proxy).

We have tackled the problem as follows. First, we have exploited the fact that growth maximization is equivalent to sequential maximization of the end of period conditional expected logarithm of wealth. Next, we approximated the log optimal portfolio by a mean-variance efficient portfolio. The specification of the conditional expected return reflected our beliefs about the profitability of contrarian strategies. In particular, we have specified the conditional expected return as a linear function of derivation from the average return in the previous period. The conditional variance-covariance matrix was specified by the Constant Conditional Correlation model from the Multivariate GARCH class.

The approach just described poses several shortcomings. In particular, the use of approximation along with so called error maximization property of the mean-variance optimization leads to overbetting in the sense of opening too large position in absolute terms (too much leverage). Realizing that the problem could be partially countered by investing some fixed fraction of the amount suggested by the original strategy (proportional reduction of all position sizes), we went on to find the optimal size of the fraction in various scenarios by means of Monte Carlo simulations.

Equipped with the simulation results, we evaluated the performance of the strategy on historical data. Based on economic criteria, we chose weekly returns of three firms active in oil mining equipment and affiliated services industry. Over the period from 1987 to 2010, we were able to generate an annual geometric mean return of about 24% when no cap on leverage was assumed even after accounting for mild transaction costs. When we capped the leverage by two, the return shrunk to 6.6%.

While the results are satisfactory, we believe that the room for improvement is still vast. Out of the long list of possible tweaks, we will mention only those we consider the most promising. First, we believe that exploitation of economic reasons behind contrarian profitability would allow us to create a better specification of the conditional expected return. As for the specification of the conditional variance-covariance matrix, Pollet & Wilson (2010) provides evidence that correlations varies over time. Hence we should consider a model that allows time variation in the conditional correlation matrix - perhaps the Variable Correlations GARCH model introduced by Tse & Tsui (1999). When it comes to the error maximization property and the resulting overbetting, the ill effects could be partially neutralized by incorporating estimation errors into the variance-covariance matrix or directly into the mean-variance optimization procedure (Ceria & Stubbs 2006). Finally, better results could be obtained by also incorporating the trading costs into the mean-variance optimization routine as well.

Bibliography

Akgiray, V. (1989): "Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts." *Vedat Akgiray* **62**: pp. 55–80.

Algoet, P. H. & T. M. Cover (1988): "Asymptotic optimality and asymptotic equipartition properties of log-optimum investment." *The Annals of Probability* **16**: pp. 876–898.

Avellaneda, M. & J.-H. Lee (2008): "Statistical arbitrage in the u.s. equities market."

Bera, A. K. & M. L. Higgins (1993): "Arch models: Properties, estimation and testing." *Journal of Economic Surveys* **7**: pp. 307–366.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* **31**: pp. 307–327.

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model." *Review of Economics and Statistics* **72**: pp. 498–505.

Boudoukh, J., M. P. Richardson, & R. F. Whitelaw (1994): "A tale of three schools: Insights on autocorrelations of short-horizon stock returns." *Review of Financial Studies* **7**: pp. 539–573.

Brennan, M., N. Jagadeesh, & B. Swiminathan (1993): "Investment analysis and the adjustment of stock prices to common information." *Review of Financial Studies* **6**: pp. 799–824.

Broadie, M. (1993): "Computing efficient frontiers using estimated parameters." *Annals of Operations Research: Special Issue on Financial Engineering* **45**: pp. 21–58.

Ceria, S. & R. A. Stubbs (2006): "Incorporating estimation errors into portfolio selection: Robust portfolio construction." *Axiom Research Papers* **3**.

Chopra, V. & W. Ziemba (1993): "The effects of errors in means, variances, and covariances on optimal portfolio chioce." *Journal of Portfolio Management* pp. 6–11.

Cont, R. (2001): "Empirical properties of asset returns: stylized facts and statistical issues." *Quantitative Finance* **1**: pp. 223 – 236.

Engle, R. (2001): "Garch 101: The use of arch/garch models in applied econometrics." *Journal of Economic Perspectives* **15**: pp. 157–168.

Franses, P. H. & D. VanDijk (1996): "Forecasting stock market volatility using (non-linear) garch models." *Journal of Forecasting* **15**: pp. 229–235.

Hakansson, N. (1971): "Capital growth and the mean variance approach to portfolio selection." *The Journal of Financial and Quantitative Analysis* **6**: pp. 517–557.

Hakansson, N. & W. Ziemba (1995): "Capital growth theory." In "Handbooks in Operations Research and Management Science," Elsevier Science.

Hansen, B. (2000): "Online econometrics textbook."

Joshua M. Pollet, M. W. (2010): "Average correlation and stock market returns." *Journal of Financial Economics* **96**: pp. 364–380.

Kelly, J. L. (1956): "A new interpretation of information rate." *Bell System Tech. Journal* **35**: pp. 917–926.

Khandani, A. & A. Lo (2007): "What happened to the quants in august 2007." *Journal of Investment Management* **5**: p. 5–54.

Lachout, P. & Z. Prášková (2005): *Základy náhodný procesů*. Karolinum.

Lamoreux, C. G. & W. D. Lastrapes (1990): "Heteroscedasticity in stock returns data: Volume versus garch effects." *Journal of Finance* **45**: pp. 221–229.

Latane, H. A. (1959): "Criteria for choice among risky ventures." *Journal of Political Economy* **67**: pp. 144–155.

LEHMANN, B. N. (1990): "Fads, martingales, and market efficiency." *Quarterly Journal of Economics* **105**: pp. 1–28.

LO, A. & C. MACKINLAY (1990a): "When are contrarian profits due to stock market overreaction?" *Review of Financial Studies* **3**: pp. 175–205.

LO, A. & C. MACKINLAY (1990b): "An econometric analysis of nonsynchronous trading." *Journal of Econometrics* **45**: pp. 181–211.

LO, A. W. (2005): "The adaptive markets hypothesis: Market effciency from an evolutionary perspective."

MARKOWITZ, H. (1952): "Portfolio selection." *The Journal of Finance* **7**: pp. 77–91.

MICHAUD, R. O. (1989): "The markowitz optimization enigma: "is 'optimized' optimal?"." *Financial Analysts Journal* **31**.

POLE, A. (2007): *Statistical Arbitrage: Algorithmic Trading Insights and Techniques.* John Wiley and Sons.

POLLET, J. M. & M. WILSON (2010): "Average correlation and stock market returns." *Journal of Financial Economics* **96**: p. 364–380.

PRÁŠKOVÁ, Z. (2007): *Základy náhodných procesů II.* Karolinum.

PULLEY, L. B. (1983): "Mean-variance approximations to expected logarithmic utility." *Operations Research* **31**: pp. 685–696.

ROY, A. (1952): "Safety first and the holding of assets." *Econometrica* **20**: pp. 431–449.

RUDIN, W. (1976): *Principles of Mathematical Analysis.* McGraw-Hill.

SILVENNOINEN, A. & T. TERASVIRTA (2008): "Multivariate garch models." SSE/EFI Working Paper Series in Economics and Finance No. 669.

THORP, E. O. (1997): "The kelly criterion in blackjack, sports betting, and the stock market." *The 10th International Conference on Gambling and Risk Taking* .

THORP, E. O. (2003): "A perspective on quantitative finance: Models for beating the market."

TSE, Y. K. & K. C. TSUI (1999): "A note on diagnosing multivariate conditional heteroscedasticity models." *Journal of Time Series Analysis* **20**: pp. 679–691.

Appendix A

Mathematical background

A.1 Taylor series

We are interested in a Taylor expansion of:

$$g_M(\mu, \sigma) = \sum_{m=1}^{M} \left(\frac{1}{2} \log(1 + \frac{\mu}{V_{\tau-1}M} + \frac{\sigma}{V_{\tau-1}\sqrt{M}}) + \right.$$
$$\left. + \frac{1}{2} \log(1 + \frac{\mu}{V_{\tau-1}M} - \frac{\sigma}{V_{\tau-1}\sqrt{M}}) \right)$$

around point. The general formula for expansion of bivariate function $f(x, y)$ around (x_0, y_0) is:

$$f(x, y) \stackrel{t}{=} \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{\partial^{r+s} f(x_0, y_0)}{\partial^r .x \partial^s y} \frac{(x - x_0)^r}{r!} \frac{(y - y_0)^s}{s!} \tag{A.1}$$

Let us compute first few derivatives:

$$\frac{\partial g_M(\mu, \sigma)}{\partial \mu} = \frac{1}{2} \sum_{m=1}^{M} \left(\frac{1}{1 + \frac{\mu}{V_{\tau-1}M} + \frac{\sigma}{V_{\tau-1}\sqrt{M}}} + \frac{1}{1 + \frac{\mu}{V_{\tau-1}M} - \frac{\sigma}{V_{\tau-1}\sqrt{M}}} \right) \frac{1}{V_{\tau-1}M},$$
$$\frac{\partial g_M(\mu, \sigma)}{\partial \sigma} = \frac{1}{2} \sum_{m=1}^{M} \left(\frac{1}{1 + \frac{\mu}{V_{\tau-1}M} + \frac{\sigma}{V_{\tau-1}\sqrt{M}}} - \frac{1}{1 + \frac{\mu}{V_{\tau-1}M} - \frac{\sigma}{V_{\tau-1}\sqrt{M}}} \right) \frac{1}{V_{\tau-1}\sqrt{M}},$$
$$\frac{\partial^2 g_M(\mu, \sigma)}{\partial^2 \sigma} = \frac{1}{2} \sum_{m=1}^{M} \left(\frac{1}{\left(1 + \frac{\mu}{V_{\tau-1}M} + \frac{\sigma}{V_{\tau-1}\sqrt{M}}\right)^2} + \frac{1}{\left(1 + \frac{\mu}{V_{\tau-1}M} - \frac{\sigma}{V_{\tau-1}\sqrt{M}}\right)^2} \right) \frac{-1}{V_{\tau-1}^2 M}.$$

Observe that higher order derivatives are $o(\frac{1}{\sqrt{M}})$ as taking additional derivatives spawn higher powers of $\frac{1}{M}$. Substitution into the A.1 then yields the formula:

$$g_M(\mu, \sigma) = \frac{\mu}{V_{\tau-1}} - \frac{\sigma^2}{2V_{\tau-1}^2} + o(\frac{1}{\sqrt{M}}).$$

A.2 Boundedness

We need to prove that the set

$$\{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = \sigma^2, \boldsymbol{w}'\boldsymbol{\iota} = 0\}$$

is bounded. Because

$$\{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = \sigma^2, \boldsymbol{w}'\boldsymbol{\iota} = 0\} \subset \{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = \sigma^2\},$$

it suffices to show that

$$S = \{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = \sigma^2\}$$

is bounded. The proof will be done by contradiction. Let us assume that S is not bounded. The rest of the proof can be done in the following steps:

(1) Because S is not bounded, there is a sequence of $\{\boldsymbol{w}_n\}^\infty$ such that for some i: $w_{i,n} \to \infty$ or $w_{i,n} \to -\infty$.

(2) $\boldsymbol{\Sigma}_\tau$ is variance-covariance matrix and hence is positive semidefinite (see theorem 1.6 in Prášková 2007). The full rank assumption along with positive semidefiniteness guarantees positive definiteness. Hence $\boldsymbol{\Sigma}$ has Cholesky decomposition. We will denote the decomposition by $\boldsymbol{U}$ and $\boldsymbol{U}'$.

(3) Because the matrix $\boldsymbol{U}$ is lower triangular it follows that for some i: $(\boldsymbol{w}_n'\boldsymbol{U})_i \to \infty$ or $(\boldsymbol{w}_n'\boldsymbol{U})_i \to -\infty$.

(4) Finally, (3) and $\boldsymbol{w}_n'\boldsymbol{\Sigma}\boldsymbol{w}_n = \boldsymbol{w}_n'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{w}_n$ gives us $\boldsymbol{w}_n'\boldsymbol{\Sigma}\boldsymbol{w}_n \to \infty$. That cannot be, because $\boldsymbol{w}_n \in S$ and therefore $\boldsymbol{w}_n'\boldsymbol{\Sigma}\boldsymbol{w}_n = \sigma^2 < \infty$. Hence the contradiction.

A.3 Efficient frontier

Our task is to maximize:

$$\boldsymbol{w}_\tau' \boldsymbol{\mu}_\tau.$$

Subject to:

$$\boldsymbol{w}_\tau' \boldsymbol{\Sigma}_\tau \boldsymbol{w}_\tau = \sigma^2,$$
$$\boldsymbol{w}_\tau' \boldsymbol{\iota} = 0.$$

The problem will be solved by means of the Lagrange multipliers method. Lagrangian function will assume the form

$$\mathcal{L} = \boldsymbol{w}_\tau' \boldsymbol{\mu}_\tau - \frac{1}{2}\lambda_1(\boldsymbol{w}_\tau' \boldsymbol{\Sigma}_\tau \boldsymbol{w}_\tau - \sigma^2) - \lambda_2(\boldsymbol{w}_\tau' \boldsymbol{\iota}).$$

Note that we have premultiplied the first multiplier by 1/2 in order to simply the derivation. Such an operation will not change the set of solutions.

Necessary conditions for a local maximum are:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_\tau} = \boldsymbol{\mu}_\tau - \lambda_1 \boldsymbol{\Sigma}_\tau \boldsymbol{w}_\tau - \lambda_2 \boldsymbol{\iota} \overset{!}{=} 0, \tag{A.2}$$
$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \boldsymbol{w}_\tau' \boldsymbol{\Sigma}_\tau \boldsymbol{w}_\tau - \sigma^2 \overset{!}{=} 0, \tag{A.3}$$
$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \boldsymbol{w}_\tau' \boldsymbol{\iota} \overset{!}{=} 0. \tag{A.4}$$

Consider the case $\lambda_1 = 0$. That would imply:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_\tau} = \boldsymbol{\mu}_\tau - \lambda_2 \boldsymbol{\iota} \overset{!}{=} 0.$$

Because all elements of vector $\lambda_2 \iota$ assume the same value, the system would only have a solution if all elements of vector $\boldsymbol{\mu}_\tau$ assume the same value as well. This is not possible as per one of our initial assumption.

Now, consider the case that $\lambda_1 \neq 0$. First, we solve the equation (A.2) for $\boldsymbol{w}_\tau$,

$$\boldsymbol{w}_\tau = (\lambda_1 \boldsymbol{\Sigma}_\tau)^{-1}(\boldsymbol{w}_\tau - \lambda_2 \boldsymbol{\iota}). \tag{A.5}$$

Invertibility is guaranteed by the full rank assumption.

The next step is to substitute (A.5) into (A.4)

$$\boldsymbol{\iota}'(\lambda_1 \boldsymbol{\Sigma}_\tau)^{-1}(\boldsymbol{\mu}_\tau - \lambda_2 \boldsymbol{\iota}) = 0,$$

and solve for λ_2

$$\lambda_2 = (\boldsymbol{\iota}'\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{\iota})^{-1}(\boldsymbol{\iota}'\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{\mu}_\tau). \tag{A.6}$$

For convenience, we denote

$$L = (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota})'\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}).$$

Our next task is to deal with λ_1. We substitute (A.5) into (A.3)

$$\left((\lambda_1\boldsymbol{\Sigma}_\tau)^{-1}(\boldsymbol{\mu}_\tau - \lambda_2\boldsymbol{\iota})\right)'\boldsymbol{\Sigma}_\tau\left((\lambda_1\boldsymbol{\Sigma}_\tau)^{-1}(\boldsymbol{\mu}_\tau - \lambda_2\boldsymbol{\iota})\right) = \sigma^2,$$

and solve for λ_1 to get

$$\lambda_1 = \pm\frac{\sqrt{(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota})'\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota})}}{\sigma}. \tag{A.7}$$

We denote

$$K = (\boldsymbol{\mu}_\tau - L\boldsymbol{\iota})'\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}).$$

Finally, substituting (A.6) and (A.7) into (A.5) yields

$$\boldsymbol{w}_\tau = \pm\frac{\sigma}{\sqrt{K}}\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}),$$

where one solution corresponds to position sizes that minimize the conditional expected return and the other corresponds to position sizes that maximize the conditional expected return.

Apparently, the solution that maximizes expected return is

$$\boldsymbol{w}_\tau = +\frac{\sigma}{\sqrt{K}}\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{\mu}_\tau - L\boldsymbol{\iota}).$$

To see that, consider increasing the conditional expected return of asset i while keeping conditional expected returns of other assets as well as the variance-covariance matrix constant and observe what happens to relative weights.

A.4 Predictive power of deviation from the average return

Our task is to show that deviation from average return in the previous period is a meaningful choice of regressor. The in sample profitability claimed by Khandani & Lo (2007) can be expressed as

$$\sum_{\tau=1}^{T} \boldsymbol{d}'_{\tau-1}\boldsymbol{r}_t > 0.$$

By assuming that that same relationship holds in the population, we obtain:

$$\mathbb{E}[\boldsymbol{d}'_{\tau-1}\boldsymbol{r}_t|\psi_t] > 0. \tag{A.8}$$

By substituting our specification into A.8, we obtain:

$$\mathbb{E}[\sum_{i=1}^{n} d_{t-1,i}(d_{t-1,i}\beta_i + \epsilon_t)|\psi_t] > 0.$$

The above inequality in turn leads to:

$$\sum_{i=1}^{n} \mathbb{E}[d^2_{t-1,i}]\beta_i > 0.$$

In other words, some of the coefficients β_i must be nonzero and $v_{t-1,i}$ has predictive power.

A.5 Fraction of explained variance

We want to derive an algebraic relationship between β as understood in Subsection 3.2.2 and the fraction of the explained variance (FEV) of returns of any individual asset. The derivation will be conducted under the assumption that the process described by the model is stationary. We will firstly present a stream of equalities and then provide comments. Before we begin, we denote:

$$\boldsymbol{N} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix}.$$

Proceeding to the derivation:

$$
\begin{aligned}
FEV_i &= 1 - \frac{\text{var}[\epsilon_{t,i}]}{\text{var}[r_{t,i}]} &\text{(A.9)}\\
&= 1 - \frac{\text{var}[\epsilon_{t,i}]}{\text{var}[\epsilon_{t,i}] + (\sum_{k=1}^{\infty}(\boldsymbol{\beta N})^k \text{var}[\boldsymbol{\epsilon}_t](\boldsymbol{\beta N})^k)_{ii}} &\text{(A.10)}\\
&= 1 - \frac{\text{var}[\epsilon_{t,i}]}{\text{var}[\epsilon_{t,i}] + (\sum_{k=1}^{\infty}\beta^{2k}\boldsymbol{N}\text{var}[\boldsymbol{\epsilon}_t]\boldsymbol{N})_{ii}} &\text{(A.11)}\\
&= 1 - \frac{1}{1 + \sum_{k=1}^{\infty}\beta^{2k}\boldsymbol{(NPN)}_{ii}} &\text{(A.12)}\\
&= 1 - \frac{1}{1 + \frac{\beta^2}{1-\beta^2}\boldsymbol{(NPN)}_{ii}} &\text{(A.13)}\\
&= \frac{\beta^2\boldsymbol{(NPN)}_{ii}}{1 - \beta_2 + \beta^2\boldsymbol{(NPN)}_{ii}}. &\text{(A.14)}
\end{aligned}
$$

Equality A.9 is simply the definition of the fraction of the explained variance.

To obtain the equality A.10, we have rewritten our model in VAR form (details are given in Section A.6). Next, we apply the law of total variance and the stationarity assumption:

$$
\begin{aligned}
\text{Var}[\boldsymbol{r}_t] &= \mathbb{E}[\text{var}[\boldsymbol{r}_t|\psi_{t-1}]] + \text{var}[\mathbb{E}[\boldsymbol{r}_t|\psi_{t-1}]]\\
&= \text{var}[\boldsymbol{\epsilon}_t] + \boldsymbol{\beta N}\text{var}[\boldsymbol{r}_t]\boldsymbol{\beta N}\\
&= \text{var}[\boldsymbol{\epsilon}_t] + \sum_{k=1}^{\infty}(\boldsymbol{\beta N})^k\text{var}[\boldsymbol{\epsilon}_t](\boldsymbol{\beta N})^k.
\end{aligned}
$$

To obtain equality A.11, we need to realize that $\boldsymbol{\beta}$ is a diagonal matrix with all the elements on the diagonal attaining identical value and hence can be replaced by a scalar of the same value. The equality follows once we rearrange the order of multiplication and realize that $\boldsymbol{N}^k = \boldsymbol{N}$.

Equality A.12 is obtained by dividing both the numerator and the denominator by $\text{var}[\epsilon_{t,i}]$. This step is possible because of the CC specification and because in this particular case $\text{var}[\epsilon_{t,i}]$ is constant across all i.

Equality A.13 is obtained by summing the geometric series $\sum_{k=1}^{\infty}(\beta^2)^k$. Equality A.14 is obtained by algebraic manipulation.

A.6 Stationarity

We want to show that the return process specified in Chapter 2 is weak-stationary under combinations of parameter values given in Subsection 3.2.2.

This can be proved along the following lines:

(1) We show that the error process $\boldsymbol{\epsilon}_\tau$ is weak-stationary.

(2) We show that $\boldsymbol{r}_\tau$ is a filtration of $\boldsymbol{\epsilon}_\tau$ with absolutely convergent coefficients.

(3) With (1) and (2) in place, the stationarity follows from theorem 5.4 in Prášková (2007).

Let us prove (1) and (2). To prove (1), we need to show that:

(a) First moments, and second moments and cross moments exist and are finite.

(b) For all $i, j \in 1, 2, 3$ and for all $t, s \in \mathbb{Z}$ it holds that $\text{cov}[\epsilon_{i,t}, \epsilon_{j,s}]$ depends only on $t - s$.

(c) For all $i \in 1, 2, 3$ and it holds that $\mathbb{E}[\epsilon_{i,t}]$ is constant for all $t \in \mathbb{Z}$.

To show (a), it is sufficient to show that $\text{var}[\epsilon_{i,t}]$ exists and is finite. If we show the existence and finiteness of $\text{var}[\epsilon_{i,t}]$, the existence and finiteness of second cross moments and first moments will follow from Holder inequality. The existence and finiteness of $\text{var}[\epsilon_{i,t}]$ in turn follow from the observation that for all $i \in \{1, 2, 3\}$, $\gamma_i + \alpha_{1,i} = 0.9$ and theorem A.1 in Bollerslev (1986). With (a) in place, (b) and (c) follow from the model construction.

To prove (2), we will:

(a) Rewrite the original model as a restricted VAR using lag polynomial.

(b) Show that $(\boldsymbol{I} - \beta \boldsymbol{N})^{-1} = \sum_{n=0}^{\infty} \beta^n \boldsymbol{N}^n$.

(c) Define operator $\sum_{n=0}^{\infty} L^n \beta^n \boldsymbol{N}^n$ and apply it to both sides

Starting with (a), the original model of the form

$$
\begin{aligned}
r_{t,1} &= \beta_1 d_{t-1,1} + \epsilon_{t,1}, \\
r_{t,2} &= \beta_2 d_{t-1,2} + \epsilon_{t,2}, \\
r_{t,3} &= \beta_n d_{t-1,n} + \epsilon_{t,n},
\end{aligned}
$$

where

$$
d_{t-1,i} = r_{t-1,i} - \frac{1}{3} \sum_{j=1}^{3} r_{t-1,j}
$$

can be rewritten as

$$\begin{pmatrix} r_{t,1} \\ r_{t,2} \\ r_{t,3} \end{pmatrix} = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} r_{t-1,1} \\ r_{t-1,2} \\ r_{t-1,3} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \epsilon_{t,3} \end{pmatrix}.$$

If we denote

$$\boldsymbol{N} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix},$$

we can write:

$$\boldsymbol{r}_t = \beta \boldsymbol{N} \boldsymbol{r}_{t-1} + \boldsymbol{\epsilon}_t.$$

Note that β is scalar and hence it commutes with matrices. Using the lag operator, we get:

$$(\boldsymbol{I} - L\beta\boldsymbol{N})\boldsymbol{r}_t = \boldsymbol{\epsilon}_t.$$

To show (b), realize that $\boldsymbol{N}^n = \boldsymbol{N}$ and for all values of β under consideration $\beta^n \to 0$. Hence

$$\beta^n \boldsymbol{N}^n \to 0$$

and (b) follows from (theorem B.2 Lachout & Prášková 2005).

Moving to (c), with (b) at hand, we know that operator $\sum_{n=0}^{\infty} L^n \beta^n \boldsymbol{N}^n$ is well defined in the sense that $\sum_{n=0}^{\infty} L^n \beta^n \boldsymbol{N}^n \boldsymbol{\epsilon}_t$ converges both almost surely and in the quadratic mean. Applying the operator to both sides yields

$$\boldsymbol{r}_t = \sum_{n=0}^{\infty} \beta^n \boldsymbol{N}^n \boldsymbol{\epsilon}_{t-n},$$

which concludes the proof of (2).

Appendix B

Stock selection and period length selection

In this section, we will explain our selection of stocks and period length in more detail. The selection process will reflect two principles: viability of the strategy and validity of the results.

B.1 Stock selection

Viability of our strategy depends on our ability to forecast conditional expected return, which in turn depends on our specification. Recall that we have specified the conditional expected return by a deviation from the average return in the previous period. In conjunction with the results presented by Lo & MacKinlay (1990a), we believe that viability of our specification can be attributed to two phenomena: overreaction and lead-lag relationship (returns of some stocks tend to lead returns of other stocks). In turn, the lead-lag relationship can be attributed to a difference in the speed of price adjustment after arrival of information that has value implication between multiple stocks (Brennan *et al.* 1993).

Let us focus on the different speed of adjustment. The different speed of adjustment can only exist if there is information with common value implication to begin with. Hence, we will be looking for a set of stocks for which arrival of information with common value implication is possible. Moreover, the higher the frequency of arrival and the stronger the value implication, the better off we are. Arrival of information with common value implications is frequent and the value implication is strong in situations when profitability of the companies is exposed to common risk factors. Perhaps the most important factor is demand for output. Demand for output of any given set of companies is a common risk factor when the companies compete on the same market (from both product and geographic perspective).

Let us shift our attention towards overreaction. The overreaction is triggered by arrival of information with value information. As in the previous case, the higher the frequency of arrival and the stronger the value implication, the better off we are. While this requirement will not point out in any particular direction, it would help us exclude certain sectors such as utilities.

The next requirement is driven by our desire to preserve validity of our results. Several authors (for example Boudoukh *et al.* 1994) have argued that the reported profitability of contrarian strategies might be caused by non-synchronous trading. If this was true, the profitability would be spurious. Non-synchronous trading is more pervasive for stocks whose traded volume is small, which tends to be the case for small companies Lo & MacKinlay (1990b). Hence, in order to mitigate the impact of non-synchronous trading on validity of our results, we will only consider companies with large capitalization.

Finally, we will require the selected stocks to provide us with a sizable sample of historical returns. (Of course, whether the sample is large enough of not will also be influenced by our choice of period length). Again, the requirement is motivated by desire to preserve validity. In particular, results of simulation of historical performance based on a short sample might be dismissed as a fluke.

Based on the criteria, we have selected the stocks of the following companies: Schlumberger N.V, Baker Hughes Incorporated, and Halliburton Company. By April 2011, the capitalization of the companies was approximately \$120, \$30 and \$40 billion respectively. With respect to the firm size, the companies belong to the first decile of all companies traded on the U.S. stock exchanges. Moreover, all the companies have been listed since 1987, which gives us 23 years of historical returns. Finally, the companies are direct competitors in the market for oil and gas mining equipment and affiliated services. Taking the foregoing facts into account, we can conclude that the companies fulfill our criteria.

B.2 Period length

Even if we select stocks that are prone to overreaction and different speed of adjustment, our specification of conditional expected return might fail if we select inappropriate period length.

Moreover, the period length influences the transaction costs. This becomes apparent once we realize that the portfolio is rebalanced in every period. Since every rebalancing carries transaction costs, short period length would result in excessive transaction costs.

Let us proceed to considerations regarding validity of our results. We have already explained that validity of our results could be undermined by non-synchronous trading. Non-synchronous trading is much more likely to occur if the rebalancing period is short Lo & MacKinlay (1990b). Hence, in order to preserve validity of our result, we should avoid rebalancing period that is too short.

Taking in the foregoing considerations into account, we have chosen period length of one week. Drawing from findings of Avellaneda & Lee (2008), our specification of the conditional mean should be viable under this selection. With regards to the sample size, 23 years of data and weekly returns guarantee us over 1200 individual observation - a number we deem sufficient. Also, taking into account the size of our companies, nonsynchronous trading in non-existent when weekly data are considered. Finally, choice of weekly returns guarantees sizable reduction of transaction costs when compared to daily returns.

UNIVERSITAS CAROLINA PRAGENSIS
založena 1348

Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

Opletalova 26
110 00 Praha 1
TEL: 222 112 330,305
TEL/FAX: 222 112 304
E-mail: ies@mbox.fsv.cuni.cz
http://ies.fsv.cuni.cz

Akademický rok 2009/2010

TEZE BAKALÁŘSKÉ PRÁCE

Student:	Marek Janča
Obor:	Ekonomie
Konzultant:	PhDr. Ladislav Krištoufek

Garant studijního programu Vám dle zákona č. 111/1998 Sb. o vysokých školách a Studijního a zkušebního řádu UK v Praze určuje následující bakalářskou práci

Předpokládaný název BP:

Construction of a market neutral portfolio in the presence of lead-lag relationship

Charakteristika tématu, současný stav poznání, případné zvláštní metody zpracování tématu:

In the late 1980´s and the early 1990´s, a huge of body literature documenting positive cross autocorrelation between returns of individual securities traded on the NYSE appeared. Lo (1990) suggested a market-neutral portfolio construction method taking advantage of these findings. The suggested method, however, possesses several shortcomings such as not being mean-variance efficient and being only dollar neutral. The aim of my bachelor thesis is to provide an alternative method – a one which will not suffer from those shortcomings – and test the method on the Prague Stock Exchange.

Struktura BP:

Chapter 1: Motivation (Statistical arbitrage as a trading strategy, the empirical evidence of the presence of lead lag relationship)
Chapter 2: Theoretical foundations (Presentation of Lo´s model, Solution of constrained optimization problem, specification of joint and marginal distribution of returns in two periods)
Chapter 3: Application on the PSE (Estimation, results)
Chapter 4: Practical limitations (Non-trading, Transaction costs, non-stationarity, misspecification of the distribution)

Seznam základních pramenů a odborné literatury:

Altay, E. (2003), Cross-autocorrelation between small and large cap portfolios in the German and Turkish stock markets, EconWPA Finance series 0308005.
Bachelier, L. (1900), Thorie de la speculation, Annales Scientifiques de l'Ecole Normale Superieure , Vol. III, pp. 21-86.
Boudoukh, J., Richardson, M. P. and Whitelaw, R. F. (1994), A tale of three schools: Insights on autocorrelations of short-horizon stock returns, Review of Financial Studies, Vol. 7, pp. 539-573.
Campbell, J., Lo, A. and MacKinlay, C. (1997), The Econometrics of Financial Markets, Princeton University Press.
Chui, A. and Kwon, C. (1998), Cross-autocorrelation between a shares and b shares in the Chinese stock market", Journal of Financial Research ,Vol. 21, pp. 333-353.
Hayashi, F. (2000), Econometrics, Princeton University Press.
Iwaisako, T. (2004), Stock index autocorrelation and cross-autocorrelations of size-sorted portfolios in the Japanese market, Hitotsubashi University, Discussion Paper Series a448.
Khandani, A. and Lo, A. (2007), What happened to the quants in august 2007, Journal of Investment Management , Vol. 5, p. 554
Lo, A. and MacKinlay, C. (1990a), When are contrarian profits due to stock market overreaction? Review of Financial Studies , Vol. 3, pp. 175-205

Datum zadání:	Červen 2010
Termín odevzdání:	Červen 2011

Podpisy konzultanta a studenta:

V Praze dne 7.6.2010